TU Delft

Delft University of Technology

Reliable timetable design for railways and connecting public transport services

Sparing, Daniel

DOI
10.4233/uuid:a2c50995-d3b0-4db8-ac63-b30e4ec88fe2

Publication date
2016

Document Version
Final published version

Citation (APA)
Sparing, D. (2016). *Reliable timetable design for railways and connecting public transport services*. [Dissertation (TU Delft), Delft University of Technology]. TRAIL Research School. https://doi.org/10.4233/uuid:a2c50995-d3b0-4db8-ac63-b30e4ec88fe2

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

TRAIL

Summary

Railways and public transport form an essential part of our transport systems: together with walking and cycling, they are the space-efficient and environmentally friendly alternatives to private cars. However, new infrastructure is costly and therefore there is a strong need that the existing network is used in an optimal manner. All research topics in this thesis therefore focus on improving railway and public transport timetabling.

About the Author

Daniel Sparing received his M.Sc. degree in Electrical Engineering from the Budapest University of Technology and Economics in 2008. He performed his Ph.D. research at the Department of Transport and Planning at the Delft University of Technology between 2010 and 2014. He currently works as a consultant in machine learning and mathematical optimization.

TRAIL Research School ISBN 978-90-5584-203-2

Erasmus University Rotterdam · Radboud University Nijmegen · TU Delft · University of Twente · TU/e Technische Universiteit Eindhoven · University of Groningen

THESIS SERIES T2016/5

Daniel Sparing

Reliable Timetable Design for Railways and Connecting Public Transport Services

TRAIL THESIS SERIES

Daniel Sparing



Reliable Timetable Design for Railways and Connecting Public Transport Services

Propositions

Pertaining to the dissertation

Reliable Timetable Design for Railways and Connecting Public Transport Services

Daniel Sparing

25 May 2016

1. Timetable optimization results and smarter use of resources should not be excuses to avoid investing in railway infrastructure. (Chapter 3)
2. If public money is going into railways, then their historical performance needs to be open data on a detailed level. (Chapter 4)
3. In a rich country, slow, low frequency transit lines are of little value: they should be upgraded to fast, high frequency services, or scrapped. (Chapter 5)
4. Timetabling shows that being on time is a question of time reserves – but life is too short for time reserves.
5. Cars take up an alarmingly disproportionate amount of urban space – and research funding.
6. One cannot "solve" traffic jams without road pricing: if a measure temporarily relieves road congestion, that only gives people incentive to move there.
7. A researcher who gives a talk on the value of time and then queues 15 minutes for free coffee to save money does not practice what they preach.
8. For some of us, the PhD research is the time to learn some modesty and explore one's intellectual limits.
9. The PhD years are a fantastic learning opportunity: one can acquire deep knowledge on a wide range of subjects while procrastinating to avoid writing.
10. Psychology and popular science should give as much attention to friendships as they do to romantic relationships.

These propositions are considered opposable and defendable and have been approved as such by the promotor Prof. Dr.- Ing. I.A. Hansen.

Stellingen

Behorend bij het proefschrift

Reliable Timetable Design for Railways and Connecting Public Transport Services

Daniel Sparing

25 mei 2016

1. Optimalisatie van de dienstregeling en een betere benutting van de infrastructuur zijn geen redenen om minder te investeren in het spoor (Chapter 3)

2. Het spoor is mede gefinancierd met belastinggeld. Dit pleit voor openbaring van historische gegevens op een gedetailleerd niveau (Chapter 4)

3. Openbaar vervoer dat zowel langzaam is en rijdt met een lage frequentie, is in een welvarend land van weinig toegevoegde waarde. Deze lijnen moeten ofwel worden versneld en de frequentie verhoogd, of worden geschrapt.

4. De kunst van dienstregelingontwikkeling toont aan dat het op tijd zijn een kwestie is van zogenaamde tijdreserves – het leven is echter te kort voor het reserveren van tijd.

5. Auto's gebruiken een alarmerend groot deel van de stedelijke openbare ruimte – alsmede financiering van onderzoek.

6. Men kan files niet "oplossen" zonder gebruik te maken van tolheffing. Als een nieuwe maatregel tijdelijk files vermindert, dan trekt dit juist meer verkeer aan.

7. Een onderzoeker die een lezing geeft over de waarde van tijd en vervolgens 15 minuten in de rij staat voor gratis koffie doet niet wat hij zegt.

8. Tijdens de periode van het promotieonderzoek leert men bescheiden te zijn en verkent men zijn intellectuele grenzen.

9. De periode van een PhD onderzoek bevatten waardevolle leermomenten: men kan diepgaande kennis opbouwen over een breed scala aan onderwerpen om zo het schrijfwerk te vermijden en uit te stellen.

10. Psychologie en populaire wetenschap moeten evenveel aandacht geven aan vriendschappen als dat ze doen aan romantische relaties.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotor Prof. Dr.- Ing. I.A. Hansen.

Reliable Timetable Design for Railways and Connecting Public Transport Services

Daniel Sparing

Delft University of Technology, 2016

This research is supported by the Netherlands Organisation for Scientific Research (NWO).

Cover photo: Daniel Sparing

Reliable Timetable Design for Railways and Connecting Public Transport Services

Proefschrift

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus prof.ir. K.C.A.M. Luyben,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
woensdag 25 mei 2016 om 10:00 uur

door

Daniel SPARING
Master of Science in Electrical Engineering
Budapest University of Technology and Economics, Hongarije
geboren te Boedapest, Hongarije

Dit proefschrift is goedgekeurd door de
promotor: Prof. Dr.-Ing. I.A. Hansen
copromotor: Dr. R.M.P. Goverde

Samenstelling van de promotiecommissie:

Rector Magnificus	voorzitter
Prof. Dr.-Ing. I.A. Hansen	promotor
Dr. R.M.P. Goverde	copromotor

Onafhankelijke leden:

Prof. Dr. ir. S.P. Hoogendoorn	Technische Universiteit Delft
Prof. Dr. ir. R.P.B.J. Dollevoet	Technische Universiteit Delft
Prof. Dr. rer. nat. habil. K. Nachtigall	Dresden University of Technology, Germany
Prof. Dr. L.G. Kroon	Erasmus Universiteit Rotterdam
Prof. Dr.-Ing. N. Nießen	RWTH Aachen University, Germany

TRAIL Thesis Series no. T2016/5, the Netherlands Research School TRAIL

TRAIL
P.O. Box 5017
2600 GA Delft
The Netherlands
E-mail: info@rsTRAIL.nl

ISBN 978-90-5584-203-2

Copyright © 2016 by Daniel Sparing

All rights reserved. No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the author.

Printed in the Netherlands

"I like the anxiety."
- Jerry Seinfeld, on coffee causing anxiety.

Preface

This 4-year Ph.D. programme was a remarkable period for me and not only because it lasted 6 years. Graduate school, maybe not too surprisingly, is a truly educational experience: I learned so much about optimization, railways, transport, and urbanism; about conducting academic research; I learned rowing, the Dutch language, social skills. Maybe even more importantly, I discovered types of work I am inefficient at, theories I failed to understand, and skill sets I had to give up trying to acquire. That said, one does not need to be good at everything, and I truly embrace both sides of this coin: I feel that I grew just as much from what I could not master as from what I could.

Ph.D. research is highly paradoxical as it is individual work, and yet I feel indebted to so many people who made this possible. I am most thankful to my supervisors, Professor Ingo Hansen, and Rob Goverde, for inviting me to work in Delft, their critical professional guidance all along, and last but lot least for their enormous patience through all the years of my stumbling through the clueless moments and the writer's blocks. I also especially appreciate that my independent committee members agreed to evaluate my work, Professors Serge Hoogendoorn, Rolf Dollevoet, Karl Nachtigall, Leo Kroon, and Nils Nießen: their critical and insightful comments helped improve the readability of this thesis.

This research project was funded by the Netherlands Organisation for Scientific Research (NWO) within their program *Sustainable Accessibility of the Randstad*, therefore I am grateful to them and ultimately to the Dutch taxpayer for their generosity. Thanks as well to our project user group, among others Suzanne Kieft, for their advice; as well as the Dutch railway infrastructure manager ProRail, the travel information company 9292 REISinformatiegroep, and Stefan de Konink of the Dutch OpenGeo Foundation for providing data for this research.

I feel fortunate to have met so many bright and fun railway researcher colleagues in Delft and beyond, many of whom became close friends of mine since, such as Francesco Corman, Pavle Kecman, Nadjla Ghaemi, Niels van Oort, Egidio Quaglietta, Nikola Besinovic, Evelien van der Hurk, Gabor Maroti, Paul Bouman, Peter Sels, Daniel Hörcher, as well as the other members of my research team: Yuval Kantor, Andrew Switzer, Ties Brands, and Gijs van Eck. It was huge pleasure to share the Ph.D. student days with my colleagues and friends Mahtab, Giselle, Mario, Olga, Thomas and Lisa, Erik-Sander, Bernat, Tamara, Mo, Xavi and Montse, Kakpo, Meng, Yufei, and many others.

I am glad for all the talks on transport and on life with Rob van Nes, Paul Wiggenraad, Robert Bertini, Jarrett Walker, and all the train enthusiasts and transport experts at the *Jonge Veranderaars*, at the Hungarian Urban and Suburban Transit Association (VEKE), and at the Centre for Budapest Transport (BKK). Furthermore I would like to thank Ymkje de Boer and our support staff at the TU Delft and at the TRAIL Research School for all the help. I would also like to express my thanks to three of my earlier supervisors and managers who helped me shape my data scientist career to this day, Csaba Gáspár, Bertalan Danko, and Keve Müller.

I got to know so many great people living in the Netherlands beyond my colleagues and university friends. Taking the risk that the disappointment of those I forget to include in this list will be larger than the joy of the ones named, *onder anderen* thanks to my first Dutch friend Merijn and Katus, Evan, Adolfo, Giacomo, Joanna, Verka, Szinti, Dave and Andy, Sára, Jurgen and all the rowers at *De Delftse Sport*. You are the reasons I am happy in retrospect with my decision of moving to the Netherlands.

I close with saying thanks to all my lovely friends from Hungary and all over the world who visited me regularly: Panka, Dávid and Jan, Dóra, Tamás and Eri, Jani and Marie, Andris and Dóri, Andris and Miki, Zsófi and Szeki, Dalma, Tomi, Kriszta, Dorci, Ábel, Roberto, Stella, Hanna, Richard, Steven, and many more. Only with you all could these years turn out to be anything better than a solitary and disconnected experience far away from home. Finally I say thanks to my family: my parents, my grandfather, my sister, for their support during the Ph.D. years, and, well, there is no less sentimental way to put it, during all my life.

Daniel Sparing
Singapore, April 2016

Contents

Preface **vii**

1 Introduction **1**

1.1 Motivation 1

1.2 Thesis objectives 2

1.2.1 Railway timetable design and timetable stability analysis . . . 3

1.2.2 Managing transfer resistance 3

1.3 Thesis contributions 3

1.3.1 A railway timetable optimization model with timetable stability as objective 4

1.3.2 Multimodal transfer modelling 5

1.4 Thesis outline 6

2 Timetabling of railways and connecting public transport lines **9**

2.1 Introduction 9

2.2 Timetabling of high capacity utilization railway networks 10

2.2.1 High capacity utilization 12

2.2.2 Heterogeneous services 16

2.2.3 Focus on reliability 20

2.3 Synchronizing trains and connecting public transport lines 21

2.4 Literature review 21

2.4.1 Railway capacity estimation 21

2.4.2 Line and frequency planning 22

2.4.3 Timetable generation and evaluation 23

2.4.4 Timetable synchronization and connection management . . . 26

2.5 Conclusions 27

3 Optimizing periodic railway timetables for stability 29

3.1 Introduction . . . 30

3.1.1 Contributions of the model . . . 30

3.1.2 Previous work . . . 31

3.2 A railway timetable optimization model focusing on stability . . . 32

3.2.1 Initial assumptions and definitions . . . 32

3.2.2 The periodic event-activity network . . . 32

3.2.3 The optimization model . . . 41

3.2.4 Interpreting the optimization model results . . . 44

3.2.5 The expanded timetable . . . 45

3.2.6 Extensions . . . 46

3.3 Dimension reduction techniques . . . 48

3.4 Computational results . . . 51

3.5 An iterative solution approach . . . 58

3.6 Conclusions . . . 64

4 Transfer time modelling with open data 65

4.1 Introduction . . . 65

4.2 Open public transport data . . . 67

4.2.1 The Dutch example . . . 68

4.2.2 Insights from AVL data . . . 69

4.3 Transfer walking times . . . 72

4.3.1 Case Study: The Schiphol-Haarlem-Amsterdam West network 73

4.4 Application examples . . . 77

4.4.1 Transfer station resistance . . . 77

4.4.2 Synchronizing timetables: Urban night buses in Amsterdam . 80

4.4.3 Enriched travel advice for transfers . . . 87

4.5 Conclusions . . . 89

5 Delay management in a multi-operator network **91**
5.1 Introduction . . . 91
5.2 Modelling a periodic pubic transport network . . . 92
5.2.1 Variables and constraints . . . 92
5.2.2 Controllability of connections . . . 94
5.2.3 Max-plus algebra representation . . . 94
5.2.4 Passenger delay estimations . . . 96
5.2.5 Estimating minimum process times . . . 97
5.3 Defining connections in a multi-operator network . . . 98
5.4 Holding advice for significant connections . . . 99
5.4.1 The optimization model . . . 99
5.4.2 Model limitations . . . 101
5.4.3 Solution approach . . . 101
5.5 Case Study: The Green Heart . . . 103
5.5.1 The example network . . . 103
5.5.2 Candidate transfers . . . 105
5.5.3 Significant connections . . . 105
5.5.4 Optimal connection control . . . 106
5.6 Conclusions . . . 107

6 Conclusions **111**
6.1 Summary of the main contributions . . . 111
6.1.1 Optimizing high-frequency railway timetables for stability . . . 112
6.1.2 Detailed modeling of intermodal transfers . . . 113
6.2 Recommendations for future research . . . 115

Bibliography **117**

Summary **126**

Samenvatting **129**

About the author **135**

TRAIL Thesis Series **136**

List of Figures

1.1 Thesis outline 7

2.1 The railway planning process 11

2.2 Yearly rail passenger kilometres per capita (a) and railway network length per capita (b) in the EU15 and selected countries (BE: Belgium, CH: Switzerland, DK: Denmark, FR: France, NL: The Netherlands, SE: Sweden) 14

2.3 Blocking time 15

2.4 The compression method: (a) uncompressed time-space diagram with blocking times pictured, (b) compressed blocking time diagram, (c) macroscopic approximation with minimum headways shown, (d) compressed macroscopic graph 16

2.5 Detail of a 2014 departure board at Amsterdam Centraal station (Source: Netherlands Railways) 17

2.6 Possibilities of scheduling local and express services on different track layouts 19

2.7 Capacity balance 19

2.8 Ratio of respondents dissatisfied with given aspect of rail travel 20

3.1 Schematic line graph and periodic event-activity network of two train lines (Line 1 in red and line 2 in blue) with infrastructure activities (in black) 35

3.2 Exemplary time-space diagram for an overtake at a station where the slower train is stopping, with (a) overtaking train not stopping, (b) overtaking train stopping 36

3.3 Example events and activities before (a) and after (b) inserting a single dummy node for a station dwell with one faster train, not stopping (dashed black and dotted black arrows are the infrastructure events related to the original arrival and departure nodes, respectively; gray arrows are infrastructure arcs not related to the dummy node extension) 38

3.4 Time-space diagram for an overtake without stopping including a station dummy node 39
3.5 Numerical example for illegal overtake (left) and added dummy node (right) 40
3.6 Example for removal of redundant infrastructure constraints: with a red cross overlay on panel (a) and removed in panel (b) (the regularity constraints are shown as dotted gray lines, station id is not shown in nodes for clarity.) 49
3.7 Two timetables that only differ in their run numbers, to illustrate symmetry-breaking constraints 50
3.8 The SAAL corridor on a geographic map and its schematic track layout in the West-East direction (thick red line: two tracks, double red line: four tracks, Map source: ProRail) 54
3.9 Maps of the different scenarios (blue - intercity train lines and intercity stations, red - local train lines, ”Asd” - Amsterdam Centraal, ”Lls” - Lelystad Centrum, ”Bd” - Breda, ”Ah” - Arnhem Centraal, ”Vl” - Venlo, ”Gn” - Groningen) 55
3.10 Maps of further different scenarios (blue - intercity train lines and intercity stations, red - local train lines, ”Asd” - Amsterdam Centraal, ”Lls” - Lelystad Centrum, ”Bd” - Breda, ”Ah” - Arnhem Centraal, ”Vl” - Venlo, ”Gn” - Groningen) 56
3.11 Schematic track layout (top), optimal compressed (middle) and expanded (bottom) timetable for the SAAL corridor (one direction pictured, red: local trains, blue: intercity trains, black dotted lines: critical infrastructure constraints) 57
3.12 Optimized expanded timetable of the NVG scenario, Leiden-Lelystad Centrum (above) and Uitgeest-Utrecht (below) corridor 61
3.13 Intermediate objective value and its bounds during the solver run for scenarios SAAL, and NL-IC, UT and NVG (dotted lines - objective bounds, dashed line - LP bound, continuous line: intermediate objective value) 62
3.14 Calculation time of a stable timetable as a function of the number of nodes and arcs in the event-activity network 63
3.15 Calculation time of an optimal timetable as a function of the number of nodes and arcs in the event-activity network (optimum not proven in the largest 3 cases) 63

4.1 Near-perfect grid networks: (a) the rectangular grid bus network of Portland, Oregon, United States (in red), (b) the radial grid light rail network of Amsterdam, the Netherlands (in blue) 67

4.2 Departure and arrival delay distribution of bus lines included in GOVI KV6 at the analysed stations 70

4.3 Vehicle delays (a) and headways (b) along a single route 71

4.4 Average delay per stop in January 2013 (green: early, yellow: on time, red: late, on all Amsterdam bus services of the operator Connexxion) . 72

4.5 A geographic map of the selected railway network Haarlem / Schiphol / Amsterdam West (source: ProRail) 74

4.6 Schematic "tube" map of railway lines and frequencies of the selected railway network Haarlem / Schiphol / Amsterdam West (source: Treinreiziger.nl) 74

4.7 3D model of platforms (yellow) and walk paths (green), station Amsterdam Lelylaan 75

4.8 Line layout of transit lines at station Amsterdam Lelylaan. (line width is proportional to frequency, blue – bus, red – metro, yellow – tram, black – train) 79

4.9 Platform layout at (a) Amsterdam Zuid, a station with spread-out platforms, and (b) Schiphol, a compact station (source: opnvkarte.de) . . 80

4.10 Average intermodal transfer walking times and hourly number of train departures at each station in the investigation area 81

4.11 Simplified time-space diagram of night train and bus services at the Dutch stations Amsterdam Centraal, Schiphol, Rotterdam Centraal, and at Zurich Stadelhofen station (2010, source: the official online timetables of the respective public transport providers) 84

4.12 Bus platforms at Amsterdam Centraal used by night buses. 85

4.13 Current and redesigned hourly urban bus and train arrivals and departures at Amsterdam Centraal station 87

4.14 Arrival punctuality of a vehicle and transfer waiting time for the passenger at Hilversum Sportpark station, measured across all Sundays in January 2013, between 10:00 and 23:00 88

4.15 Enriched travel advice including probabilities of early departure, transfer feasibility and on time arrival 90

5.1 Connection classes depending on different amounts of arrival delay. . 100

5.2 Network of four bus lines and connecting train lines in the Netherlands 104

5.3 Connection classifications for the 20 random initial delay scenarios . . 106

5.4 Vehicle delays if all connections cancelled 109

5.5 Passenger delays if all connections cancelled 109
5.6 Passenger delays in case of optimal control 109

List of Tables

2.1 Characteristics of high capacity utilization and synchronization-based networks . . . 10

2.2 Basic railway network data of the EU15 and selected countries (2009, except Japan: 2000) . . . 13

2.3 Average commercial speeds of selected train services using the Willemsspoortunnel, Rotterdam (IC: Intercity, HST: High speed train, data source: Netherlands Railways timetable) . . . 18

3.1 Interpretation of optimizer results . . . 44

3.2 List of input data types used from the DONS/PETER timetable format 53

3.3 Graph size for different scenarios and optimal timetable calculation times 58

3.4 Stable and optimal timetable calculation times with the basic and the iterative method for different scenarios . . . 60

4.1 Example data output from BISON interface KV6 . . . 69

4.2 Walking speed on flat surface and stairs . . . 73

4.3 Distance and walking time between platform pairs at station Amsterdam Lelylaan . . . 76

4.4 Average and maximum intermodal walking distance and time at different stations . . . 77

4.5 Value of time for different passenger types and selected modes (EUR/h) 78

4.6 Number of lines, hourly frequency and commercial speed of different modes at station Amsterdam Lelylaan . . . 78

4.7 Number of train and feeder lines, hourly frequencies and commercial speed . . . 79

4.8 Urban night buses in Amsterdam (2010) . . . 82

4.9 Bus timetable at Amsterdam CS in the nights after Thursday–Saturday (2010) . . . 83

4.10 Average waiting time in minutes between lines as Amsterdam Centraal 87

4.11 Percentage of intermodal connections between 5 and 15 minutes . . . 87

5.1 List of bus lines 104

5.2 Characteristics of the example *controllable subnetwork* 104

5.3 Transfer stations 105

5.4 Results for a single delay scenario and different passenger weights . . 107

5.5 Optimization result statistics for 100 delay scenarios, $a = 2$ 107

Chapter 1

Introduction

This thesis introduces two new aspects of railway and public transport timetabling research that can contribute to the design of more reliable pubic transport networks: the stability-optimized railway timetable and the exploitation of open data in transfer modelling. For the first time, in this research a periodic timetable optimization problem is defined for heterogeneous railway networks with variable train running times, where the objective function of the optimization problem is the minimum cycle time of the network, which is an indicator of timetable stability. This optimization model can improve various parts of the timetabling processes, from infrastructure planning to line planning to the design of the actual daily timetable; and it was implemented as a software package that provides clear visual outputs of both the optimized timetable and the progress of the optimization process based on standard line planning data structures currently used in other timetable planning tools in the Netherlands.

The second set of contributions focuses on a key part of the public transport journey that has a pivotal role in satisfying or upsetting passengers: the transfer, or connection, especially between lines of different modes, run by different transport operators. We recognize that for the purposes of timetable planning and line synchronization, the accurate modelling of transfer nodes and transfer times is vital. We provide an approach to utilize open data for public transport that recently became available to improve the accuracy of transfer modelling and therefore the accuracy of timetable planning and line synchronization.

In the remaining of this chapter, we describe in detail the motivation behind this research, the two main thesis objectives, and the related two main sets of contributions of this thesis.

1.1 Motivation

The railway and public transport industry is under pressure from the regulating governments and the wider public to increase its ridership while facing limited financial

resources. This expected ridership increase is partly fueled by expected demographic increase in the future, and partly from the societal desire for a modal shift from the private car to public transport and other less environmentally harmful and more space-efficient transport modes. Limited financial sources, on the other hand, mean that simply building new infrastructure, buying new vehicles and hiring new staff is not a viable option in itself: improvements in the efficiency of public transport operations are necessary, such as improving the reliability and punctuality of scheduled train and bus operations without increased costs.

1.2 Thesis objectives

A key ingredient to increasing public transport ridership is to offer an attractive timetable and transport service. In detail, a timetable can be considered attractive if it offers a short total travel time at high frequencies, with high reliability. Another way the timetable affects the attractiveness of public transport is whether the offered capacity satisfies the demand. A high frequency operation is therefore desirable both from the capacity and the total travel time point of view, but attention is necessary to reliability due to the high capacity utilization. It is, however, not justified to run vehicles at high frequency in case of low demand: in case of low frequencies, the differences in total travel time between different network timetable options are dominated by the transfer waiting times. Therefore, in the following, we focus on two different timetabling problems that appear in practice in different locations and times: maximizing throughput and stability of high frequency operations, and minimizing transfer resistance in case of low frequency operations.

For the design of reliable high-frequency operations, we continue to focus on railway networks. As many railway lines, such as in the Netherlands, have already a high capacity utilization rate, even higher train frequencies can lead to an unreliable network where small disturbances have wide and long-lasting effects. The notion of timetable stability is used to describe the resilience of the timetable to small disturbances, and this needs to be taken into account in the design process.

The transfer resistance of lines operating at low frequencies is mostly caused by the too long transfer waiting times, which can be improved by synchronized timetables. If, for example, the buses or other vehicles of connecting public transport lines depart just after transferring passengers arrive from a train, then the transfer waiting time is eliminated, which can be half the headway on average in an uncoordinated case. On the other hand, in case of a short planned transfer waiting time and frequent arrival delays, the transfer waiting time can be as large as the headway of the departing vehicle. Therefore both timetable synchronization and holding strategy of departing vehicles are necessary.

1.2.1 Railway timetable design and timetable stability analysis

Traditionally, railway timetable design and timetable stability analysis have been considered as independent problems that were solved sequentially. The timetable design problem aims to generate a feasible timetable based on the desired line pattern, frequencies and the available infrastructure. When formulated as an optimization problem, the objective function can be e.g. total vehicle running time, total passenger travel time, total waiting time, costs, or a combination of these measures. The existing network timetable stability analysis method, on the other hand, takes a defined timetable as input, and outputs a measure of timetable stability, such as an estimate of the capacity utilization ratio. It is necessary, however, to include the notion of timetable stability already in the design phase to ensure reliable timetables via finding a balance between variations of train speeds, headways and buffer times.

- Research objective 1: Develop an optimization model to maximize the stability of periodic railway network timetables.

1.2.2 Managing transfer resistance

Several gaps are identified in current practice that account for suboptimal timetables with respect to transfers, especially in case of changing between different modes or operators and lines with low frequencies. Current timetable planning methods often use simple norms for minimum design headways between arrivals and departures of interconnected lines at transfer stations, as well as rough estimates of passenger transfer walk and waiting times, while an accurate estimate of transfer walk times is desirable, based on a detailed model of vehicle platforms and passenger routes of transfer stations, especially in case of connections between different modes or different operators. When synchronizing timetables, systematic deviations from the timetable during operations, i.e. the delay distributions of the lines in question have to be taken into account. Finally, in case of synchronized timetables, dispatchers need more accurate real-time information and decision support identifying important connections at risk based on actual delays. All of these objectives can be facilitated by the recent availability of open or freely accessible transit data.

- Research objective 2: Detailed modelling of intra- and intermodal passenger transfers using open transit data.

1.3 Thesis contributions

This section highlights the main contributions of the research documented in Chapters 3–5 of the thesis. Following the structure of the research objectives, we group the

contributions into the ones related to high-frequency railway timetable design and the ones related to multi-modal networks and the detailed analysis of transfers.

1.3.1 A railway timetable optimization model with timetable stability as objective

We developed a new, stability-focused periodic timetable optimization model for busy railway networks. This approach allows for a quick evaluation of whether a required train line pattern is feasible on known infrastructure, and provides an optimized timetable if possible. Such a model can be used for a variety of timetabling purposes: supporting the design of the actual railway timetable; in experimenting with new, innovative stop patterns of train lines and evaluating their feasibility; and the evaluation of infrastructure bottlenecks to identify which infrastructure improvement could yield the best results in increasing capacity.

The main contribution of this part of the research is the idea to directly use the timetable stability as the main objective of the mathematical optimization, and its application on a heterogeneous train network with flexible train speeds and train orders. Periodic timetable stability, as we shall see in the later chapters, is quantified for this purpose by the minimum cycle time of the timetable: if this minimum cycle time is less than the nominal timetable period (such as one hour), then and only then the timetable is stable, and the bigger this gap is, the higher the stability is, at the expense of unused capacity. This key idea allows for the integration of previously separate timetable design steps: the choice of train orders and the evaluation of timetable stability; thus avoiding the need for a feedback loop of several iterations of timetable design.

This timetable optimization model is also highly flexible: while it does assume a fixed line pattern and routing, the running and dwell times are only constrained from above by sensible business rules of what is considered an acceptable running and dwell time reserve, and similarly, train orders and overtake locations are flexible.

The limited railway infrastructure, however, is explicitly modelled, headways separating trains and overtake limitations are taken into account. When defining overtake constraints at stations, we introduce a new method to work only with headway constraints and dummy nodes, in order to avoid the definition of a large number of new train order constraints that would otherwise be necessary for flexible train order models.

Our main contributions towards reducing the problem size of the optimization problem are twofold. First, we use a number of reduction techniques to reduce the problem definition, taking advantage of the symmetry of the periodic timetable and applying so-called symmetry-breaking constraints. Second, we developed an iterative solution method using a flexible range of the cycle time, that adaptively re-adjusts the cycle time range. This ensures that intermediate solutions are found fast, that there is feedback on

the solution process, and finally, simply to speed up the calculation time to the optimal solution.

In summary, the main contributions are the following:

- a new railway timetable optimization model that offers great flexibility: run times and supplements, dwell times, overtake locations and train orders are fully flexible within the predefined business and infrastructure constraints;
- using a flexible cycle time as objective function of the timetable optimization, in order to optimize the timetable directly for stability;
- an implicit modelling of train overtake constraints using headways and dummy nodes, to allow for flexible train orders;
- an iterative solving method of the optimization problem in order to provide fast intermediate solutions and feedback on the solving progress for large instances;
- dimension reduction methods taking advantage of fixed-interval timetables in real life in order to reduce the solution space of the mathematical optimization problem.

1.3.2 Multimodal transfer modelling

In the second part of our research, the contributions focus on improving the modelling of transfers to help transit planners minimize transfer waiting time, synchronize timetables, and therefore substantially improve the connection experience for passengers, which can in turn have a key effect on the general attractiveness of public transport.

We connect the recent availability of open transit data to the needed more accurate measure of transfer times at large stations and provide a method to use free geospatial data to improve previously crude estimates on the required transfer walking time. This walking time calculation can largely be automated provided that a simple three dimensional model of a transfer node and the related platform assignments of lines are captured. With an accurate distance and time estimate between each pair of platforms, taking into account details like the platform length, stairs, and escalators, it becomes possible to more accurately assess which desired transfer time is feasible for commuters or less experienced passengers, and which rescheduling or platform reassignment options can improve the connection. The model is applied to multimodal transfer nodes of railways, tram, bus, and metro lines.

We also contribute three worked out case studies of the detailed walking time calculations. The first is a quantification of the transfer resistance of a whole station, which allows for comparison of the performance of different transfer nodes and the identification of bottlenecks in a larger network. The second is an example of timetable synchronization of a low-frequency multi-modal network, the night bus and night train

network timetable in Amsterdam. In this case, accurate walking time calculations are essential, as otherwise a badly synchronized timetable can actually lead to a large increase of transfer waiting times, if the scheduled connections are missed frequently. The third and final case study is an extension of passenger information systems, where historical arrival and departure time data, as well as transfer walking time data is used to provide the passengers with a data-driven estimate on the probability of assuring a connection, as well as an estimate on departure and arrival time delays, again, based on historical operations performance.

Finally, we propose a generic method for the description and solution of the delay management problem by using the max-plus algebra technique. The two main contributions of this chapter are the following. First, we show how to use the results of a max-plus algebra-based delay propagation algorithm to filter out connections at risk. Second, we propose an optimization approach using exhaustive search on this small set of shortlisted connections, to propose a connection management decision based on the minimization of total passenger waiting time.

The summary of the main contributions is the following:

- a method for using open data to accurately estimate transfer walking times;
- a set of case studies for the above, including the estimation of transfer station resistance, timetable synchronization, and improved passenger information;
- a max-plus algebra-based reformulation of the delay management problem and a simple calculation of network delay propagation;
- a fast method for automatic delay management of low-frequency timetable networks that filters out connections at risk and provides advice for dispatchers on holding measures.

1.4 Thesis outline

The six chapters of this thesis are organized as follows. The current, introductory chapter gives an overview of the motivations, objectives, and contributions of the thesis. Chapter 2 reviews the existing literature on railway timetable design algorithms and timetable stability analysis methods, as well as on multimodal scheduling and delay management. Chapter 3 presents a railway timetable optimization method focusing on the stability of the timetable, based on Sparing & Goverde (2013b). Chapter 4 focuses on timetable synchronization between railways and connecting public transport lines, based on my contributions in the joint journal paper van Oort et al. (2015), and based on Sparing & Goverde (2011). A new methodology including example applications is presented to estimate transfer station resistance based on the physical layout of transfer stations and the timetable. Chapter 5 presents a methodology to identify

important connections at risk in case of a multimodal network with efficient delay propagation calculation by max-plus algebra modelling, based on the journal paper Sparing & Goverde (2013a). Finally, Chapter 6 summarizes the findings of the thesis and provides suggestions for future research.

A graphical outline of the thesis is represented in Figure 1.1. The reader interested in railway scheduling is advised to read Chapters 2–3 & 6, while the reader interested in multimodal timetable synchronization is advised to follow the Chapters 2 & 4-6.

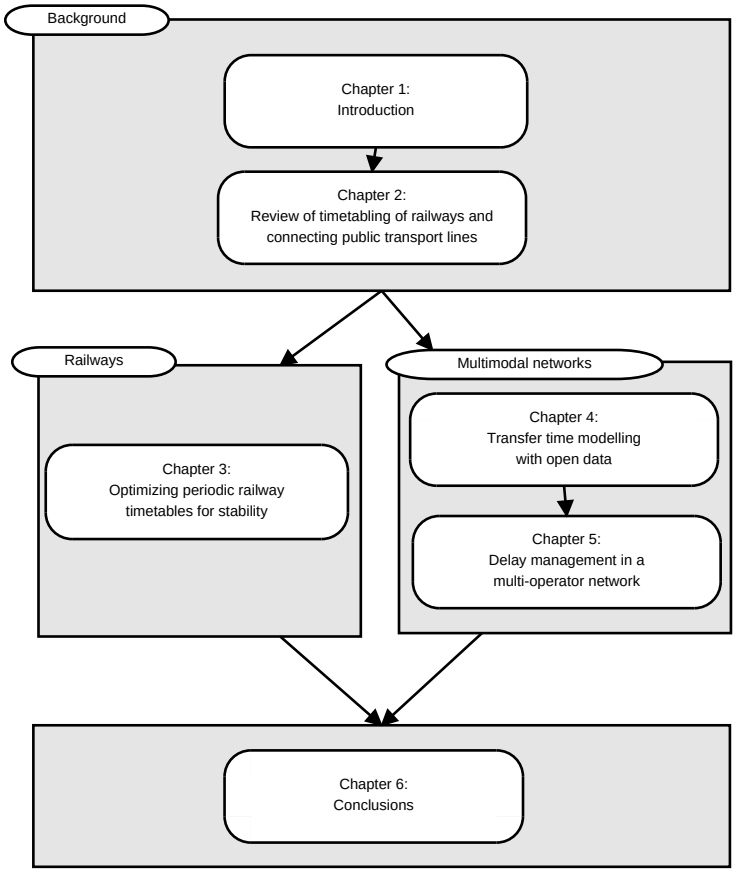


Figure 1.1: Thesis outline

Chapter 2

Review of timetabling of railways and connecting public transport lines

2.1 Introduction

The timetabling process for a railway or public transport network consists of defining the planned departure and arrival times of vehicles, based on the available resources and the expected demand. Resource availability concerns the available infrastructure defining possible routes, speeds and capacity; as well as the available fleet of vehicles and staff. Timetabling can also point out where infrastructure bottlenecks are or whether the available resources are sufficient, in other words, where investment might be necessary or beneficial. Expected demand describes the expected amount of passengers using the transport service, estimated by trip origin-destination measurements, transport assignment models and in case of existing networks, vehicle occupation measurements. As the timetable itself can also influence demand, matching the timetable to the demand can also be seen as an interactive process.

The timetabling of railways and public transport faces many requirements, such as infrastructure capacity limitations, financial constraints, overcrowding or fulfilling a minimum service requirement despite low demand. Fortunately, these challenges usually do not appear at the same time and networks can be classified or decomposed into very different types of systems based on their capacity utilization and the relationship between supply and demand. Walker (2008) classifies the purposes of public transport into *patronage goals* and *coverage goals*, where the former seek to maximize ridership for financial and environmental reasons, while the latter strive for a minimum service quality at all locations based on social reasons and geographic equity. Inspired by this classification, we propose to divide railway and public transport systems into high frequency, *high capacity utilization networks*, where demand is high and the goal of timetabling is to maximize capacity, and low frequency, *synchronization-based networks*, where the low demand does not justify high frequencies and therefore the synchronization of transfer connections is essential to provide an attractive service. Table 2.1

Table 2.1: Characteristics of high capacity utilization and synchronization-based networks

	high capacity utilization networks	**synchronization-based networks**
demand vs. supply	demand $>$ supply (overcrowding)	demand $<$ supply (room for higher ridership)
goal	meet existing high demand	provide basic service
frequency	high	low
frequency determined by	infrastructure capacity	operational costs
typical headway (min)	$\leq$5 (urban), $\leq$15 (regional)	10$\leq$ (urban), 30-60 (regional)
passenger arrival rates	uniform	according to timetable
timed/guaranteed transfers	not necessary	desirable
infrastructure capacity	important	important only at synchronization nodes
way to increase ridership via timetable	increase capacity to meet existing high demand	provide timed and guaranteed transfers to minimize transfer resistance
Examples	metro networks, railway networks in dense urban regions	local trains and regional buses outside dense urban regions

provides an illustration to the characteristics of the two types of networks.

Based on the classification of networks above, we divide the literature review of railway and public transport timetabling as follows. In Section 2.2 the timetabling of high capacity utilization networks is explored, where the goal is to estimate the infrastructure capacity, check the feasibility of line plans and timetables, and evaluate the reliability of timetables. We restrict the scope here to railway networks. In Section 2.3 on the other hand, we consider synchronization-based networks, where the frequencies are limited because of cost reasons and low demand, and the design goal is to minimize transfer wait times by synchronization. Here we focus on the intermodal synchronization of low-frequency train lines and connecting public transport services.

2.2 Timetabling of high capacity utilization railway networks

The railway timetable is the essential product of a passenger train operator: it defines the service offered including sequence of stations any train line is serving, the travel times between these stations, the frequencies of the train services and implicitly the

possible connections and connection waiting times in case a direct train service is not offered. The railway timetable of a train operator can be compared to a menu of a restaurant: the possible clients of the railway companies, the travelers judge the railway system and decide to use the service or not based on the attractiveness of the timetable, and on the ability of the operator to execute the timetable with at most acceptable deviations.

The timetable is also central in the planning process of the train operator (see Figure 2.1, based on Caimi (2009)). Once passenger demand is known, infrastructure planning, line planning and timetabling can be seen as forming a loop in the planning process. While in traditional railway planning infrastructure planning came first and timetabling later, in the Swiss railway development program *Rail 2000* timetabling has explicitly been defined as prior to infrastructure planning (Caimi, 2009). In any case, understanding these steps as part of a loop is sensible given their mutual dependency. Finally, vehicle (rolling stock) and crew rosters are driven by the required timetable. Once the planning process is complete, the performance of the operator is most commonly evaluated by some measure of deviation from the timetable.

On a societal level, the timetable influences the mode choice of travelers and therefore has an societal effect via the different externalities of transport modes, such as traffic safety, pollution, livability, and travel time. Furthermore, if a desired timetable requires rail infrastructure investment, then it has an effect on public budgets, as most infrastructure projects are publicly financed regardless of transport mode.

In case of railway networks facing high demand, such as the main railway lines of Western Europe, the goal of timetabling can be informally stated as running as many trains as possible on the given infrastructure. The timetabling process can deliver insights into where the infrastructure should be extended in an efficient way: a notable case where the timetable process predominantly drives the infrastructure planning process is the philosophy of the Swiss Rail 2000 project (Caimi, 2009).

If we further on assume a fixed infrastructure, then the timetabling process consists of the following steps. Estimating the capacity of a railway line or network consists of calculating this maximum possible number of trains independent of the timetable, but with the assumption of the required train types and their frequencies. Line planning is the process of determining the routes of trains, their stop patterns and frequencies,

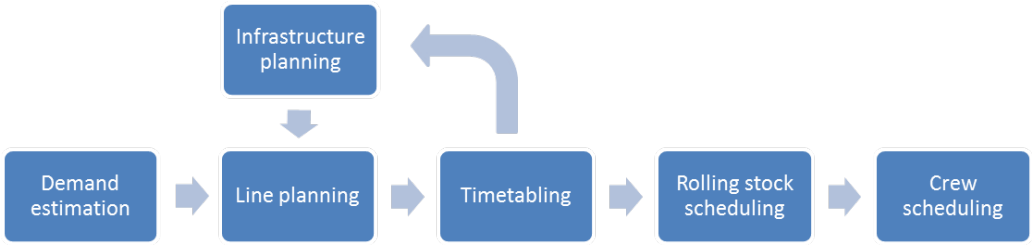


Figure 2.1: The railway planning process

based on the demand and possibly the results of the capacity estimation. Timetable generation is the core process of timetabling, trying to find a feasible schedule on the given infrastructure according to the desired line and frequency plan. Finally, timetable evaluation mostly consists of methods to analyse the reliability of timetables.

The extent of computational challenge railway timetabling provides can also be seen in the substantial academic and industry effort it attracts: in 2008, the Franz Edelman award of the operations research society INFORMS was awarded to Kroon et al. (2009), the developers of the new Dutch railway timetable, while in 2009, IBM opened a Global Rail Innovation Center in Beijing, China (IBM, 2009). In an effort to explain this complexity, the following we point out three characteristics of the timetabling problem that prove to be challenging in practice: high capacity utilization, heterogeneity of services, and focus on reliability.

2.2.1 High capacity utilization

Many railway corridors are close to saturation, meaning that new train services can only be added if another service is cancelled or causing substantial delays. See Table 2.2 for an international comparison of population, railway network length, yearly rail passenger kilometers, and railway modal share of different countries (data sources: European Commission (2013a); World Bank (2014); Ministry of Land Infrastructure Transport and Tourism (2014); East Japan Railway Company (2002)). The Netherlands stands out as a country with a particularly high yearly passenger kilometers versus network length ratio: Denmark has less than half the Dutch passenger traffic on a similar network size, and Sweden has comparable traffic on a much longer network. On the other end of the spectrum, the Japanese network is ten times larger but the passenger traffic is more than 20 times higher.

A practical example and testimony of this saturation is the controversy around the intercity service between the Amsterdam and Brussels, as the following. The hourly Amsterdam–Brussels intercity service, also called the *Beneluxtrein* (Benelux train), was cancelled at the December 2012 timetable change together with the introduction of high speed trains between the same two cities. The rolling stock used for the high speed service, however, lost its license on 17 January, 2013, to operate on the Belgian rail network (De Standaard, 2013). Therefore an intercity service every two hours between The Hague and Brussels was restored one month later (Netherlands Railways, 2013). This new service, however, provides inferior coverage and frequency to its predecessor, as the 2013 timetable was already planned with the cancelled Beneluxtrain in mind and left little space for ‘new’ services.

If we observe the trends in railway traffic and network size, we can see that the capacity utilization is steadily increasing in Western Europe. On Figure 2.2 the yearly rail passenger kilometres (a) and the rail network size (b) is plotted – both normalized to per capita –, for the EU15, selected countries within the EU15, Switzerland and

Table 2.2: Basic railway network data of the EU15 and selected countries (2009, except Japan: 2000)

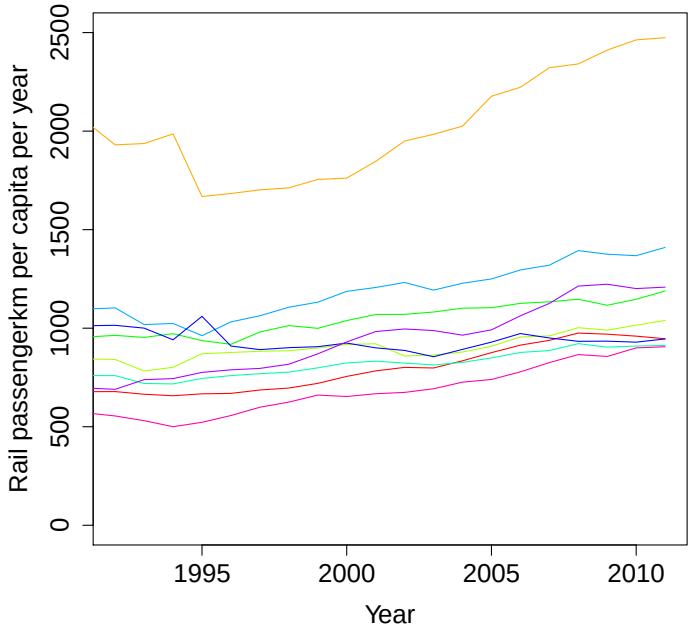
Country	Population	Network length (km)	Passenger kms (million pkm/y)	Modal share
Belgium	10,753,000	3,578	10,427	7.4%
Germany	82,002,000	33,706	81,206	7.8%
Denmark	5,511,000	2,131	6,152	9.5%
France	62,466,000	33,778	85,914	9.0%
Netherlands	16,486,000	2,886	15,400	8.8%
Sweden	9,256,000	9,946	11,321	8.7%
UK	61,595,000	16,173	52,765	6.8%
EU15	394,456,000	152,491	356,557	7.1%
Switzerland	7,702,000	3,544	18,571	17.3%
Japan	126,870,000	20,165	393,765	27.0%

Japan, in the 20-year time window of 1991–2011. While the railway network size is stagnating in all countries in question, there is a clear and consistent increase in passenger traffic, that seems to be unaffected even by the global financial crisis since 2009.

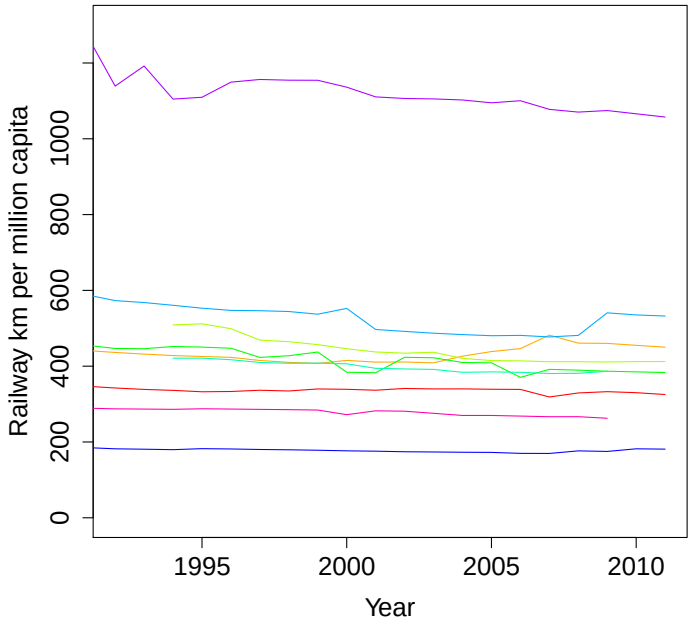
Returning to our running example of the Netherlands, two possibly related phenomena are apparent in Figure 2.2 (data sources: European Commission (2013a); World Bank (2014)). First, contrary to all other countries the comparison and the EU15 average, Dutch passenger traffic did not increase in this time window. Second, the size of the Dutch network is surprisingly small on a per capita basis in comparison. These two facts put the common marketing motto of the Dutch railway network *"drukst bereden spoor"* (most dense railway) (ProRail, 2014) in international perspective: this is not caused by unusually many passenger kilometers but by the rather small network size per capita. We note that the Dutch Central Agency for Statistics came to the same conclusion (Ramaekers et al., 2009).

From the timetable planning perspective, the high capacity utilization means that timetable optimization problems in practice often degenerate to timetable feasibility problems, i.e. in practice typical questions are whether a new train service, a new stop on an existing train line, or a frequency increase is possible on the given infrastructure.

A more formal definition of measuring *infrastructure utilization* and *capacity consumption* is provided in the International Union of Railways (UIC) leaflet on capacity (International Union of Railways (UIC), 2013) via the compression method applied to sections of the railway network. Intuitively, the UIC capacity consumption is defined as the ratio of the minimum time needed for a practically representative set of



(a)



(b)

Figure 2.2: Yearly rail passenger kilometres per capita (a) and railway network length per capita (b) in the EU15 and selected countries (BE: Belgium, CH: Switzerland, DK: Denmark, FR: France, NL: The Netherlands, SE: Sweden)

trains in a representative order to traverse any given point of the section to the time scheduled for all these trains. In more detail, given a time-space diagram of the given set of trains in given order on the particular section, maximal *compression* of the train paths is applied in the time dimension while keeping their orders so that the trains can still follow each other without any speed limitation for traffic reasons (*conflict-free*). For fixed-block signalling, as used on virtually all railway lines (as opposed to moving block signalling used on a few urban metro lines), the compression method is further detailed as follows.

The *blocking time* (also called *occupancy time* in International Union of Railways (UIC) (2013)) of a given train on a given block section is defined as time difference between the train driver is able to see a signal and that the same signal can become clear (most often showing a green aspect) for a following train (see Figure 2.3, source: International Union of Railways (UIC) (2013)). Then the time-space diagram is maximally compressed so that the train speeds remain unchanged and the blocking times do not overlap. Finally, a new instance of the *first* train is added virtually as *last* to the compressed stack. The infrastructure utilization then is defined as the ratio of the headway between this first and last train (equal along the line) and the scheduled time duration for the trains in question; and capacity consumption is the same measure additionally including time supplements, but without any buffer time between the blocking time diagrams. See Figure 2.4 for an example of a time-space diagram with blocking times pictured (a) and its compressed version, as well as a macroscopic approximation of the train paths with minimum follow-up times (headways) defined based on the blocking time calculation (c) and its compressed version (d).

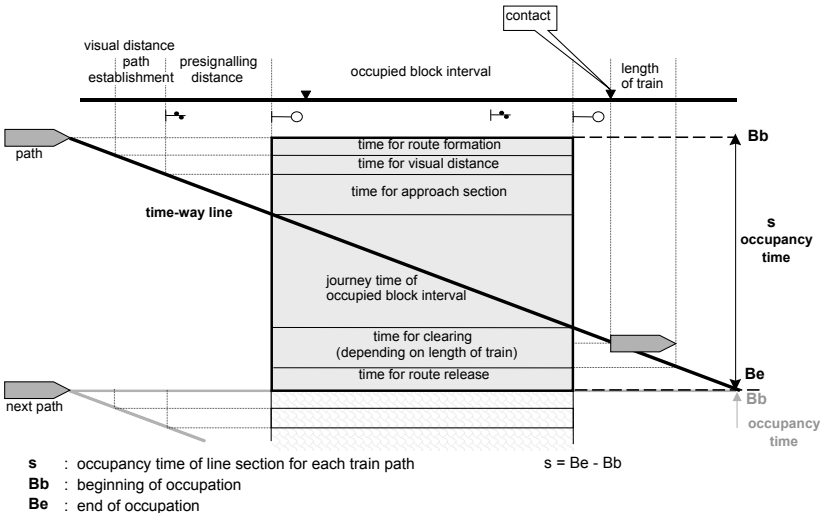


Figure 2.3: Blocking time

The high capacity utilization of some railway networks differentiates the planning challenges from other timetabled transport modes. For example, the capacity bottleneck of commercial aviation is runway capacity (Smith, 2013), with ample capacity available

in the air between the origin and destination. Urban and long-distance bus and coach services also most often can assume sufficient public road capacity and therefore the planning constraints remain available vehicles, crew, and profitability. Metro (rapid transit) and some light rail networks, however, also often experience saturated capacity along the service line. An example evidence of metro line capacity saturation is the retrofitting of Paris Metro line 1 with automatic train control to decrease headways to 85 seconds (RATP, 2010). What does differentiate railway networks from urban metro networks, however, is the wide variety of services using the same infrastructure. This is explored in the following section.

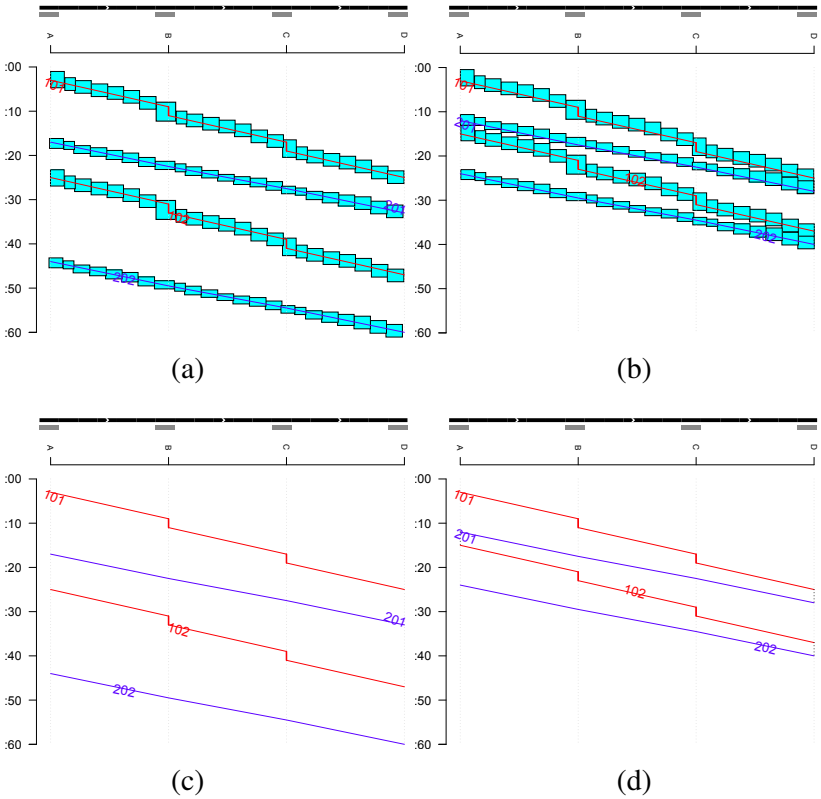


Figure 2.4: The compression method: (a) uncompressed time-space diagram with blocking times pictured, (b) compressed blocking time diagram, (c) macroscopic approximation with minimum headways shown, (d) compressed macroscopic graph

2.2.2 Heterogeneous services

On railway networks, the trains using the same infrastructure most often have significantly different origins, destinations, stopping patterns, and commercial speeds. Passenger train services are organized in most countries into train classes of different typical stop distances and commercial speeds from local trains calling at all stations to intercity to high speed services. The underlying reasons are both offering higher capacity and shorter travel time between the more important station pairs. If the time-



Figure 2.5: Detail of a 2014 departure board at Amsterdam Centraal station (Source: Netherlands Railways)

table allows for convenient transfers between local and long distance trains, this system also means a shorter travel time between smaller stations than with only local trains, via transfers to a long distance service and/or vice versa. Furthermore, freight trains using the shared infrastructure add to the complexity. While the maximum speeds of freight trains are lower than that of passenger trains and their braking distance longer, freight trains may also be faster than passenger trains around stations where the latter are expected to stop.

See Figure 2.5 for an example showing a detail of a departure board at Amsterdam Centraal station. The topological map shows the railway line from Amsterdam Centraal via Schiphol airport to Rotterdam and further. While local trains stop at all listed stations and more, Intercity trains only stop at the *IC*-labeled stations and high speed trains only stop at *HST*-labeled stations. These latter only share physical infrastructure with the other trains between Amsterdam Centraal and Hoofddorp, as well as from Rotterdam Centraal via Rotterdam Blaak to Rotterdam Lombardijen (not pictured). In reality, the diversity in stopping patterns is even more complex: some Intercity trains do not stop in Schiedam while some Intercity trains do stop at Amsterdam Lelylaan.

Another example from the same line is shown on Table 2.3, listing the average commercial speeds of selected Southbound train services all using the same infrastructure, the Willemsspoortunnel under the Maas river in Rotterdam. This tunnel contains the station Rotterdam Blaak present on Figure 2.5. These train services range from around an hour to more than three hours in duration, from local train lines to international high speed trains. There are furthermore frequent freight trains using this line, not included in Table 2.3.

Much of the challenges and options of scheduling multiple services of different line

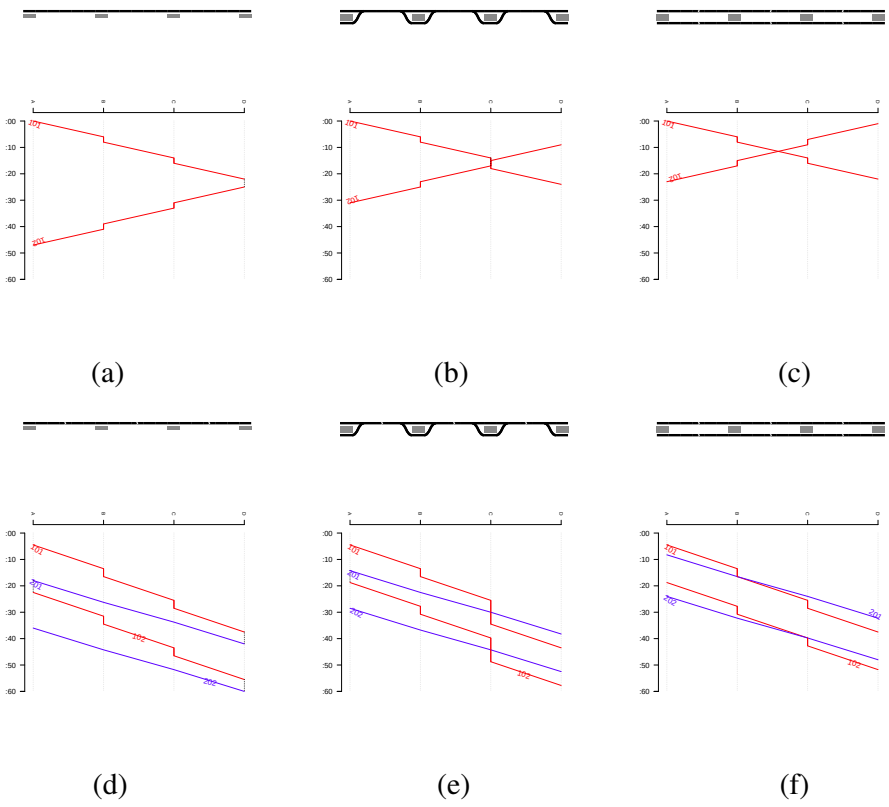
Table 2.3: Average commercial speeds of selected train services using the Willemspoortunnel, Rotterdam (IC: Intercity, HST: High speed train, data source: Netherlands Railways timetable)

Train class	From	To	Line length (km)	Duration (HH:MM)	Mean line speed* (km/h)
Local train	The Hague	Roosendaal	78	01:26	54.4
Intercity	The Hague	Venlo	168	02:20	72.0
Int'l IC	The Hague	Brussels	156	02:25	64.6
HST	Amsterdam	Breda	102	01:08	90.0
Int'l HST	Amsterdam	Paris	472	03:27	136.8

*Including time loss due to deceleration at intermediate stops

patterns can be illustrated by a simple example of scheduling a periodic local and a periodic express train service on a four-stop line (Figure 2.6). Assuming that only one track is available in one direction with no possible overtake location (siding) between end stations *A* and *D*, the only option available is to schedule the trains sequentially (d), with possibly reducing the speed of the express train service if necessary. In case overtake is possible at a siding at intermediate stations *B* and *C*, a revised timetable is possible with an overtake at station *B* (e), resulting in much lower capacity utilization. Finally, in case at least two tracks are available along the line, the two train services can run independently without concerns about overtaking. The example shows that the sidings and the second track increases the flexibility and the capacity the railway line – at the cost of the additional infrastructure. We note that these simple choices are similar to the problem of scheduling (local) trains in two direction on a single-track line: while no sidings or second track result in very large headways (a), sidings (b) and finally multiple track (c) provide flexibility and low headways, at the cost of new infrastructure.

The heterogeneity of services sharing infrastructure sets railway networks apart from other timetabled transport modes. In case of urban public transport, express metros, trams and buses are rather exceptions to the rule (New York City, Lyon and Budapest are examples of such services, respectively). An earlier edition of the UIC leaflet on capacity also pointed out the wide difference in heterogeneity between metro and heavy rail networks, as reproduced in Figure 2.7 (International Union of Railways (UIC), 2004). While long-distance buses and commercial flights can have very different service patterns, they tend to not share the same limited infrastructure, as we have discussed in the previous section.



(a) (b) (c)

(d) (e) (f)

Figure 2.6: Possibilities of scheduling local and express services on different track layouts

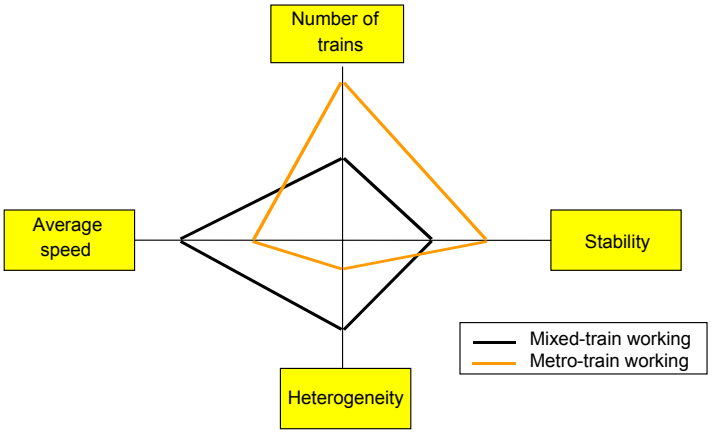


Figure 2.7: Capacity balance

2.2.3 Focus on reliability

The customers of railway and public transport services have very high expectations on on-time performance and reliability. For example, according to the yearly report of the Netherlands Railways (2012), while 74% of their customers were in general satisfied with the service provided, only 49% were satisfied with on-time performance—the lowest value of seven criteria. This customer dissatisfaction exists despite the fact that the same railway network has a World-class on-time performance of 94.2% of the trains arriving within 5 minutes of the planned arrival time (same source). According to an Eurobarometer report (European Commission, 2013b), if we ignore cleanliness of train cars, the only aspect respondents were more dissatisfied with than delays (*"Punctuality and reliability"*) was with what happens to them *during* those delays (*"Information provision [...] in particular in case of delay"*) (Figure 2.8, data source: European Commission (2013b)).

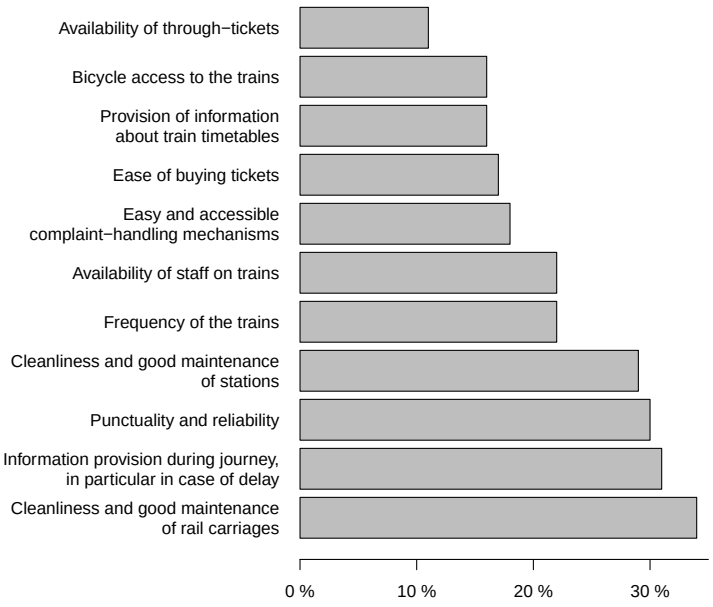


Figure 2.8: Ratio of respondents dissatisfied with given aspect of rail travel

In travel behavior research, it appears that travel time reliability has both a high importance and a high dissatisfaction level for potential passengers (Van Oort, 2011), and therefore reliability is a decisive factor in modal choice (König & Axhausen, 2002).

All this evidence on the importance of reliability implies that in railway timetabling particular attention has to be given to the practical reliability of the generated timetables. This can be in the form of an *a posteriori* sensitivity analysis of generated timetables or ideally the incorporation of some measure of reliability already in the timetable optimization process—as in case of the model presented later in Chapter 3.

2.3 Synchronizing trains and connecting public transport lines

In the previous section describing scheduling in a typically high capacity utilization setting, we assumed that demand justifies the highest possible capacity given the infrastructure and other resources, and the computational task was to find such an optimal schedule. This scenario is indeed realistic in many urban rapid transit networks and intercity railway lines around dense urbanized areas. In contrast, in case of lower demand, such as many regional bus networks or regional railway lines, the maximum utilization of capacity cannot be met and the level of service is constrained by the operating costs, rather than infrastructure constraints. In these cases, the potential of improving the quality of the timetable lies in speeding up the lines and reducing the transfer waiting times between these services without increasing their frequencies.

A sequence of arriving and departing services at the same stop area, where passengers might alight and board (transfer, connect) as part of a meaningful journey, is called a *connection* or *transfer*. It is a a *timed transfer* if the timetable is designed such that the transfer waiting time before boarding a low frequency service is substantially less than its headway, and adjusting the timetable to introduce timed transfers can be called *synchronizing* the timetable. If the operation of the vehicles in a timed transfer is organized in a way that the departing vehicle waits in case of a slight delay of the arriving vehicle, the transfer is a *guaranteed transfer* and the policies and dispatching actions deciding the thresholds for waiting is called *connection management*.

In section 2.4.4 we refer to earlier work in synchronizing timetables and in connection management.

2.4 Literature review

2.4.1 Railway capacity estimation

The most widely accepted standard to estimate railway capacity is defined in the Capacity leaflet of the International Union of Railways (UIC) (2004, 2013), as described in Section 2.2.1. This compression method based on blocking time theory counts as a microscopic model of a single line section, given the train orders and speeds.

Further works extend the railway capacity calculation to nodes and complex networks and methods independent of the timetable. Lindner (2011) investigate the applicability of the compression method and point out that it is not applicable for evaluating station capacity, and therefore insufficient for a network where the limiting factor for overall capacity is capacity at stations, not between stations. Mussone & Wolfler Calvo (2013) calculate railway capacity without using the timetable, based on the work of Burdett &

Kozan (2006), by dividing the network into nodes and paths for which the occupation ratio is calculated given the number of trains and blocking times for each train and resource. An optimization model is then defined to maximize the number of trains, while respecting the capacity limits. As no actual timetable is used, in effect their method can be seen as calculating an upper bound for the maximum utilization: while it is not certain that a feasible timetable exists with the calculated number of trains (headway conflicts may still exist, and they do not include the set-up time of the route, driver reaction time and approach time), certainly no feasible timetable can include more trains.

De Kort et al. (2003) uses max-plus algebra to estimate railway capacity independent of the timetable in a probabilistic way. The proposed method allows for the modelling of complex infrastructures, for example a line with mixed single-track and double-track sections as in the example, however it is not possible to differentiate between different train types or different train speeds and stop patterns on the same infrastructure.

2.4.2 Line and frequency planning

The problem of line and frequency planning is to determine the geographical layout of train services and their frequencies given the available infrastructure and other resources, and given passenger demand. Approaches typically can be classified as minimizing the number of passenger transfers (or, as an approximation, maximizing the number of direct connections), and minimizing operational cost. A detailed review of line and frequency planning literature is given by Guihaire & Hao (2008), see also Schöbel (2012).

The following works optimize the line and frequency structure of the Dutch railway network. Bussieck et al. (1997) uses mixed-integer linear programming (MILP) with cutting planes and relaxation to calculate line and frequency plans maximizing the number direct passenger connections on the Dutch InterRegio (medium and long distance) railway network. Bussieck et al. (2004) optimizes both direct connections and cost, and uses non-linear optimization, additionally to MILP, resulting in faster calculations of the optimal timetable. Goossens et al. (2006) uses branch and cut to find lower bounds for the same Dutch line planning instances. Goerigk et al. (2013) compares the effect of four different line planning methods on timetabling, robustness, and delay management, confirming that more direct trips lead to shorter travel times. They also observe that the proportion of missed connections does not only depend on the buffers but also on the line planning algorithm type.

Other works on timetable optimization use urban public transport networks as case studies. Guan et al. (2006) simultaneously calculates a line and frequency plan and the passenger assignment, illustrated on a case study on the Hong Kong rapid transit network. Nachtigall & Jerosch (2008) combines the optimization problems of line planning of an urban bus network with traffic assignment. Hadas & Shnaiderman (2012)

assumes a fixed urban transit line and calculates the frequencies minimizing both the number of empty seats and passengers missing the vehicle due to overcrowding.

2.4.3 Timetable generation and evaluation

Cacchiani & Toth (2012) present a survey of timetabling methods, classifying them into *nominal* and robust methods, based on whether they take into account timetable stability or not. The methods are further classified, among others, whether they are used for periodic or aperiodic timetables, valid only on a corridor or on a network level, and different objective functions. While Borndörfer & Liebchen (2008) provide a particular example when a cyclic, regular-interval timetable is suboptimal with respect to the number of vehicles required; in the following we consider regular-interval timetables unless stated otherwise.

Another classification aspect can be whether a timetabling model has a level of detail of the infrastructure including all switches, signals and block sections, a *microscopic* model, or only models the network as a set of stations and important junctions and the connecting lines, a *macroscopic* model. Using a microscopic model for a large network is impractical, and while a microscopic model is necessary to construct a conflict-free timetable for a large station or a complex junction. The approaches are therefore split into local microscopic models and macroscopic models capable of handling large networks with a lower level of detail (Schlechte et al., 2011). A further distinction can be made between models that consider different routing options and models that only focus on the timing and ordering of fixed-route trains. Note that a routing model is not necessarily microscopic.

In the following we first review nominal timetable generation methods, where the goal is to find a feasible timetable regardless of its reliability; then methods to evaluate the reliability of the timetable are considered. The third part, robust timetable generation can be seen as the simultaneous combination of these two former problems.

Nominal timetable generation

One way timetabling trains differs from general scheduling problems in industrial engineering is that in many countries the train timetable is periodic, repeating usually every hour. Therefore the Periodic Event Scheduling Problem (PESP), defined by Serafini & Ukovich (1989) building on the *Program Evaluation and Review Technique* (PERT) (Malcolm et al., 1959), was applied to railways on a macroscopic level by numerous approaches. Schrijver & Steenbeek (1993) developed the algorithm named CADANS for solving the network timetable problem. Nachtigall (1996) models the train timetable with a periodic event network where nodes are train arrivals and departures and arcs are dwell, run or change activities. Based on the CADANS network timetable design model, the constraint programming system DONS was introduced (Schrijver, 1998; Kroon et al., 2009). DONS can find feasible solutions to the railway network of

the Netherlands if it exists under the given initial parameters, or points to the critical constraints if a feasible solution does not exist.

Goverde (1999) reformulated PESP with buffer times as decision variables and exploiting the graph structure of the network to reduce the number of variables. Lindner (2000) extends the PESP formulation to include operational costs in the objective. Liebchen (2008) optimized a homogeneous, high-frequency metro network, implemented in practice. Kroon & Peeters (2003) extend the infrastructure constraints to variable running times, by inserting further dummy nodes into the graph where necessary. Another problem type successfully applied to timetable planning is the Quadratic Semi-Assignment Problem (QSAP), e.g. by Schuele et al. (2009), who found that for the particular case of exploring the effect of small changes in the departure times on transfer time, QSAP had better results than PESP. A new MILP model scheduling extra freight trains into an existing passenger train timetable is presented by Cacchiani et al. (2010), with a Lagrangian heuristic that enables finding a feasible timetable in a few hours for a large instance. These models focus on timing with fixed train routing. On the other hand, Caprara et al. (2011) includes flexible routing solving the platform assignment problem with only small deviations allowed related to an existing macroscopic timetable. While applied for delay management and not initial timetable generation, the timetabling model used by Dollevoet et al. (2012) also takes into account passenger flows.

Besides macroscopic modelling on the network level, microscopic timetabling is necessary for complex station areas and junctions where routes are explicitly taken into account to check the feasibility of the network timetable and further optimize it. Caimi et al. (2008) focus on the microscopic nominal timetable problem for large station areas, as they propose the division of the train network into larger station areas with complex structure and saturated traffic, *condensation zones*, and connecting long lines of simple structure, *compensation zones*. They argue that because of the traffic saturation, no time reserve should be included in the condensation zones, therefore their modelling becomes simpler as train speeds are fixed at their maximum, including the boundaries of the zones. Furthermore, the possible values of passing times are discretized to further reduce the search space. Then a feasible schedule is calculated for the condensation zones using heuristics based on a conflict graph representation of train paths.

Besides stations, complex junctions also need microscopic verification of timetable feasibility. Lusby et al. (2011) route trains through a same-level junction by modelling the trains and their possible paths as a set-packing problem. As for a given train and route, multiple possible speed profiles through the junction are considered, this approach in fact does not necessarily output *conflict-free* train paths: when following this model the trains might have to reduce their speeds below the maximum possible. This in practice may be applicable if an appropriate driver assistance system is available providing an appropriate speed advice.

Timetable evaluation

One metric that is used to approximate the concept of macroscopic network-level *stability* of a periodic timetable is the minimum cycle time of the timetable (Braker, 1993; Heidergott et al., 2005; Goverde, 2007, 2008), this is because modelling the periodic railway system as a periodic even-activity network, the difference between the minimum cycle time and the timetable period equals to the mean delay reduction per hour viewed over many hours, depending on the timetable. The relationship between the nominal and the minimum cycle time describes the capacity utilization of the timetable: the timetable is stable exactly if the minimum cycle time T is less than the nominal cycle time T_0, i.e. $T < T_0$, and the larger $T_0 - T$ is, the more time reserve there is available. This ratio T/T_0 is defined in Goverde (2007) as *network throughput*. Therefore there is a strong relationship between the capacity of the physical network and the stability of the timetable: infrastructure capacity determines the pace at which the timetable can be executed, therefore the minimum cycle time T, and the stability of the timetable can be described as the relationship between T and T_0. The max-plus algebra technique (Heidergott et al., 2005) has been successfully applied for timetable evaluation using the cycle time by Braker (1993) and Goverde (2007).

On the microscopic level, Delorme et al. (2009) evaluates the stability of a timetable for a given station using graph theory by using the amount of buffer time available between each consecutive train pair. Then the delay propagation given a set of initial primary delays is quickly calculated. The limitation of this model is the assumption that the delay one train indirectly propagates to a third train equals to the sum of the buffer times between two successive train pairs, while in practice this calculation might be too conservative because of the available time supplements.

Robust timetable generation

Robust timetable generation consists of timetable design methods where the reliability of the timetable is taken into account already during the design process.

Looking at macroscopic network-level models, Kroon et al. (2008) expanded the PESP approach to a stochastic case. Fischetti et al. (2009) compare four different methods, including the *light robustness* method proposed in Fischetti & Monaci (2009), to improve the robustness of a timetable where robustness is represented by the average cumulative delay for a set of minor disruption scenarios, assuming no train cancellation or reordering. All four methods assume an existing nominal timetable and the train orders are fixed during the robust optimization. Liebchen et al. (2010) extend the light robustness approach concentrating on connection management during the distribution of timetable slack. Cacchiani et al. (2011) solve a robust timetable optimization problem using a Lagrangian heuristic. The robustness is represented by the amount of dwell time supplement for each train, and flexible running times are modelled by a set of available train paths for each train run.

A common characteristic of the above approaches is that they treat the period of the timetable as a fixed parameter and they minimize a certain sum of travel time, waiting

time or transfer waiting time. For example, Nachtigall & Voget (1996), (Nachtigall & Voget, 1997), Wong et al. (2008), Schuele et al. (2009), and Liebchen et al. (2010) minimize transfer waiting times, Nachtigall (1996) minimize the total travel time, and Kroon & Peeters (2003) define a general objective function dependent on any of the variables, but still using fixed period length. Lindner & Zimmermann (2005), however, optimizes for cost. One known work where the objective function of the timetable optimization is the cycle time is Heydar et al. (2013), based on Bergmann (1975). The authors assess the capacity of a single track, unidirectional railway line with passing loops by calculating the minimum period for given number of local and express trains using Mixed Integer Linear Programming (MILP). The dwell times at intermediate stations are variable and therefore the passing locations for trains are flexible. The main limitation of this model is that the train speeds are constant during the optimization and equal for all train classes.

Caimi et al. (2010) improve the mathematical formulation for the microscopic train scheduling problem. Flexibility in train departure and arrival times is modeled by sets of train paths for each train, and the robustness of the calculated timetable is estimated by the buffer times available between consecutive train runs. Dewilde et al. (2013) generate a local microscopic timetable, iteratively solving station area routing, platform assignments, and train timetabling including time shifts with fixed train orders and swapping train pairs, while maximizing the buffer time between concecutive trains to take into account robustness. They also define a second robustness indicator based on the deviation of realized passenger travel time from the planned value. However, this indicator is not used during the optimization models, but only in their evaluation using simulation.

2.4.4 Timetable synchronization and connection management

Early works in synchronizing timetables include Domschke (1989) proposing using branch-and-bound algorithm to solve a relaxation of the quadratic semi-assignment problem (QSAP), and Bookbinder & Desilets (1992) proposing a variant of the related quadratic assignment problem (QAP) to optimize the time-shift of otherwise fixed bus routes to minimize transfer waiting times. See also the timetabling approaches focusing on minimizing transfer waiting time in the previous Section 2.4.3, for example Schuele et al. (2009).

Connection management has already received significant attention in literature, with different modelling assumptions and target functions. Knoppers & Muller (1995) examine when timed transfers are beneficial with respect to frequency and reliability conditions. Goverde (1998) focuses on train networks with predefined connections and makes a distinction between waiting time in vehicles and on platforms. Heidergott & De Vries (2001) also describe interconnected train networks and introduce some heuristics to reduce the solution space created by multiple connections. Ginkel & Schöbel (2007) use a hybrid indicator of passenger volumes for a missed connection

and vehicle delay for a maintained connection. Andersson et al. (1998) describes the crew pairing problem and the best practice in the European airline industry and how it relates to the tightness of the connections. For a further review of connection management literature, refer to Ginkel & Schöbel (2007).

Several pieces of previous work also focus on comparing different transfer stations by investigating their physical layouts or the available transport supply. Brändli & Berg (1979) analyse the effect of a new pedestrian underpass on pedestrian behavior at Zurich main station, with emphasis on commuters. Söngen (1979) introduces a connection matrix of types of pedestrians such as transferring, boarding, alighting passengers, and people walking through the station, taking into account the multiple levels of a station such as underground, ground level and elevated platforms. Guo & Wilson (2011) model the transfer penalty of London Underground stations taking into account route choice through the network and detailed variables describing conditions like ramps, stairs and escalators. Waterson et al. (2010) compare railway stations on a given line focusing on travel time taking into account rail services of different stopping patterns. Kirchoff (1992) describes the requirements for large urban transfer stations, including the capacities of stairs and escalators and the sizing of railway platforms. Weidmann (1993) describes the modelling of pedestrian traffic which makes walking time estimation possible based on the station layouts. Kruse (1995) classifies transfer nodes according to the geometric design of the interconnected stations, public transport modes and lines. He identifies timetable optimization and minimization of transfer walking distance as the most promising measure for reducing the transfer resistance. Debrezion et al. (2009) include the frequency and the travel time of local public transport, among others, in their service quality index.

2.5 Conclusions

This chapter described the problem of railway timetabling by dividing the research challenges into timetabling models for highly occupied railway networks on the one hand, and timetable synchronization approaches for typical low frequency, low demand feeder and distributor lines on the other hand.

We identified three typical properties of dense railway networks that make related timetabling problems unique: a very high capacity utilization, heterogeneous service patterns, and pressure to focus on reliability. These attributes make this timetabling problem different from related problems such as rapid transit (metro), low frequency bus network or airline scheduling. We also covered connecting public transport lines, that form part of the majority of passenger journeys by train and therefore are seen as an integral part of the timetabling problem. We reviewed available literature addressing these topics and identified the following gaps that are then answered through the contributions of this thesis.

While there are numerous railway timetable stability evaluation methods in the literature, as well as railway timetable generation models that take some notion robustness into account, there is no known railway timetable optimization model yet that takes a measure of network-level robustness as the objective function during optimization, using flexible train orders, run and dwell times. Chapter 3 presents such a railway timetable optimization model.

Timetable synchronization models usually assume a single fixed minimum transfer time and do not take the detailed layout of the transfer stations into account. In general, many types of public transport models, including many travel advice systems, have a simplified view of a transfer node, that can now be enriched using recently opened up public data. Chapter 4 therefore presents the detailed modelling of walking distance, time and resistance at transfer stations and three application examples.

Finally, Chapter 5 presents a delay management formulation that explicitly takes into account that only a subset of the network is controllable by a dispatcher, and also using a max-plus algebra based reformulation reduces the optimization problem size. The results of this work also provide a decision support tool for traffic control where a dispatcher can still make the important connection management decisions, but prior to that the model identifies the critical subset of connections that requires attention.

Chapter 3

Optimizing periodic railway timetables for stability[1]

This chapter introduces a railway timetable optimization model that focuses on maximizing the stability of the timetable already during the optimization. In Section 2.4 we reviewed the state-of-the-art of estimating railway network capacity, generating a stable timetable and assessing its stability. In Section 2.4.3 in particular, we presented robust timetable generation methods, which do not only focus on the feasibility but also the stability of the timetable, either by *ex post* stability analysis or maximizing buffer times between trains during the optimization. Hardly any paper takes into account, however, the *network-level stability* during the optimization process. We therefore present a model to handle this shortcoming.

This chapter is organized as follows. Section 3.1 details the contributions of the presented model and the related previous works. In Section 3.2 we define an optimization model that generates a timetable with maximum stability using flexible train orders and running and dwell times. We here also give an interpretation of different results returned by an optimization solver. Section 3.3 describes techniques to improve the running time of the solution process, including treating the multiplication of variables, dimension reduction and dynamic frequencies. Section 3.4 reports computational results on real-life networks. Section 3.5 extends the solution approach with an iterative optimization method and includes further computational results proving the applicability of the iterative method on large networks. Finally, Section 3.6 concludes the chapter.

[1]This chapter is an edited and extended version of the conference proceeding: Sparing, D., & Goverde, R.M.P. (2013). An Optimization Model for Periodic Timetable Generation with Dynamic Frequencies. In *Proceedings of the 16th International IEEE Annual Conference on Intelligent Transportation Systems (ITSC 2013)*. The Hague, The Netherlands.

3.1 Introduction

In Section 2.2 we identified the following three challenges of railway timetabling: high capacity utilization, heterogeneous services, and focus on reliability. In the following we introduce the contributions of our model reflecting to these challenges, as well as the relevant previous work.

3.1.1 Contributions of the model

The macroscopic periodic timetable optimization model described in this chapter focuses on high-frequency, high capacity-utilization networks. It takes the desired stop patterns and frequencies of train lines, as well as the structure of the railway network and minimum process times, and generates a stable timetable optimized for stability, if it exists. The running and dwell times of trains are flexible up to predefined boundaries, and the train orders are fully flexible. The overtake locations are also flexible, given a priori, within a predefined list of stations with enough capacity for overtakes. While the model is designed for periodic timetables, an aperiodic timetable can also be calculated using a single period.

The model takes into account limited railway infrastructure by respecting predefined minimum headways between trains at stations and junctions. The capacity limits of small stations are modelled by forbidding overtakes. Large stations, however, are assumed to have enough capacity. Train orders on the open track are always preserved despite flexible running times. We do not consider re-routing and assume that in case of multiple tracks in a direction, the train lines are pre-assigned to one of the tracks.

The optimization model presented solves the timetable feasibility problem according to the given line structure and process duration bounds. In case a feasible timetable exists, it also generates stable timetables, and returns one optimized for stability. Even in case of infeasibility, a timetable with a cycle time larger than the nominal cycle time is generated, which is useful for finding bottlenecks.

The line patterns and activity duration bounds can be represented by a graph called a periodic event network (PEN) (Nachtigall, 1996; Großmann et al., 2012). The limitations of railway infrastructure capacity at stations or junctions can be modeled by infrastructure arcs in the PEN. On the other hand, the infrastructure limitations not occurring at a station or junction, but *between* two such points can be modelled implicitly by the duration bounds, or where necessary by using additional nodes.

Based on the PEN, an optimization model can be formulated with the cycle time as the objective function to be minimized, as the minimum cycle time is a useful descriptor of periodic timetable stability (Goverde, 2007). This model can be reformulated to become a Mixed-Integer Linear Programming (MILP) instance and it can therefore be efficiently solved by available MILP solvers. Note that minimizing the cycle time

does not mean that the train running and dwell times are also minimized, therefore a secondary objective can be added to ensure minimal running and dwell times for a given cycle time. Once the MILP formulation is solved, the results can be interpreted as follows. A feasible set of the optimization model with the cycle time not larger than the nominal cycle time corresponds to a stable timetable, while an optimal solution of the optimization model corresponds to a stable timetable with minimal cycle time.

While the MILP model can be directly solved for small instances, for larger networks further preprocessing is useful to limit the number of variables and constraints. A periodic timetable often has further symmetries than just the main period: for example, in an hourly fixed-interval timetable, there might be many, if not most, trains, following a 30-minute or even 15-minute regular pattern. These symmetries can heavily be exploited to only optimize for a core period. In case the PEN has multiple components, these can be calculated separately and appended after optimization.

Predefined lower and upper bounds for the minimum cycle time can substantially speed up the solving time. In this case, if the optimal solution is at a boundary or the model is infeasible, the lower bound has to be decreased or the upper bound increased, respectively, and the model recalculated. Furthermore, pre-calculating the possible range of number of period crossings for each train line reduces solution time. Note that this method is highly sensitive of the cycle time bounds set. An iterative execution of the optimization model with re-setting the cycle time bounds and period crossing range after each run is a very effective method to extend the reach of the model to large networks.

The proposed macroscopic model can be connected to a microscopic model or microscopic simulation in an intuitive way: we use infrastructure headway times as inputs, these have to be generated by a microscopic model or microscopic simulation. In turn, the results of the model can be verified by microscopic means, and the model input parameters (e.g. headways) can be corrected if necessary. The model can also be seen as a train order and overtake location optimizer, that can be fed in a second, macroscopic model that can handle even larger networks, with fixed (or initially defined) train orders and overtake locations.

3.1.2 Previous work

Section 2.2 presents a relevant literature review. In particular, we heavily use concepts introduced by the timetable generation methods based on the Periodic Event Scheduling Problem (PESP) (Serafini & Ukovich, 1989), described in Section 2.4.3.

In Section 2.2.1 we saw that a useful concept capturing the capacity utilization of a given timetable on a given infrastructure is the *infrastructure utilization* (International Union of Railways (UIC), 2013), the ratio of the time period needed to execute a schedule and the nominal time period it is scheduled in. A closely related concept applied to periodic timetables is the *minimum cycle time* as introduced by Goverde

(2007). In a periodic timetable, the planned cycle time needs to exceed the maximum cycle mean over all circuits in the periodic event-activity network. In other words, for a given planned cycle time, a timetable is possible exactly if its minimum cycle time is not larger than the planned cycle time. The ratio of the minimum cycle time and the planned cycle time can be seen as a generalisation of infrastructure utilization for periodic timetables to networks.

The minimum cycle time of these macroscopic models is therefore conceptually similar to the microscopic compression method. The macroscopic models use minimal departure and arrival time headways as input parameters, and these values can be obtained using the microscopic compression method. The macroscopic models, on the other hand, can include other time constrains between event pairs besides the limited infrastructure capacity, such as passenger connections, crew and rolling stock-related constraints, and any constraints linking multiple corridors together.

3.2 A railway timetable optimization model focusing on stability

3.2.1 Initial assumptions and definitions

Without loss of generality we further assume that the planned cycle time is 60 minutes: the approach can be easily generalized for other periods such as 30 or 120 minutes. While the following model can also be extended to single track lines, for simplicity we assume that all railway tracks are one directional. This limitation still allows the modelling of large high capacity utilization railway networks of 2-4 track lines, such as the central part of the Dutch railway network.

We define *stability* as the degree of capability of the periodic timetable to return to its schedule from disturbance causing delays, after this disturbance is removed: a timetable that can return to schedule faster from the same delay disturbance is considered more stable. The *minimal cycle time* of a set of departure and arrival event orders defined by a timetable is the minimal period length in which the execution of these events is feasible using up running time supplements and buffer times if necessary, but without re-ordering events dependent on each other or violating any minimal time duration constraints. We define this more concretely later in Section 3.2.3.

3.2.2 The periodic event-activity network

Following the notation in Schöbel (2006) for event-activity networks and applying it for the periodic case just as in case of a PEN (Nachtigall, 1996), we model a railway

system as the *periodic event-activity network* $\mathcal{N} = (\mathcal{E}, \mathcal{A}, T)$, where T is the common cycle time for all events. An event $i \in \mathcal{E}$ is a tuple

$$i = (Line_i, Station_i, EventType_i), \tag{3.1}$$

where *Line* is the train line identifier, unique for each one directional train run per one cycle time period: i.e. in case of the cycle time of one hour, if a bidirectional train service runs twice per hour, that corresponds to four lines. While in general, the cycle time can be different from one hour, all events use the same cycle time. *Station* is a train station, junction or other timetable point; with a binary flag assigned to each station describing whether overtake is possible at the given station. *EventType* can take values from the set $\{dep, arr, thr\}$, representing departure, arrival, and through events, respectively. Furthermore, for each station, a binary variable is set whether overtakings between two trains are possible or not. Additional information on the station locations and line types can be given for visualization purposes.

We also define the subsets $\mathcal{E} = \mathcal{E}_{\text{dep}} \cup \mathcal{E}_{\text{arr}} \cup \mathcal{E}_{\text{thr}}$ corresponding to each $EventType$, to simplify notation.

An activity $a_{ij} \in \mathcal{A}$ is a tuple

$$a_{ij} = (i, j, ActivityType_{ij}, L_{ij}, U_{ij}), \tag{3.2}$$

where $\{i, j\} \subset \mathcal{E}$ are respectively the start and end events,

$$ActivityType_{ij} \in \{run, dwell, infra\}$$

where $infra$ stands for for minimum infrastructure headways, and the allowed range of the activity duration is

$$(L_{ij}, U_{ij}).$$

Again, we define the subsets $\mathcal{A} = \mathcal{A}_{\text{dwell}} \cup \mathcal{A}_{\text{run}} \cup \mathcal{A}_{\text{infra}}$ corresponding to the $ActivityType$ values to simplify notation. A run activity connects a departure or through event to an arrival or through event, a dwell activity always connects an arrival event to a departure event, while an infrastructure activity can connect any two activities.

Then, for a given cycle time t and ensuring that for all processes a_{ij}, $0 \leq L_{ij} \leq U_{ij} < t$ (see below), a timetable is an assignment $i \to \tau_i$ of periodic scheduled times to each event $i \in \mathcal{E}$ so that for each $a_{ij} \in \mathcal{A}$,

$$L_{ij} \leq (\tau_j - \tau_i \mod t) \leq U_{ij}. \tag{3.3}$$

Note the ordering of events in an infrastructure activity does not symbolize an ordering of events, as an infrastructure activity with lower and upper bounds of l, u, respectively is equivalent to an edge between the same two events but directed in the inverse direction and with lower and upper bound $T - u$, $T - l$ respectively. In other words, in the graph representation of the periodic event-activity network, an infrastructure edge can be traversed in both forward and backward direction.

Modelling a single train line

A single train line run can be modeled in an intuitive way as a time-ordered directed chain of departure, arrival, and optionally through events, connected by run and dwell activities. The departure and arrival events are unambiguously defined by the stations where the train stops at for boarding or alighting passengers, while the through events need to be defined at any intermediate station, junction, or other timetable point where infrastructure activities need to be defined, as described in the next section.

The run and dwell activity durations are defined as follows. The minimum run and dwell times can be set based on predefined norms, measurements or simulations. For example, minimum dwell times might be defined by a train operator for each train type and station type; while minimum running times can be a lower percentile value of a series of running time measurements, optionally with an extra proportional slack time added as required.

The maximum run and dwell times, on the other hand, can be much more freely set: after all, a large run or dwell duration in this model can be directly translated to a large running or dwell time on the real train network. However, there might be business norms in effect, that require that the train run meets expectations by maintaining a limited deviation from its minimum possible running and dwell times. These norms or expectations can then either directly be translated into the maximum durations of running and dwell times, or, as we will see later in this chapter, defined as a higher level constraint for the whole train line, allowing for larger local deviations and therefore more flexibility.

Modelling limited infrastructure capacity

Following Odijk (1996), we also introduce infrastructure activities representing minimum headway constraints between consecutive trains because of shared infrastructure resources. On the macroscopic level, most headway constraints can be modelled as a minimum time separation between two events of two trains, and this minimum duration can be measured or simulated with microscopic tools. This type of infrastructure constraint can conveniently be modeled in our periodic event-activity network with infrastructure activities connecting train events. In case a minimum time separation is necessary at a location where previously no train events were defined, it is necessary to define through events for all affected trains by splitting the run activities into two, and consequently the infrastructure activities can be defined between all these events.

We define a *timetable point* as a station, a junction or other location where trains can have scheduled times. In case there are more than two trains using the same timetable point sharing some infrastructure, then in general a full graph between all these events needs to be defined. It is possible, however, to avoid pairs of activities between each two events and just define one activity, according to Goverde (1999), as follows. Let the activities be i and j, and the directed minimum headway times L_{ij} and L_{ji}, for which $L_{ij}+L_{ji} \leq T$ holds. Then a single infrastructure activity a_{ij} can be defined with the bounds $(L_{ij}, T-L_{ji})$ periodic with cycle time T according to (3.3).

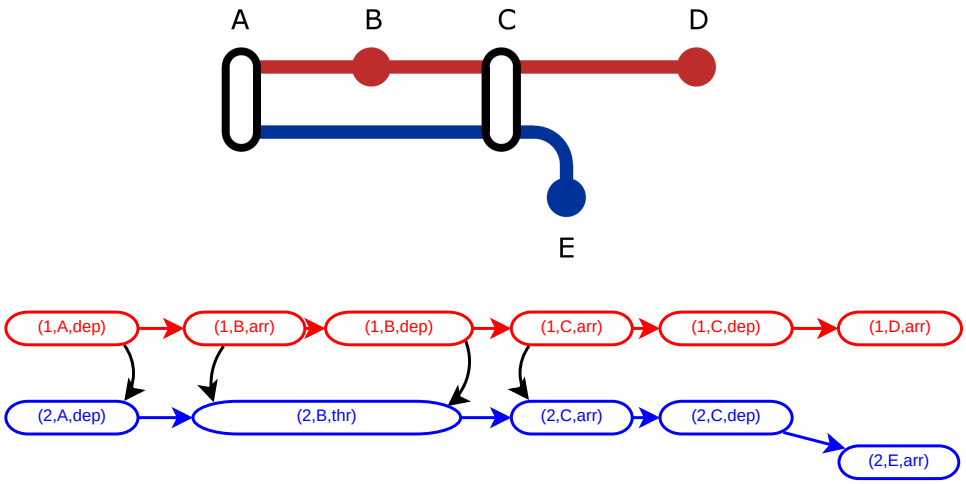


Figure 3.1: Schematic line graph and periodic event-activity network of two train lines (Line 1 in red and line 2 in blue) with infrastructure activities (in black)

See Figure 3.1 for an example periodic event-activity network of two train lines, where "(1, A, dep)" is the *dep* (departure) event of Line 1 at station *A*, etc. Train line 1 is a local train stopping at stations A, B, C, and D. Train line 2 is an express train line stopping at station A, running through station B without stopping, and stopping at stations C and E. Each train line is represented by departure, arrival and through nodes, and run and dwell activities in between. Infrastructure constraints at the shared resources, leaving station A, entering and leaving station B, and entering station C, are represented by infrastructure activity arcs.

In case of flexible run and dwell times, however, the model above is not sufficient anymore to ensure the separation of trains using shared infrastructure, as pointed out in Kroon & Peeters (2003). In particular, the following two scenarios are valid according to the above model, but impossible in reality:

1. On the open track, one train overtakes another one, while still observing the relevant departure-departure and arrival-arrival headways.

2. At a station without overtake facility, one train overtakes another one, while still observing the relevant arrival-arrival and departure-departure headways.

Kroon & Peeters (2003) give the solution to the first problem by introducing further dummy nodes and infrastructure activities where necessary. We use the same approach, and extend it in a similar fashion for stations where train overtake is not possible.

Let us consider such an "illegal" station overtake as the following: at station A, train line i is overtaken by train line j running in the same direction, while both trains stop at the station. In another case, again at station A, train line i is overtaken by train line

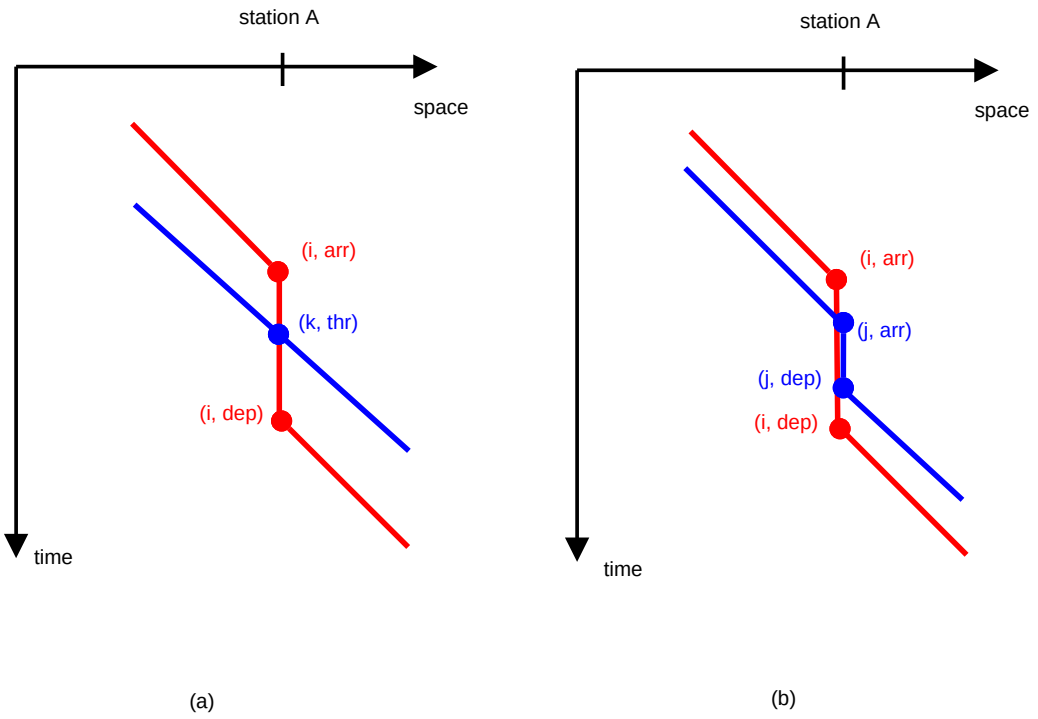


Figure 3.2: Exemplary time-space diagram for an overtake at a station where the slower train is stopping, with (a) overtaking train not stopping, (b) overtaking train stopping

k running in the same direction, but in this case while train i stops at station A, train k does not. The following approach can be iteratively extended for further overtaking trains that may either stop of not. In case of train k, for events to take place in the violated order of (i,arr), (k,thr), (i,dep), where we omitted the station index for clarity, the following have to be true (see Figure 3.2(a)):

$$L_{(i,arr),(k,thr)} + L_{(k,thr),(i,dep)} \leq U_{(i,arr),(i,dep)}. \tag{3.4}$$

Similarly, for train j, the following needs to be true for an invalid overtake in the form of the event sequence (i,arr), (j,arr), (j,dep), (i,dep) (see Figure 3.2(b)):

$$L_{(i,arr),(j,arr)} + L_{(j,arr),(j,dep)} + L_{(j,dep),(i,dep)} \leq U_{(i,arr),(i,dep)}. \tag{3.5}$$

Therefore the following preprocessing is necessary to ensure that no invalid overtakes take place: for each station where overtake is not possible, all trains stopping need to be checked for the above inequalities if any other train can overtake them and in case of violation, dummy node(s) need to be inserted in the dwell process of the overtaken train to ensure that the overtaking is not permitted by the constraints.

The detailed method to insert these dummy dwell nodes is the following: for every such dwell activity of a potentially slower train where an illegal overtake would be possible according to one of the above two inequalities, among all potentially overtaking trains k we take one where $L_{(i,arr),(k,thr)} + L_{(k,thr),(i,dep)}$ or $L_{(i,arr),(k,arr)} + L_{(k,arr),(k,dep)} +$

$L_{(k,dep),(i,dep)}$ is minimal, and let this minimal sum of lower bounds be L_{fast} for the faster train. We also denote $U_{slow} = U_{(i,arr),(i,dep)}$ for brevity. Then this dwell process needs to be split by $n = \lfloor U_{slow}/L_{fast} \rfloor$ additional nodes into $n+1$ sections.

We now introduce *dummy* as a new possible value of $EventType$ and insert the dummy event nodes $(Line_i, dummy_1)$, ..., $(Line_i, dummy_n)$ (omitting the station ID in the notation for clarity as before). To simplify notation, let us rename (i, arr) to $(i, dummy_0)$ and (i, dep) to $(i, dummy_{n+1})$ so that we can refer to the dwell activities with dummy nodes as $((i, dummy_x), (i, dummy_{x+1}))$ regardless if they include the arrival or departure node or not. Then replace the activity

$$\left\{(i, arr), (i, dep), dwell, L_{(i,arr),(i,dep)}, U_{(i,arr),(i,dep)}\right\}$$

(where L and U stand for the initial bounds of this activity) with the ordered chain of activities

$$\left\{(i, dummy_x), (i, dummy_{x+1}), dwell, L_{(i,arr),(i,dep)}/(n+1), U_{(i,arr),(i,dep)}/(n+1)\right\}, \tag{3.6}$$

for $0 \leq x \leq n$.

Finally, we create n copies of any infrastructure constraint involving events $(i, arr) = (i, dummy_0)$ and $(i, dep) = (i, dummy_{n+1})$ and in the replicated versions we replace the applicable original event with each of the dummy nodes $(i, dummy_x)$ for $1 \leq x \leq n$ (excluding, thus, the first and last dummy node, i.e. the original arrival and departure nodes, $(i, arr) =\equiv i, dummy_0)$ and $(i, dep) \equiv (i, dummy_{n+1})$). An example is shown in Figure 3.3(b) for a faster stopping and a faster not stopping train and $n = 1$, the situation is analogous for more faster trains, and multiple dummy nodes. If this results in multiple infrastructure activities between the same two events, the multiple arcs can be merged into a simple arc using the maximum of the required infrastructure headways across the multiple arcs, such as in the case of the parallel dashed and dotted arcs in Figure 3.3(b) between nodes $(i, A, dummy1)$, and (k, A, thr).

Having inserted the dummy node(s), we show why an illegal overtake is now disabled for the case of $n = 1$ and the overtaking train not stopping, the same can be shown similarly for the other cases. An overtake would mean an order of events

$$(i, dummy_x), (k, thr), (i, dummy_{x+1}),$$

for some x, as in Figure 3.4. (Recall that as train orders are flexible and infrastructure headways are modelled by a single one-directional arc with lower and upper bounds between two points, the direction of the infrastructure arc does not symbolize the ordering of trains within the hour. Therefore a path of consecutive events within the hour may traverse an infrastructure arc backwards.) As the constraints are satisfied, for the sequence of events $(i, dummy_x), (k, thr), (i, dummy_{x+1})$, the upper bound of the duration of activity $((i, dummy_x), (i, dummy_{x+1}))$ has to be at least the sum of the lower bounds of the durations of activities $((i, dummy_x), (k, thr))$ and $((k, thr), (i, dummy_{x+1}))$:

$$L_{(i,dummy_x),(k,thr)} + L_{(k,thr),(i,dummy_{x+1})} \leq U_{(i,dummy_x),(i,dummy_{x+1})}.$$

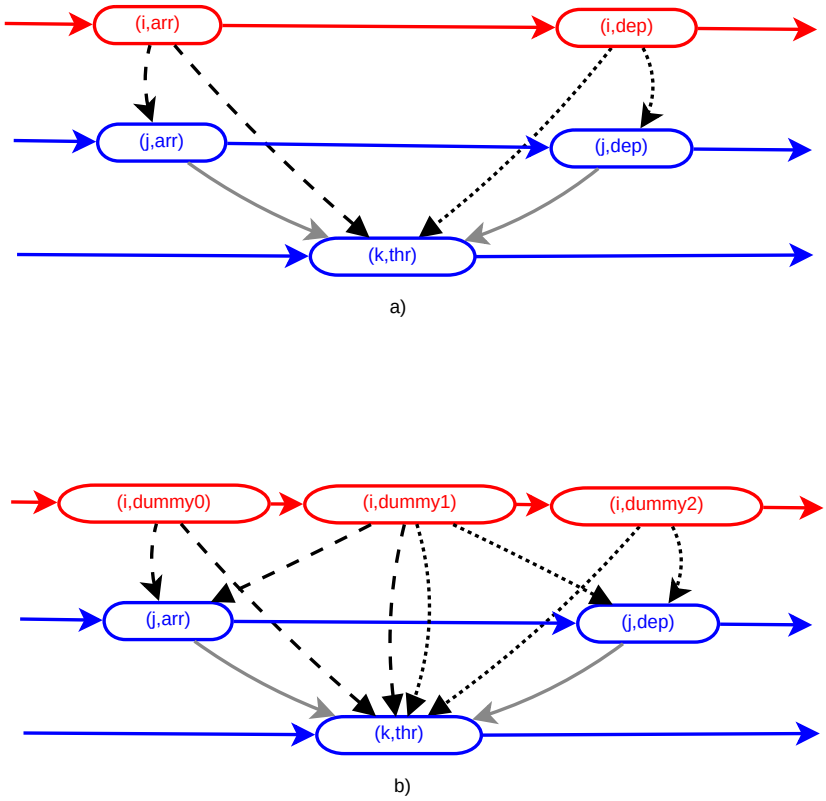


Figure 3.3: Example events and activities before (a) and after (b) inserting a single dummy node for a station dwell with one faster train, not stopping (dashed black and dotted black arrows are the infrastructure events related to the original arrival and departure nodes, respectively; gray arrows are infrastructure arcs not related to the dummy node extension)

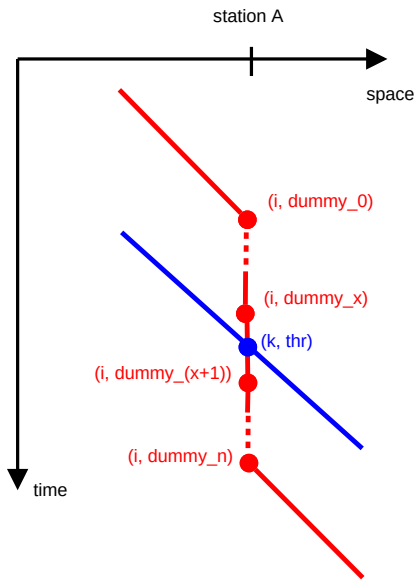


Figure 3.4: Time-space diagram for an overtake without stopping including a station dummy node

On the one hand, $U_{(i,dummy_x),(i,dummy_{x+1})}$ was defined in (3.6) a $U_{(i,arr),(i,dep)}/(n+1)$. On the other hand, as activities $((i,dummy_x),(k,thr))$ and $((k,thr),(i,dummy_{x+1}))$ are copies of $((i,arr),(k,thr))$ and $((k,thr),(i,dep))$ respectively, their lower bounds are equal to $L_{(i,arr),(k,thr)}$ and $L_{(k,thr),(i,dep)}$ respectively. Therefore it follows that

$$L_{(i,arr),(k,thr)} + L_{(k,thr),(i,dep)} \leq U_{(i,arr),(i,dep)}/(n+1).$$

Rearranging for n, using the definition of n and the earlier notations of U_{slow} and L_{fast},

$$U_{slow}/L_{fast} - 1 \geq n = \lfloor U_{slow}/L_{fast} \rfloor$$

which is not possible, therefore we proved that the illegal overtake now is disabled by the constraints including the dummy node.

To give a numerical example, the left side of Figure 3.5 shows an illegal overtake. This is in line with the original constraints, as the maximum dwell time is 10, while both minimum infrastructure headways are 3, therefore a timetable of 5, 10, and 15 for the arrival, pass, and departure respectively satisfies the activity bounds. In case that this station does not allow for overtakes, $\lfloor 10/(3+3) \rfloor = 1$ dummy node is added to split the dwell process into two, and the two infrastructure arcs are duplicated to connect to the dummy node as well. After this, an overtake is not possible anymore, as follows. Let us assume without loss of generality that it happens between the dummy and the departure node (the argument is identical for between the arrival and dummy node). This on the one hand would mean that the dummy node and the overtake point is separated by at least 3 minutes and the overtake and the departure node are separated by at least 3 minutes, therefore the dummy and departure nodes are separated by at

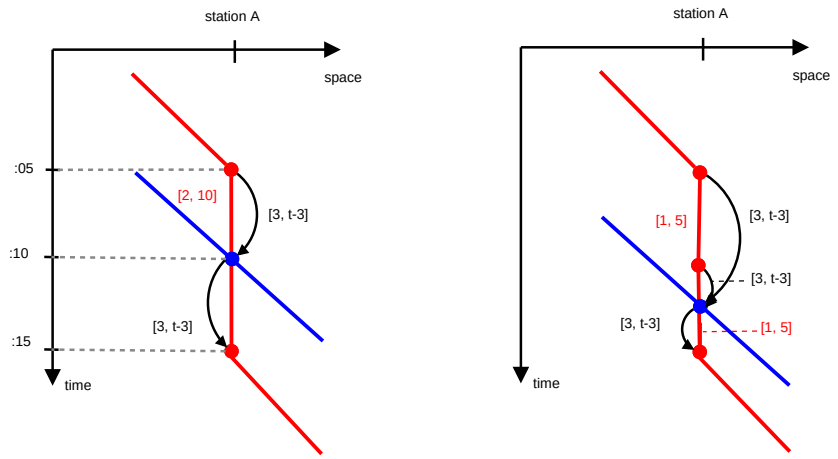


Figure 3.5: Numerical example for illegal overtake (left) and added dummy node (right)

least 6 minutes. On the other hand, the maximum separation between the dummy and the departure node is at most $10/(n+1) = 5$ minutes, leading to an infeasible solution.

The usefulness of the above approach lies in the fact that in practice, there are relatively few locations in the original graph where the assumptions above are violated and dummy nodes need to be inserted, and even within these cases the vast majority needs only one dummy node per case.

Modelling processes longer than the cycle time

Above we assumed that all activities have duration bound ranges smaller than the cycle time. It can happen though, that this is not immediately the case, e.g. in case of long running times. This would lead to a modelling issue as we will require that for all activities, $U_{ij} < T$ in order to be able to define constraints including the binary variables $z_{ij} \in \{0,1\}$, as we will see later. Therefore in case of processes initially longer than the cycle time, these need to be split into shorter sections with dummy events connecting them until this condition is satisfied. Concretely, if for an activity (i,j) it holds that $U_{ij} \geq T$, then let $n = \lfloor U_{ij}/T \rfloor + 1$ and then we split (i,j) to n segments by introducing dummy events $dummy_1$, ..., $dummy_{n-1}$ and replacing activity (i,j) with activities $(i, dummy_1)$, $(dummy_1, dummy_2)$, ..., $(dummy_{n-1}, j)$ where the lower and upper bounds of each are L_{ij}/n and U_{ij}/n, respectively. For the bounds L, U of these new activities it will all hold true that $L \leq U < T$.

Later the cycle time will become a variable: then the exact process is to define a lower bound L for the variable cycle time $t \geq L$ and ensure that all duration bounds are lower than this cycle time lower bound, i.e., $L_{ij} \leq U_{ij} < L, \forall (i,j) \in \mathcal{A}$. This can be achieved in an identical manner as above for the case of T.

3.2.3 The optimization model

Building on the periodic event-activity network as defined earlier, finding a stable timetable given cycle time T can be defined as the feasibility model

$$L_{ij} \leq (\tau_j - \tau_i \mod T) \leq U_{ij} \quad \forall (i,j) \in \mathcal{A}_{\text{run}} \cup \mathcal{A}_{\text{dwell}}, \tag{3.7}$$

$$L_{ij} \leq (\tau_j - \tau_i \mod T) \leq T - L_{ji} \quad \forall (i,j) \in \mathcal{A}_{\text{infra}}, \tag{3.8}$$

$$0 \leq \tau_i < T \quad \forall i \in \mathcal{E}, \tag{3.9}$$

where for the bounds it holds that $0 \leq L_{ij} \leq U_{ij} < T \ \forall (i,j) \in \mathcal{A}_{\text{run}} \cup \mathcal{A}_{\text{dwell}}$, and $0 \leq L_{ij} < T$ and $0 \leq L_{ji} < T \ \forall (i,j) \in \mathcal{A}_{\text{infra}}$. Also note that from the leftmost and rightmost sides of (3.8) it follows that $L_{ij} + L_{ji} \leq T, \forall (i,j) \in \mathcal{A}_{\text{infra}}$.

We saw earlier that it is possible to ensure that all activity duration bounds are within $[0,T)$, therefore the leftmost and rightmost sides of inequalities (3.7)-(3.8) are also in the range $[0,T)$ and so

$$(\tau_j - \tau_i \mod T)$$

can be rewritten unambiguously as

$$\tau_j - \tau_i + z_{ij} T$$

where $z_{ij} \in \{0,1\}$. Note that then $z_{ij} = 0$ if $\tau_j \geq \tau_i$, and $z_{ij} = 1$ if $\tau_j < \tau_i$.

Consider a solution to the above feasibility model, with $\tau_i \ \forall i \in \mathcal{E}$ being the event time of each event, and the binary $z_{ij} \ \forall (i,j) \in \mathcal{A}$ variables describing for each activity the order of events within the timetable period. We now state the earlier definition (see Section 3.2.1) of the minimal cycle time more concretely: the minimal cycle time of this timetable, more precisely of this set of event orders, is the minimal value of T for which a solution with the same values of $z_{ij} \ \forall (i,j) \in \mathcal{A}$ exists (but possibly with different values of $\tau_i \ \forall i \in \mathcal{E}$) .

Intuitively, the minimal cycle time is the following. Take a railway network in a periodic timetable, and modify the dispatching rules such that the trains do not need to wait for their scheduled departure and arrival events, however, they do respect all minimal process times defined between any event pairs, and therefore also the order or trains is not changed. Make the trains depart and arrive, within these rules, as soon as possible. In this system, any periodic event will occur with a period not longer than the minimal cycle time.

Recall that stability is the degree of capability of the periodic timetable to return to its schedule from disturbance causing delays, after this disturbance is removed: a timetable that can return to schedule faster from the same delay disturbance is considered more stable. Therefore the minimum cycle time of a periodic timetable is a valuable estimator of its stability (Braker, 1993; Heidergott et al., 2005; Goverde, 2007): the difference between the period and the minimal cycle time is the minimum mean

hourly delay reduction that is certainly possible given any initial delay and no further disturbance, with no train re-ordering or cancellation, regardless of which event or how many events are delayed.

We argued that for otherwise comparable timetables of identical train stopping patterns and frequencies, the one with lower minimum cycle time is the more stable, so we transform the above feasibility problem into the following optimization problem with variable cycle time t:

minimize

$$t$$

subject to

$$L_{ij} \leq \tau_j - \tau_i + z_{ij}t \leq U_{ij} \quad \forall (i,j) \in \mathcal{A}_{\text{run}} \cup \mathcal{A}_{\text{dwell}}, \tag{3.10}$$

$$L_{ij} \leq \tau_j - \tau_i + z_{ij}t \leq t - L_{ji} \quad \forall (i,j) \in \mathcal{A}_{\text{infra}}, \tag{3.11}$$

$$0 \leq \tau_i < t \quad \forall i \in \mathcal{E}, \tag{3.12}$$

$$L \leq t, \tag{3.13}$$

$$z_{ij} \in \{0,1\} \quad \forall (i,j) \in \mathcal{A} \tag{3.14}$$

where $L > 0$ is a lower limit on t, $0 \leq L_{ij} \leq U_{ij} < L \; \forall (i,j) \in \mathcal{A}_{\text{run}} \cup \mathcal{A}_{\text{dwell}}$, and $0 \leq L_{ij} < L$ and $0 \leq L_{ji} < L$ for all infrastructure activities (i,j). Also note that from the leftmost and rightmost sides of (3.11) it follows that $L_{ij} + L_{ji} \leq t, \forall (i,j) \in \mathcal{A}_{\text{infra}}$.

Note that as the variable cycle time t and the event times τ are real numbers, we do not assume that these values are in full minutes, and support a precision to 1 second as well. However, in the post-processing of the optimization results (see Section 3.2.5), if necessary it is possible to limit the final event times to full minutes.

In the following, modifications of the above model definition are explained in order to transform it into an MILP model. Because the cycle time is a variable in this model, constraints (3.10) and (3.11) contain the product $z_{ij}t$ of two variables, which violates the linearity conditions. Let U be a suitably large upper bound for the objective value t. A product of the binary variable z and a bounded continuous variable $0 \leq t \leq U$ can be reformulated as the following four linear constraints using the new variable $y = zt$ (Williams, 1990):

$$y \leq Uz, \tag{3.15}$$

$$y \leq t, \tag{3.16}$$

$$y \geq t - U(1-z), \tag{3.17}$$

$$y \geq 0. \tag{3.18}$$

Constraint (3.12) includes a strict inequality relating to cycle time t. This can be replaced by a non-strict inequality with $t - \delta$ where δ is a suitably small value, such as 1 second.

Consequently, the rewritten MILP formulation is as follows:

minimize

$$t$$

subject to

$$L_{ij} \leq \tau_j - \tau_i + y_{ij} \leq U_{ij} \quad \forall (i,j) \in \mathcal{A}_{\text{run}} \cup \mathcal{A}_{\text{dwell}}, \tag{3.19}$$

$$L_{ij} \leq \tau_j - \tau_i + y_{ij} \leq t - L_{ji} \quad \forall (i,j) \in \mathcal{A}_{\text{infra}}, \tag{3.20}$$

$$0 \leq \tau_i \leq t - \delta \quad \forall i \in \mathcal{E}, \tag{3.21}$$

$$L \leq t \leq U, \tag{3.22}$$

$$z_{ij} \in \{0,1\} \quad \forall (i,j) \in \mathcal{A}, \tag{3.23}$$

$$y_{ij} \leq U z_{ij} \quad \forall (i,j) \in \mathcal{A}, \tag{3.24}$$

$$y_{ij} \leq t \quad \forall (i,j) \in \mathcal{A}, \tag{3.25}$$

$$y_{ij} \geq t - U\left(1 - z_{ij}\right) \quad \forall (i,j) \in \mathcal{A}, \tag{3.26}$$

$$y_{ij} \geq 0 \quad \forall (i,j) \in \mathcal{A}, \tag{3.27}$$

that can be equivalently rewritten in the conventional form of MILP conditions:

minimize

$$t$$

subject to

$$\tau_i - \tau_j - y_{ij} \leq -L_{ij} \quad \forall (i,j) \in \mathcal{A}, \tag{3.28}$$

$$\tau_j - \tau_i + y_{ij} \leq U_{ij} \quad \forall (i,j) \in \mathcal{A}_{\text{run}} \cup \mathcal{A}_{\text{dwell}}, \tag{3.29}$$

$$\tau_j - \tau_i + y_{ij} - t \leq -L_{ji} \quad \forall (i,j) \in \mathcal{A}_{\text{infra}}, \tag{3.30}$$

$$\tau_i - t \leq -\delta \quad \forall i \in \mathcal{E}, \tag{3.31}$$

$$y_{ij} - U z_{ij} \leq 0 \quad \forall (i,j) \in \mathcal{A}, \tag{3.32}$$

$$y_{ij} - t \leq 0 \quad \forall (i,j) \in \mathcal{A}, \tag{3.33}$$

$$t + U z_{ij} - y_{ij} \leq U \quad \forall (i,j) \in \mathcal{A}, \tag{3.34}$$

$$L \leq t \leq U, \tag{3.35}$$

$$\tau_i \geq 0 \quad \forall i \in \mathcal{E}, \tag{3.36}$$

$$y_{ij} \geq 0 \quad \forall (i,j) \in \mathcal{A}, \tag{3.37}$$

$$z_{ij} \in \{0,1\} \quad \forall (i,j) \in \mathcal{A}. \tag{3.38}$$

Table 3.1: Interpretation of optimizer results

		Interpretation		
Solver status	Variable ranges	A timetable exists?	A timetable generated?	Optimal timetable generated?
Intermediate, no solution	$t_{LP} \leq T$	unknown	no	no
	$T < t_{LP}$	no	n/a	n/a
Infeasible		no	n/a	n/a
Intermediate solution	$t \leq T$	yes	yes	unknown
	$T < t \cap t_{LP} \leq T$	unknown	no	no
	$T < t_{LP}$	no	n/a	n/a
Optimal solution	$t \leq T$	yes	yes	yes
	$T < t$	no	n/a	n/a

T is the nominal cycle time, t is the best solution value for the minimum cycle time, and t_{LP} is its best LP lower bound for the given optimizer run.

3.2.4 Interpreting the optimization model results

Although the optimality of the generated timetable can only be guaranteed if the solver reached the optimal solution, information about the timetable can be inferred in several of the other cases as well. Table 3.1 shows how the different solver statuses and related variable value ranges can be interpreted, including t_{LP}, the objective value of the LP relaxation of the model, which is therefore a lower bound on t. On the one hand, if an intermediate feasible solution exists with $t \leq T$, we can already conclude that a stable timetable exists and it is generated by the model. On the other hand, irrespective of whether a feasible solution is known or not, if we know that $T < t_{LP}$ then we already know that a stable timetable is not possible. This of course is also the case if the MILP model is infeasible or if for the optimal solution, $T < t$.

While the core idea of the model is to keep the cycle time flexible, we have seen above that the model definition restricts the range of the cycle time to $[L,U]$. This in practice does not restrict the flexibility of the cycle time variable if the optimization is executed in an iterative fashion. In case that the calculation returns a cycle time equal to this predefined lower bound, the solver can be restarted with a decreased lower bound. Likewise, in case of infeasibility a higher upper bound can be used in a subsequent run. Furthermore, if the sole goal of an optimization run is to decide feasibility, both cycle time bounds can be set to the nominal cycle time and the model degenerates to a feasibility problem generating a stable timetable if possible, without taking stability into account.

3.2.5 The expanded timetable

As mentioned earlier, a feasible solution to the optimization problem with t corresponds to a stable periodic timetable with cycle time t. However, what in practice really is required, is a periodic timetable according to the nominal cycle time T. There are a few ways to transform and further optimize the obtained timetable to end up with a timetable of cycle time T.

The simplest method, that might not be sufficient in all practical purposes, is rescaling, resulting in a valid but potentially too slow timetable (if $t \ll T$). As long as $t < T$, a feasible set of the optimization problem can be transformed to have cycle time T in a straightforward way by substituting each event time τ_i with $\tau_i T/t$ and the expanded timetable would still obey all minimum running, dwell time and infrastructure headway constraints while keeping the same optimized minimum cycle time. To see an example of such a timetable expansion, see Figure 3.11 in Section 3.4. Note that the minimum cycle time is equal for the expanded and the compressed timetable because of the definition of the minimal cycle time: both timetables have the same event orders, and the minimal cycle time is the shortest period for which a feasible timetable exists for the given event orders. As the compressed timetable was optimized to have the minimal cycle time for the given line pattern inputs, this means that the expanded timetable has maximal gap between the minimal cycle time and the nominal timetable period, and therefore maximal stability.

This expanded timetable fixes the departure event times and includes some running time supplements and headway buffer times due to the expansion. In case of initial delays, the delay absorption is thanks to delayed trains making use of these supplements and buffer times and their delay being reduced at each such event. Trains that are not delayed by an initial delay or a by a knock-on delay do not alter their planned event times, but with no train re-ordering this does not cause further delays: if the timetable is feasible, then because of the linearity of the model an event either is on time, has primary delay, or has at most the delay amount of one of the events it depends on via a time constraint.

In case that the rescaled timetable is not yet satisfactory, it is possible to further improve as follows. If some run or dwell process times become unacceptably long by the expansion, the original model can be re-executed with lower upper bounds. The model can be also re-run with the constants $L = U = T$ and using the sum of run and dwell times as the objective function to be minimized, see also the ”Line-level activity upper bounds” in the next section.

Finally, the output of this model can serve as an initial solution to another timetable optimization model that may improve travel times or other objectives but needs preset train orders. This second stage optimization model can be an LP problem with all binary variables fixed, as the optimal train order for maximal throughput is determined by the current model. For more advanced approaches for this follow-up optimization step of buffer time allocation, see (Kroon et al., 2008; Burdett & Kozan, 2015).

3.2.6 Extensions

While the optimization model above is able to capture the scheduling problem of a periodic train network with flexible train orders, running and dwell times, taking into account infrastructure limitations and optimizing for timetable stability, there are a few timetable requirements that are often present in practice and need to be incorporated into the model to ensure its usefulness. In the following, we describe such extensions, addressing train services at frequencies higher than once per timetable period, parallel train lines, connection constraints, and line-level duration upper bounds.

High frequency lines

Previously, we assumed that all train lines depart from their first station once in each period. In practice, train lines can have different frequencies, with the lowest frequency lines defining the main period length and other trains running e.g. twice or four times per period. In our model we can easily model such high frequency train line as a set of multiple train lines at the basic frequency, however, in this case it is not guaranteed that these trains will have a regular interval service. In fact, the optimal solution will often be "batching" all these trains together: running them directly one after the other. Therefore, a new type of constraint is necessary to ensure that these train lines are separated in a regular fashion.

We extend the tuple definition of events to include a fourth item, $RunNr$, therefore an event $i \in \mathcal{E}$ becomes the tuple

$$i = (Line, Station, EventType, RunNr),$$

where $RunNr$ is the counter of the train run within the period. Therefore for all trains running once per period, $RunNr = 1$. We further define the set of *regularity activities*

$$\mathcal{A}_{reg} := \{(i,j) \times \mathcal{E}x\mathcal{E} | (Line_i = Line_j) \cap (Station_i = Station_j) \cap$$

$$(EventType_i = EventType_j) \cap (RunNr_i + 1 = RunNr_j)\},$$

Then using the notation F_{Line_i} for the frequency of the line of event i for the given cycle time, we introduce the constraint

$$(\tau_j - \tau_i \mod t) = t/F_{Line_i}, \quad \forall (i,j) \in \mathcal{A}_{reg}, \tag{3.39}$$

that can be converted to the MILP-compatible form

$$\tau_j - \tau_i + y_{ij} - (1/F_{Line_i})t = 0 \quad \forall (i,j) \in \mathcal{A}_{reg}, \tag{3.40}$$

$$y_{ij} - Uz_{ij} \leq 0 \quad \forall (i,j) \in \mathcal{A}_{reg}, \tag{3.41}$$

$$y_{ij} - t \leq 0 \quad \forall (i,j) \in \mathcal{A}_{reg}, \tag{3.42}$$

$$t + Uz_{ij} - y_{ij} \leq U \quad \forall (i,j) \in \mathcal{A}_{reg}, \tag{3.43}$$

$$0 \leq y_{ij} \quad \forall (i,j) \in \mathcal{A}_{reg}. \tag{3.44}$$

Note that while equations (3.41)-(3.44) are identical to equations (3.32)-(3.34) and (3.37) and they can be conveniently merged by extending $\mathcal{A}$ with $\mathcal{A}_{reg}$, equation (3.40) is a new sort as it includes the cycle time multiplied by some factor.

Parallel train lines

We saw before how to include train lines in the model having a higher frequency than once per period. Another common market requirement concerns groups of train lines that have a different stop sequence, but they share their route on a substantial part of their trip. In this case, it is often desired that these lines are evenly placed on the common, parallel section, providing a regular service. Such a constraint can easily be added using the same method as high frequency train lines above, except that the events connected by such a regularity activity belong in this case to different train lines.

Connection constraints

Further constraints can be added reflecting certain business requirements. One such a requirement can be a constraint on the duration between the arrival of one train and the departure of another, to ensure passenger connections. Such requirements can be added as new activities in the graph using the structure of inequalities used for run and dwell times: we can define a new activity type $\mathcal{A}_{transfer}$ that connects the arrival of one train to the departure of another train and its bounds are the acceptable minimum and maximum transfer times including walking time.

Line-level activity upper bounds

The values of running and dwell activity upper bounds are decided as a compromise between timetable flexibility and the amount of acceptable time allowances: too little allowed time allowance restricts the model in finding feasible solutions, while too much time allowance can result in running times unacceptably high in practice. However, these two bounds for time allowances are meaningful in different scale: flexibility in running time is more important in a local level planning through a bottleneck, while a practical threshold for maximum time allowance is more applicable regarding the whole line or a longer segment. In other words, the model can be improved by allowing for more flexibility locally, given that the time allowances are not too high on a global level. To achieve this, the running and dwell activity upper bounds can be increased, and to limit the extension of running times of each line, new upper bounds are defined for the sum of run and dwell times of each line.

Let $\mathcal{A}_l$ be all run and dwell activities of a single run of line l (note that in case of multiple runs, it is sufficient to consider only the first), i.e.

$$\mathcal{A}_l = \left\{ (i,j) \in \mathcal{A}_{\text{run}} \cup \mathcal{A}_{\text{dwell}} | Line_i = Line_j = l, RunNr_i = RunNr_j = 1 \right\}, \tag{3.45}$$

and be *Lines* the set of all lines. Then if for line l such suitable upper bound is U_l, then the new constraints are as follows:

$$\sum_{(i,j)\in\mathcal{A}_l} \left(\tau_j - \tau_i + y_{ij}\right) \leq U_l \quad \forall l \in Lines. \tag{3.46}$$

Note that the model only stays feasible if

$$\sum_{(i,j)\in\mathcal{A}_l} L_{ij} \leq U_l \quad \forall l \in Lines \tag{3.47}$$

and for line l these constraints can only become an equality if

$$U_l \leq \sum_{(i,j)\in\mathcal{A}_l} L_{ij}. \tag{3.48}$$

3.3 Dimension reduction techniques

The following pre-processing steps aim at reducing the search space of the MILP model.

Connected components

In the graph representation, different train lines are connected to each other because of the infrastructure, regularity and connection constraints. However, it can still be that the graph is separable to multiple connected components, for example if a train line is operated independently of other lines. As one connected component has no effect on the others, the optimization model could be executed separately for each connected component.

Greatest common divisor of frequencies

In case of regular-interval timetables, often many train services have a headway time of less than the full hour, namely 30, 20, 15 or sometimes even 10 minutes. If the greatest common divisor g of all line frequencies is larger than 1, the timetable can be calculated with updated frequencies $F' = F/g$. This reduces the number of events by a factor of $1/g$ and reduces the number of processes even to a larger extent, because of the structure of the infrastructure processes connecting all events related to a given resource. The new calculated timetable with cycle time t can then be simply scaled back to the original frequencies by duplicating the events at times τ_i to new times $\tau_i + t$, $\tau_i + 2t$, ..., $\tau_i + (g-1)t$ and the re-scaled timetable has a minimum cycle time $g \cdot t$.

Single train line with frequency $F = 1$

It is possible to go one step further in the spirit of the greatest common divisor reduction described above, if we consider the following. If there exists exactly one train line l_1 with frequency equal to 1, while all other frequencies have a greatest common divisor $g > 1$, then the optimal timetable is equivalent to also l_1 having frequency g.

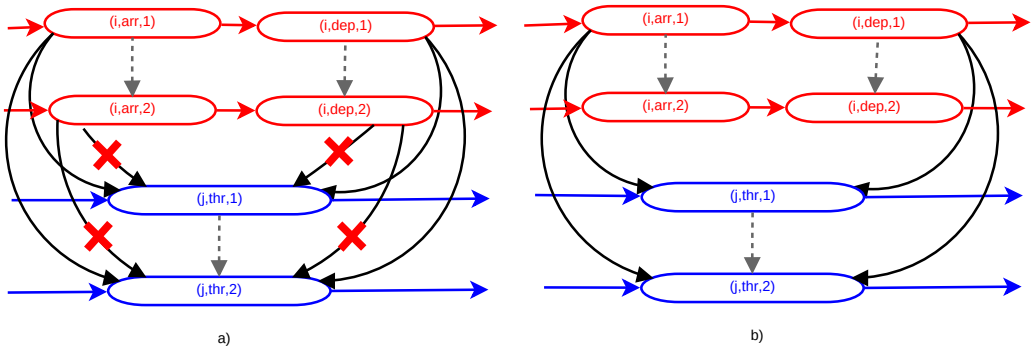


Figure 3.6: Example for removal of redundant infrastructure constraints: with a red cross overlay on panel (a) and removed in panel (b) (the regularity constraints are shown as dotted gray lines, station id is not shown in nodes for clarity.)

This is because the whole timetable is symmetric to a time shift of t/g, except l_1, therefore given a stable timetable, line l_1 can be multiplied to all other $g-1$ time slots. To take advantage of this symmetry, we first increase the frequency of l_1 to g, calculate the timetable taking advantage of the greatest common divisor reduction, and finally remove $g-1$ instances of l_1 to get back to the required frequency of 1.

Redundant infrastructure constraints

If after the previous two steps there still is a set of train lines with equal frequency that is larger than one, it is possible to remove a substantial amount of infrastructure constraints. Take each pair, e.g. lines l_1 and l_2, and while keeping all constraints between the first run of l_1 and any run of l_2, remove all constraints between any further runs of l_1 and l_2. Formally, let us define any total strict ordering $\prec$ of the train lines l_i (e.g. based on the ordering of the line numbers), and then for each pair of train lines $l_1 \prec l_2$ with $1 < F_{l_1} = F_{l_2}$, remove all infrastructure constraints for event pairs (i, j) where $Line_i = l_1$, $Line_j = l_2$ and $RunNr_i > 1$ (recall that F_l is the frequency of line l and $RunNr$ is an attribute of an event referring to the run number which can be larger than one for frequencies larger than one). An example is shown on Figure 3.6. Note that because of the symmetry of the train lines, this reduced set will ensure that all trains are separated accordingly: for example, in Figure 3.6, if there was an infrastructure conflict between the second run of line i and a run of line j, because of the symmetry enforced by the regularity constraints this would also mean an infrastructure conflict between the first run of line i and a run of line j, which is not possible as by definition we did not remove any infrastructure constraints related to the first run.

Symmetry-breaking constraints

For many line patterns there are a larger number of symmetric solutions possible, such as two solutions that only differ in a shift of all events in time or swapping the first and the second run of a train line with frequency 2. The optimization run can be speeded up substantially by ”breaking” these symmetries and fixing one of many symmetrical choices. Therefore these constraints are also called *symmetry-breaking constraints*

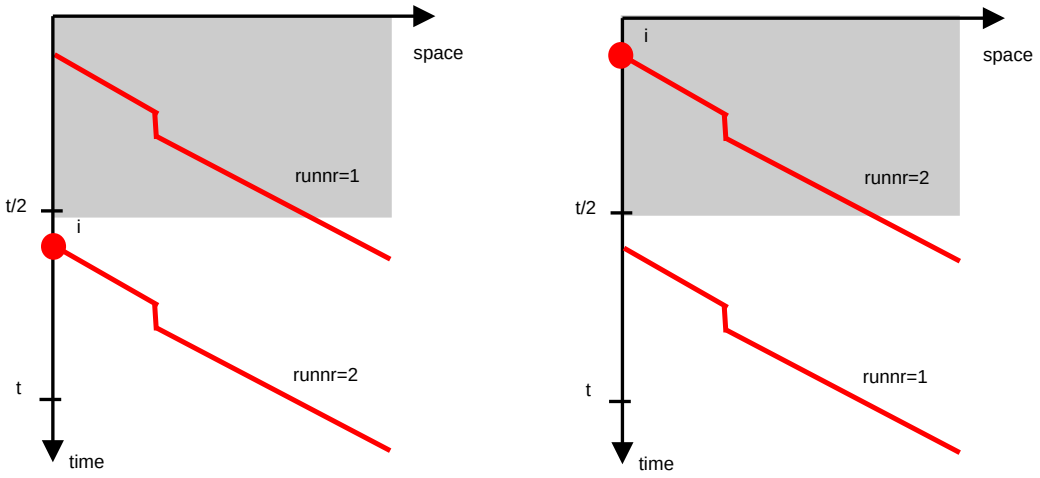


Figure 3.7: Two timetables that only differ in their run numbers, to illustrate symmetry-breaking constraints

(Liberti, 2008).

Two constraints are added to avoid multiple practically identical solutions. Firstly, if two solutions are identical except a uniform shift in time for all event times, they also represent timetables with identical characteristics and minimum cycle time. Therefore, it is possible without loss of generality to choose a single event and fix its event time. Therefore let *init* be this one initial event, and we set $\tau_{\text{init}} = 0$.

The other constraint considers train lines with frequencies larger than one. In this case, a new timetable created by shifting the event times of this line with a multiple of the headway time t/F is identical to the original timetable. Therefore if $\mathcal{F}$ is a set containing a single (arbitrary) event for each train line and F_i is the frequency of the line of event i,

$$\tau_i < \tfrac{t}{F_i} \quad \forall i \in \mathcal{F}\,. \tag{3.49}$$

Similarly to constraint (3.12), the strict inequality can be replaced by a non-strict one by replacing t with $t - \delta$. Note that there is no need to restrict constraint (3.49) to the lines with frequencies larger than one: if $F_i = 1$ then this constraint degenerates to the existing (3.12).

For example, on Figure 3.7, the timetables on the left and the right of the red train with frequency 2 are identical, except that the run numbers are shifted. If we require an arbitrary point i to have an event time less than t/F_i, in this case $t/2$, then only the right side version stays valid and the solution space is reduced.

Marking limits

We call the number of times a process (i, j) crosses the cycle time boundary its *marking*, equal to the related variable z_{ij}. As we earlier ensured that all processes are shorter than the cycle time, a marking is in $\{0, 1\}$. In practice, as the majority of dwell and run

activities are substantially shorter than the cycle time, this means that a chain of run-dwell processes of a certain train line crosses the cycle time boundary much less often than the number of its arcs, meaning that the majority of the arcs have marking zero. To capture this observation, we can define valid inequalities for the sum of markings for each train line by calculating the sum of minimum and maximum process durations and comparing them to the maximum and minimum cycle time bound, respectively. More concretely, recall the earlier definition of $\mathcal{A}_l$ (3.45) for the set of all run and dwell times of the first run of line l, and *Lines* as the set of all lines. Note that if a process (or chain of processes) has duration d and the cycle time is T, then the number of times that this process can cross the cycle time boundary is within $[\lfloor d/T \rfloor, \lceil d/T \rceil]$. Applying this for variable durations and cycle time, for the first run of all lines (note that second and further runs are not necessary to include as they are constrained by the regularity constraints):

$$\left\lfloor \sum_{(i,j)\in\mathcal{A}_l} L_{ij}/U \right\rfloor \leq \sum_{(i,j)\in\mathcal{A}_l} z_{ij} \leq \left\lceil \sum_{(i,j)\in\mathcal{A}_l} U_{ij}/L \right\rceil \forall l \in \mathit{Lines}$$

Note that the effectiveness of this reduction technique is closely related to the size of the cycle time bounds defined.

Run and dwell durations as secondary objective

While the main objective of our model is minimizing the cycle time, in general many different timetable versions can exist with equal train orders and equal minimum cycle time, but slightly different running and dwell times. To break this symmetry, we introduce the sum of running and dwell times in the objective function as a secondary objective (with an appropriately small weight). Note that this reduction method has the practical advantage of making the model choose a timetable with the lowest time allowances among multiple timetables with identical minimum cycle time.

3.4 Computational results

We tested the optimization on four network scenarios of increasing size, using all the dimension techniques implemented from Section 3.3. All data is from the timetable database used by the timetable stability evaluation tool PETER (Goverde, 2007) that is in turn based on the DONS database (Hooghiemstra & Teunisse, 1998) describing the 2007 Dutch periodic timetable for the morning peak. In the PETER timetable file, the national Dutch hourly timetable is included in the detail of stations, junctions and other timetable points together with the minimum running and dwell times, as well as the default minimum infrastructure headways between all train event pairs sharing the same infrastructure.

Table 3.2 contains a detailed description of the used data format. Note that as we store process durations and the calculated event times as continuous (floating point) vari-

ables, and treated as real values by the optimization model, the model supports minimum process time constraints of seconds precision. We included all constraints and minimum process times for run, dwell, and infrastructure headways form the PETER timetable file, however we did not include connection constraints to ensure flexibility in the train orders. For the minimum process times of infrastructure constraints we used the normative default values (such as three minutes) as only norms were given.

During the preprocessing, the network is filtered to the desired network area and subset of lines, as we describe the scenarios later. Maximum durations of run processes are set as $U_{ij} = 1.3L_{ij} + 20$ seconds for each run process (i, j) where L_{ij} is the minimum duration as read from the PETER file and the bounds are expressed in seconds. For infrastructure headways, the actual minimum processing values from the PETER file were used, without adding a running time supplement (e.g. 5%) – if such running time supplement is needed, it can be included in the post-processing step once the train orders are set (see Section 3.2.5): note that if the ratio of the actual and the nominal cycle time is below 95%, a 5% running time supplement certainly fits (before rounding). Furthermore, it is also possible to include the 5% running time supplement at the time of the optimization model input preprocessing if needed. The cycle time for the scenarios based on this timetable was 60 minutes, reduced to 30 minutes in the preprocessing according to the greatest common divisor of frequencies preprocessing step described in Section 3.3, as all lines included in the scenarios have a 30 minute symmetry. We also verify at this time whether for all line pairs using the same station or pair of stations illegal overtakes at a station or between stations are forbidden by the existing constraints, and if not, the dummy nodes are added as described in Section 3.2.5, both for too large dwell times and too large running times. The preprocessing of converting the PETER files to an input file to the MILP solver takes at most a few minutes on a standard PC.

The first scenario consists of the single railway corridor Schiphol-Amsterdam-Almere-Lelystad ("SAAL"), as shown on a geographic map and as a schematic track layout in Figure 3.8. This railway line consists of two-track and four-track sections, all unidirectional, and two junctions with flyovers at Diemen and Weesp as pictured. Including all local and intercity train lines on this corridor, the model consists of 18 train lines and 13 stations, that translate to 156 event nodes and 455 activity arcs in the periodic event-activity network. Recall that the reason for the high number of event nodes is that the model includes events not just at stations but also at other timetable points, such as junctions, where the separation of trains is necessary.

The further scenarios are local trains stopping at Amsterdam Centraal ("AMS-R"), all intercity trains stopping at Utrecht Centraal or Amsterdam Sloterdijk ("UT-ASS-IC"), and all Intercity trains in the Netherlands ("NL-IC"). Scenario "UT" consists of all intercity and local trains calling at Utrecht Centraal station. Scenario "NL-IC+SAAL" is a combination of the earlier two scenarios of the same name. Finally, scenario "NVG" consists of all local and intercity trains running in the large urban region (*Noordvleugel*, "North Wing") around Amsterdam, bounded by stations Leiden Centraal,

Table 3.2: List of input data types used from the DONS/PETER timetable format

Entity	Attribute	Example	Note
Cycletime	Cycle Time	1 hour	Stored as floating point
Station	*Station Abbr.*	Asdz	
	Station Type	IC	Intercity station, other station or other timetable point
	Station Name	Amsterdam Zuid	For visualization purposes
	Longitude	4.873337603	Retrieved from OpenOV*
	Latitude	52.33886371	Retrieved from OpenOV
Line	*Line ID*	270	Stored as integer
	Direction	0	0 or 1
	Line type	IC	Intercity or local train
	Nr. of runs	2	Per cycle time, stored as integer
Event	*Event ID*	l270_r2_d0_arr_Asdz	Generated unique identifier from the other variables
	Line ID	270	Matches *Line ID* of *Line*
	Direction	0	Matches *Direction* of *Line*
	Run Number	2	Integer between 0 and Nr. of runs of *Line*
	Event Type	*arr*	dep, *arr* or thr
	Station Abbr.	Asdz	Matches *Station Abbr.* of *Station*
Process	*From Event ID*	l270_r2_d0_arr_Asdz	Matches *Event ID* of *Event*
	To Event ID	l270_r2_d0_dep_Asdz	Matches *Event ID* of *Event*
	Process Type	*dwell*	*run*, *dwell* or $infra$
	Min. Duration	60 seconds	Stored as floating point, i.e. at least seconds precision

*(Stichting OpenGeo, 2013)

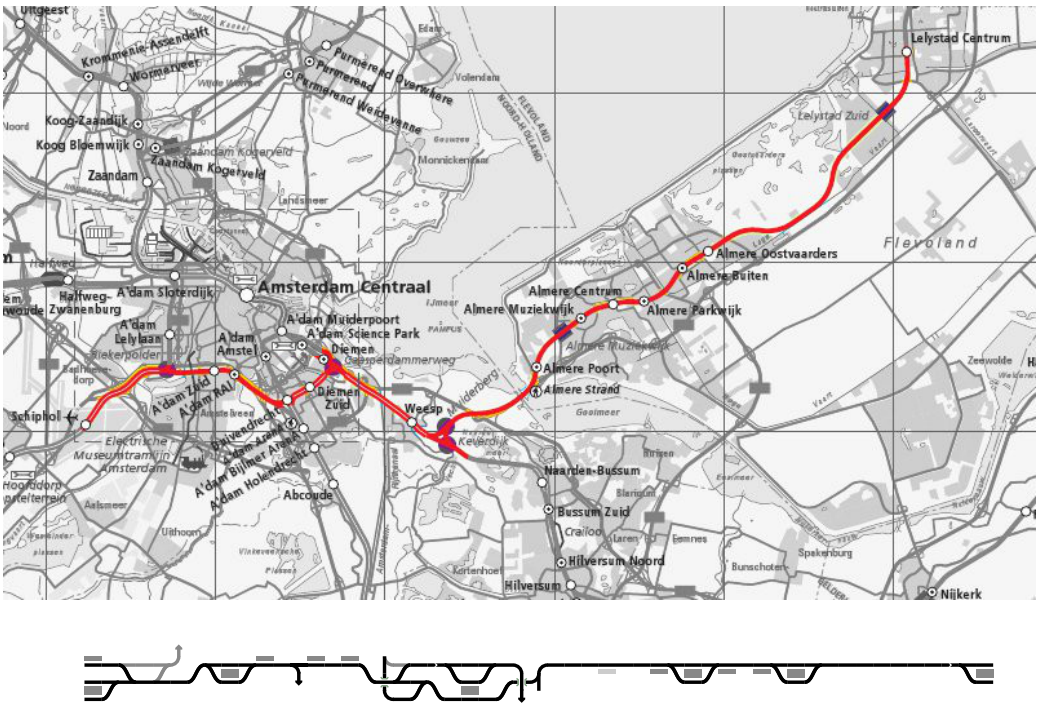
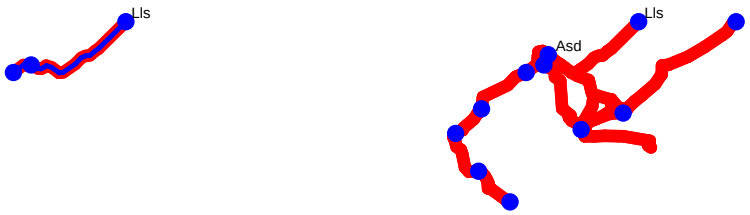


Figure 3.8: The SAAL corridor on a geographic map and its schematic track layout in the West-East direction (thick red line: two tracks, double red line: four tracks, Map source: ProRail)

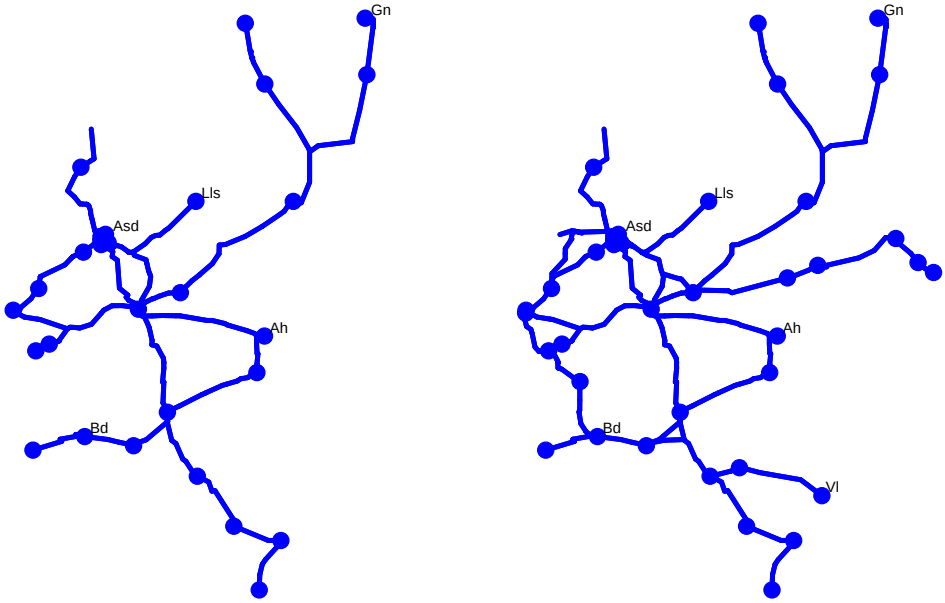
Gouda, Utrecht Centraal, Amersfoort and Lelystad Centrum. See Figure 3.9 and Figure 3.10 for a comparative map of all four scenarios, and see Table 3.3 for the number of lines, stations, and periodic event-activity network dimensions of all scenarios.

For solving the MILP problem we used IBM ILOG CPLEX version 12.4 on a generic PC with 12 GB RAM and a six-core 3.47 GHz CPU with all cores used during the CPLEX run. Figure 3.11 shows the optimization results for the SAAL corridor in the West-East direction. On the compressed timetable, an hourly pattern compressed into the cycle time of 36 minutes and 36 seconds is pictured, and the critical infrastructure headways are pictured (black dotted lines) that determine the minimum cycle time. The expanded timetable represents a stable hourly timetable with minimal cycle time. As mentioned earlier, the expanded timetable can be further optimized with another optimization model using fixed train orders and cycle time, if necessary.

The results of the other three scenarios and calculation times for all scenarios are reported in Table 3.3. All preprocessing techniques from Section 3.3 were used. We can conclude that the optimization model finds the optimal solution in a few minutes for smaller networks represented by event-activity networks with less than a thousand arcs or nodes. However, one of the two larger intercity networks takes 22 minutes to run and this model also proved ineffective in quickly calculating stable timetables for larger scenarios. Therefore in the next section we introduce a method to iteratively run the optimization model solver to address this scalability problem.

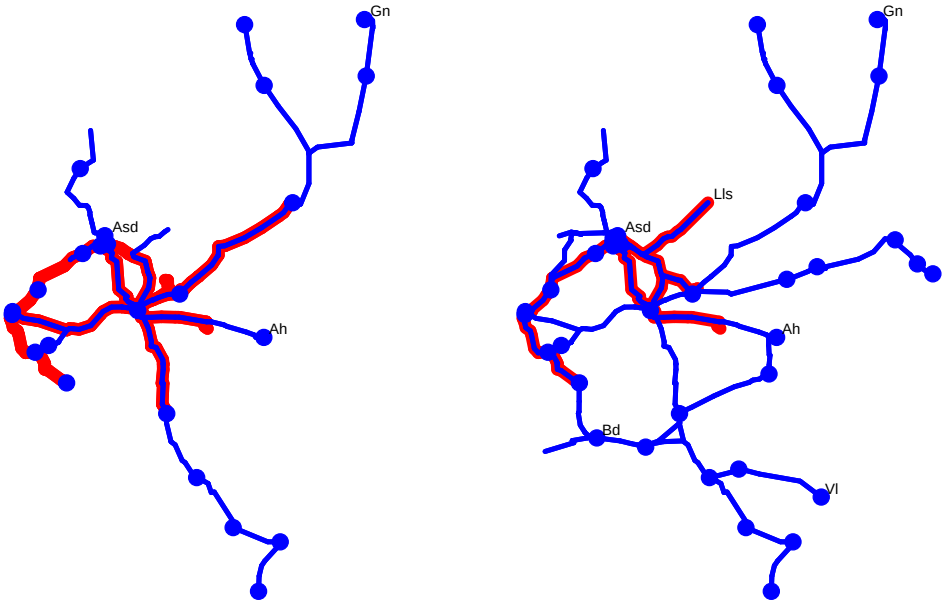


(a) SAAL (b) AMS-R

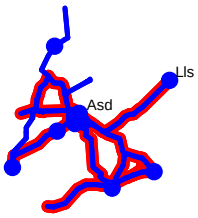


(c) UT-ASS-IC (d) NL-IC

Figure 3.9: Maps of the different scenarios (blue - intercity train lines and intercity stations, red - local train lines, ”Asd” - Amsterdam Centraal, ”Lls” - Lelystad Centrum, ”Bd” - Breda, ”Ah” - Arnhem Centraal, ”Vl” - Venlo, ”Gn” - Groningen)



(e) UT (f) NL-IC+SAAL



(g) NVG

Figure 3.10: Maps of further different scenarios (blue - intercity train lines and intercity stations, red - local train lines, ”Asd” - Amsterdam Centraal, ”Lls” - Lelystad Centrum, ”Bd” - Breda, ”Ah” - Arnhem Centraal, ”Vl” - Venlo, ”Gn” - Groningen)

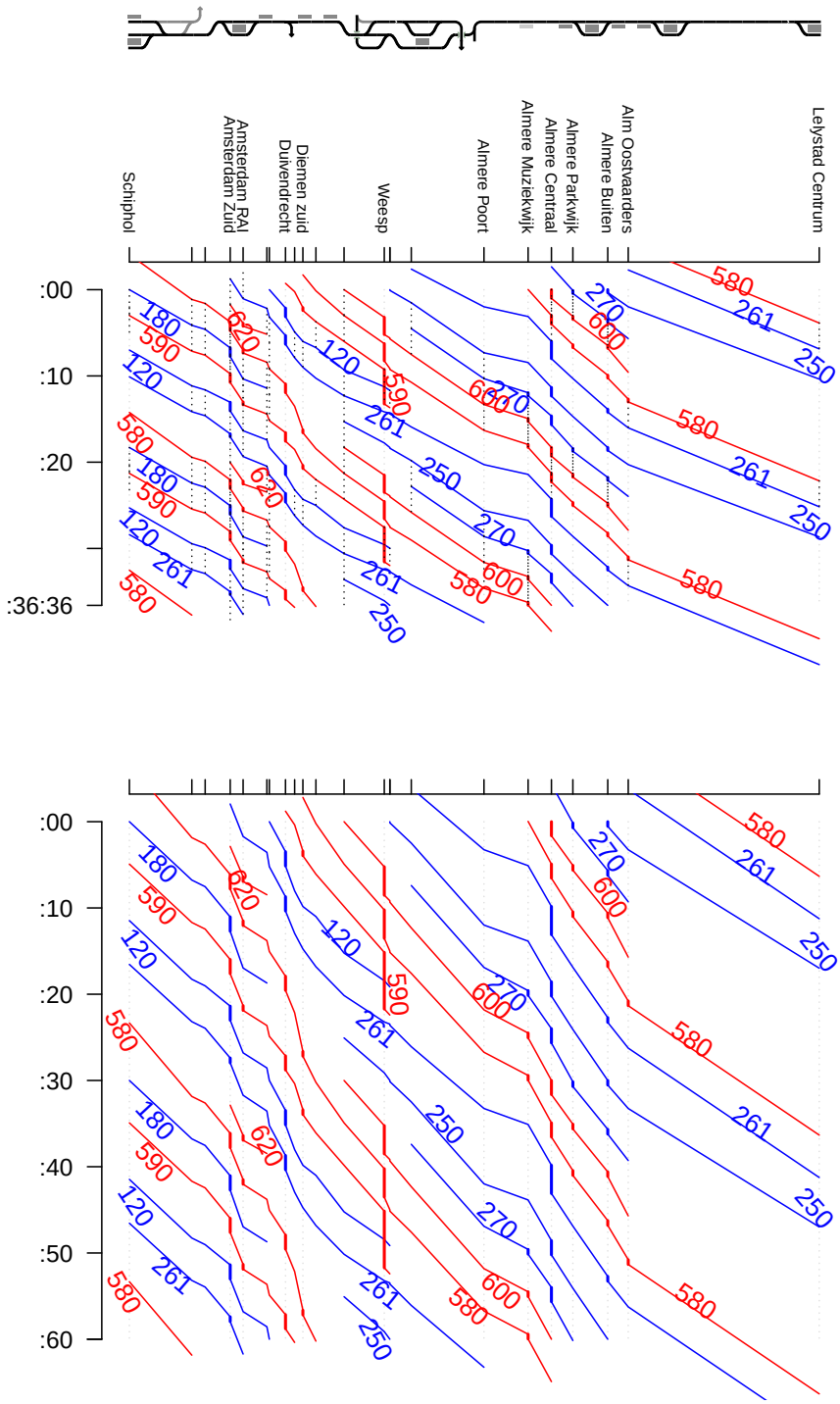


Figure 3.11: Schematic track layout (top), optimal compressed (middle) and expanded (bottom) timetable for the SAAL corridor (one direction pictured, red: local trains, blue: intercity trains, black dotted lines: critical infrastructure constraints)

Table 3.3: Graph size for different scenarios and optimal timetable calculation times

Scenario	Lines	Stations	Nodes	Arcs	Solution time	$^t/_T$
SAAL	18	13	156	455	00:00:45	61.0%
AMS-R	16	42	605	791	00:02:30	51.4%
UT-ASS-IC	20	142	1033	1473	00:22:33	40.0%
NL-IC	30	193	1515	2175	00:05:12	40.0%
UT	28	140	1352	2552	04:00:00<	n/a
NL-IC+SAAL	35	195	1837	3103	04:00:00<	n/a
NVG	65	82	1594	3917	04:00:00<	n/a

3.5 An iterative solution approach

In the previous section we have seen that the presented optimization model is able to calculate a stable intercity timetable for the whole Dutch network in a couple of minutes, with flexible train orders and running and dwell times. The model, however, is not suitable to calculate larger timetable instances, such as a large regional timetable with both local and intercity trains, in one optimization solver run in a suitable time (we used 4 hours of computation time as a timeout). Therefore, in this section we present an iterative method to calculate large stable timetables very fast, in just a few seconds, and optimize these timetables for stability in a few hours.

The key idea of this method is that while the cycle time t is a variable, nevertheless it is bounded by the predefined limits (L,U) and the smaller this range is, the faster the optimization solver can advance within the reduced search space. Hence, we introduce an iterative calculation where the cycle time range (L,U) is periodically adjusted and the calculation resumed.

We take advantage of the fact that many MILP solvers, including CPLEX, allow a definition of an initial feasible set during the declaration of the MILP problem, that guarantees feasibility and may further speed up the optimization process: we use the intermediate solution of one solve iteration to initialize the following iteration. We define the function name $MILPSolve(\mathcal{E},\mathcal{A},L,U,k)$ that represents solving the MILP problem defined earlier given the events $\mathcal{E}$ and activities $\mathcal{A}$, as well as the cycle time bounds (L,U), and the optional variable k containing an intermediate solution.

Algorithm 1 describes the structure of the iterative process. Given the timetable input values $(\mathcal{E},\mathcal{A})$, the nominal cycle time T, and the iteration configuration parameters $F > 1, 0 < g_{min} < g_{init} < 1$, we first run $MILPSolve()$ with fixed cycle time $t = L = U = T$ to obtain a feasible solution. Then the range (L,U) is initialized such that $L < t = U$ according to the initial relative gap parameter g_{init}. Consequently, we iteratively run the solver $MILPSolve()$ until optimality within the bounds, or until a timeout is reached since the last change in the objective value. The bounds are re-set after each iteration according to the factor F: if optimality was found at the lower bound then the gap $(U-L)/U$ is increased by a factor of F, and if new solutions were found and the

cycle time variable is still at the current upper bound then the gap is decreased by a factor of F. And otherwise the relative gap does not change. In all three cases, the bounds are re-set as $U = t$ and L according to the defined gap. The iterations stop when optimality is reached or the relative difference between L and U reached the minimum relative gap g_{min} when we return the intermediate solution. We used the values $F = 2$, $g_{init} = 0.01$, and $g_{min} = 0.001$. For the timeout of $MILPSolve()$ we used the internal iteration counter of CPLEX and the timeout value of 800 000 CPLEX iterations, which is approximately between 2-4 minutes depending on the problem size. The reason we used a timeout based on the iteration count and not on clock time is that this way the solver runs are fully reproducible.

Algorithm 1 Iteratively optimize timetable

function ITERATIVELYSOLVE($\mathcal{E}$, $\mathcal{A}$, T, F, g_{init}, g_{min})
  k $\leftarrow$ MILPSOLVE($\mathcal{E}$, $\mathcal{A}$, T, T) // initial solution for solver
  $U \leftarrow T$
  $L \leftarrow (1 - g_{init})T$
  while $g_{min} < (U - L)/U$ **do**
    (SolverStatus, t, k) $\leftarrow$ MILPSOLVE($\mathcal{E}$, $\mathcal{A}$, L, U, k)
    if SolverStatus = 'Optimal' and $L < t$ **then**
      return (t, k) // optimal solution
    (L, U) $\leftarrow$ RESETBOUNDS(L, t, U, F)
  return (t, k) // intermediate solution

For obtaining an attractive calculation time, it is instrumental that the reset of the cycle time bounds (L,U) is performed in an adaptive fashion. In particular, if the difference $U - L$ between the cycle time bounds is too large, a single iteration can take too long time, or reach timeout. On the other hand, if the cycle time bound is too small, the number of iterations can be too high, again causing a long total calculation time. Therefore we define the adaptive $ResetBounds()$ function as described in Algorithm 2. As a default rule, we set $L := (1 - g)t$ and $U := t$ where g is the relative gap of the previous run. However, if the previous run was perceived too fast (t actually reached the lower bound L), or too slow (t did not improve from $t = U$ until timeout), we increase or decrease relative gap g, respectively, using the relative gap factor variable F set as a parameter of the function.

Using the previously described iterative process, we were able to extend the scope of our optimization model to quickly calculate stable timetables of large regions with mixed traffic. Figure 3.12 shows the expanded timetables of two representative corridors of the NVG scenario, in both directions. To illustrate the effect of the iterative calculation with varying cycle time ranges, Figure 3.13 shows the objective value of the optimization solve runs as functions of the calculation time. Figure 3.13(b) illustrates the mechanism of the iterative calculation the best: the long horizontal lines of the objective value signify situations when the optimization solver was unable to improve the solution until a timeout. After a re-set of the objective bounds, however, the

Algorithm 2 Reset search bounds given intermediate solution

function RESETBOUNDS(L, t, U, F)
  $g \leftarrow (U - L)/U$ // gap
  if $L = t$ **then** // optimality within bounds: increase relative gap
    $g \leftarrow F * g$
  else if $t = U$ **then** // no new solutions: decrease relative gap
    $g \leftarrow g/F$
  $L \leftarrow (1 - g) * t$
  $U \leftarrow t$
  return (L, U)

solver was able to continue and find optimality.

Table 3.4 shows calculation times for all scenarios both using the basic solve method seen in the previous section and the current iterative solve. All preprocessing techniques from Section 3.3 were used both in the basic and the iterative case. In the latter, we make the distinction between the calculation time until a stable timetable and to optimality. In some cases, it was not possible to prove mathematical optimality before timeout, in these cases calculation time to the best attained value is reported.

Table 3.4: Stable and optimal timetable calculation times with the basic and the iterative method for different scenarios

	Basic solve	Iterative solve			
Scenario	Optimal	Stable	Optimal	Iterations	$^t/_T$
SAAL	00:00:45	00:00:00	00:01:01	7	61.0%
AMS-R	00:02:30	00:02:02	00:02:02	7	51.4%
UT-ASS-IC	00:22:33	00:00:06	00:05:09	8	40.0%
NL-IC	00:05:12	00:00:08	00:25:34	10	40.0%
UT	04:00:00<	00:00:13	03:57:30*	72	70.5%*
NL-IC+SAAL	04:00:00<	00:00:11	01:47:39*	27	61.0%*
NVG	04:00:00<	00:00:02	03:35:00*	42	68.1%*

* Optimality not proven, values related to stabilized value reported

Figure 3.14 and Figure 3.15 plots calculation times to feasibility and optimality on a logarithmic scale as a function of the size of the graph. Note that for the three largest cases, only the iterative approach is able to find a stable and optimal timetable, this is represented by the three points of the basic solution method in the upper right corner of the charts corresponding to the ”4h¡” mark of the vertical axis. We can conclude that the iterative solution method enables a significant reduction in computation times and allows stable timetable calculations of large mixed traffic networks in a few minutes, and stability optimization in a few hours.

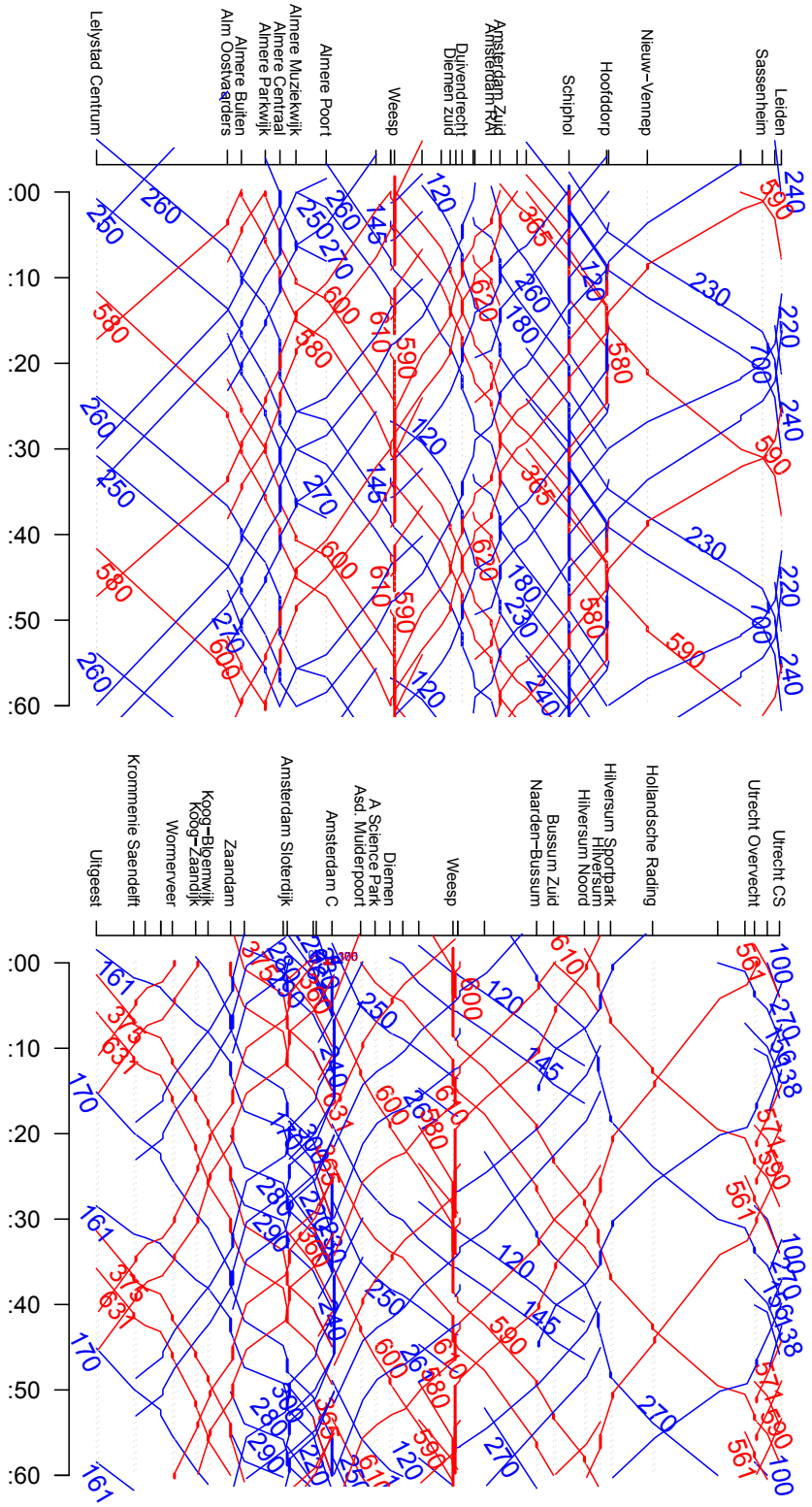
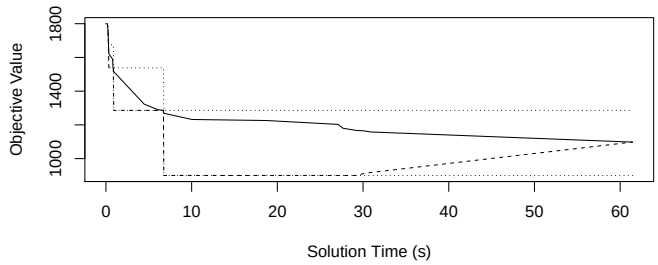
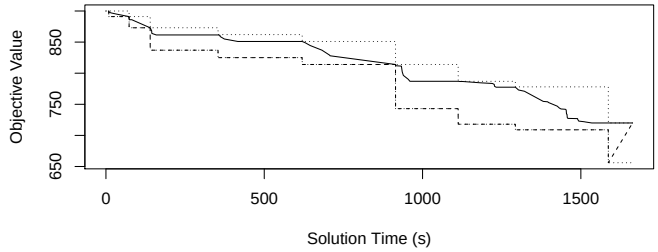


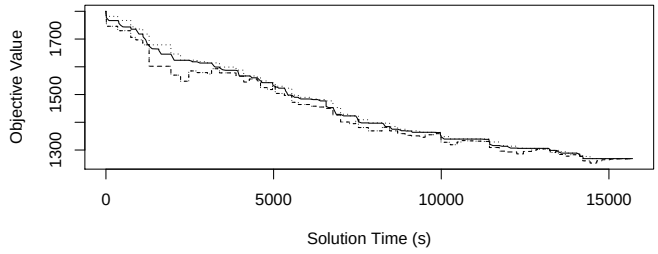
Figure 3.12: Optimized expanded timetable of the NVG scenario, Leiden-Lelystad Centrum (above) and Uitgeest-Utrecht (below) corridor



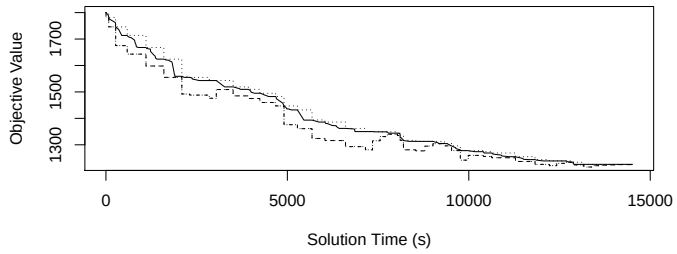
(a) Scenario SAAL



(b) Scenario NL-IC



(c) Scenario UT



(d) Scenario NVG

Figure 3.13: Intermediate objective value and its bounds during the solver run for scenarios SAAL, and NL-IC, UT and NVG (dotted lines - objective bounds, dashed line - LP bound, continuous line: intermediate objective value)

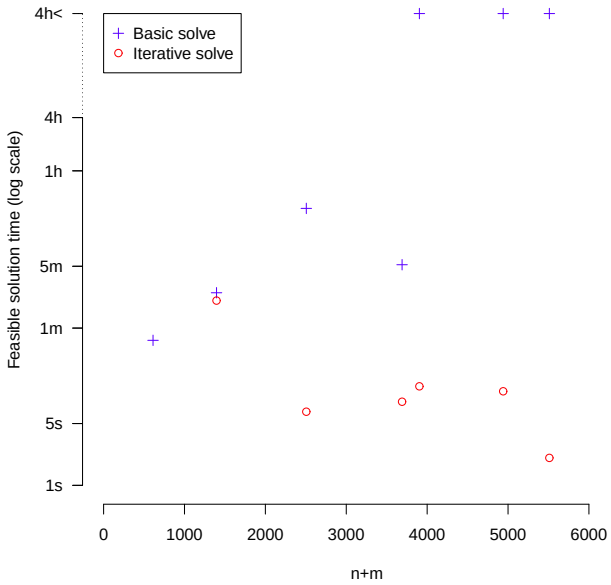


Figure 3.14: Calculation time of a stable timetable as a function of the number of nodes and arcs in the event-activity network

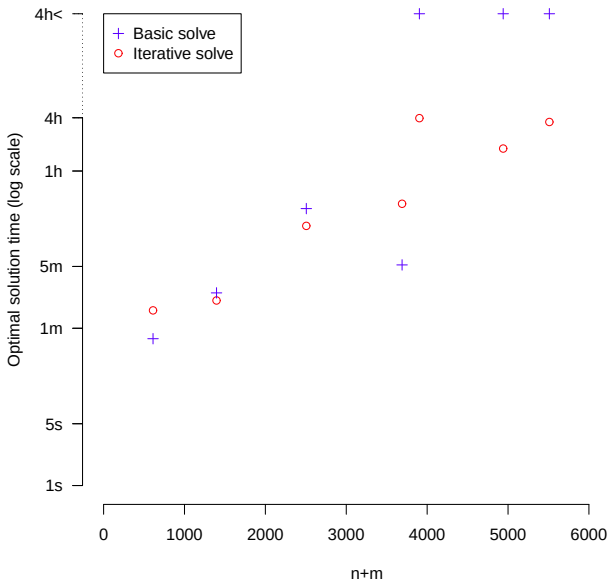


Figure 3.15: Calculation time of an optimal timetable as a function of the number of nodes and arcs in the event-activity network (optimum not proven in the largest 3 cases)

3.6 Conclusions

In this chapter we presented a periodic railway timetable optimization model that can handle flexible train orders, running and dwell times, and uses a measure of timetable stability as the objective function. We introduced several dimension reduction techniques, as well as an iterative optimization approach, to ensure that the model is applicable to large mixed traffic networks. Using available timetable planning data for the Dutch national railway network, we applied the optimization model to several networks of different sizes and train classes to illustrate the usefulness of the model.

The results show that our approach is able to calculate a stable timetable for a large mixed traffic network in just a few seconds, and optimize the timetable for stability in 2-3 hours calculation time on a generic computer. This makes the model highly suitable both for quick comparison of many different timetable scenarios of different frequencies and line patterns, and for improving existing timetables to increase their stability by the appropriate reordering of trains and reallocation of buffer times according to the stability-oriented optimization.

Chapter 4

Transfer time modelling with open data[1]

4.1 Introduction

While Chapter 3 focused on increasing capacity in high-demand railway networks via a method to design a high-frequency railway timetable, in this chapter we investigate another common aspect of most public transport trips: transfers. Remaining at the example of the high-frequency railway network, not only transfers between two train services are important for the passenger experience and therefore the demand and revenue of the train operator company, but also transfers between trains and the local feeder or distributor public transport lines, such as local or regional buses, light rail or metro lines.

In the case of feeder and distributor lines, the *transfer station resistance*, a measure of the additional inconvenience a passenger experiences when compared to a direct service, is determined by many factors such as the synchronization of the services and therefore the transfer wait time, the layout of the transfer node and the related transfer walking time, as well as facilities at the station and the passenger safety. Ths transfer station resistance is therefore a more general concept than the *transfer time*, that is the time spent during a transfer between the arrival and the consequent departure. The transfer time can be further divided into *transfer walking time* and *transfer waiting time*, where the former is largely dependent on the layout of the station and the second on timetable synchronization. In this chapter we introduce a new method for the estimation of the transfer station resistance based on open source data, using an average transfer walking time on all services at a given station as a proxy for transfer station resistance. In particular, we model station layouts in detail and measure transfer waiting times more accurately than current timetable planner tools do, as many of

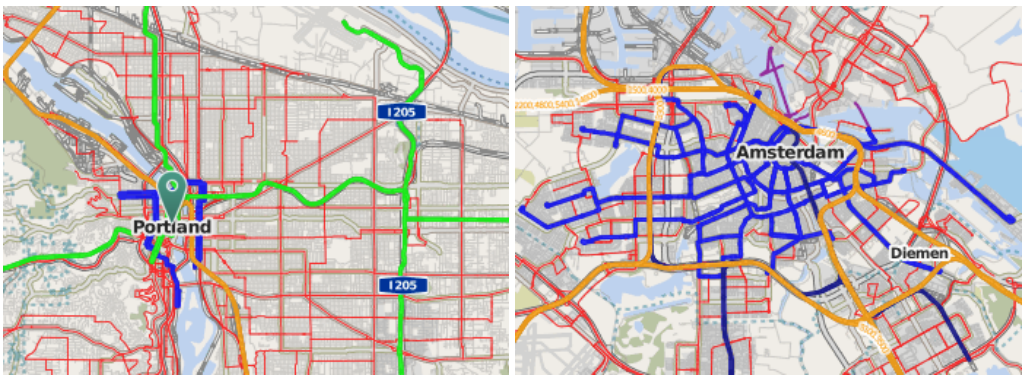
[1]This chapter contains an edited version of parts of the journal paper: Van Oort, N., Sparing, D., Brands, T., & Goverde, R. M. P. (2015). Data driven improvements in public transport: the Dutch example. *Public Transport*.

those only can work with rough estimates of walking times, especially in case of operators other than the incumbent main railway operator, who might not have access to detailed station plans. Consequently, we provide an approximation of transfer station resistance as the relationship between the typical transfer walking time and the amount of transport supply at a station. These detailed transfer walking time values can also be used to improve real-time passenger information, as illustrated in the final section of this chapter.

Transfers, or *connections*, are inevitable characteristics of public transport networks. This despite the fact the transfers are often one of the strongest inconveniences passengers experience, who most often prefer direct services. Transfers are necessary, because in practice it is financially and operationally impossible to design an attractive line service layout that provides a direct service between any arbitrary pair of locations. (Without the criterion of 'attractiveness', such a layout is certainly possible: a service that stops at each address of the area in question one by one, in some arbitrary order.)

It is possible, however, to mitigate the inconveniences caused by transfers. The first approach is to limit the number of transfers for the typical journey. Taking into account spatial and usage patterns, it is often possible to design an urban network where *most* trips are transfer-free, e.g. if most non-home destinations are in a central business district (CBD). Relaxing the requirement of no transfers, it is also very well possible to design an urban line network layout where *any* two locations are connected with at most one transfer. This is true for a rectangular grid or for a radial grid, if the grid is dense enough to assume that from any origin or destination, a station of either orientations (such as North-South and East-West, or radial and tangential) is reachable, see Figure 4.1 for approximate real-life examples. A second, complementary approach is to limit the inconvenience of transfers, by reducing the walking time between platforms and the transfer waiting times, and improve passenger information, passenger safety and then further amenities at the transfer location. In the following of this chapter, we will focus on reducing transfer inconvenience by accurately estimating the transfer walking time, reduce the transfer waiting time, and improving the provided passenger information. For previous work on detailed analysis or physical station layouts, see the last paragraph of Section 2.4.4 of the literature review.

The remaining of this chapter is organized as follows. Section 4.2 describes the state of open public transport data, with focus on the Netherlands. Section 4.3 shows the measurement transfer walking time using the available high-resolution open spatial data on transfer stations. Section 4.4 describes three application examples of high quality transfer walking time: transfer station resistance, timetable synchronization, and enriched real-time passenger information. Section 4.5 concludes the chapter.



(a) (b)

Figure 4.1: Near-perfect grid networks: (a) the rectangular grid bus network of Portland, Oregon, United States (in red), (b) the radial grid light rail network of Amsterdam, the Netherlands (in blue)

4.2 Open public transport data

Public transport companies have always dealt with large amounts of data when designing timetables, scheduling vehicles and staff, collecting fares and more recently tracking vehicle locations. However, it has only recently become possible to store large amounts of historic vehicle location and fare collection data, and therefore to analyse this data. Furthermore, in line with other Open Data initiatives in the public sectors, data related to public transport is currently becoming publicly available in more and more areas, notably in North America and more recently in certain European cities. The first type of public transport data that became publicly available is timetable information. Besides supplying public transport route planners with timetable data, computer-readable timetable information also allows for efficient analysis and comparison of public transport networks, describing spatial coverage, commercial speeds, frequencies, and connections to adjacent public transport networks. Timetable information provides no insight yet, however, in the performance of the timetable realization and hence the service reliability of public transport and the real-time timetable information. Accurate real-time Automatic Vehicle Location data (AVL) has become available for public transport operators with the wide availability of GPS and GSM devices. AVL data has also become publicly available in many areas in the recent years, albeit often with the condition that it is only used for passenger information. Early examples include the transit agencies of Washington, Boston and some other US bus companies. We note that these days most Western public transport operators provide some kind of real-time vehicle location (or expected vehicle arrival time) information to the public, but often this information is still not technically or legally available for storing or further processing by third parties.

Another type of data that became widely available at no cost or as open data is geographic data. Services like *Google Maps* and *OpenStreetMap* allow a wide range of

geospatial applications. While some, like Google Maps, only provide satellite images at no cost that might not be fully open to use for any purpose, open data services like OpenStreetMap make detailed geotagged data free to download and use for any purpose. Application areas of such geospatial data for railway and public transport operations research range from better measurement of line layouts to supporting travel demand analysis to data visualization. Later in of this chapter our focus of using this data will be the detailed modelling of transfer locations.

4.2.1 The Dutch example

In the Netherlands, most public transport operators are on board with the initiative called Borderless Public Transport Information (Grenzeloze Openbaar Vervoer Informatie, GOVI (GOVI, 2013), aiming at making a wide range of public transport information available and processable from timetables to fares, vehicle location and punctuality. The data exchange interfaces (*koppelvlakken*) are defined by the set of standards called BISON (BISON, 2013). Another source of open public transport information, such as a GTFS feed on the national level, is the company 9292 REISinformatiegroep BV (9292 REISinformatiegroep, 2013), a company specialized in passenger information owned by Dutch operators. GOVI was designed to facilitate data communication between vehicles and the land side to enable dynamic passenger information. As an additional benefit, the actual and scheduled vehicle positions and times are logged in a database. Although this database was not the objective of the GOVI system, it is extremely helpful to monitor and analyze public transport performance. In particular, in 2012, the first Dutch public transport operator agreed to legally release AVL data via GOVI for storage and analysis by third parties, such as researchers and developers (Webwereld, 2012). Since then several other operators joined. Such data streams are publicly available via the Dutch OpenGeo Foundation (Stichting OpenGeo, 2013). The source of the planned timetable and AVL data presented later in this chapter is the transit authorities and OpenGeo.

One substantial gap in open transit data for the Netherlands at the moment is open realized arrival and departure data for the railways. In fact, two different kinds of data sources exist with different limitations: train describer data (TROTS) and publicly available actual arrival times data from the Netherlands Railways (NS), the via the NS API. TROTS is a train describer system that records the train numbers occupying a track section including their timestamps, as well as changes of signal aspects and switch directions. However, while there has been substantial research on TROTS data (Kecman, 2014), it cannot openly be obtained in general. The NS API, on the other hand, containing expected arrival times in real time, is openly available only for the purpose of passenger information, and it is not permitted to store this data or use it for performance evaluation purposes. In conclusion, unlike most bus and urban public transport operators contributing to GOVI/BISON, actual train arrival and departure data is unfortunately not available as open data for a posteriori analysis.

Table 4.1: Example data output from BISON interface KV6

Time	Message type	Operator	Line	Journey	Stop	Punctuality
08:29:00	INIT	...	B120	7001	99990140	
08:29:00	ONSTOP	...	B120	7001	99990140	60
08:29:22	DEPARTURE	...	B120	7001	99990140	82
08:31:28	DEPARTURE	...	B120	7001	99990290	88
08:51:04	ONROUTE	...	B120	7001		
08:52:37	ARRIVAL	...	B120	7001	99990500	-202
08:52:37	END	...	B120	7001	99990500	

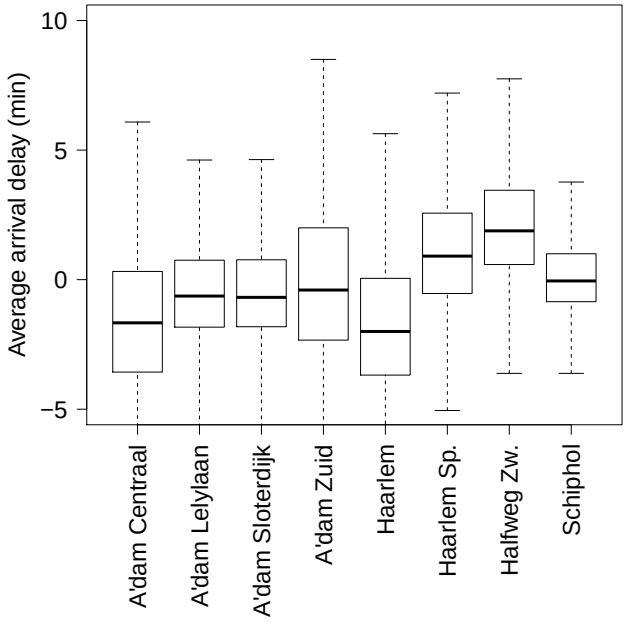
4.2.2 Insights from AVL data

As a first step in an open data analysis such as AVL data, it is important to understand the structure and the quality of the data source. In our case, AVL data was available for several months from multiple operators in the format described by interface KV6 of the BISON standard mentioned earlier. An example extract of the most important data attributes and the first and last few records related to a single public transport vehicle trip is presented in Table 4.1.

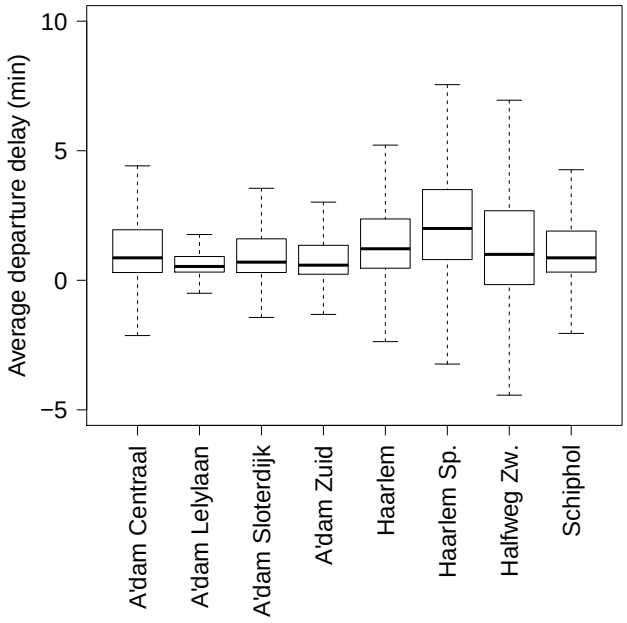
This data table consists of timestamped messages of important events of the vehicle trip. In particular, a trip starts with an INIT initial message and ends with an END message, and all departures are logged with a DEPARTURE message. In case of some stops an ARRIVAL message is recorded too, allowing for an estimate of the dwell time. Furthermore, in case that there is no departure and arrival event taking place for a longer time duration (about a minute), an ONSTOP or an ONROUTE message is recorded, including exact location. Our data source already includes a value for delay, which equals to the difference of the message timestamp and the planned arrival or departure time. Note that the planned arrival and departure was also available in the BISON interface KV1 for the related lines.

Realized departure and arrival times allow for the measurement of service punctuality: see Figure 4.2 for the distributions of arrival and departure delays at certain stations around Amsterdam West and Haarlem. The distribution is measured for all buses of the operators Connexxion and Arriva at the given stations in the period of 19th November 2012 to 4th February 2013. Note that the box plots shown in this figure plot the median value, first and third quartile, as well as the range of the remaining values after the removal of outliers.

A commonly used visualization (Furth et al., 2006; Van Oort & Van Nes, 2009) of the performance of a transit line is plotting each trip as a line chart in a coordinate system of stops versus delay. Figure 4.3(a) shows one month of bus trips of a certain line, as well as the median and 15th and 85th percentiles. Such a chart is useful to see both the level of variations in the execution of the timetable and the systematic deviations.

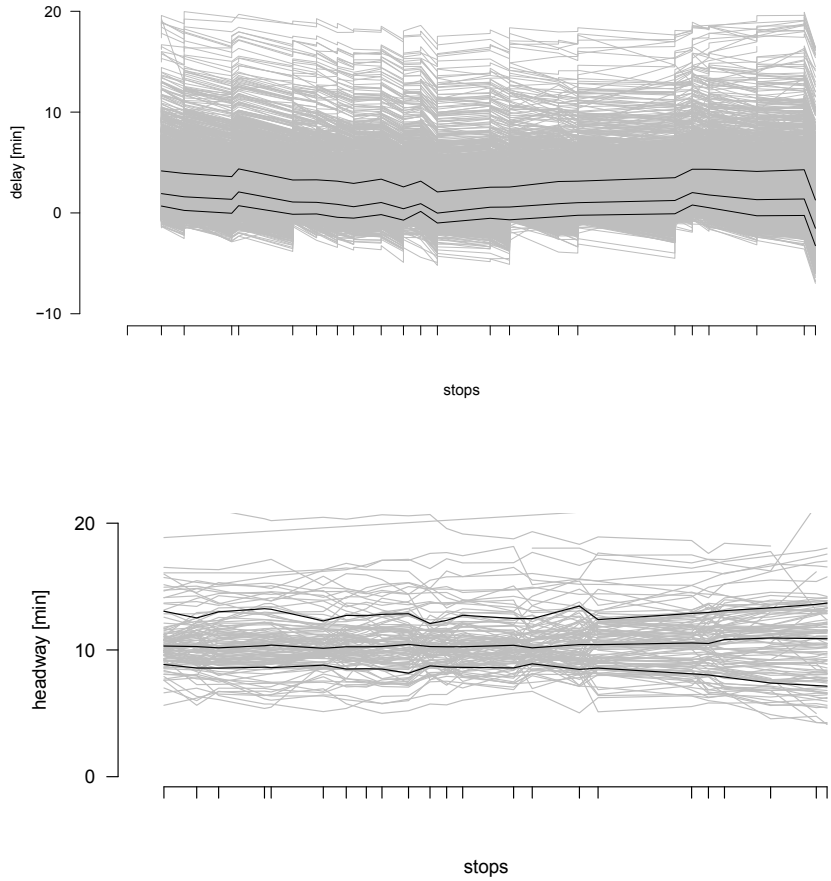


(a)



(b)

Figure 4.2: Departure and arrival delay distribution of bus lines included in GOVI KV6 at the analysed stations



(a)

(b)

Figure 4.3: Vehicle delays (a) and headways (b) along a single route

Other phenomena that are shown by this particular chart are the ample time reserve used just before the last stop and the use of some holding points during the trips.

Another way to look at the same data is to plot vehicle headways instead of delays. A high frequency line with a high level of delays but regular headways remains attractive to the passengers. The chart is shown on Figure 4.3(b), the scheduled headway is 10 minutes. This location-headway chart points out the regularity of high-frequency services along the line, as well as possible bus bunching.

The ubiquitous availability of vehicle locator devices allows one to take a step further from line-based performance evaluation and investigate patterns at the network level. Phenomena only visible on the network-level are the reliability of transfers, area-related issues and possible bunching or interference on multiple lines with shared sections. An example of a network-level data visualization is the average delay at each stop, including stops with several transit lines, shown on Figure 4.4.

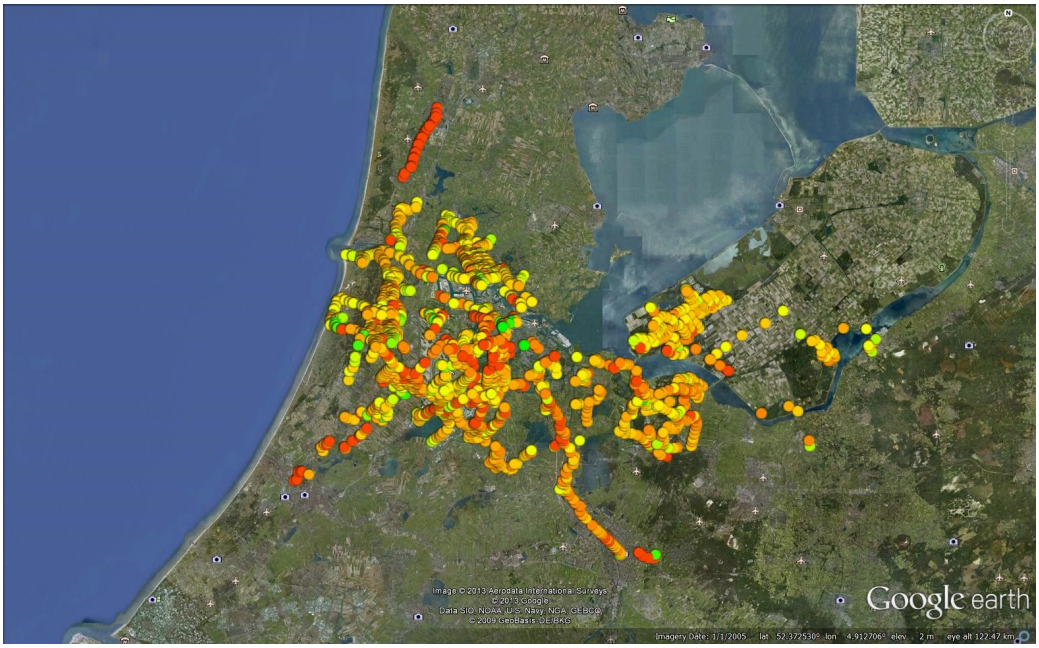


Figure 4.4: Average delay per stop in January 2013 (green: early, yellow: on time, red: late, on all Amsterdam bus services of the operator Connexxion)

4.3 Transfer walking times

The detailed geospatial data described in Section 4.2 makes it possible to measure and model transfer station layouts in detail. Measuring the walking distance between any pair of platforms enables the estimation of transfer walking times at high accuracy. We expect that stations with relatively high transfer walking times will be substantially less attractive for passengers than other, otherwise similar stations. This detailed modelling of transfer walking times is a useful input for further analyses such as transfer station resistance, timetable synchronization, and enriched real-time passenger information, as we will see in Section 4.4.

The freely or openly available geographic tools such as Google Maps allow to build a model of a station by building a three-dimensional *walking grid* of recorded spatial points that represent all platforms and walking paths between them at a suitable resolution. Such a spatial point is a *(longitude, latitude, altitude)* tuple and we suggest a resolution of at most 100 meters – this means that in case of platforms longer than 100 meters, multiple points need to be recorded such that adjacent points are at most 100 meters from each other. Note that while collecting longitude and latitude using available mapping tools is very easy, altitude within a station building might be harder to obtain: in this case one can use suitable approximations knowing typical design heights of passenger tunnels and bridges under and above railway tracks, and observe that only the relative height is important in walking distance calculations.

Once such a 3D walking grid is collected including walking routes, stairs and escala-

Table 4.2: Walking speed on flat surface and stairs

environment	speed
flat surface	1.34 m/s
stairs, down	0.694 m/s
stairs, up	0.610 m/s

tors based on the openly available satellite images, it is possible to measure the distance between each pair of adjacent points. Note that the WGS64 (”GPS”) coordinate system of latitudes and longitudes is not Euclidean and therefore distances are not trivial to calculate, this nevertheless can be solved using a dedicated geographic library, such as the *PostGIS* geospatial extension to the open source database *PostgreSQL*. Walking time between point pairs can be measured by taking into account representative walking speed values on flat surface, ascending and descending stairs, available from literature (Weidmann, 1993), see Table 4.2.

Once the walking distance and time between any pair of adjacent points is known, including the detailed walking path possibly containing detours due to stairs, escalators and general station layout, the walking distance between any two platforms was calculated by Dijkstra's algorithm. In case of long platforms, such as mainline trains, a distinction has to be made in the walking times from and to the platform: while the shortest walking time to the platform is the shortest walking time to any of the platform's representative points, for the other direction it is more accurate to take the average or maximum walking time from all points of the platform taking into account the fact that passengers alight from the train at the whole length of the platform.

Using walking times between all platform pairs, a suitable statistic such as the mean, median or maximum walking time among all platform pairs can be used to represent the typical walking time at the station.

4.3.1 Case Study: The Schiphol-Haarlem-Amsterdam West network

To illustrate the above methodology to represent transfer station resistance, we apply the method to the train network between Schiphol Airport, Haarlem and West Amsterdam as follows.

A geographic map of the selected network of seven stations is shown on Figure 4.5 and a schematic map of the 2011 services is on Figure 4.6. This small network includes stations of all scale from stations with only local trains stopping to international and high speed train stations. These stations are served by local public transport, in particular local and regional buses, bus rapid transit lines, trams, metro/light rail and river ferries.

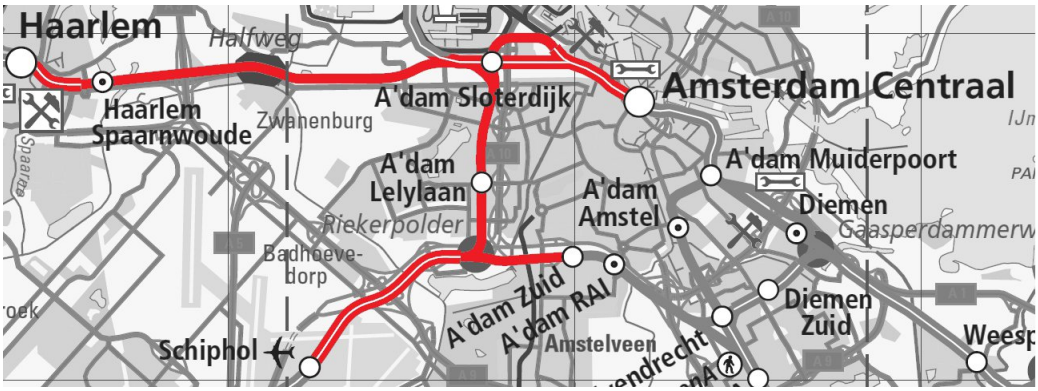


Figure 4.5: A geographic map of the selected railway network Haarlem / Schiphol / Amsterdam West (source: ProRail)

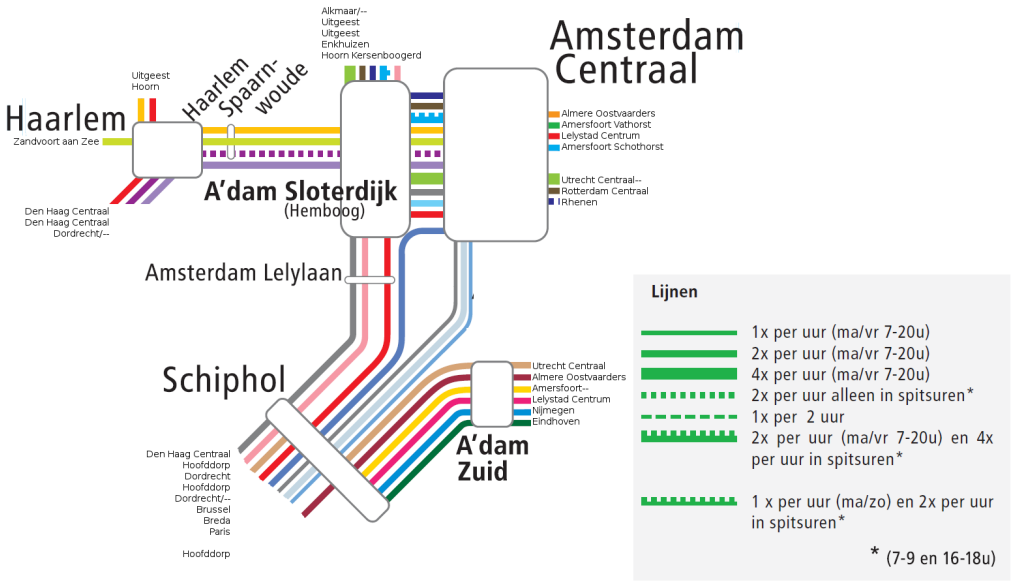


Figure 4.6: Schematic ”tube” map of railway lines and frequencies of the selected railway network Haarlem / Schiphol / Amsterdam West (source: Treinreiziger.nl)

In order to estimate the transfer station resistance, the walking distance and time between different platforms is estimated. Therefore, 3D geographical models of the stations are built.

Furthermore, special attention is given to long platforms. In particular, while buses and trams (here) are relatively short vehicles, metros and trains can be as long as about 150 and 340 m, respectively. Therefore, in case of the long metro and train platforms, there are 3 and 7 points recorded respectively along the platform so that the distance between consecutive points is not larger than 100 m.

These platform locations and the walking paths connecting them are recorded in the WGS84 (”GPS”) coordinate system, i.e. latitude, longitude and altitude, using Google Maps. As an example, platforms and walking paths of station Amsterdam Lelylaan are shown in Figure 4.7.

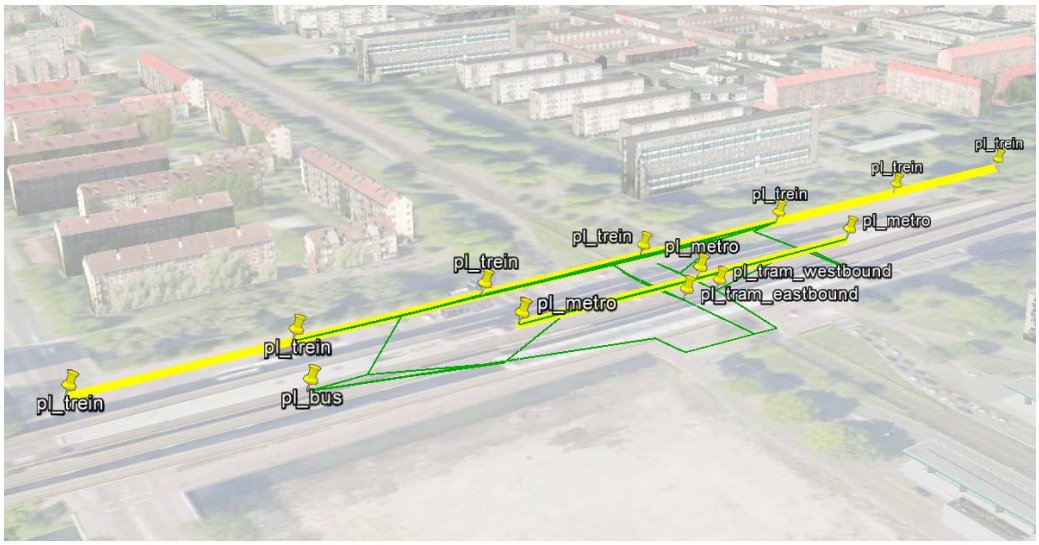


Figure 4.7: 3D model of platforms (yellow) and walk paths (green), station Amsterdam Lelylaan

Arrival and departure platforms are considered differently as follows. For an arrival platform of longer than 100 m, thus modeled by multiple points, *all* of these nodes are considered. Then the walking distance will be not a single value but a range of values, appropriately describing the different walking times different passengers experience. For departure platforms, however, only the node with the lowest walking time is calculated, again in line with the assumptions on passenger behavior, that this part of the destination platform will be used.

To measure distances between two platforms, a 3D walking mesh within the station is recorded, including the estimate of a reasonably direct walking path between all platforms. In practice, such a grid can be created with reasonable manual work based on 10-30 points for a given station. Using such a walking grid, and using the different

walking speeds v on flat surface, v_d and v_u on stairs walking down and up, respectively to convert walking distances to walking times, Dijkstra's algorithm was used to calculate shortest walking times and the related routes between platform pairs.

Therefore, the walking time t is

$$t = \sum \frac{1}{v} s + \sum \frac{1}{v_u} s_u + \sum \frac{1}{v_d} s_d, \tag{4.1}$$

where s is the distance through flat surface, s_d and s_u are the horizontal distances walking down and up stairs, respectively. To model the additional inconvenience of using stairs besides the increased walking time, it is also possible to use higher weights for walking on stairs to derive a generalized transfer walking cost function. Note that the existence different up- and downwards speeds is another reason, besides the effect of long platforms, why the walking time between two platforms is different in the two directions.

Based on the multi-point modelling of long platforms the average and maximum (worst-case) walking distance between two platforms can be calculated. Using walking speed estimates, distance can be converted to average and maximum walking time. The average distance is more useful for capacity calculations, while the worst-case for robustness analyses. These values are shown on Table 4.3, again, for the example station Amsterdam Lelylaan.

Table 4.3: Distance and walking time between platform pairs at station Amsterdam Lelylaan

From platform	To platform	Avg. dist.	Avg. time	Max. dist.	Max. time
pl_bus	pl_metro	117.9 m	87.9 s	117.9 m	87.9 s
pl_bus	pl_tram_eastbound	169.2 m	126.3 s	169.2 m	126.3 s
pl_bus	pl_tram_westbound	178.6 m	133.3 s	178.6 m	133.3 s
pl_bus	pl_trein	80.0 m	59.7 s	80.0 m	59.7 s
pl_metro	pl_bus	151.8 m	113.3 s	203.0 m	151.5 s
pl_metro	pl_tram_eastbound	87.5 m	65.3 s	118.9 m	88.7 s
pl_metro	pl_tram_westbound	96.6 m	72.1 s	141.4 m	105.5 s
pl_metro	pl_trein	138.8 m	103.6 s	154.1 m	115.0 s
pl_tram_eastbound	pl_bus	169.2 m	126.3 s	169.2 m	126.3 s
pl_tram_eastbound	pl_metro	61.8 m	46.1 s	61.8 m	46.1 s
pl_tram_eastbound	pl_tram_westbound	74.3 m	55.4 s	74.3 m	55.4 s
pl_tram_eastbound	pl_trein	66.5 m	49.7 s	66.5 m	49.7 s
pl_tram_westbound	pl_bus	178.6 m	133.3 s	178.6 m	133.3 s
pl_tram_westbound	pl_metro	75.6 m	56.4 s	75.6 m	56.4 s
pl_tram_westbound	pl_tram_eastbound	74.3 m	55.4 s	74.3 m	55.4 s
pl_tram_westbound	pl_trein	99.7 m	74.4 s	99.7 m	74.4 s
pl_trein	pl_bus	167.9 m	125.3 s	306.7 m	228.9 s
pl_trein	pl_metro	179.9 m	134.2 s	254.3 m	189.8 s
pl_trein	pl_tram_eastbound	151.0 m	112.7 s	235.5 m	175.8 s
pl_trein	pl_tram_westbound	173.7 m	129.7 s	275.5 m	205.6 s

Having calculated walking distances and times for different stations and aggregating these values, the average and maximum intermodal walking distances and times can

be compared between different stations (Table 4.4), where "intermodal" means that we only include transfers between train and non-train and vice versa in the aggregation. As expected, we can see a substantial spread in walking times between different stations as larger stations such as Amsterdam Centraal have larger distances than small stations with less services, such as Amsterdam Lelylaan. Whether these differences are proportional to the provided service level or not is the subject of analysis later in the next section. The ratio of average and maximum values can also differ between stations, with the largest differences at stations like Haarlem Spaarnwoude with long train platforms and a single access point at one end of the platform.

Table 4.4: Average and maximum intermodal walking distance and time at different stations

Station	Avg. dist.	Avg. time	Max. dist.	Max. time
Amsterdam Centraal	300.3 m	224.1 s	674.7 m	503.5 s
Amsterdam Lelylaan	132.2 m	98.6 s	306.7 m	228.9 s
Amsterdam Zuid	356.3 m	265.9 s	655.8 m	489.4 s
Amsterdam Sloterdijk	338.8 m	252.8 s	642.0 m	479.1 s
Haarlem	204.3 m	152.5 s	385.2 m	287.5 s
Haarlem Spaarnwoude	192.0 m	143.3 s	446.9 m	333.5 s
Schiphol	193.3 m	144.2 s	363.2 m	271.0 s

The geospatial model used for the above calculations can be also exploited for further pedestrian flow modelling analyses. For example, as Söngen (1979) differentiates pedestrians of a transit node into transferring, boarding, alighting passengers, and people walking through the station, these other flows other than transferring passengers can also be taken into account if appropriate data on passenger volumes for each entry and exit location are available.

Once the representative transfer walking time is known, it is possible to convert it into monetary values if the value of time for the passengers is available. The report from Significance et al. (2012) publishes surveyed value of time values separately for private car, train, and urban public transport, and for the separate passenger groups commuters, business, and other, see Table 4.5 according to Significance et al. (2012). If the composition of passenger groups at a given station is available, then the walking time values in Table 4.4 can be converted into passenger walking costs.

4.4 Application examples

4.4.1 Transfer station resistance

Earlier we saw that different stations can have substantially different typical transfer walking times. It is understandable, however, because of physical constraints, that sta-

Table 4.5: Value of time for different passenger types and selected modes (EUR/h)

	Private car	Train	Urban public transport
Commuters	9.25	11.50	7.75
Business	26.25	19.75	19.00
Other	7.50	7.00	6.00

tions with a much larger transport supply will inherently have larger transfer walking times. Therefore we propose in this section to observe transport supply and transfer walking time together: a station with relatively low transfer walking time is a good candidate for a station with synchronized timetables, see Chapter 4 for timetable synchronization. On the other hand, a station with a relatively high transfer walking time compared to other stations with similar transport supply is a good candidate for a redesign of platform layouts – this, however, is out of the scope of the remaining chapters of this thesis.

In the following we quantify the transport supply of the stations in our case study network. Some of these stations connect the train lines to a complex multi-modal public transport network, as an example see Table 4.6 listing the number of lines, hourly frequency and the commercial speed of the different lines at station Amsterdam Lelylaan and see a map of all direct services from this station on Figure 4.8. The number of train and feeder lines, hourly frequencies and commercial speed for all stations is shown in Table 4.7. The data source for these analyses is the earlier mentioned KV1 interface of BISON containing the planned timetable and the geographical coordinates of stops, available from (Stichting OpenGeo, 2013). The visualization in Figure 4.8 is prepared in Google Earth directly from the KV1 data.

Table 4.6: Number of lines, hourly frequency and commercial speed of different modes at station Amsterdam Lelylaan

Mode	No. of lines	Hourly freq.	Avg. speed (km/h)
Bus	6	40	20.0
Intercity train	2	8	58.5
Local train	1	4	54.6
Metro	1	12	33.7
Tram	2	32	16.1

Table 4.7: Number of train and feeder lines, hourly frequencies and commercial speed

Station	Train lines	freq.	speed	Feeder lines	freq.	speed
Haarlem Spaarnwoude	1	4	47.5	3	24	17.6
Amsterdam Zuid	8	25	66.5	10	88	21.4
Haarlem	5	20	58.0	15	105	20.4
Schiphol	16	49	68.3	17	100	25.0
Amsterdam Lelylaan	3	12	57.2	9	84	20.5
Amsterdam Sloterdijk	15	48	59.7	12	106	21.7
Amsterdam Centraal	21	65	62.7	47	485	20.5

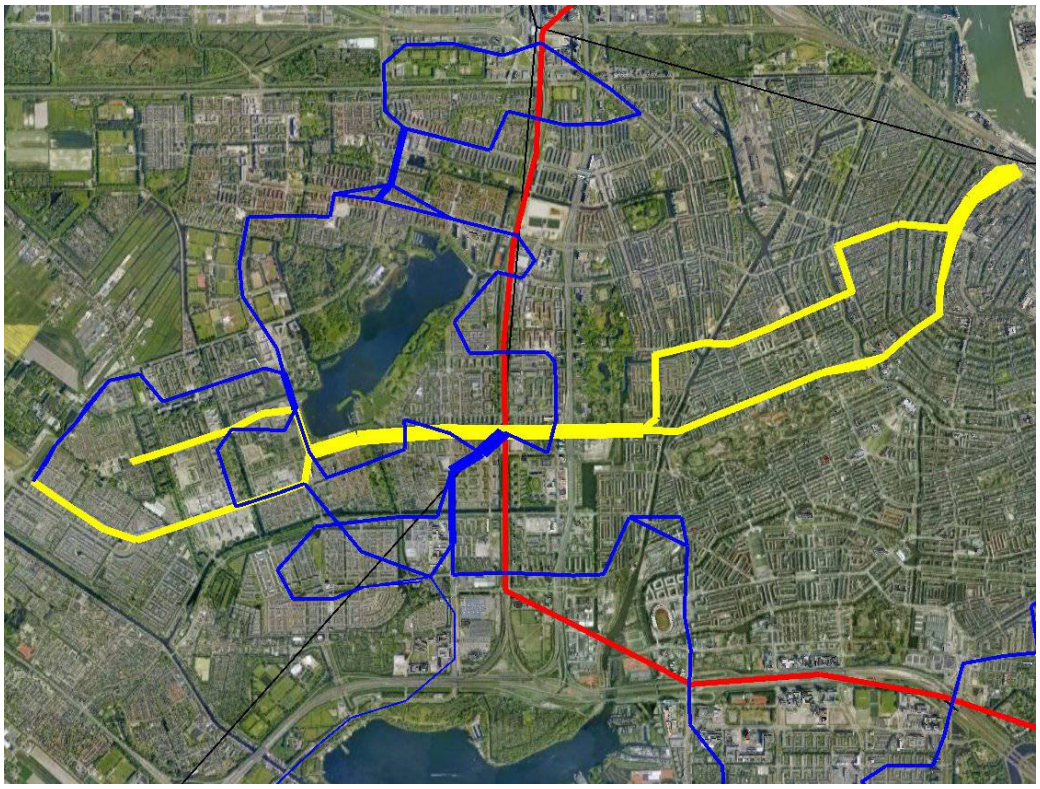
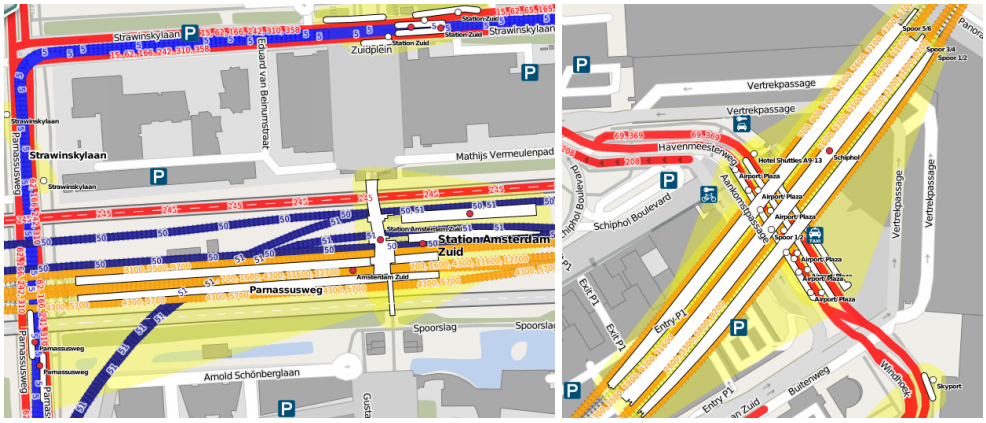


Figure 4.8: Line layout of transit lines at station Amsterdam Lelylaan. (line width is proportional to frequency, blue – bus, red – metro, yellow – tram, black – train)

To visualize the relationships between the different transfer station resistances and service levels of the stations in question, we plot the average intermodal walking time and the hourly number of train departures at the given station on Figure 4.10. Note that Leiden Centraal and Den Haag HS stations are also included here, to increase the number of data points for the regression analysis. While the number of measurements is relatively low for statistical analysis, we nevertheless added a linear regression trendline, and we can notice the slight tendency we expected that stations with more lines also have longer walking times. Indeed, while Amsterdam Centraal has the largest mean intermodal walking time, this is not unexpected given the very large number of lines



(a) (b)

Figure 4.9: Platform layout at (a) Amsterdam Zuid, a station with spread-out platforms, and (b) Schiphol, a compact station (source: opnvkarte.de)

all stopping at that station. There are, however, several stations that do substantially better or worse than other stations of similar size. In particular, the stations Haarlem, Schiphol and Amsterdam Lelylaan have very low walking times. On the other hand, Amsterdam Zuid and the new station Halfweg Zwanenburg has relatively long walking times.

We investigated the layouts of these stations and we observed the cause of these differences. Some stations are very compact, such as Lelylaan or Schiphol, where there are platforms directly above each other, taking advantage of the multi-level station buildings (see Figure 4.9(a), note the tram platforms at the bottom left and top right corners at a distance from the train and metro platforms). Some other stations are more spread out with no stops stacked on top of each other, such as Amsterdam Zuid, where tram 6 was removed from the railway station area because of the construction of a new metro line and therefore necessitates a relatively long walk (see Figure 4.9(b)). Another layout aspect that has an effect is whether there are multiple accesses to the long train platforms, including the middle, or only a single access at one end of the platform. This latter layout of Amsterdam Zuid can explain the relatively long walking times at these stations.

4.4.2 Synchronizing timetables: Urban night buses in Amsterdam

The detailed modelling of transfer walking time described above can be used not only for transfer station resistance estimation but also as essential input in better synchronizing timetables. The timetable synchronization problem is as follows. While a high-frequency network always provides low transfer waiting times, high frequencies are not justified everywhere and at every time In case of low frequency lines, if the timetables are not synchronized, this leads to long transfer waiting times and therefore an

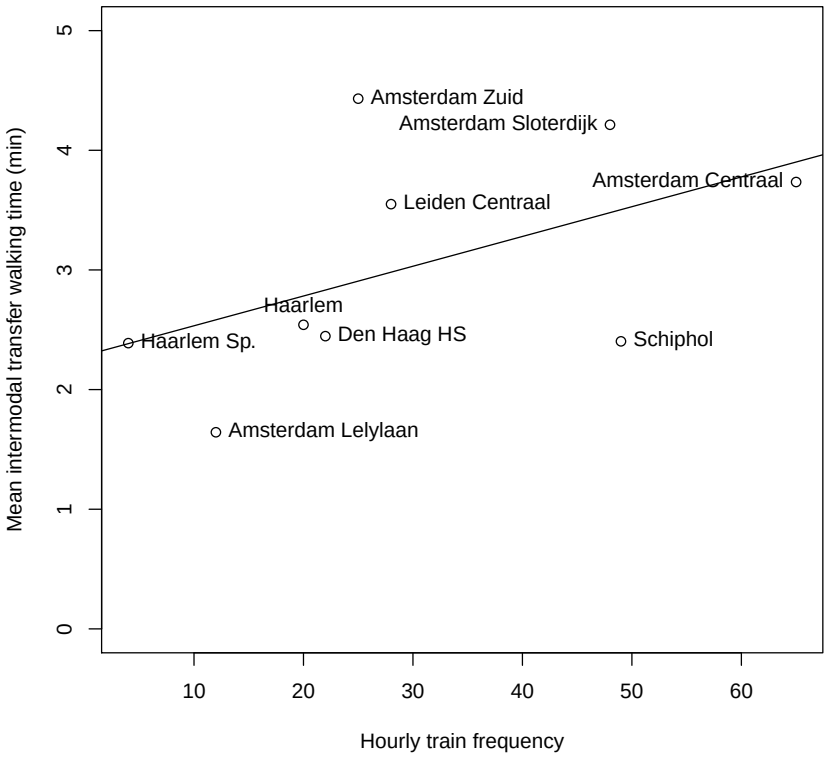


Figure 4.10: Average intermodal transfer walking times and hourly number of train departures at each station in the investigation area

Table 4.8: Urban night buses in Amsterdam (2010)

line number	route	headway on Thu–Sat (min)	headway on Sun–Wed (min)
348/392	CS–Station Sloterdijk/Schiphol Plaza	30	60
352	CS–Geuzenveld	30	60
353	CS–Osdorp de Aker	30	60
354	CS–Amstelveen	30	60
355	CS–Gein	30	60
357	CS–Bijlmermeer	30	60
358	CS–Badhoevedorp	30	60
359	CS–IJburg	30	60
360*	CS–Buikslotermeerplein	30	–
361	CS–Banne Buiksloot	30	60
363	CS–Molenwijk	30	60
392	Station Sloterdijk–Schiphol Plaza	60	60

* only in summer

unattractive public transport offer. Therefore it is desirable to *synchronize* the low-frequency connecting lines by ensuring that all relevant services arrive and depart at a transfer node around the same time to ensure low waiting times respecting the transfer walking times.

In Sparing & Goverde (2011) we presented a comparison of the night timetables of Zurich, Switzerland, and selected Dutch cities, and a timetable synchronization approach that is also presented in this section and that depends on accurate transfer walking times.

The Amsterdam urban operator *GVB* operates 12 lines at mostly 30 minute headways in the nights following Thursday, Friday and Saturday and 11 lines at 60 minute headways in the other four nights, see Table 4.8 where "CS" is the central station. Although the night trains also operate at a 60 minute headway, synchronization between the train and bus networks is missing, so attractive transfer times are not available between most train and bus directions. We show that it is possible to significantly improve train-bus transfer times at Centraal Station (the the only transfer station of the network) even with minor and cost-neutral modifications of the current bus network.

The departure, arrival times and return trip times of the bus lines are given in Table 4.9, where return trip time is the difference between the departure and arrival time of a given vehicle, including outbound and inbound trip time and layover time at the distance station but not including layover time at the central station. Note again that as headways are exactly 60 minutes, it is enough to list the minutes within the hour. The departure and arrival times of the night trains are also shown, where *Ut, Asd* and *Rtd* are abbreviations for Utrecht, Amsterdam Centraal and Rotterdam, respectively. The

Table 4.9: Bus timetable at Amsterdam CS in the nights after Thursday–Saturday (2010)

line	arrival at CS (min)	departure at CS (min)	return trip time (h:mm)
348/392	:10, :40	:00, :30	1:10/2:10
352	:14, :44	:21, :51	0:53
353	:26, :56	:14, :44	1:12
354	:11, :41	:05, :35	1:36
355	:00, :30	:23, :53	1:37
357	:07, :37	:08, :38	1:29
358	:19, :49	:29, :59	1:20
359	:02, :32	:03, :33	0:59
361	:00, :30	:13, :43	0:47
363	:23, :53	:28, :58	0:55
Asd–Ut–Asd	*:42*	*:17*	*1:25*
Asd–Rtd–Asd	*:14*	*:45*	*2:29*

trains in fact operate on the line Rotterdam–Amsterdam–Utrecht, so the train arriving at :42 from Utrecht is the same train as the one departing at :45 for Rotterdam and vice versa.

Arrival and departure times for Amsterdam and selected other cities are graphically shown in Figure 4.11. These diagrams are hourly time-space diagrams simplified so that the horizontal axis shows one hour and all trains arrive from and depart towards the bottom of the chart and buses towards the top. Our assumption is that it is possible to redesign this network to better synchronize with the train services, even keeping the same costs. For example, comparing the graph for Amsterdam to the similar graph of Zurich Stadelhofen/Bellevue in Figure 4.11, we find that the hourly Amsterdam bus arrivals and departures are more or less evenly distributed in time. In contrast, in Zurich, most of the buses arrive before and leave after the train times, providing convenient transfers. In this case, we are only comparing two isolated, albeit the most important, nodes of the two networks, which is by no means a comprehensive analysis. If we consider, however, that the Amsterdam network has no other stops with advertised connections, as we mentioned above, while the Zurich area has multiple nodes with advertised and sometimes guaranteed train-bus connections, this only increases the difference.

Redesigned timetable

As we would like to provide a superior alternative to the current bus network with no additional costs and without a redesign of the line and stop patterns, the first approach keeps all line frequencies and relative dwell times, but shifts bus departure and arrival times to provide attractive connections to and from the trains. As the network has only one transfer node, this transformation would not decrease the service quality anywhere

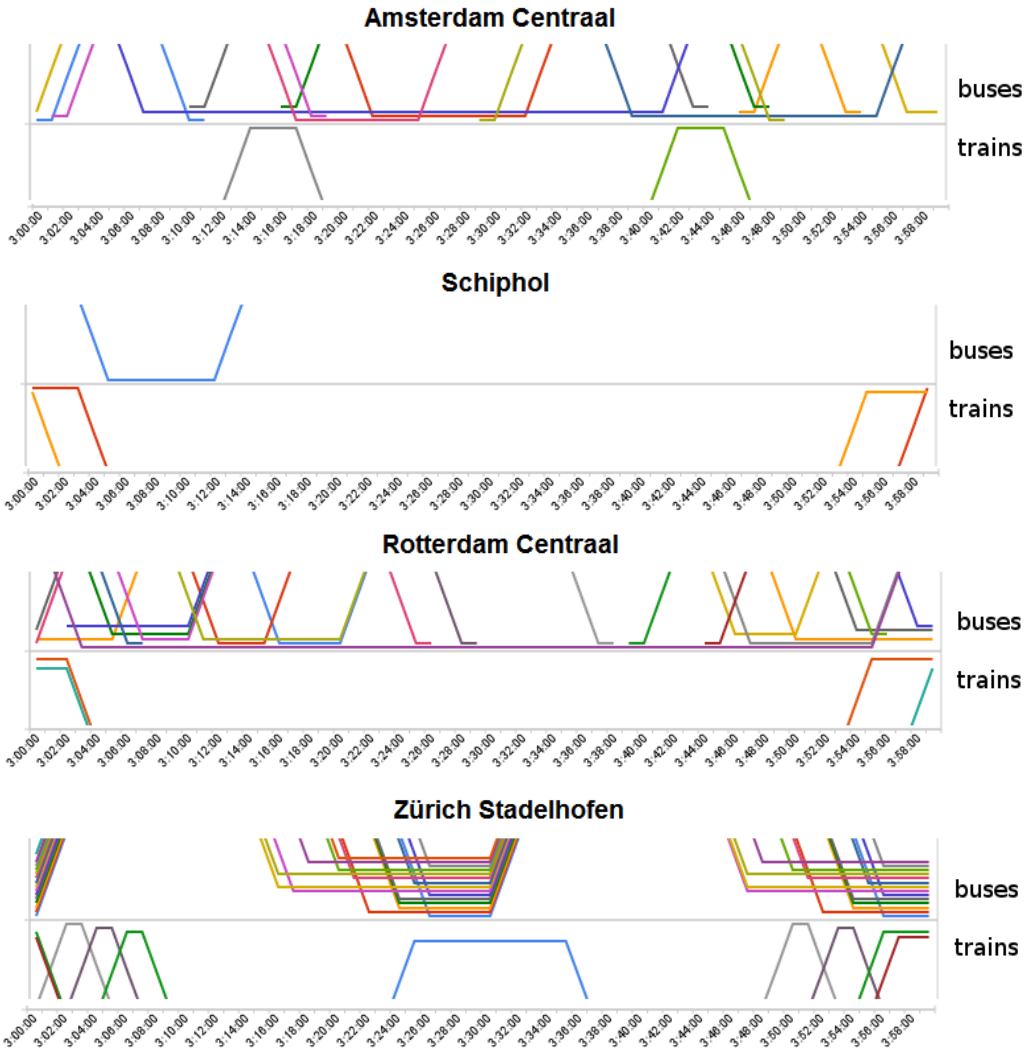


Figure 4.11: Simplified time-space diagram of night train and bus services at the Dutch stations Amsterdam Centraal, Schiphol, Rotterdam Centraal, and at Zurich Stadelhofen station (2010, source: the official online timetables of the respective public transport providers)



Figure 4.12: Bus platforms at Amsterdam Centraal used by night buses.

else in the network.

Consider train arrival and departure times A_1, A_2, D_1, D_2 for the two trains, and bus arrival and departure time variables a_i and d_i for all the bus lines, as well as a minimum bus-train transfer time for this station T_{min}, that can be obtained using the detailed transfer walking time analysis as described earlier in this chapter. Currently, the transfer time at Amsterdam Centraal to the night buses is officially 8 minutes. This is due to the fact that night buses use bus platforms far away from the train station building, while the close platforms are unused, see Figure 4.12. A relocation of night buses to the closest platforms is highly recommended to reduce transfer time, increase comfort and make bidirectional transfers possible. Taking this into account, we will consider a 5 minute minimum transfer time in our further calculation examples.

As stop patterns and the layover time at the distant terminus are to be left unchanged, we know that the layover time $d_i - a_i$ at the central station also remains constant, so $d_i = a_i + L_i$, where L_i is the layover time at CS which can be calculated from the original times, see Table 4.9. As these calculations consider a 30 minute periodic schedule, all arithmetic operations are to be interpreted in minutes, modulo 30, unless explicitly noted otherwise.

Now for a given bus B_i, the waiting times to and from train T_j are the following:

$$t_{B_i \to T_j} = D_j - (a_i + T_{min}) = D_j - a_i - T_{min} \pmod{30}, \tag{4.2}$$

$$t_{T_j \to B_i} = d_i - \left(A_j + T\right) = (a_i + L_i) - \left(A_j + T_{min}\right) \tag{4.3}$$

$$= a_i + L_i - A_j - T_{min} \pmod{30}. \tag{4.4}$$

We hereby use the lower feasible transfer time of the two buses in an hour to a certain train as the transfer time for that line to the given train. Without using estimated passenger data, it is not straightforward to prioritize between the two train directions. Although the Rotterdam direction serves a few more stations, Utrecht is also an important destination. We would like to maximize potential new passengers by offering connections with low waiting times for a given bus line. Therefore we define a utility function $\mathcal{F}(t)$ for waiting time t excluding minimum transfer walking time. Following

the similar, but stepwise constant, utility function definition of Schröder & Solchenbach (2006), we define a continuous version, the monotonous non-increasing function defined for positive waiting times t, that is concave at small transfer times and convex at large transfer times, to capture that the marginal utility for a 1 minute improvement is largest for medium waiting times: for very short waiting times a time improvement increases the risk of missing the transfer, while for very long waiting times a time unit of improvement has no material effect.

In the following, we will use the logistic function

$$\mathcal{F}(t) = \frac{1}{1+\exp(T_c - T_{min} - t)} \tag{4.5}$$

as the monotonous non-increasing function which is concave at small and convex at large values, $T_c = 15$ minutes and $T_{min} = 5$ minutes for the critical and the minimum transfer time, respectively. To maximize passenger utility, for each bus we maximize

$$\sum_j \mathcal{F}\left(t_{B\to T_j}\right) + \eta \sum_j \mathcal{F}\left(t_{T_j\to B}\right), \tag{4.6}$$

where

$$t_{B\to T_j} = D_j - a - T_{min} \pmod{30}, \tag{4.7}$$

$$t_{T_j\to B} = a + L - A_j - T_{min} \pmod{30}, \tag{4.8}$$

$0 \le a < 30$ are the variable arrival times, L are the fixed layover times for the given bus, A_j and D_j are the constant arrival and departure times for train T_j and η is a positive relative weight of the train-to-bus transfers to the bus-to-train transfers, that can be obtained from measuring the demand for both types of transfers. In the following we assume $\eta = 0.9$, i.e. prioritizing bus-to-train transfers.

The recalculated arrival and departure times are shown in Figure 4.13b. We can notice from the obtained timetable that the main improvement is an almost symmetrical timetable pattern around the train arrivals and departures. This became possible as the difference between the two hourly train times is almost 30 minutes, the same as the headway of all bus lines. Due to short turnover times at central station, two-way connections are only possible for six of the ten bus lines (eight if we accept 3 minute transfer times), but bus-to-train transfer is guaranteed to 19 directions of the 20 total (10 bus lines to two train directions). Arriving by train from the direction of Utrecht or Rotterdam, there is a bus in 3–15 minutes to all lines except two.

To conclude, while more information on the relative transport demands of the bus lines and train directions would be needed to select the optimal combination of these timetables, we have shown that it is possible to significantly increase the number of possible transfers within 15 minutes with very simple changes in the current timetable. The new timetables require slightly more bus platforms as the current one, but as we have already seen before, there is sufficient unused capacity at the Central Station.

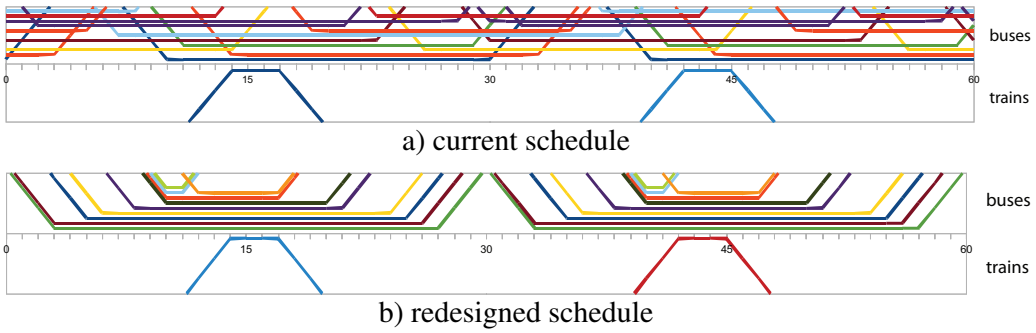


Figure 4.13: Current and redesigned hourly urban bus and train arrivals and departures at Amsterdam Centraal station

Table 4.10: Average waiting time in minutes between lines as Amsterdam Centraal

	bus-to-train	train-to-bus	bus-to-bus
current timetable	18.3	19.4	20.2
redesigned timetable	9.8	15.9	15.5

Evaluation

Table 4.10 shows the average waiting times for all intermodal and bus-to-bus connections. Furthermore, the percentage of connections within 5 to 15 minutes are shown in Table 4.11. In case of recalculating optimal departure times on a line-by-line basis, we were able to significantly increase the number of intermodal connections of attractive transfer times. Meanwhile, the offered connections between bus lines also increased.

This approach, however, contains a risk if not executed properly: namely, if the transfer walking times are underestimated and this leads to missed connections, average transfer wait times actually will be *worse* than in the unsynchronized case, i.e. almost the full headway, instead of half the headway. This is why the accurate measuring of transfer walking times, with a method such as we described earlier in this chapter, is essential to timetable synchronization.

4.4.3 Enriched travel advice for transfers

The previously investigated transfer times become particularly valuable in case of inter-operator transfers. An aspect of public transport travel that was previously invisible to the public and to each operator, but of substantial importance to the passenger, is the reliability of transfers between multiple operators, such as between a long-distance

Table 4.11: Percentage of intermodal connections between 5 and 15 minutes

	bus-to-train	train-to-bus	bus-to-bus
Amsterdam, current timetable	45%	30%	34%
Amsterdam, redesigned timetable	95%	70%	54%

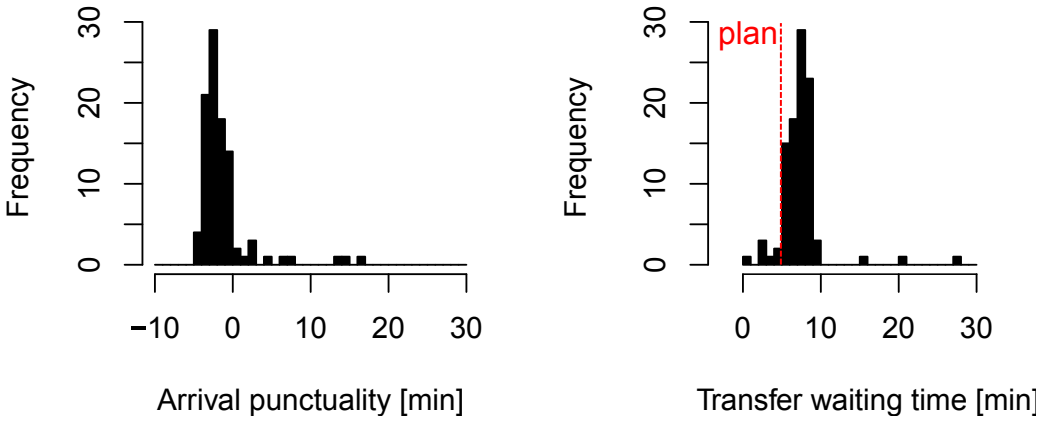


Figure 4.14: Arrival punctuality of a vehicle and transfer waiting time for the passenger at Hilversum Sportpark station, measured across all Sundays in January 2013, between 10:00 and 23:00

train and a local bus. With open data, it is possible for anyone (so also to any operator) to investigate the actual reliability of inter-operator transfers and for the operator to take steps if necessary.

Figure 4.14 shows a discrepancy between vehicle punctuality and passenger experience, for an example transfer that is scheduled to take five minutes excluding walking time. Note that on the chart *arrival punctuality* is the signed difference between planned and realized arrival times in the given time frame at the given station, and *transfer waiting time* is the difference between the departure time of the first train and the actual arrival time of the feeding bus line plus a constant transfer walking time. The planned and actual times are all retrieved from the BISON KV1/KV6 data streams described earlier. As we saw earlier, actual departure and arrival times for trains were not available at the time of our analysis, we can assume that the delays of buses and trains are correlated and that the trains have a reasonably high punctuality, making this distribution chart an appropriate estimate. It is common that a public transport timetable includes a substantial time reserve before an important stop, and therefore as the left part of the figure shows, the vehicles are consistently early at the transfer stop. However, this means that the passengers structurally have to wait much longer at the transfer stop than they can expect from the timetable. As waiting time on the platform is perceived much less comfortable than in-vehicle time (Van der Waard, 1988), this means that trips including this transfer are perceived of a less quality than expected from the timetable.

The relevance of open AVL data with regard to improving transfers is the following: open information on the reliability of inter-operator transfers makes it possible for any operator and the transport authorities to gain insight into the reliability of these transfers and take steps if necessary, such as synchronizing timetables, holding vehicles in case of minor delays and informing passengers. See Chapter 5 for identifying transfers of interest and choosing which vehicles to hold in a multi-operator setting.

As we covered in our introduction to open transit data, the original reason for opening up actual departure and arrival time information was to provide widespread real-time passenger information. Collecting historical data, on the other hand, used to be against most terms and conditions of data providers, considered sensitive business data. With the new developments, however, making transit data available also for historical analysis, it is possible to calculate distributions of running times, arrival times, and transfer waiting times, as we have seen in the previous paragraphs. Building on these distributions, and returning to passenger information, it is possible to enrich travel advices to provide estimations on the reliability of the public transport trip as follows.

Let us assume that probability distributions of departure and arrival times are available for all transit lines at all stops at all times. For example, fitting a relevant distribution on all arrivals and departures in the past 3-6 months on the given type of day (working day/Saturday/Sunday) at the given time period (morning peak/midday/afternoon peak/evening) for the given line, stop and direction can be a good estimate.

Then for a travel advice with no transfers, we can provide the passengers with the distribution of the departure and arrival times. In fact, considering what really matters to the passenger, we can summarize these two distributions in two indicators: (1) the probability that the vehicle departs too early and (2) the probability that the vehicle is on time at the destination stop (within a certain time tolerance). If necessary, we can further simplify these indicators to a ”traffic light” approach: colour coding the favourable, medium and low-reliability cases.

In case of the travel advice including transfers, we can add one single indicator for each transfer: the probability that the transfer is possible. (Note that for the passenger's perspective, independent arrival and departure time distributions are irrelevant during a transfer, except for whether the transfer is possible or not.) Recall that in case of transfers to low-frequency lines, passengers would even pay for accurate travel information (Molin et al., 2009).

See Figure 4.15 for an example for such an enriched travel advice. Here the transfer feasibility is calculated as the proportion of times when the mean actual departure time of the second service is larger than the mean arrival time of the first vehicle plus a minimum transfer walking time, in this case 3 minutes; measured on an appropriate time window, in this case one month. In case where it cannot be assumed that the arrival and departure time distributions of the two services are independent, then the calculation has to be modified as the proportion of days when the difference between the actual arrival and departure times *on the same day* is larger than the minimum transfer walking time.

4.5 Conclusions

In this chapter we focused on the detailed modelling of public transport transfers in order to support station design, timetable planning and passenger information systems,

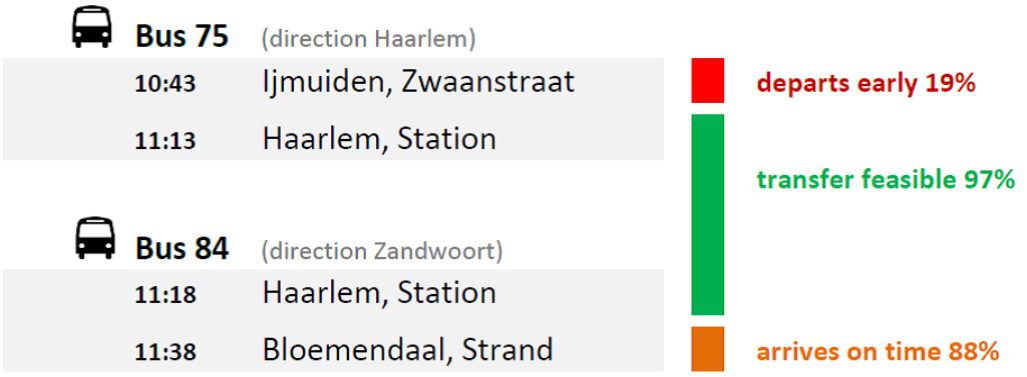


Figure 4.15: Enriched travel advice including probabilities of early departure, transfer feasibility and on time arrival

measures that can in turn improve the attractiveness of a public transport system. In particular, in Section 4.1 we described why transfers are a vital and unavoidable part of most public transport journeys and in Section 4.2 the developments in open public transport data that allow for better transfer modelling.

The contribution of this chapter is to present a methodology to use open transit data to improve public transport models, in particular to improve the accurate estimation of transfer walking times and transfer resistance based on specific station design information. These accurate transfer walking times, in turn, serve as essential input to many other public transport models: we provide the three examples of quantifying transfer station resistance with mean walking times, timetable synchronization with minimizing transfer waiting times taking into account transfer walking times, and enriched real-time passenger information with historical reliability on transfers also including historic realized arrival and departure times.

All in all, these contributions all highlight how open data enables improvements to the public transport system that were previously not possible. We noted that it is unfortunate that similar railway data is not available as open data, and that the usefulness of this research, timetable synchronization, and passenger information could be improved it train data opened up for analysis just like most bus and urban public transport data already is.

Chapter 5

Delay management in a multi-operator network[1]

5.1 Introduction

This chapter introduces a method based on *max-plus algebra* to classify potential public transport connections based on their feasibility for given initial delays, with the objective to help operational decisions. In Chapter 4 we argued for the necessity of *timed transfers* in case of low frequency lines. However these timed transfers can in fact *increase* the average transfer wait time in case of frequent delays and no holding of the departing vehicle. Therefore a strategy to hold departing vehicles in case of moderate arrival delays is necessary, also called a *guaranteed transfer*. In this chapter we describe a three-way classification of potential transfers given actual delays, based on whether the connection is in order, at risk, or "hopeless". Given such a classification, dispatchers can focus on the *significant connections at risk*. We describe the delay propagation given initial delays in a max-plus algebra formulation that makes this classification of connections possible. Further, we define the holding decision problem as an optimization model to give advice on keeping or cancelling these connections.

Public transport users prefer direct, short, high-frequency services in order to minimize travel time, waiting time and inconvenience. On the other hand, it is neither possible nor efficient to provide direct connections between all origin and destination pairs. Similarly, high-frequency services are not justified by demand on all public transport routes. Therefore it is inevitable for any public transport network to include some transfer connections between low frequency lines.

A measure to tackle the issue of long transfer waiting times is *timed transfers*: a timetable which synchronizes arrival and departure times of different lines in order to make

[1]This chapter is an edited and extended version of the journal paper: Sparing, D. & Goverde, R.M.P. (2013). Identifying effective guaranteed connections in a multimodal public transport network. *Public Transport*, 5(1-2), 79–94.

the waiting time significantly lower than the headway. However, in case of arrival delays, timed transfers can actually result in a higher overall waiting time than in the case of uncoordinated timetables. Therefore, a timed transfer can additionally be *guaranteed* when a departing vehicle waits for moderately delayed arriving vehicles. A guaranteed transfer significantly improves the travel experience of transferring passengers while causing only relatively mild delays for other passengers. Nonetheless, there is a cost to guaranteed transfers and this cost also depends on the scheduled slack times in the timetable of the departing vehicle; in other words, on how fast the departure delay can be absorbed. This *delay management problem* has already received significant attention in literature, see Section 2.3 for related works.

The steps explored in this chapter to ensure guaranteed transfers are the following: (1) identify which transit service pairs constitute a possible connection based on the proximity of stops and a maximum transfer wait time; (2) the classification of these connections given a delay scenario into maintained, cancelled connections, and connections at risk (*significant connections*), and (3) an optimization model to give a holding advice in case of the significant connections.

The contributions of this paper are the following. Using the delay propagation algorithm based on max-plus algebra we effectively reduce the problem size before optimization has to take place. We define a max-plus algebra–based reformulation of the optimization problem, taking into account passenger delays caused by both missed connections and arrival delays. Finally, we explicitly make a difference between the whole public transport network and a controllable subnetwork, to reflect practice and reduce the problem space.

The chapter is organised as the following. Section 5.2 describes the mathematical model for periodic public transport networks, the max-plus algebra representation, and the estimation of minimum process times given scheduled and realized running times. Sections 5.3–5.4 describe the identification of connections, the classification of transfers given a delay scenario, and the optimization of connections at risk, respectively. Section 5.5 describes a case study based on a real-life network and Section 5.6 concludes the chapter.

5.2 Modelling a periodic pubic transport network

5.2.1 Variables and constraints

A public transport network operating according to a periodic schedule can be modelled as a discrete-event dynamic system (Goverde, 2010) as follows. Let T be the timetable period, often $T = 60$ min, and i a certain periodic timetable event such as a departure, arrival or passage at a station. Then $x_i(k)$ is the event time of event i scheduled in period k. Hence, if $k = 0$ represents the initial period $[0, T)$, then $k = 1$ represents the

next period $[T, 2T)$, and in general $x_i(k)$ is scheduled in period $[kT, (k+1)T)$. Let D, A and P denote a departure, arrival, and passage event, respectively. Then an event i can be identified by a tuple of attributes $i = (d_i^0, S_i, L_i, T_i)$, where d_i^0 is the initial scheduled event time, S_i is the station of the event, L_i the line, and $T_i \in \{A, D, P\}$ is the event type. Note that d_i denotes either a scheduled departure, arrival or passage time. The set of all events is indicated as N.

Schedule constraints

The public transport timetable defines scheduled event times $d_i(k)$ for the departure, arrival and passing events i. While early arrival and passing events are usually allowed, early departures are typically forbidden in most public transport networks. This results in the constraints

$$x_i(k) \geq d_i(k), \quad i \in \{i \in N | T_i = D\}.$$

Recall that $\{i \in N | T_i = D\}$ is the set of departure events.

Precedence constraints

The successive events on a given vehicle journey are connected by activities with given minimum process times like minimum running, dwell, and layover times. Moreover, infrastructure restrictions – especially on railway infrastructure – may imply that events of different lines using the same piece of infrastructure can only take place with sufficient minimum headway time elapsed between them, and moreover, the timetable fixes an order of the vehicles. Such constraints can apply to following trains on the same railway track, merging or crossing trains at a railway junction, or trains of opposite direction passing at loops on single track lines.

Both journey and infrastructure constraints can be defined between any departure, arrival, or passage event as a precedence constraint of the form

$$x_i(k) \geq a_{ij}(k) + x_j\left(k - \mu_{ij}\right), \tag{5.1}$$

where $a_{ij}(k) \geq 0$ is the minimum process time from event j to i in period k, and $\mu_{ij} \in \mathbb{N}_0$ is the *period shift*, meaning that event i is scheduled μ_{ij} periods later than j, with $\mu_{ij} = 0$ thus implying that events i and j are in the same period. Based on the scheduled event times, μ_{ij} can be calculated as (Goverde, 2007)

$$\mu_{ij} = \frac{a_{ij}^0 + d_j^0 - d_i^0}{T},$$

where a_{ij}^0 is the *scheduled* process time given in full minutes and $d_i^0, d_j^0 \in [0, T)$ are the scheduled event times in the initial period. Note that $a_{ij}^0 > a_{ij}(k)$ in case of a positive time reserve for the given process. For any given network it is possible to transform the model, by adding dummy events, so that $\mu_{ij} \in \{0, 1\}$ for all (j, i) (Goverde, 2005).

5.2.2 Controllability of connections

Connection constraints describe the interdependency of an arrival event of one line and a departure event of another line because of a shared resource (vehicle or crew) or because of a guaranteed passenger transfer. Vehicle or crew constraints are hard connections that cannot be relaxed, so they are also modelled as a precedence constraint (5.1) from an arrival to a departure event. On the other hand, some passenger transfers can be seen as 'soft' connections, or *controllable*, that can be activated or not based on the actual vehicle delays.

These soft connections can be modelled using the decision variables $\delta_{ij}(k) \in \{-\infty, 0\}$, which mean that the controllable connection from event $j\left(k-\mu_{ij}\right)$ to $i(k)$ is *active* if $\delta_{ij}(k) = 0$ and *inactive* if $\delta_{ij}(k) = -\infty$. This is modelled as

$$x_i(k) \geq a_{ij}(k) + \delta_{ij}(k) + x_j\left(k-\mu_{ij}\right).$$

Note that an inactive soft connection is broken while an active connection implies a guaranteed transfer.

We assume that a list C of controllable connections is available. Note that a controllable connection is a pair (j,i) of an arrival event j and a departure event i with a given minimum transfer time $a_{ij} \geq 0$ and a decision variable $\delta_{ij} \in \{-\infty, 0\}$. Not all arrival/departure pairs of different lines at a given station are generally defined as a connection. In particular, a minimum transfer time must be available in the timetable to enable the transfer in punctual operations and moreover a maximum waiting time applies up to which we still speak of a (controllable) connection. If a list of possible connections is not available then we may generate a list from the timetable. Assume that the minimum transfer time $t_{\min}$ for a given arrival and departure platform pair is given, and that the maximum acceptable waiting time w_{max} is defined. Then, the set of possible soft connections is given as

$$C = \{(j,i) \in E | T_j = A, T_i = D, S_j = S_i, L_j \neq L_i, (d_i^0 - d_j^0 - t_{\min}) \mod T < w_{\max}\}.$$

In case of such a generated connection list, however, extra care has to be taken to filter out potential connections that are not attractive to the passengers. This is possible by taking into account passenger flows.

5.2.3 Max-plus algebra representation

Discrete event systems as described above can be formulated and analysed effectively in max-plus algebra (Heidergott et al., 2005). In particular, max-plus models have been applied successfully to the evaluation of periodic railway timetables (Braker, 1993; Goverde, 2007).

Max-plus algebra is an algebraic structure defined on $\mathbb{R}_{\max} := \mathbb{R} \cup \{-\infty\}$ with the max operator and addition instead of addition and multiplication in conventional algebra,

respectively. This structure satisfies the properties of an *idempotent semiring*, see e.g. Heidergott et al. (2005). That is, max-plus addition and multiplication are defined for $a, b \in \mathbb{R}_{\max}$ as

$$a \oplus b := \max(a, b) \text{ and } a \otimes b := a + b.$$

Matrix addition $\oplus$ and matrix multiplication $\otimes$ are defined analogous to conventional algebra, i.e., for matrices $A, B \in \mathbb{R}_{\max}^{n \times n}$

$$\begin{aligned} (A \oplus B)_{ij} &:= a_{ij} \oplus b_{ij} = \max(a_{ij}, b_{ij}) \\ (A \otimes B)_{ij} &:= \bigoplus_{l=1}^{n} (a_{il} \otimes b_{lj}) = \max_{l=1,\ldots,n} (a_{il} + b_{lj}). \end{aligned}$$

The precedence and controllable connection constraints above can be rewritten using max-plus algebra notation as follows. Let Π denote the set of all precedence constraints, including hard connections, and C the set of all controllable connections. Then the model becomes in max-plus algebra

$$x_i(k) = \bigoplus_{(j,i) \in \Pi} \left(a_{ij} \otimes x_j(k - \mu_{ij}) \right) \oplus \bigoplus_{(l,i) \in C} (a_{il} \otimes \delta_{il}(k) \otimes x_l(k - \mu_{il})) \oplus d_i(k).$$

This model can be written in vector notation as follows. Collect all event times $x_i(k)$ in the event time vector $x(k) = (x_1(k), \ldots, x_n(k))' \in \mathbb{R}^n$, and define the (uncontrollable) matrices $A_0, A_1 \in \mathbb{R}_{\max}^{n \times n}$ as

$$(A_l(k))_{ij} = \begin{cases} a_{ij}(k) & \text{if } (j,i) \in \Pi \text{ and } l = \mu_{ij} \\ -\infty & \text{otherwise,} \end{cases}$$

and the controllable matrices $B_0, B_1 \in \mathbb{R}_{\max}^{n \times n}$ as

$$(B_l(u(k),k))_{ij} = \begin{cases} a_{ij}(k) + \delta_{ij}(k) & \text{if } (j,i) \in C \text{ and } l = \mu_{ij} \\ -\infty & \text{otherwise.} \end{cases}$$

The vector $u(k) \in U$ is the control vector in period k which determines which of the connections in period k are canceled and which are not, i.e., $u(k)$ sets $\delta_{ij}(k) = -\infty$ for all connections $(i, j) \in C$ that are broken and $\delta_{ij}(k) = 0$ otherwise. U is as set of control vectors specifying all feasible combinations of δ_{ij}. The model can be written in matrix notation as

$$x(k) = A_0(k)x(k) \oplus A_1(k)x(k-1) \oplus B_0(u(k),k)x(k) \oplus B_1(u(k),k)x(k-1) \oplus d(k). \tag{5.2}$$

Alternatively, this can be expressed as

$$x(k) = [A_0(k) \oplus B_0(u(k),k)]x(k) \oplus [A_1(k) \oplus B_1(u(k),k)]x(k-1) \oplus d(k).$$

The model (5.2) can now be used to calculate the delay propagation from any given time point for any given past delays (Goverde, 2010). Without loss of generality, we

may assume a given initial condition $x(0) = x_0$, where $x_0 = d(0) + z(0)$ with $z(0) \geq 0$ the vector of delays in the initial period. From this given initial condition, the event time estimates $x(k)$ can now be calculated for any $k \geq 1$ up to a suitable time horizon $K \geq 1$. In particular, a lower bound of all (secondary) delays can be obtained by braking all controllable connections, i.e., $\delta_{ij} = -\infty$ for all $(j,i) \in C$, or equivalently, $B_0(u(k),k)$ and $B_1(u(k),k)$ having all entries equal to $-\infty$. Then (5.2) reduces to

$$x(k) = A_0(k)x(k) \oplus A_1(k)x(k-1) \oplus d(k), \quad x(0) = x_0. \tag{5.3}$$

By guaranteeing any additional connection $(j,i) \in C$ the total vehicle delay may increase, but the passenger waiting time might decrease depending on the transferring and onboard passenger volumes and the amount of delay recovery.

5.2.4 Passenger delay estimations

Modelling or forecasting passenger flows for disturbed operations is difficult, especially taking into account the mode choice which is also affected, see Van Eck (2011) for a multimodal traffic model taking into account simultaneous mode choice. On the other hand, simply counting the number of maintained and broken connections is not sufficient as an objective since the demand and attractiveness of different connections can vary significantly. Therefore, at least a rudimentary model for passenger demand is required.

Assume that an estimate of the number of boarding and alighting passengers is known for every stop of every service. Denote by $w_i(k)$ the number of boarding or alighting passengers in period k, depending on the event type T_i, i.e., $w_i(k)$ is the number of alighting passengers for $\{i : T_i = A\}$ and the number of boarding passengers for $\{i : T_i = D\}$, while by convention $w_i(k) = 0$ for passage events $\{i : T_i = P\}$. Furthermore, for each controllable connection $(j,i) \in C$ let $w_{ij}(k)$ be the number of transferring passengers from arrival event j to departure event i in period k. These passenger count estimates can be provided by previous passenger counts or traffic flow models. See Hilderink et al. (2010) for a traffic flow model that models public transport passenger flows at this detail.

If detailed passenger counts are not available or not available over different periods over a day, then we may resort to fixed weights w_i and w_{ij} measuring the relative importance of stops and connections. Such passenger counts allow for a reasonable comparison of total passenger delays in case of maintained or broken connections. In case of a maintained connection with a delayed departure, the total arrival passenger delay on the given line is the sum of vehicle arrival delays weighted by the number of alighting passengers per stop. The total arrival delay in the network using an appropriate time horizon K is given by

$$\sum_{k=1}^{K} \sum_{j \in \{j:T_j=A\}} w_j(k) \cdot \max\left(0, x_j(k) - d_j(k)\right),$$

where the maximum of delay and zero is required so that early arrivals do not count. Moreover, the delay due to the missed connections can be estimated as

$$\sum_{k=1}^{K} \sum_{(j,i)\in M(k)} w_{ij}(k) c_j,$$

with c_j the scheduled headway of line i and $M(k)$ the set of missed connections in period k defined as

$$M(k) = \{(j,i) \in C : x_i(k) - x_j(k-\mu_{ij}) - a_{ij} < 0\}. \tag{5.4}$$

These delays can be compared for different connection controls $u(k)$.

Note that this model relies on at least moderately accurate passenger weights, as any two vehicle journeys with consecutive arrival and departure can constitute a defined connection, as long as the passenger transfer weight is significant. In case of lines where such a connection is deemed useless for real trips (e.g. because of being a large detour), this has to be reflected in the passenger transfer numbers. The accuracy of the ratio of arrival and transfer counts also are important for candidate connections where the volume of transferring passengers is substantially larger than of in-vehicle passengers, or vice versa. On the other hand, when the delays caused by a connection maintained or cancelled become comparable, moderate errors in passenger counts has a lower effect, as in such cases neither of the decisions can cause a big imbalance in the total delay.

5.2.5 Estimating minimum process times

When calculating estimated arrival times in the future, we are using a delay propagation algorithm assuming known minimum process times, as described in Section 5.2.1. While for railway systems and rapid transit systems for segregated right-of-way the minimum running times can be estimated using simulations based on acceleration and braking characteristics, for bus and tram lines in mixed traffic such calculations might not be feasible or accurate. Therefore for mixed traffic lines the following possibilities remain:

- We can use a fixed percentage of the scheduled running time, such as 85%: this is especially promising if we can assume that the scheduled running times were obtained by extending the measured running times with a proportional time reserve, and we can use the inverse of this operation.
- Another possibility is to use recorded vehicle journeys from automatic vehicle location (AVL) or train describer data, if available.

In the following we describe the approach using AVL data. If realized vehicle times are available from an AVL system, it is possible to measure minimum running times accurately. At first, a substantial set of AVL measurements can reveal systematic deviations

from the timetable. See for example Figure 4.3(a) in the previous chapter displaying all bus trips (except severe outliers) on a single line in a one month period, with the 15th, 50th and 85th percentile in black: the buses are systematically early at their last stop. In an example related to an intermodal transfer, Figure 4.14 in the previous chapter displays the distribution of the arrival times of a bus at a transfer station (left), and the resulting distribution of the transfer waiting times when changing from the bus to a train. Note that as the departing train has a headway of 30 minutes, the transfer waiting time is within 0 and 30 minutes as well. In this case, the bus is systematically arriving early at the stop, and because both services are low frequency (the headways are larger than typical delays, this results in a larger transfer wait time than according to the timetable.

To estimate minimum process times, we can use one of the following methods: (1) an average or median of running times in a time period with low traffic and low demand, such as on Sunday, or (2) a low percentile of the running times, such as the 15th percentile.

In our calculations in the following we used two methods: applying 85% of the scheduled running time, and using the 15th percentile of the realized running times over one month of data (January 2013).

5.3 Defining connections in a multi-operator network

In a single operator network, defining connections can be trivial: approximately all combinations of different lines stopping at the same stop. In case of multiple operators, however, it can become more difficult as the stops can be further away and the size of the network and frequent changes to the timetable might necessitate an automatic approach. In our case we used very simple criteria to define connections: all combinations where the stops are at most 250 meters from each other and given the timetable the planned transfer waiting time is at most 6 minutes.

Given the identified connections above, for a certain delay scenario and delay propagation model estimating future delays, we can classify connections into 3 classes:

1. connections maintained without any dispatching action,
2. significant connections: connections at risk requiring the delaying of the departing vehicle, and
3. connections cancelled, where delaying the departing vehicle is undesirable.

See Figure 5.1 for an illustration: no or minor arrival delay belongs to case (1), while medium and large delay corresponds to cases (2) and (3), respectively. The decision boundary between (1) and (2) is simply when the estimated transfer waiting time is

zero, while the boundary between (2) and (3) depends on the maximum acceptable delaying of the second vehicle. In the following calculations we assume that this maximum delay is 5 minutes.

5.4 Holding advice for significant connections

5.4.1 The optimization model

The goal of the optimization model is to find the optimal decision parameters $\delta_{ij}(k)$ for each controllable connection (j,i) and period $k \geq 1$ so that the estimated total passenger delay caused by either maintaining or braking a controllable connection is minimal. Input parameters are the scheduled line headways c_j, the passenger weight $w_i(k)$ for each arrival event i, and the transfer weight $w_{ij}(k)$ for each connection $(j,i) \in C$, as well as the initial condition including delays x_0, and a time horizon $K \in \mathbb{N}$. Then the optimization problem becomes

$$\min \sum_{k=1}^{K} \left(\sum_{j \in \{j:T_j=A\}} w_j(k) \cdot \max\left(0, x_j(k) - d_j(k)\right) + \sum_{(j,i) \in M(k)} w_{ij}(k) c_j \right),$$

subject to

$$\begin{array}{ll}
x_i(k) \geq d_i(k) & i \in \{i : T_i = D\}, k = 1 \ldots, K \\
x_i(k) \geq a_{ij}(k) + x_j(k - \mu_{ij}) & (j,i) \in \Pi, k = 1 \ldots, K \\
x_i(k) \geq a_{ij}(k) + \delta_{ij}(k) + x_j(k - \mu_{ij}) & (j,i) \in C, k = 1 \ldots, K \\
\delta_{ij}(k) \in \{-\infty, 0\} & (j,i) \in C, k = 1 \ldots, K \\
x(0) = x_0, &
\end{array}$$

where $M(k)$ is the set of missed connections as defined in (5.4).

The problem is solved using the max-plus interpretation where the delay propagation can be computed very quickly. The constraint set of the above optimization problem can therefore be formulated equivalently as

$$\begin{aligned}
x(k) = {} & A_0(k)x(k) \oplus A_1(k)x(k-1) \oplus \ldots \\
& B_0(u(k),k)x(k) \oplus B_1(u(k),k)x(k-1) \oplus d(k) \qquad k = 1, \ldots, K \\
x(0) = {} & x_0,
\end{aligned}$$

where $u(k) \in \{-\infty, 0\}^{|C|}$ is a $|C|$-dimensional control vector encoding the connection decisions. Note that $|C|$ denotes the size of C, i.e., the number of controllable connections.

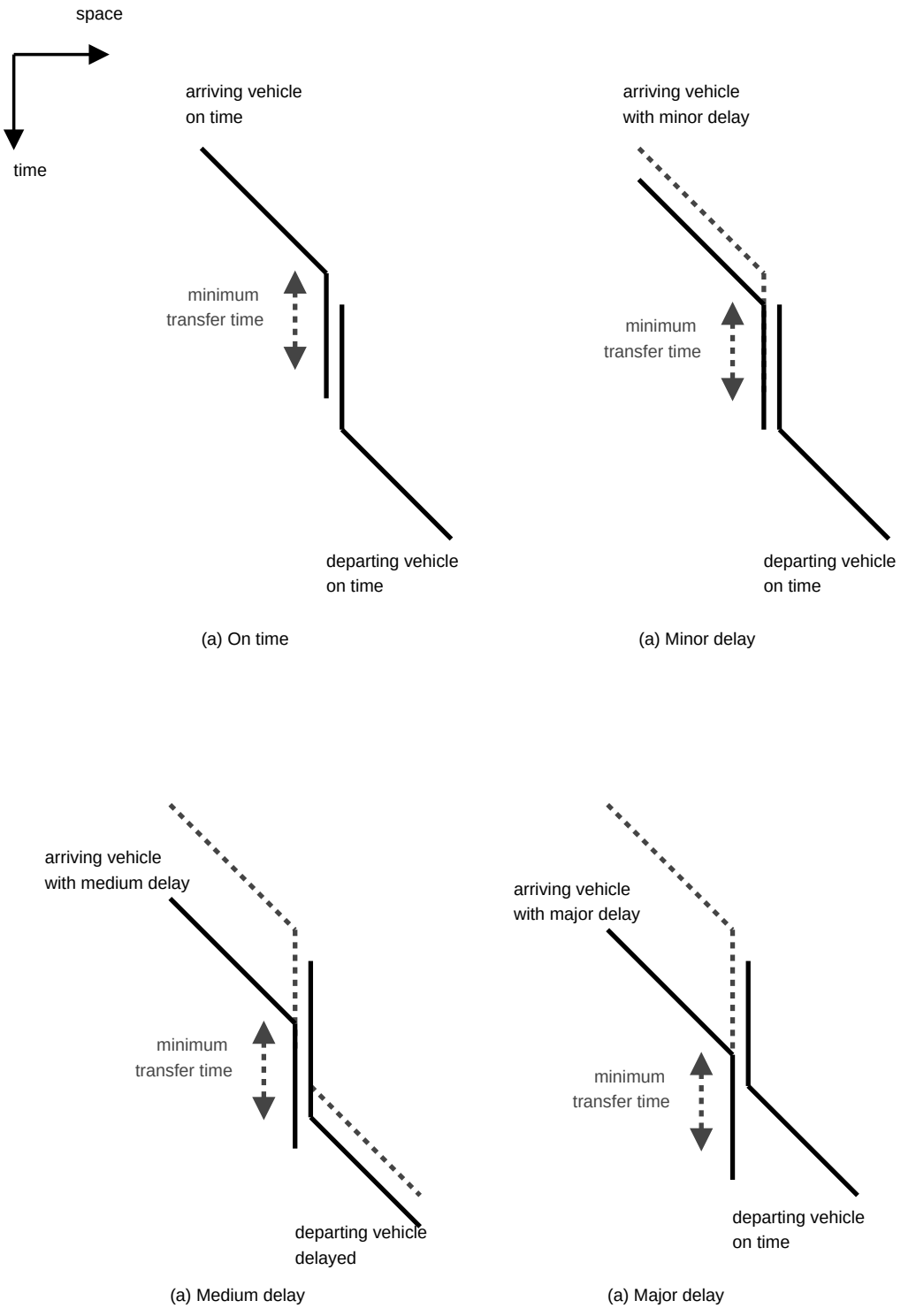


Figure 5.1: Connection classes depending on different amounts of arrival delay.

5.4.2 Model limitations

The model described above is based on the assumptions that the public transport system at question is a network of individual low-frequency lines and the disturbances are moderate such that they do not yet lead to passenger rerouting. The limitations of this model can also be derived from these assumptions as follows. (1) The model is only applicable for low-frequency lines where the headways are larger than convenient waiting times; in case of high-frequency lines, timed transfers become less important, and the assumption that the passenger delay in case of a missed transfer is the headway also doesn't hold. (2) We assume that there are no two lines that run together on a shared section, this would again invalidate the assumption on the missed transfer delay. In case there are such lines, they can be decomposed into a single line on the shared section and several lines connecting to this line outside the shared section. (3) We assume that there is no re-routing of passengers, as the disturbances are moderate and the network is too sparse to offer many alternative routes. The model might still be useful though if only a minority of the passengers are re-routed, but this has to be validated with a traffic model for a given network.

5.4.3 Solution approach

The solution to the optimization problem may be found using a standard mixed-integer programming solver. For problems with many controllable connections we can explore the (max-plus) structure of the problem and use a branch-and-bound procedure to find the optimal combination of decision variables. In both cases, reducing the dimension of the problem will help solving it more quickly. In this section we give two preprocessing steps that can be used for this aim.

Identifying significant connections

The decision variables are the 'binary' variables $\delta_{ij}(k)$ corresponding to the controllable connections. In a preprocessing step we may reduce this set to only those connections that are significant to the optimization problem. For this we look at the output of the initial delay propagation for the given initial delay scenario and all connections broken. Then the controllable connections can be categorized depending on the initial delay scenario and a maximum allowable departure delay $d_{\max}$ as follows (see Section 5.3):

- Automatically maintained connections: those arrival-departure pairs where a transfer is possible despite the delays, without further delay of the departure vehicle.
- Significant connections: Those connections that are maintained by delaying the departing vehicle by at most $d_{\max}$.

- Automatically broken connections: Those connections that must be cancelled otherwise the departing vehicle will be delayed by more than $d_{\max} \geq 0$.

The set of significant connections C_s is thus defined for each period $k = 1, \ldots, K$ (taking into account the period of the departure event, if different from the arrival event) as

$$C_s(k) = \left\{ (j,i) \in C : 0 \leq x_i(k) - a_{ij}(k) - x_j(k - \mu_{ij}) \leq d_{\max} \right\},$$

where the event times $x(k)$ are computed using the delay propagation model (5.3) with all connections broken. Note that C_s contains the significant controllable connections for each period separately. The parameter $d_{\max}$ is called the synchronization margin in Knoppers & Muller (1995). In general, these synchronization control margins may be different for different connections (Goverde, 1998). However, in our model it is just a fixed upper bound on the allowed delay, while the outcome of the optimization will determine how long a connecting vehicle will wait for a delayed feeder vehicle. For instance, a departing vehicle with ample running time supplement and buffer time ahead may wait longer than a departing vehicle with a tight schedule or a capacity bottleneck nearby.

Once the set of significant connections are identified based on an initial delay scenario and given the parameter $d_{\max}$, the scope of the delay management problem is significantly reduced from a large set of connections to a small subset of connections at risk. This shortlist can then be used either directly by a dispatcher, or can be further processed to provide recommendations to maintain or cancel a connection by holding the second vehicle or not, such as using max-plus based control as described by Heidergott & De Vries (2001), or by solving the MILP problem described in Section 5.4.1.

Hence, a shortlist of significant connections can be obtained rapidly from combining the delay propagation algorithm with a (generated) controllable connection list and using appropriate thresholds. In real-time online usage, the significant connections can be retrieved given the actual delays, thus helping drivers and controllers focus on those transfers where decisions have to be made. In off-line timetable evaluation usage, the significant connections, for given realistic delay scenarios, represent the most vulnerable transfers. A timetable planner may then try to modify the timetable to improve the reliability of these transfers.

The controllable subnetwork

Another way to decrease the problem size is a partitioning of the network in a controllable and an uncontrollable subnetwork. The controllable subnetwork is a set of lines on which dispatching actions can be made, such as delaying a departure. The uncontrollable subnetwork consists of the remaining lines on which the expected user of the decision support model has no influence. Typically, no vehicle or crew member operates on both subnetworks, while passenger transfers can be defined between the two subnetworks.

As an illustrative example, the controllable subnetwork can be a bus network controlled by a bus dispatcher, while the uncontrollable subnetwork is a train network. In this

case, it is realistic to assume that the dispatchers of the bus company have influence on the bus network but not on the train network.

The advantage of this separation is that the optimization problem only has to take the controllable subnetwork into account. Only the initial delay propagation computation with all controllable connections broken must be run for the whole network to find the delays at the boundaries of the controllable network. In the next step, the effect of the different combinations of broken or maintained controllable connections can be evaluated much faster by only calculating delays on the smaller, controllable subnetwork.

5.5 Case Study: The Green Heart

5.5.1 The example network

The example network consists of a regional bus network of 4 lines in the West of the Netherlands as listed in Table 5.1 and the Dutch national railway network timetable which have several stations in common, see Figure 5.2. In this case study, the bus network is the controllable subnetwork, while the train network is the external, non-controllable subnetwork. The timetables and minimum process times are obtained from different sources and are based on the years 2007 and 2013 in case of the train and the bus network, respectively, as the most recent available sources at the time of the analysis. In particular, the minimum process times for the train network were obtained from the data used by the DONS system of the Dutch rail infrastructure operator ProRail, after processing by the timetable stability analysis tool PETER (Goverde & Odijk, 2002). The minimum process times of bus lines are calculated according to Section 5.2.5 based on publicly available data through the Dutch GOVI system (*Grenzeloze Openbaar Vervoer Informatie*, Public Transport Information Without Borders) (van Oort et al., 2015).

As accurate measured or modelled passenger count data was not available, example passenger flow values were generated for testing reasons as the following. The number of alighting passengers for an arrival event i is defined as $w_i = r_1^2$ and the number of connecting passengers as $w_{ij} = ar_2^3$, where $0 < r_1, r_2 < 1$ are uniformly distributed independent random variables. Note that the absolute value of these weights do not matter, only their ratio. The convex quadratic and cubic functions of uniformly random variables were used to reflect the high inhomogeneity in demand in real networks due to geography and network structure. For numerical calculations, weights with different $a > 0$ values have been used to represent relative numbers of transferring and in-vehicle passengers and to investigate the applicability of the model. Therefore, the goal of the case study network is not to reflect a real life situation on a given day perfectly, but to provide an example of realistic size and complexity, see Table 5.2.



Figure 5.2: Network of four bus lines and connecting train lines in the Netherlands

Table 5.1: List of bus lines

line	from	to	Headway
370	Schiphol	Alphen aan den Rijn	15 min
380	Alphen aan den Rijn	Den Haag Centraal	60 min
382	Boskoop	Den Haag Centraal	30 min
383	Capelle a/d IJssel	Den Haag Centraal	60 min

Table 5.2: Characteristics of the example *controllable subnetwork*

Vehicle journeys	16/h
Stops	57
Transfer stations	6
Events	440
Controllable connections	95
Model graph size	
Vertices	11487
Edges	97120
Vertices, controllable subnetwork	424
Edges, controllable subnetwork	440

Table 5.3: Transfer stations

station	transfer time (min)
Boskoop	4
Alphen aan den Rijn	5
Schiphol	4
Den Haag Centraal	5
Nieuwekerk a/d IJssel	5

Initial delay scenario

To investigate a concrete network status including delay, assume that the delays are known over a full timetable period $k = 0$. Without loss of generality, we assume that the moment for the following calculations is the end of period $k = 0$: initial delays within period $k = 0$ are known, while consecutive delays in the successive periods $k \geq 1$ can be calculated using the max-plus model, taking into account different dispatching actions. For our calculation purposes, 100 different random initial delay vectors are calculated, based on a delay probability $p = 10\%$ for each departure event and a uniform delay distribution between 0 and 10 minutes.

5.5.2 Candidate transfers

We assume that initially no guaranteed connections are offered between the bus and the train network, i.e., no transfer constraints are defined. We therefore define a list of candidate transfers between arriving trains and departing buses.

As a first step, we expect that there is a list of stop pairs available between bus stops and train stations between which a transfer is physically possible in reasonable time, as well as the estimated minimum transfer time t_l required at station S_l between alighting from a train to boarding a bus (Table 5.3).

Furthermore, we define $w_{\max} = 6$ min as the maximum acceptable waiting time for a connection. Therefore the connection list C includes all (j,i) arrival-departure pairs at transfer station S_l, with minimum transfer time t_l, where $t_l \leq d_i - d_j \mod 60 \leq t_l + w_{max}$.

5.5.3 Significant connections

The set of significant connections C_s is determined using $d_{\max} = 4$ and the delay scenarios generated as described in Section 5.5.1. Figure 5.3 shows the classification of connections in the first period for the first 20 delay scenarios. The results show that although there are 95 controllable connections in the model, only about $10-30\%$ of them are significant and require more attention.

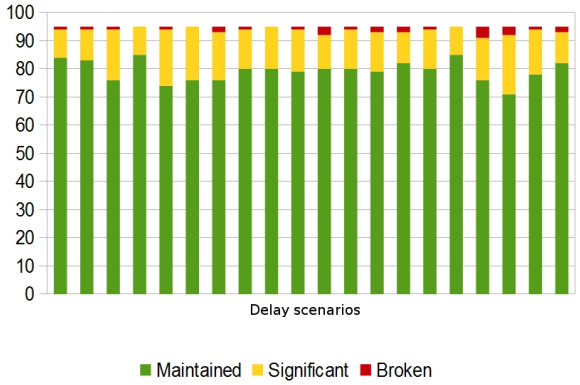


Figure 5.3: Connection classifications for the 20 random initial delay scenarios

Another insight that can be gained from such results, assuming that not random, but realized or simulated delay distributions are used, is a statistic of how often a given connection becomes significant or broken beyond the maximum departure delay limit. Such results, together with information on the demand, can help timetable planners and operating staff to identify vulnerable transfers requiring more attention.

5.5.4 Optimal connection control

For a given delay scenario and set of significant connections, the solution of the optimization problem (Section 5.4.1) returns the combination of maintained and cancelled transfers with the lowest total passenger delay. In the MATLAB environment on a PC with 12 GB RAM and a quad-core 3.47 GHz CPU, this calculation takes around 2 minutes for the problem size described in Table 5.2. Using generated example passenger count data, Table 5.4 shows the total passenger arrival delay and the total passenger transfer delay in case of all significant connections cancelled or maintained, for three different sets of passenger weights. Furthermore, Figure 5.4 and Figure 5.5 show an example of the vehicle and passenger delays, respectively, in case of all connections cancelled. It is visible that even in case of moderate vehicle delays, passenger delays can be significant because of missed transfers.

In case of all connections maintained, the total passenger transfer delay is significantly smaller, as the passengers of the maintained connections do not experience transfer delay anymore. This comes at the relatively lower cost of increased arrival delays. Note that there can still be arrival delays for all significant connections cancelled, because of the initial delays within the controllable network; while there is some transfer delay even for all significant connections maintained, due to the automatically broken connections.

The optimal control strategy consists of a mixture of some significant connections maintained and the others cancelled. The third row for each set of passenger weights

Table 5.4: Results for a single delay scenario and different passenger weights

pass. weights	control strategy	arrival delay (hh:mm:ss)	transfer delay (hh:mm:ss)	total delay (hh:mm:ss)
$a = 0.5$	all connections cancelled	49:54:20	51:30:00	101:24:20
	all connections maintained	73:18:40	00:15:00	73:33:40
	optimal control	62:30:34	06:45:00	69:15:34
	improvement to all cancelled			31.7%
	improvement to all maintained			5.9%
$a = 1$	all connections cancelled	53:32:23	58:30:00	112:02:23
	all connections maintained	70:09:43	00:00:00	70:09:43
	optimal control	63:04:15	01:45:00	64:49:15
	improvement to all cancelled			42.1%
	improvement to all maintained			7.6%
$a = 2$	all connections cancelled	51:14:13	83:15:00	134:29:13
	all connections maintained	65:36:03	06:00:00	71:36:03
	optimal control	58:07:51	07:15:00	65:22:51
	improvement to all cancelled			51.4%
	improvement to all maintained			8.7%

Table 5.5: Optimization result statistics for 100 delay scenarios, $a = 2$

	min.	mean	max.	std. dev.
total delay in optimal control	35:55:13	112:17:54	68:01:15	16:02:24
improvement to all cancelled	18.7%	58.6%	79.7%	11.3 pp
improvement to all maintained	0.0%	3.5%	20.6%	3.7 pp

pp = percentage points

in Table 5.4 shows this control strategy, with the lowest total delay. Figure 5.6 shows passenger delay for optimal control, where the extent of total passenger delay is reduced as compared to the first case without guaranteed connections (Figure 5.5). Finally, Table 5.5 shows how much the total delay is improved using 100 different delay scenarios on a single passenger demand weight set.

5.6 Conclusions

In this chapter we presented a max-plus algebra-based approach to tackle the delay management problem. We used a fast delay propagation calculation to enable quick identification of connections at risk. We introduced a max-plus algebra reformulation of delay management as an optimization problem. Finally, we exploited the fact that often any public transport operator is only controlling a subset of the network

its passengers are using, therefore the scope of the optimization can be limited to the controllable part of the network.

A real-life network example was used to present a case study that proves the applicability of our approach on realistic problem sizes. The model can be used both as a decision support tool for public transport dispatchers to identify connections at risk, or even as an optimization approach to provide advice on holding or cancelling connections. Future work should focus on better including estimated passenger volumes in the optimization function. In particular, as smart card data is becoming available, this data source can strongly enrich the effectiveness of the delay management optimization.

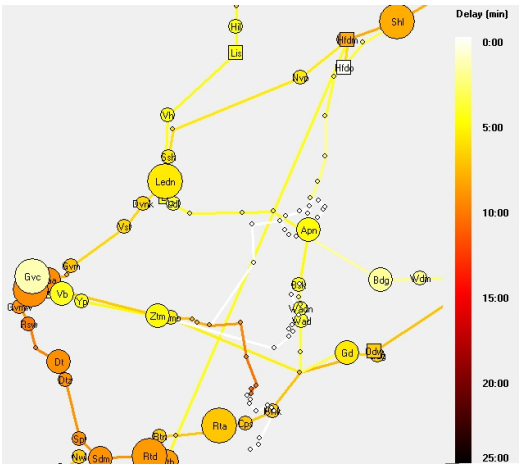


Figure 5.4: Vehicle delays if all connections cancelled

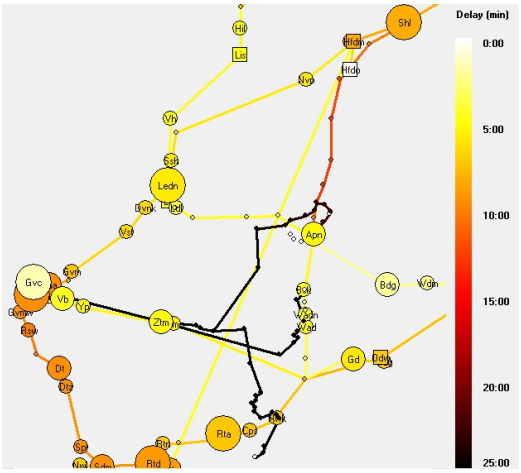


Figure 5.5: Passenger delays if all connections cancelled

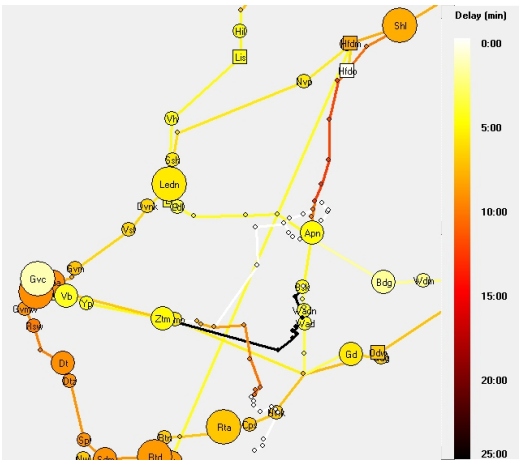


Figure 5.6: Passenger delays in case of optimal control

Chapter 6

Conclusions

6.1 Summary of the main contributions

This research focused on contributing to improving the reliability and therefore attractiveness of public transport systems by focusing on two key aspects: the stability of the high capacity utilization railway timetable, and the accurate measurement and modelling of transfers, and synchronization of lines, especially in the case of intermodal or inter-operator transfers. Mathematical optimization is used in the tackling of both problems: in the first case, we focused on the challenge of defining the appropriate objective function to describe railway timetable stability, in the second, we focused on the more accurate measurement of optimization parameters and solution space reduction in order to enable accurate transfer modeling and delay management.

Both railway timetabling and delay management has already received a very substantial research attention. With regard to railway timetable optimization,we identified in the literature a gap between timetable design and stability analysis on a network scale and developed an integrated optimization model. While the end-to-end railway timetable planning is a mathematically very complex and therefore typically iterative process from passenger demand estimation to route and platform allocation, we did argue for the value of combining the steps of planning departure and arrival times, including train orders, and the timetable stability estimation. With removing the need for iterations, this allows for arriving at a superior timetable than the one that would be inherently limited to the practical maximum number of planning iterations.

For multimodal transfer modelling and delay management, in our literature review we identified the need to more accurately model transfer walking and waiting times at large, multimodal transfer stations. Many timetable planning tools, including the journey planners available to the public, typically include crude transfer time norms even for large and complex stations, sometimes adjusted based on passenger feedback, as in the case of journey planners. A multimodal transfer station of heavy rail, urban public transport and long distance buses, however, can have a complex structure with

numerous platforms, several access corridors, stairs and escalators, and ticket gates. Therefore transfer walking time at such a station can vary from a convenient one or two minute cross-platform transfer to a 10-15 minute walk to a remote platform via egress/access gates. When planning a multi-leg journey, and especially when optimizing a timetable for line synchronization, it is therefore essential, to improve the crude existing estimates of transfer walking times, and we propose to take advantage of recent open geographic and open transit data to do this in an automated fashion.

The following of this chapter gives a summary of our main findings, closing with recommendations for future research.

6.1.1 Optimizing high-frequency railway timetables for stability

The main idea elaborated in Chapter 3 is a railway timetable optimization model for periodic schedules that directly optimizes the minimum cycle time, which is a good approximate indicator of timetable stability. We showed that this approach can be used with flexible running times and flexible train orders for large networks and examine the possibility of certain service patterns performed on the given infrastructure, and for possible line patterns also determine the optimal train order. In another view, the model can also be used to evaluate possible infrastructure extensions and the extent to which they increase line capacity. Finally, ignoring timetable stability, this optimization model is also simply just an efficient way to calculate a stable timetable for a given infrastructure and line pattern.

The key idea behind and contribution of this part of the research is the combination of the following two observations. When using mixed integer linear programming for timetable optimization, even if for the goal of finding a feasible schedule, one needs an objective function to guide an optimization solver towards the desired result. This objective function is best to capture the quality and attractiveness of the timetable, and therefore running times, transfer waiting times, and costs of operation or other generalized costs are often used. The other observation is that when evaluating the stability of a periodic timetable, a good proxy to capture the available buffer times on a network level is the notion of the minimum cycle time, and its ratio to the nominal timetable period. Therefore we combined these two aspects and we used the cycle time as an objective function to optimize a heterogeneous railway timetable for the first time.

The optimization model we provided is also very flexible in terms of the set of different timetables it can consider at once. The whole timetable optimization process is inherently an iterative process due to its complexity, and therefore our model also needs several parameters to be fixed such as the line plan including stopping patterns, routing, and track choice. However, we do allow for flexible train running times and dwell times and flexible train orders, as well as train overtake locations, which increases the flexibility compared to the majority of existing optimization models. Of

course, just like with any other model as well, this flexibility can be further increased using iterative runs, such as the evaluation of infrastructure changes, or changes to the line pattern or track choice.

Another technical contribution of this optimization model is related to its above flexibility: we extend existing results on how to model overtaking of trains in flexible train order models. The challenge here was to make sure that trains can overtake each other only at stations and at open track sections where the infrastructure is available in reality, without having to define a very large number of constraints due to the flexibility of train orders. We obtain results for the open track overtake problem using dummy nodes and extend those results to make them applicable to stations, as well as provide a detailed procedure to automatically modify a periodic event network with flexible train orders and flexible running and dwell times to take into account these infrastructure constraints related to overtakes.

We also introduce several dimension reduction techniques to reduce the solution space, with the common characteristic of taking advantage of the existing symmetries of the periodic timetable. The periodic, fixed-interval timetable used throughout many countries has several operational and passenger attractiveness advantages; we introduce a large number of symmetry-breaking constraints in the optimization model. The common idea behind these constraints is that if two train lines provide a wholly symmetric service, e.g. alternating each 30 minutes, then it is irrelevant, which of these two trains depart first or second in an hour.

A key contribution related to the solution procedure of the optimization model is a dynamic iterative approach for breaking the optimization run into smaller, more manageable chunks. The advantage of this approach is not only to reduce the solution time, but also to provide feedback on the optimization progress, and obtain intermediate results, that helps timetable designers to draw conclusions earlier and re-configure the line plan if needed, to speed up the timetable planning process. We discover that introducing the flexible cycle time, which is the key feature of this model, increases the solution space, by affecting the pre-processing of overtake limitations, and the dimension reduction processes. Therefore we actually take a temporary step back and limit the cycle time to an assumed bound interval, and attempt solve the problem with a given timeout. After the intermediate solution is found or the timeout is reached, we adjust the bounds appropriately in order to arrive at the final, optimal timetable. We have shown that it is easier to control the solution process with this method and arrive at useful intermediate results, such as a feasible, but not necessarily optimal timetable.

6.1.2 Detailed modeling of intermodal transfers

In Chapter 4 we gave an example for how the open data that has recently become available can help public transport timetabling and measure aspects that previously were hidden from each transit operator. In particular, we describe how openly available

transfer station layouts can be used to estimate transfer walking times at high precision, and we also show how the more accurate transfer walking times can be combined with AVL data, i.e. realized arrival and departure time information. The available open data in turn is a necessary input for transfer station resistance estimation, timetable synchronization, and passenger information. Chapter 5 introduces a max-plus algebra-based approach for delay management. We use a max-plus delay propagation method to calculate expected future delays in an efficient manner, and we introduce a method to identify a subset of connections that are at risk, to focus both the optimization problem and support decision-making by dispatchers.

The key contributions of this part of the research include several ways to exploit open, or freely available geographic and transit data to improve existing transfer time and timetable performance estimates. Online, free geographic data from several providers has become so detailed, that it now includes not only street maps, but detailed layout of e.g. shopping malls, and also railway stations and multimodal transfer stations, to the level of detail of levels, stairs and escalators, and shops, ticket gates, etc. This level of detail in practice exceeds even the data quality available from e.g. railway infrastructure managers, where such information might be outdated, or paper-based. We take advantage of this new data source to show that it can improve a key input parameter for journey planning and timetable planning: the accurate calculation of transfer walking distance and transfer walking times. We provide a clear approach that can use a simple model of such a transfer station based on a few dozen collected data points, and connecting these with a timetable database in order to enhance the quality of transfer time estimates.

Besides the above, we also developed three applications that take advantage of enhanced transfer time estimates. The first application is an evaluation of the transfer resistance at a given station, as a weighted aggregate of transfer walking times compared to the passenger transport and traffic volume of the station. This evaluation can be used by a station manager, transport operator, or government agency, to quickly evaluate the performance of this part of the public transport facility, and identify transfer nodes that need further improvement of better passenger information, timetable synchronization, platform reallocation or redesign.

Finally, we extended the application scope of max-plus algebra to a delay management problem with multiple operators. A key practical observation is that a transport provider in real operations can only influence the dispatching and holding of their own network, and has very limited operational influence on the other, connecting network. Therefore we explicitly pay attention to this aspect and provide a delay propagation and delay management method that focuses on helping a smaller operator in their delay management by identifying transfers at risk in real-time and providing holding advice.

6.2 Recommendations for future research

In this final section we provide two suggested research directions to expand the results of this thesis. First, we describe the potential extensions to our railway timetable optimization model. Second, we describe the expected directions of new types and sources of open transit data, and the new areas of research that they make possible.

The periodic railway timetable optimization model defined in this thesis could be further improved in two ways: expanding the types of railway networks for which the model is applicable for, and finding new methods to speed up the optimization process. While it was out of scope for this thesis, we believe that the extension of this model to include bi-directional, single track lines should be relatively straightforward. While the core periodic event network is applicable just as well for single track operations, it is the modelling of infrastructure limitations via dummy nodes that needs to be extended to include the constraint of no passing trains on a single track line with no sidings. Often, but not always related to single-track networks, constraints describing synchronized connections might need to be introduced, if we move away from our initial assumption used in this thesis, that these networks are very high frequency. Finally, freight train paths should be included in the model – this should be straightforward once a predefined route, track choice and macroscopic speed profile of such freight train paths are available. The second area to improve is to find methods to speed up the optimization model solution process. We already used several hyperparameters on top of the simple optimization model to influence the solution time such as objective bounds and timeouts, a more systematic investigation of the best hyperparameters, including the analysis of the logs of the optimization solver, and experimenting using other available MILP solvers, might be fruitful. Results in this area indirectly relate back to the previous topic of the applicability of this optimization model: of course, a faster solution method means that a larger railway network can be analyzed in the same time.

Regarding our research on intermodal transfers an open transit data, our recommendation for future research is to continue taking advantage of new types of measurements and new data sources as they are released. During the writing of this thesis, for example, the open availability of Automatic Vehicle Location (AVL) data in the Netherlands increased from minimal to covering the majority of the larger urban public transport and regional bus lines. A still missing key puzzle piece is open data from the mainline train network – while such data might be available for researchers under certain conditions according to individual agreements, this information is still not openly available to everyone, despite the fact that the train operator and infrastructure manager are companies owned and subsidized by the government. The other continuously improving data source is geographic information system (GIS) data, both available at no cost from commercial sources and open GIS data, that includes much more accurate spatial information needed for pedestrian routing than just a few years ago. Finally, smartphone data can contain highly valuable information on driver and transit passenger

journeys, and this data is already used by companies to improve their journey planning applications, but analysis of smartphone data, for research as much as for profit, raises understandable concerns of privacy. Once accurate realized train arrival and departure data, spatial information, and maybe anonymized smartphone data is available, the scientific analysis of them can only lead to much more applicable practical recommendations than based on any simulated of heavily extrapolated data set.

Bibliography

9292 REISinformatiegroep. (2013). *9292 Open Data (in Dutch).* Retrieved 2013-07-18, from http://9292opendata.org/

Andersson, E., Housos, E., Kohl, N., & Wedelin, D. (1998). Crew Pairing Optimization. In G. Yu (Ed.), *Operations research in the airline industry* (Vol. 9, pp. 228–258). Springer US.

Bergmann, D. R. (1975). Integer programming formulation for deriving minimum dispatch intervals on a guideway accommodating through and local public transportation services. *Transportation Planning and Technology*, *3*(1), 27–30.

BISON. (2013). *BISON (in Dutch).* Retrieved 2013-07-18, from http://bison.connekt.nl/

Bookbinder, J. H., & Desilets, A. (1992). Transfer Optimization in a Transit Network. *Transportation science*, *26*, 106–118.

Borndörfer, R., & Liebchen, C. (2008). When Periodic Timetables Are Suboptimal. In S. Nickel & J. Kalcsics (Eds.), *Operations research proceedings 2007* (pp. 449–454). Berlin: Springer Verlag.

Braker, J. G. (1993). *Algorithms and Applications in Timed Discrete Event Systems* (PhD Thesis). Delft University of Technology.

Brändli, H., & Berg, W. (1979). Einfluß von neuen Bahnhofzugängen auf das Fahrgastverhalten. *Verkehr und Technik*, *32*(11), 480–482.

Burdett, R., & Kozan, E. (2006). Techniques for absolute capacity determination in railways. *Transportation Research Part B: Methodological*, *40*(8), 616–632.

Burdett, R., & Kozan, E. (2015, September). Techniques to effectively buffer schedules in the face of uncertainties. *Comput. Ind. Eng.*, *87*(C), 16–29. Retrieved from http://dx.doi.org/10.1016/j.cie.2015.04.024 doi: 10.1016/j.cie.2015.04.024

Bussieck, M. R., Lindner, T., & Lübbecke, M. E. (2004). A fast algorithm for near cost optimal line plans. *Mathematical Methods of Operations Research*, *59*(2), 205–220.

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

Cacchiani, V., Caprara, A., & Fischetti, M. (2011). A Lagrangian Heuristic for Robustness, with an Application to Train Timetabling. *Transportation Science*, *46*(1), 124–133.

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

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

Caimi, G. (2009). *Algorithmic decision support for train scheduling in a large and highly utilised railway network* (PhD Thesis). ETH Zürich.

Caimi, G., Burkolter, D., Herrmann, T., Chudak, F., & Laumanns, M. (2008). Design of a Railway Scheduling Model for Dense Services. *Networks and Spatial Economics*, *9*(1), 25–46.

Caimi, G., Chudak, F., Fuchsberger, M., Laumanns, M., & Zenklusen, R. (2010). A New Resource-Constrained Multicommodity Flow Model for Conflict-Free Train Routing and Scheduling. *Transportation Science*, *45*(2), 212–227.

Caprara, A., Galli, L., & Toth, P. (2011). Solution of the Train Platforming Problem. *Transportation Science*, *45*(2), 246–257.

De Kort, A., Heidergott, B., & Ayhan, H. (2003). A probabilistic (max, +) approach for determining railway infrastructure capacity. *European Journal of Operational Research*, *148*(3), 644–661.

De Standaard. (2013). *Fyra voorlopig verboden (in Dutch).* Retrieved from `http://www.standaard.be/cnt/dmf20130118_00438028`

Debrezion, G., Pels, E., & Rietveld, P. (2009). Modelling the joint access mode and railway station choice. *Transportation Research Part E: Logistics and Transportation Review*, *45*(1), 270–283.

Delorme, X., Gandibleux, X., & Rodriguez, J. (2009). Stability evaluation of a railway timetable at station level. *European Journal of Operational Research*, *195*(3), 780–790.

Dewilde, T., Sels, P., Cattrysse, D., & Vansteenwegen, P. (2013). Robust Railway Station Planning: an Interaction Between Routing, Timetabling and Platforming. In *Proceedings of the 5th International Seminar on Railway Operations Modelling and Analysis (RailCopenhagen).* Copenhagen.

Dollevoet, T., Huisman, D., Schmidt, M., & Schöbel, A. (2012). Delay Management with Rerouting of Passengers. *Transportation Science*, *46*(1), 74–89.

Domschke, W. (1989). Schedule synchronization for public transit networks. *OR Spektrum*, *11*(1), 17–24.

East Japan Railway Company. (2002). *Social and Environmental Report 2002* (Tech. Rep.). Retrieved from http://www.jreast.co.jp/e/environment/pdf_2002/all_e.pdf

European Commission. (2013a). *EU Transport in Figures – Statistical Pocketbook 2013*. Luxembourg: Publications Office of the European Union. Retrieved from http://ec.europa.eu/transport/facts-fundings/statistics/doc/2013/pocketbook2013.pdf

European Commission. (2013b). *Flash Eurobarometer 382a: European's Satisfaction with Rail Services* (Tech. Rep.). European Commission. Retrieved from http://ec.europa.eu/public_opinion/flash/fl_382a_en.pdf

Fischetti, M., & Monaci, M. (2009). Light robustness. In *Robust and online large-scale optimization* (pp. 61–84).

Fischetti, M., Salvagnin, D., & Zanette, a. (2009). Fast Approaches to Improve the Robustness of a Railway Timetable. *Transportation Science*, *43*(3), 321–335.

Furth, P. G., Hemily, B., Muller, T. H. J., & Strathman, J. G. (2006). *TCRP Report 113: Using Archived AVL-APC Data to Improve Transit Performance and Management* (Tech. Rep.). Washington, D.C.: Transportation Research Board.

Ginkel, A., & Schöbel, A. (2007). To Wait or Not to Wait? The Bicriteria Delay Management Problem in Public Transportation. *Transportation Science*, *41*(4), 527–538.

Goerigk, M., Schachtebeck, M., & Schöbel, A. (2013). Evaluating line concepts using travel times and robustness: Simulations with the LinTim toolbox. *Public Transport*, *5*, 267–284.

Goossens, J.-W., van Hoesel, S., & Kroon, L. G. (2006). On solving multi-type railway line planning problems. *European Journal of Operational Research*, *168*(2), 403–424.

Goverde, R. M. P. (1998). Synchronization Control of Scheduled Train Services to Minimize Passenger Waiting Times. In *Proceedings of the 4th TRAIL Year Congress, Part 2*. Delft: TRAIL Research School.

Goverde, R. M. P. (1999). Improving Punctuality and Transfer Reliability by Railway Timetable Optimization. In *Proceedings of the 5th TRAIL Annual Congress.* Delft: TRAIL Research School.

Goverde, R. M. P. (2005). *Punctuality of Railway Operations and Timetable Stability Analysis* (PhD thesis). Delft University of Technology.

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

Goverde, R. M. P. (2008). Stability Analysis. In I. A. Hansen & J. Pachl (Eds.), *Railway Timetable & Traffic: Analysis - Modelling - Simulation.* Eurailpress.

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

Goverde, R. M. P., & Odijk, M. A. (2002). Performance evaluation of network timetables using PETER. In J. Allan, R. Hill, C. Brebbia, G. Sciutto, & S. Sone (Eds.), *Computers in railways, vol. viii* (pp. 731–740). Southampton: WIT Press.

GOVI. (2013). *Grenzeloze Openbaar Vervoer Informatie (GOVI) (in Dutch).* Retrieved 2013-07-18, from `http://govi.nu`

Großmann, P., Hölldobler, S., Manthey, N., Nachtigall, K., Opitz, J., & Steinke, P. (2012). Solving periodic event scheduling problems with SAT. In H. Jiang, W. Ding, M. Ali, & X. Wu (Eds.), *Advanced research in applied artificial intelligence* (pp. 166–175). Springer Berlin Heidelberg.

Guan, J., Yang, H., & Wirasinghe, S. (2006). Simultaneous optimization of transit line configuration and passenger line assignment. *Transportation Research Part B: Methodological*, *40*(10), 885–902.

Guihaire, V., & Hao, J.-K. (2008). Transit network design and scheduling: A global review. *Transportation Research Part A: Policy and Practice*, *42*(10), 1251–1273.

Guo, Z., & Wilson, N. H. M. (2011). Assessing the cost of transfer inconvenience in public transport systems: A case study of the London Underground. *Transportation Research Part A: Policy and Practice*, *45*(2), 91–104.

Hadas, Y., & Shnaiderman, M. (2012). Public-transit frequency setting using minimum-cost approach with stochastic demand and travel time. *Transportation Research Part B: Methodological*, *46*(8), 1068–1084.

Heidergott, B., & De Vries, R. (2001). Towards a (Max,+) Control Theory for Public Transportation Networks. *Discrete Event Dynamic Systems: Theory and Applications*, *11*, 371–398.

Heidergott, B., Olsder, G. J., & van der Woude, J. (2005). *Max Plus at Work: Modeling and Analysis of Synchronized Systems: A Course on Max-Plus Algebra and Its Applications*. Princeton University Press.

Heydar, M., Petering, M. E. H., & Bergmann, D. R. (2013). Mixed integer programming for minimizing the period of a cyclic railway timetable for a single track with two train types. *Computers and Industrial Engineering*, *66*(1), 171–185.

Hilderink, I., Kieft, S. C., & Wilgenburg, J. (2010). Koken met VENOM : de bereiding van een verkeersprognosemodel voor de Metropoolregio Amsterdam. In *Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk.* Roermond.

Hooghiemstra, J. S., & Teunisse, M. J. G. (1998). The Use of Simulation in the Planning of the Dutch Railway Services. In D. Medeiros, E. Watson, J. Carson, & M. Manivannan (Eds.), *Proceedings of the 1998 winter simulation conference* (pp. 1139–1145).

IBM. (2009). *New Rail Innovation Center in Beijing.* Retrieved 2014-03-19, from `http://asmarterplanet.com/blog/2009/06/new-rail-innovation-center.html`

International Union of Railways (UIC). (2004). *UIC Code 406: Capacity.* Paris: International Union of Railways (UIC).

International Union of Railways (UIC). (2013). *UIC Code 406: Capacity.* Paris: International Union of Railways (UIC).

Kecman, P. (2014). *Models for Predictive Railway Traffic Management* (PhD thesis). Delft University of Technology.

Kirchoff, H.-H. (1992). Plannung und Dimensionierung von Verknüpfungsanlagen im ÖPNV. *Der Nahverkehr*(1), 38–45.

Knoppers, P., & Muller, T. (1995). Optimized Transfer Opportunities in Public Transport. *Transportation Science*, *29*(1), 101–105.

König, A., & Axhausen, K. W. (2002). The reliability of the Transportation System and its influence on the Choice Behaviour. In *Proceedings of the 2nd Swiss Transportation Research Conference.* Monte Verità.

Kroon, L. G., Huisman, D., Abbink, E., Fioole, P.-J., Fischetti, M., Maróti, G., … Ybema, R. (2009). The New Dutch Timetable: The OR Revolution. *Interfaces*, *39*(1), 6–17.

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

Kroon, L. G., & Peeters, L. W. P. (2003). A variable trip time model for cyclic railway timetabling. *Transportation Science*(1994).

Kruse, B. (1995). Gestaltungsgrundsätze für Verknüpfungspunkte im ÖPNV. *Der Nahverkehr*, *13*(11), 59–62.

Liberti, L. (2008). Automatic generation of symmetry-breaking constraints. In *Lecture notes in computer science* (Vol. 5165 LNCS, pp. 328–338).

Liebchen, C. (2008). The First Optimized Railway Timetable in Practice. *Transportation Science*, *42*(4), 420–435.

Liebchen, C., Schachtebeck, M., Schöbel, A., Stiller, S., & Prigge, A. (2010). Computing delay resistant railway timetables. *Computers & Operations Research*, *37*(5), 857–868.

Lindner, T. (2000). *Train Schedule Optimization in Public Rail Transport* (PhD thesis). Technische Universität Braunschweig.

Lindner, T. (2011, nov). Applicability of the analytical UIC Code 406 compression method for evaluating line and station capacity. *Journal of Rail Transport Planning & Management*, *1*(1), 49–57. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S2210970611000059 doi: 10.1016/j.jrtpm.2011.09.002

Lindner, T., & Zimmermann, U. T. (2005). Cost optimal periodic train scheduling. *Mathematical Methods of Operations Research*, *62*(2), 281–295.

Lusby, R., Larsen, J., Ryan, D., & Ehrgott, M. (2011). Routing Trains Through Railway Junctions: A New Set-Packing Approach. *Transportation Science*, *45*(2), 228–245.

Malcolm, D. G., Roseboom, J. H., Clark, C. E., & Fazar, W. (1959). Application of a Technique for Research and Development Program Evaluation. *Operations Research*, *7*(5), 646–669.

Ministry of Land Infrastructure Transport and Tourism. (2014). *Monthly Statistical Report on Railway Transport.* Retrieved 2014-03-07, from http://www.mlit.go.jp/k-toukei/60/monthly/railwaydata.xls

Molin, E., Chorus, C., & van Sloten, R. (2009). The need for advanced public transport information services when making transfers. *European Journal of Transport and Infrastructure Research*, *9*(4), 397–410.

Mussone, L., & Wolfler Calvo, R. (2013). An analytical approach to calculate the capacity of a railway system. *European Journal of Operational Research*, *228*(1), 11–23.

Nachtigall, K. (1996). Periodic network optimization with different arc frequencies. *Discrete Applied Mathematics*, *69*(1-2), 1–17.

Nachtigall, K., & Jerosch, K. (2008). Simultaneous Network Line Planning and Traffic Assignment. In *8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS).* Karlsruhe.

Nachtigall, K., & Voget, S. (1996). A genetic algorithm approach to periodic railway synchronization. *Computers and Operations Research*, *23*(5), 453–463.

Nachtigall, K., & Voget, S. (1997). Minimizing waiting times in integrated fixed interval timetables by upgrading railway tracks. *European Journal Of Operational Research*, 610–627.

Netherlands Railways. (2012). *NS Jaarverslag 2012 (in Dutch)* (Tech. Rep.). Netherlands Railways. Retrieved from http://www.ns.nl/binaries/content/assets/NS/over-ns/jaarverslagen-pdf/nsjaarverslag2012.pdf

Netherlands Railways. (2013). *Vanaf 18 februari rechtstreekse Intercity-verbinding tussen de Randstad en België (in Dutch).* Retrieved 2014-03-20, from http://www.ns.nl/over-ns/nieuwscentrum/persberichten?frommonth=2&fromyear=2013&tomonth=2&toyear=2013&subject=&q=#vanaf-18-februari-rechtstreekse-intercity-verbinding-tussen-de-randstad-en-belgie

Odijk, M. A. (1996). A constraint generation algorithm for the construction of periodic railway timetables. *Transportation Research Part B: Methodological*, *30*(6), 455–464.

ProRail. (2014). *ProRail in cijfers (in Dutch).* Retrieved 2014-03-21, from http://www.prorail.nl/over-prorail/wat-doet-prorail/prorail-in-cijfers

Ramaekers, P., De Wit, T., & Pouwels, M. (2009). *Hoe druk is het nu werkelijk op het Nederlandse spoor? (in Dutch)* (Tech. Rep.). The Hague, The Netherlands: Centraal Bureau voor de Statistiek (Statistics Netherlands).

RATP. (2010). *Le 16 novembre, la RATP présente une nouvelle étape de l'automatisation de la ligne 1 du métro (in French)* (Tech. Rep.). Retrieved from http://www.ratp.fr/fr/upload/docs/application/pdf/2010-11/dp_pcc-_ligne-1.pdf

Schlechte, T., Borndörfer, R., Erol, B., Graffagnino, T., & Swarat, E. (2011, nov). Micromacro transformation of railway networks. *Journal of Rail Transport Planning & Management*, *1*(1), 38–48. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S2210970611000047 doi: 10.1016/j.jrtpm.2011.09.001

Schöbel, A. (2006). *Optimization in Public Transportation: Stop location, delay management and tariff planning from a customer-oriented point of view* (Vol. 3). Springer.

Schöbel, A. (2012). Line planning in public transportation: Models and methods. *OR Spectrum*, *34*(3), 491–510.

Schrijver, A. (1998). Routing and timetabling by topological search. *Documenta Mathematica*, 687–695.

Schrijver, A., & Steenbeek, A. (1993). *Spoorwegdienstregelingontwikkeling (in Dutch)* (Tech. Rep.).

Schröder, M., & Solchenbach, I. (2006). *Optimization of Transfer Quality in Regional Public Transit* (Vol. 84; Tech. Rep. No. 84). Kaiserslautern: Frauenhofer-Institut für Techno- und Wirtschaftsmathematik.

Schuele, I., Schroeder, M., & Kuefer, K. H. (2009). Synchronization of regional public transport systems. In *Urban Transport XV* (Vol. 107, pp. 301–311).

Serafini, P., & Ukovich, W. (1989). A Mathematical Model for Periodic Event Scheduling Problems. *SIAM Journal on Discrete Mathematics*, *2*(4), 550–581.

Significance, VU University Amsterdam, & John Bates Services. (2012). *Values of time and reliability in passenger and freight transport in The Netherlands: Report for the Ministry of Infrastructure and the Environment* (Tech. Rep. No. November). Significance.

Smith, P. (2013). *Cockpit Confidential: Everything You Need to Know About Air Travel: Questions, Answers, and Reflections*. Sourcebooks.

Söngen, J. (1979). *Beurteilung von Verknüpfungsanlagen für Personenverkehrssysteme (in German)* (PhD thesis). Technische Hochschule Darmstadt.

Sparing, D., & Goverde, R. M. P. (2011). Evaluation of Multimodal Timetable Synchronization in the Randstad. In *Proceedings of the 4th International Seminar on Railway Operations Modelling and Analysis (RailRome).* Rome.

Sparing, D., & Goverde, R. M. P. (2013a, jul). Identifying effective guaranteed connections in a multimodal public transport network. *Public Transport*, *5*(1-2), 79–94. Retrieved from http://link.springer.com/10.1007/s12469-013-0068-6 doi: 10.1007/s12469-013-0068-6

Sparing, D., & Goverde, R. M. P. (2013b). An Optimization Model for Periodic Timetable Generation with Dynamic Frequencies. In *Proceedings of the 16th international ieee annual conference on intelligent transportation systems (itsc 2013).* The Hague, The Netherlands.

Stichting OpenGeo. (2013). *OpenOV*. Retrieved 2013-07-18, from http://www.openov.nl/

Van der Waard, J. (1988). The relative importance of public transport trip time attributes in route choice. In *Proceedings of the 16th PTRC Summer Annual Meeting.* Bath, United Kingdom.

Van Eck, G. (2011). Dynamische toedeling en toetsing van grootschalige multimodale vervoerssystemen (in Dutch). In *Bijdrage aan het Colloquium Vervoersplanologisch Speurwerk.* Antwerpen.

Van Oort, N. (2011). *Service Reliability and Urban Public Transport Design* (PhD Thesis). TU Delft.

Van Oort, N., & Van Nes, R. (2009). Line Length Versus Operational Reliability. *Transportation Research Record: Journal of the Transportation Research Board*, *2112*(-1), 104–110.

van Oort, N., Sparing, D., Brands, T., & Goverde, R. M. P. (2015). Data driven improvements in public transport: the Dutch example. *Public Transport*. Retrieved from http://link.springer.com/10.1007/s12469-015-0114-7 doi: 10.1007/s12469-015-0114-7

Walker, J. (2008). Purpose-driven public transport: creating a clear conversation about public transport goals. *Journal of Transport Geography*, *16*(6), 436–442.

Waterson, B., Box, S., & Armstrong, J. (2010). Determining Rail Network Accessibility. In *5th IMA Mathematics in Transport Conference.* London.

Webwereld. (2012). *GVB maakt real-time reisinformatie open data (in Dutch).* Retrieved 2013-07-18, from http://webwereld.nl/cloud/56325-gvb-maakt-real-time-reisinformatie-open-data

Weidmann, U. (1993). *Transporttechnik der Fussgänger, Transporttechnische Eigenschaften des Fussgängerverkehrs (Literturauswertung) (in German).* ETH IVT.

Williams, H. P. (1990). *Model Building in Mathematical Programming Fourth Edition* (3rd ed.). John Wiley & Sons.

Wong, R. C. W., Yuen, T. W. Y., Fung, K. W., & Leung, J. M. Y. (2008). Optimizing Timetable Synchronization for Rail Mass Transit. *Transportation Science*, *42*(1), 57–69.

World Bank. (2014). *World Bank Databank.* Retrieved 2014-03-18, from http://data.worldbank.org/indicator/IS.RRS.TOTL.KM

Summary

This thesis focuses on railway and public transport timetable planning and optimization. The two main areas of contributions are: (1) mathematical optimization of busy railway timetables with a direct focus on timetable stability, and (2) modelling of multimodal connections between bus and railway lines. All research topics in this thesis are part of the timetabling problem of public transport, that is, choosing the arrival and departure times for train and bus services, given a certain infrastructure, available vehicles, and line structure plan, that lead to a convenient and reliable service, with extra attention to minimizing delay propagation and transfer inconvenience.

The development of a railway timetable is inherently a complex, and therefore a multistep process, from passengers demand estimation through line planning and macroscopic timetabling to platform and track allocations and driver advisory systems. Macroscopic timetabling models in particular tend to focus either on the search for feasible timetables, or the stability analysis of existing timetable plans. These two steps result in multiple iterations of timetable generation and analysis, therefore a long planning process and possibly suboptimal timetable. A combination of these two steps into one optimization process can lead therefore for faster planning and more stable schedules. Therefore our first research objective is to develop an optimization model to maximize the stability of periodic railway network timetables.

As both railway and urban public transport networks are complex systems, most timetabling and public transport models focus on just one modality, such as railways or a metro or bus network. Furthermore, for models and journey planners that *are* multimodal, the transfer times used are often only simple norms, even in case of large and complex transfer station nodes. A large proportion of passenger journeys, however, are multimodal journeys using urban public transport as an access or egress mode for a long-distance train journey, and hence attention to intermodal transfers is crucial to improve the entire transit experience. On the other hand, in the recent years new open transit and geographical data sources became available that allow transport modelling in previously unprecedented detail. Therefore our second research objective is to model intermodal transfers in detail using open data.

The stability-optimized railway timetable

In the first part of this research we presented a formulation of the railway timetabling optimization problem that uses the minimum cycle time, a proxy for the stability of the timetable, as its objective. This is the first time that this objective was used for the optimization of a mixed-traffic railway system. This combination of the timetable planning and timetable stability evaluation problem means that as the output of the optimization problem is a timetable optimized for network stability, the need for a large number of iterations including a timetable optimization step and a stability evaluation step is removed, improving the speed and the efficiency of the scheduling. This model is useful in practice in several steps of timetable planning process, from the actual planning of the current timetable to long term planning identifying possible line patterns and most beneficial infrastructure upgrades, to ad-hoc analysis of bottlenecks.

In order to be able to improve the stability by adjustments of train orders, running and dwell times, the mathematical optimization model uses flexible train orders, flexible running and dwell times, and flexible overtaking locations for fast and slow trains. This means that within a given line plan and stopping patterns, all these aspects of the timetable are optimized to improve timetable stability. In order to accurately model the possibility or impossibility of overtaking at certain stations or track segments, we introduced a method to expand the original set of timetable events and event pairs to allow overtaking only at the appropriate locations, without the need to explicitly model each possible consecutive train pair.

To successfully solve the optimization model for a large network in reasonable time, we used two key techniques: a set of dimension reduction techniques, and an iterative optimization method using dynamically adjusted objective bounds. We reduce the solution space by a set of techniques that take advantage of the existing symmetry of the periodic timetable to eliminate different solutions that are identical for all practical purposes. The iterative optimization is achieved by temporarily restricting the bounds of the cycle time that we are minimizing, introduce further dimension reducing constraints that take into account these temporary bounds, and after a successful or timed-out run re-adjusting the cycle time bounds accordingly until an optimal or sufficient quality timetable is reached, or the infeasibility established. With these methods we were able to define a stability-optimized timetable for the intercity network in the Netherlands, including local trains in the core, busiest area of the network, on a regular computer.

To implement this optimization model, we developed a software tool that reads common timetable formats already used by other timetable planning tools in the Netherlands, can work with a number of commercially available Mixed Integer Linear Programming solvers, and outputs the optimal timetable in the same format, as well as the related visualizations on both the planned timetable and the progress of the optimization process.

Multimodal transfer modelling

In the second part of this research, we showed how to exploit newly available open transit data to improve the modelling and therefore the timetable planning of complex transfer stations; and how to use max-plus algebra in the context of intermodal delay management. Both transit operations, and railway and transit research, often focuses at one network or one transit mode only, such as a railway network or a bus network. Passenger trips, on the other hand, are most often multimodal, combining e.g. a local bus or tram trip with an intercity rail journey; and therefore the attractiveness or inconvenience of these multimodal transfers are key in the perceived quality of the whole public transport system. Thanks to new, open geographic databases, and publicly available transit feeds such as timetable data and Automatic Vehicle Location data, there is now data and information available at a previously unprecedented scale, that used to be hidden even from transit operators, especially concerning the network of another operator.

We focus on the detailed modelling of transfer walking times at large multimodal transfer stations, as the accurate values for transfers are essential for journey planning, timetable synchronization, and in the evaluation of the station transfer resistance. Openly available, detailed geographic data on transfer nodes can be used, with related timetable and platform allocation data connected, to build a 3-dimensional model of a transfer station including access paths, stairs, ticket gates, to improve the crude transfer time values currently used in many journey planner and timetable planning systems. These refined transfer models can in turn be used for more accurate timetable synchronization, transfer station resistance calculations, and improving dynamic, real-time passenger information, as we show in the thesis.

Finally, we use max-plus algebra to provide a formulation of the delay management problem. We combine delay propagation, the selection of important connections, and optimized holding advice in a way that is applicable even for a smaller operator, that has only influence on a smaller regional network connecting to other main networks, such as intercity train lines. The max-plus algebra-based delay propagation algorithm allows for a reduction of the solution space of the optimization problem, that is defined with arrival delays and missed connections contributing to the cost function. The output of this approach is therefore not only a holding advice for each departure at a transfer node, but also a shortlist of connections at risk, that helps dispatchers focus on the actionable transfer directions, even overriding the advice of the optimization model if they see fit.

Samenvatting

Dit proefschrift richt zich op het plannen en optimaliseren van dienstregelingen voor het spoor en het openbaar vervoer. De twee belangrijkste gebieden waaraan deze scriptie een bijdrage levert zijn: (1) wiskundige optimalisatie van drukke spoordienstregelingen met een directe focus op stabiliteit van deze dienstregeling en (2) het modelleren van multimodale verbindingen tussen bus- en spoorlijnen. Alle onderzoeksonderwerpen in deze scriptie maken deel uit van het dienstregelingprobleem van openbaar vervoer. Dit wil zeggen: de aankomst- en vertrektijden voor trein- en busdiensten kiezen, rekening houdend met een bepaalde infrastructuur, beschikbare voertuigen en een lijnstructuurplan, wat resulteert in een handige en betrouwbare dienstverlening, waarin extra aandacht is voor het minimaliseren van vertragingen en ongemakken bij het overstappen.

De ontwikkeling van een dienstregeling voor het spoor is een inherent complex proces, waar meerdere stappen voor nodig zijn. Van een inschatting van de vraag van passagiers naar lijnplanning en macroscopische dienstregelingen, tot platform- en spoortoewijzingen en adviessystemen voor bestuurders. Macroscopische modellen voor de dienstregeling richten zich met name op de zoektocht naar haalbare dienstregelingen, of de stabiliteitsanalyse van bestaande dienstregelingen. Deze twee stappen resulteren in meerdere iteraties van het genereren en analyseren van dienstregelingen en zodoende tot een lang planproces en een mogelijke suboptimale dienstregeling. Een combinatie van deze twee stappen in één optimalisatieproces kan daarom leiden tot snellere planning en stabielere dienstregelingen. Ons eerste onderzoeksdoel is daarom ook het ontwikkelen van een optimalisatiemodel om de stabiliteit van periodieke dienstregelingen voor het spoornetwerk te maximaliseren.

Gezien het feit dat zowel het netwerk voor het spoor als voor het openbaar vervoer complexe systemen zijn, richten de meeste modellen voor dienstregelingen en openbaar vervoer zich op slechts één modaliteit, zoals het spoor of een metro- of busnetwerk. Voor modellen en reisplanners die wél multimodaal zijn, worden vaak eenvoudige normen gebruikt voor de overstaptijden, zelfs als het gaat om grote en complexe knooppuntstations. Een groot deel van de passagiersreizen zijn echter multimodale reizen, waarbij stedelijk openbaar vervoer gebruikt wordt als toegangs- of uitgangsmodus voor een lange reis. Aan de andere kant: in de afgelopen jaren is er een veelvoud aan openbare vervoers- en geografische data beschikbaar gekomen, waarmee het mogelijk is om vervoer tot op ongekend detailniveau te plannen. Daarom is ons

tweede onderzoeksdoel het modelleren van intermodale transfers. Met behulp van open data kan dit tot in detail.

De spoordienstregeling met geoptimaliseerde stabiliteit

In het eerste deel van dit onderzoek presenteerde we een formulering van het optimalisatieprobleem rondom de spoordienstregeling, waarbij gebruik wordt gemaakt van een minimale cyclustijd, die iets zegt over de stabiliteit, als het doel. Dit is de eerste keer dat dit doel gebruikt werd voor de optimalisatie van een spoorsysteem met gemengd verkeer. Deze combinatie van het plannen van de dienstregeling en het evaluatieprobleem van de dienstregelingstabiliteit betekent dat het resulteert in een dienstregeling die geoptimaliseerd is op netwerkstabiliteit. Er wordt een groot aantal iteraties opgenomen, waaronder een optimalisatiestap voor de dienstregeling en een stap voor stabiliteitsevaluatie wordt verwijderd, waardoor het plannen sneller en efficiënter kan zijn. Dit model is in de praktijk handig in verschillende stappen van het dienstregelingplanproces, van de daadwerkelijke planning van de huidige dienstregeling tot langetermijnplanning, waarbij mogelijke lijnpatronen en de meest gunstige upgrades van de infrastructuur worden geïdentificeerd, evenals een ad-hoc analyse van knelpunten.

Om in staat te zijn om de stabiliteit te verbeteren door treinvolgorden en rij- en halteertijden aan te passen, gebruikt het wiskundige model flexibele treinvolgorden, flexibele rij- en halteertijden en flexibele inhaallocaties voor snelle en langzame treinen. Dit betekent dat binnen een bepaald lijnplan en stoppatronen al deze onderdelen van een dienstregeling geoptimaliseerd worden, om te zorgen voor meer stabiliteit in de dienstregeling. Om de mogelijkheid of onmogelijkheid van het overnemen op bepaalde stations of delen van het spoor mogelijk of onmogelijk te maken, hebben we een model geïntroduceerd waarmee het aantal oorspronkelijke gebeurtenissen op de dienstregeling en het aantal combinaties van gebeurtenissen kan worden uitgebreid, om inhalen alleen mogelijk te maken op de passende locaties, zonder dat het hierbij nodig is om uitdrukkelijk elke mogelijke volgende treincombinatie te modelleren.

Om het optimalisatiemodel voor een groot netwerk binnen een redelijke termijn te realiseren, hebben we twee belangrijke technieken gebruikt: een aantal technieken voor dimensiereductie en een iteratieve optimalisatiemethode, waarvoor dynamisch aangepaste begrenzingen van de doelfunctie gebruikt zijn. We verkleinen de oplossingsruimte door een aantal technieken die gebruik maken van de bestaande symmetrie van de periodieke dienstregeling, om verschillende oplossingen die voor alle praktische doeleinden identiek zijn te elimineren. De iteratieve optimalisatie wordt gerealiseerd door de grenzen aan de cyclustijd die we minimaliseren tijdelijk te beperken en meer dimensiereductie te introduceren die rekening houden met deze tijdelijke grenzen. Na een succesvolle berekening of tijdsoverschrijding passen we de grenzen van de cyclustijden aan tot we komen tot een optimale dienstregeling of één van voldoende kwaliteit, of tot is vastgesteld dat dit onmogelijk is. Met deze methoden kunnen we

nu een voor stabiliteit geoptimaliseerde dienstregeling opstellen voor het intercity-netwerk in Nederland, waarbij we ook lokale treinen meenemen in de drukste gebieden van het netwerk.

Om het optimalisatiemodel te implementeren, hebben we een softwaretool ontwikkeld die in staat is om de gebruikelijke dienstregelingformats die al gebruikt worden in andere plantools voor dienstregelingen in Nederland in te lezen. De tool kan ook samenwerken met een aantal commercieel beschikbare Mixed Integer Linear Programming-tools en resulteert in een optimale dienstregeling die in hetzelfde formaat wordt uitgegeven, inclusief de bijbehorende visualisaties rondom zowel de geplande dienstregeling als de voortgang van het optimalisatieproces.

Multimodale transfermodellering

In het tweede deel van dit onderzoek laten we zien hoe openbaar beschikbare gegevens over vervoer het modelleren kunnen verbeteren en daarmee ook het plannen van de dienstregeling voor complexe overstapstations. Ook laten we zien hoe max-plus algebra gebruikt kan worden in de context van intermodaal vertragingsbeheer. Zowel de openbaarvervoerpraktijk als het onderzoek in spoor- en ander openbaar vervoer richt zich meestal op één netwerk of één vervoermodus, zoals de spoornetwerken of een busnetwerk. Personenreizen zijn echter meestal multimodaal en combineren bijvoorbeeld een lokale bus of tram met een reis met de intercity. Daarom zijn de aantrekkelijkheid of het ongemak van deze multimodale overstapmomenten essentieel in de waargenomen kwaliteit van het volledige systeem van openbaar vervoer. Dankzij nieuwe, open geografische databases en openbaar beschikbare vervoersinformatie, zoals dienstregelingdata en Automatic Vehicle Location-data, is er nu data en informatie beschikbaar op ongekende schaal. Deze informatie was eerder zelfs verborgen voor vervoersbedrijven, zeker als het ging om het netwerk van een andere dienstverlener.

We richten ons op de gedetailleerde modellering van overstaplooptijden bij grote, multimodale overtapstations, aangezien nauwkeurige waarden hier essentieel zijn voor het plannen van een reis, het synchroniseren van dienstregelingen en dus de evaluatie van overstapweerstand van een station. Gedetailleerde geografische data van overstapknooppunten, die openbaar beschikbaar is, kan worden gebruikt, gekoppeld aan de bijbehorende data van dienstregeling en perrontoewijzing. Hiermee kan een 3D-model worden gebouwd van een overstapstations, inclusief toegangswegen, trappen, poortjes, om zo te komen tot een verbetering van de geschatte overstaptijden die op dit moment gebruikt worden in veel reisplanners en systemen voor dienstregelingen. Deze verfijnde overstapmodellen kunnen weer gebruikt worden voor een nauwkeurigere synchronisatie van dienstregelingen, het berekenen van overstapstationweerstand en het verbeteren van dynamische, real-time passagiersinformatie.

Tot slot gebruiken we max-plus algebra voor de modellering van vertragingsvoortplanting. We combineren vertragingsvoortplanting, selectie van belangrijke verbindin-

gen en geoptimaliseerd wachtadvies op een manier die ook voor kleine vervoerders toepasbaar is en die alleen invloed hebben op een kleiner regionaal netwerk dat aansluit op de grote netwerken, zoals de intercity treinen. Het vertragingsalgoritme dat gebaseerd is op max-plus algebra biedt de mogelijkheid om de oplossingsruimte van het optimalisatieprobleem, dat wordt bepaald door aankomstvertragingen en gemiste aansluitingen die bijdragen aan de kostenfunctie, te verkleinen. Het resultaat van deze benadering is daarom niet alleen een advies voor elk vertrek van een overstapknooppunt, maar ook een shortlist van aansluitingen die risico's lopen. Hiermee kunnen verkeersleiders zich richten op overstaprichtingen waarop actie kan worden genomen en zelfs het advies van het optimalisatiemodel afwijzen als zij denken dat dat nodig is.

About the author

Daniel Sparing was born in 1985 in Budapest, Hungary. He obtained his M.Sc. in Electrical Engineering in 2008 at the Budapest University of Technology and Economics. After two years of work experience in data mining consulting, he joined the Department of Transport and Planning at the Delft University of Technology as a Ph.D. candidate in April 2010, within research project *Strategy Towards Sustainable and Reliable Multi-modal Transport in the Randstad*, funded by the Netherlands Organisation for Scientific Research (NWO). He currently works as a consultant in machine learning and mathematical optimization.

TRAIL Thesis Series

A series of The Netherlands TRAIL Research School for theses on transport, infrastructure and logistics.

Sparing, D., *Reliable Timetable Design for Railways and Connecting Public Transport Services*, T2016/5, May 2016, TRAIL Thesis Series, the Netherlands

Rasouli, S, *Uncertainty in Modeling Activity-Travel Demand in Complex Urban Systems*, T2016/4, March 2016, TRAIL Thesis Series, the Netherlands

Vries, J. de, *Behavioral Operations in Logistics*, T2016/3, February 2016, TRAIL Thesis Series, the Netherlands

Goi-Ros, B., *Traffic Flow at Sags: Theory, Modeling and Control*, T2016/2, March 2016, TRAIL Thesis Series, the Netherlands

Khademi, E., *Effects of Pricing Strategies on Dynamic Repertoires of Activity-Travel Behaviour*, T2016/1, February 2016, TRAIL Thesis Series, the Netherlands

Cong, Z., *Efficient Optimization Methods for Freeway Management and Control*, T2015/17, November 2015, TRAIL Thesis Series, the Netherlands

Kersbergen, B., *Modeling and Control of Switching Max-Plus-Linear Systems: Rescheduling of railway traffic and changing gaits in legged locomotion*, T2015/16, October 2015, TRAIL Thesis Series, the Netherlands

Brands, T., *Multi-Objective Optimisation of Multimodal Passenger Transportation Networks*, T2015/15, October 2015, TRAIL Thesis Series, the Netherlands

Ardi, zgl, *Road Pricing Policy Process: The interplay between policy actors, the media and public*, T2015/14, September 2015, TRAIL Thesis Series, the Netherlands

Xin, J., *Control and Coordination for Automated Container Terminals*, T2015/13, September 2015, TRAIL Thesis Series, the Netherlands

Anand, N., *An Agent Based Modelling Approach for Multi-Stakeholder Analysis of City Logistics Solutions*, T2015/12, September 2015, TRAIL Thesis Series, the Netherlands

Hurk, E. van der, *Passengers, Information, and Disruptions*, T2015/11, June 2015, TRAIL Thesis Series, the Netherlands

Davydenko, I., *Logistics Chains in Freight Transport Modelling*, T2015/10, May 2015, TRAIL Thesis Series, the Netherlands

Schakel, W., *Development, Simulation and Evaluation of In-car Advice on Headway, Speed and Lane*, T2015/9, May 2015, TRAIL Thesis Series, the Netherlands

Dorsser, J.C.M. van, *Very Long Term Development of the Dutch Inland Waterway Transport System: Policy analysis, transport projections, shipping scenarios, and a new perspective on economic growth and future discounting*, T2015/8, May 2015, TRAIL Thesis Series, the Netherlands

Hajiahmadi, M., *Optimal and Robust Switching Control Strategies: Theory, and applications in traffic management*, T2015/7, April 2015, TRAIL Thesis Series, the Netherlands

Wang, Y., *On-line Distributed Prediction and Control for a Large-scale Traffic Network*, T2015/6, March 2015, TRAIL Thesis Series, the Netherlands

Vreeswijk, J.D., *The Dynamics of User Perception, Decision Making and Route Choice*, T2015/5, February 2015, TRAIL Thesis Series, the Netherlands

Lu, R., *The Effects of Information and Communication Technologies on Accessibility*, T2015/4, February 2015, TRAIL Thesis Series, the Netherlands

Ramos, G. de, *Dynamic Route Choice Modelling of the Effects of Travel Information using RP Data*, T2015/3, February 2015, TRAIL Thesis Series, the Netherlands

Sierzchula, W.S., *Development and Early Adoption of Electric Vehicles: Understanding the tempest*, T2015/2, January 2015, TRAIL Thesis Series, the Netherlands

Vianen, T. van, *Simulation-integrated Design of Dry Bulk Terminals*, T2015/1, January 2015, TRAIL Thesis Series, the Netherlands

Risto, M., *Cooperative In-Vehicle Advice: A study into drivers ability and willingness to follow tactical driver advice*, T2014/10, December 2014, TRAIL Thesis Series, the Netherlands

Djukic, T., *Dynamic OD Demand Estimation and Prediction for Dynamic Traffic Management*, T2014/9, November 2014, TRAIL Thesis Series, the Netherlands

Chen, C., *Task Complexity and Time Pressure: Impacts on activity-travel choices*, T2014/8, November 2014, TRAIL Thesis Series, the Netherlands

Wang, Y., *Optimal Trajectory Planning and Train Scheduling for Railway Systems*, T2014/7, November 2014, TRAIL Thesis Series, the Netherlands

Wang, M., *Generic Model Predictive Control Framework for Advanced Driver Assistance Systems*, T2014/6, October 2014, TRAIL Thesis Series, the Netherlands

Kecman, P., *Models for Predictive Railway Traffic Management*, T2014/5, October 2014, TRAIL Thesis Series, the Netherlands

Davarynejad, M., *Deploying Evolutionary Metaheuristics for Global Optimization*, T2014/4, June 2014, TRAIL Thesis Series, the Netherlands