Synchronization Control of Perturbed Passenger and Freight Operations

Qu, W.

DOI
10.4233/uuid:e3b965d3-ffa4-4576-9e0a-55d6140fe61a

Publication date
2021

Document Version
Final published version

Citation (APA)
Qu, W. (2021). *Synchronization Control of Perturbed Passenger and Freight Operations*. [Dissertation (TU Delft), Delft University of Technology]. TRAIL Research School. https://doi.org/10.4233/uuid:e3b965d3-ffa4-4576-9e0a-55d6140fe61a

Synchronization Control of Perturbed Passenger and Freight Operations

WENHUA QU

Wenhua Qu: *Synchronization control of perturbed passenger and freight operations*,

Synchronization Control of Perturbed Passenger and Freight Operations

Proefschrift
ter verkrijging van de graad van doctor aan de
Technische Universiteit Delft,
op gezag van de Rector Magnificus
Prof.dr.ir. T.H.J.J. van der Hagen,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op
woensdag, 7 juli, 2021 om 15:00 uur

door

Wenhua QU

Master of Engineering in System Engineering
Beijing Jiaotong University, China
geboren te Shandong, China

Dit proefschrift is goedgekeurd door de promotoren.

Samenstelling promotiecommissie bestaat uit:

Rector Magnificus	voorzitter
Prof.dr.ir. L.A. Tavasszy	Delft University of Technology, promotor
Dr. J. Rezaei	Delft University of Technology, copromotor

Onafhankelijke leden:

Prof. dr. E. Ballot	Mines ParisTech
Prof. dr. R.A. Zuidwijk	Erasmus University Rotterdam
Prof. dr. R.M.P. Goverde	Delft University of Technology
Prof. dr. R.R. Negenborn	Delft University of Technology

Overig lid:

Prof. dr. F. Corman	ETH Zurich

This thesis is the result from a project funded by the China Scholarship Council [Grant number 201407090053], the department of Engineering Systems and Services, and the department of Maritime and Transport Technology at the Delft University of Technology.

TRAIL Thesis Series no. T2021/23, the Netherlands TRAIL Research School

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

ISBN: 978-90-5584-298-8

Cover design: Marco ten Eikelder

Printed in the Netherlands

To my parents

CONTENTS

1

INTRODUCTION

1.1 BACKGROUND

The transportation industry has seen a massive expansion during the last decades. For example, the passenger transport (passenger-kilometres) in Europe has increased by 27%, meanwhile the freight transports (tonne-kilometres) experiences a 28% increment from 1995 to 2016 (EuroStat, 2017, Fig 2.1.2: Transport Growth EU-28). Consequently, transportation networks are more sensitive to perturbations such as vehicle breakdowns, late releases of shipments and fluctuations of passenger or freight flows. These perturbations repeatedly occur in daily operations and cause cumulatively problems to the clients and the operators. The resulting consequences to clients can decline the service level such as extra travel time for passengers and delayed delivery of shipments. At the same time, service providers generally experience higher operating costs to deal with the perturbations and maintain the service.

Researchers, policymakers and transport planners put effort to improve the performance of passenger and freight traffic via various types of planning. There are three levels of transport planning, i.e., the strategical, the tactical and the operational planning, regarding the length of planning time (Cacchiani et al., 2014, SteadieSeifi et al., 2014). The first deals with long-term (years) investment decisions such as location choices and infrastructure arrangements. The tactical planning carries out monthly or at least weekly arrangements based on strategic planning. It concerns the routes, the service choices and possibly the corresponding frequencies and capacity. The operational level planning is generally conducted on a daily basis. It requires to consider the details of the network such as real-time available capacity and flows. Common examples of strategic planning are congestion toll and carbon tax, which are used as price leverages to promote public transport. From a tactical level, the railway industry schedules passenger-centred timetables (Robenek et al., 2018) to improve service quality. Those strategies and tactical plannings are advantageous when properly executed.

In the real-time operations, perturbations inevitably occur and causing fluctuations to the pre-designed planning. Those perturbations not only rise from the external environment (extreme weather caused hazards, unexpected additional or

Table 1.1: Three levels of transport planning.

	passenger transport	freight transport
strategical planning	carbon tax, congestion toll	carbon tax
tactical planning	scheduling	stochastic planning, robustness planning
operational planning	rescheduling	synchromodality

cancelled freight demands, delays of the shipment release), but can also from the internal system (congestion, breakdown and human-caused malpractices). As a result, the passenger experience late arrivals, broken connections and unexpected additional transfer time at intermediate stations. Freight clients suffer late delivery or very expensive last-minute switch to trucks, which will lead to more operating costs and emissions. Thus it is crucial to improve flexibility and synchronization of all the involved parts of the network at the operational level to maintain both passenger and freight flows. The focus of this thesis will be on the synchronization of flows to deal with perturbations for both passenger and freight transport networks. Synchronization is one of the key concepts in operational level management of transport flows – in innovative freight systems like synchromodality and physical internet, and in traffic management for passenger transport.

In this introductory chapter, we explain our approach to investigate the synchronization of flows in scheduled services for passenger and freight transport. Section 1.1 introduces the research scope and the objective in the corresponding networks. Section 1.2 proposes the research questions, which followed by the model approaches in Section 1.3. In the end, Section 1.4 gives an outline of the thesis.

1.2 RESEARCH SCOPE AND OBJECTIVE

This thesis aims at developing mathematical models and solving methodologies to support the synchronization control of both passenger and freight transport networks at the operational level. In real practice, these two flows co-exist in one transport network. However, the existing studies in this field explore only either passenger (Schöbel, 2009, Dollevoet et al., 2012) or freight flows (Hu et al., 2018) individually. The few studies motioned both flows just treat the freight trains as a lower priority yet ignore their attributes and the requirements of the served clients (Törnquist and Persson, 2007, Godwin, Gopalan, and Narendran, 2007, Corman et al., 2011). One of the contributions of this work is that we take into account both flows in the transport replanning while considering their characteristics. Our objective is to design synchronization approaches for dealing with perturbation of operations, for different types of networks. The thesis includes a literature review and considers three different contexts: urban railways, general mainline railways and a multimodal freight transportation network, where freight shipments can also move by road and inland waterways, as an alternative to rail transport.

In the following, we introduce the topics as also briefly listed in Table 2.

Table 2: Thesis layout

Chapter	Research Scope	Served Clients		Transport Modes		
		P*	F*	R**	B**	T**
2	literature review	✓	✓	✓	✓	✓
3	urban railways (metro)	✓		✓		
4	mainline railways	✓	✓	✓		
5	multimodal freight		✓	✓	✓	✓

P*: Passenger flows; F*: Freight flows

R**: Railways; B**: Barge; T**:Trucks or private cars

1.2.1 *Wait or depart decision about the transfer connection of the urban railway network*

The urban railway network (metro) plays a crucial role in the public transport system. It benefits from high punctuality rate, large volume, and low emission. However, it cannot provide a point-to-point service for all users. As a result, the passengers have to make several transfers between different lines. Thus providing a smooth transfer connection is important to engage more passenger flows.

The transfer connection is also considered in real-time traffic control. When delays happen, the pre-planned connections can be broken. The dispatcher then has to decide to maintain or cancel the connection. In the first situation, the connecting trains will be held until the passengers get on board. As a result, other trains (passengers) will get extra waiting delays. Besides, the delays will further propagate to other trains (passengers) and cause a domino effect of wait-depart decisions. In the second situation, the currently delayed passengers have to wait for the next connecting train, while the other trains(passengers) will not be affected. Thus the wait depart decision is a trade-off between travel time of the currently delayed trains (passengers) and the others in the network. The objective can be minimizing the total train delays or the total passenger delays. We study this optimization problem with a metro network which serves massive passenger flows. In such a way, the effect of wait or depart decisions to the overall level of service can be tested.

1.2.2 *Simultaneous traffic control of passenger and freight railway network*

An general mainline railway network consists of both passenger trains and freight trains. The two kinds of vehicles are operated based on a pre-designed schedule. When delays occur (or a delay occurs) to the network, the original plan can not be maintained due to the limited capacity of the infrastructure and strict safety requirements. Real-time dispatch methods consist of re-timing (re-planing the departure time), re-ordering (re-planing of the departure or arrive orders), rerouting, local rerouting and cancellation of services.

The two traffic flows have different characteristics, which impacts the corresponding traffic control targets. The passengers experience the whole trip as a service throughout the transport process. Evaluation criteria include the transfer connection, the waiting time at immediate stations, the degree of crowdedness, and so forth. Meanwhile, the most important task for freight transport is to accomplish the delivery before the due time. For the simultaneous real-time traffic control, it is important to balance and integrate these two different flows. We address this optimization of heterogeneous flows in a core part of the Dutch railway network which is engaged with both passenger and freight traffic. In this way, we test the benefits of a higher level of flexibility to the freight trains.

1.2.3 *Synchromodality*

A synchromodal network is a freight transportation network which allows flexible and adaptive deployment of different modes (at least two) of vehicles, organized by a centralized central coordinating agent. There are various stakeholders participate in the transport process, i.e., the shippers, the freight forwarders, the intermodal freight transport operators, the terminal operators and port authorities. If these service providers do not cooperate, each member has its own goal and policy and shares limited information with others. In that situation, the communication can be slow in reaction to cope with the uncertainties.

The centralized coordinate agent is often an innovative logistic service provider (LSP). The shippers can decide the due date but not the transport mode. Thus, the LSP can react independently based on real-time operational circumstances. The operation objective consequently is to make the optimal solution for the overall network. The objective is to minimize the operation cost to finish the already known freight tasks using pre-planned vehicles. We study such a network in the pressing circumstance of perturbations, which force a replanning of freight shipments and tests the adaptivity of the system.

1.3 RESEARCH QUESTIONS

The main objective of this thesis is to formulate models of the real-time traffic control to deal with perturbations in transport networks in which passenger and/or freight traffic are involved. The serviced entities are distinct within different networks, the involved modes of vehicles also diverse. Thus the corresponding research questions and model approach will vary with each network. Below we consider the following research questions (RQ) for each study.

General RQ: *How can the multi-class transport networks best respond to flow perturbations with real-time flow management?*

When perturbations inevitably happen to a transport network, a provisional plan is needed to continue to complete the transportation tasks. The provision plan can be affected by what the control objective is, who the involved stakeholders are and what dispatch methods can be used.

The control objective is the primary concern, which can be operator-oriented or client-oriented. For the first situation, the operators investigate how to minimize the operating costs or how to decrease the recovery time. For the second situation, the objective can be how to finish the transportation tasks (transit of passengers or/and delivery of shipments), or how to maintain the level of the service (e.g., less congestion, higher punctuality rate).

The second aspect concerns the stakeholders involved in the transportation network. Some networks serve one type of clients independently, for example, a metro network; some networks simultaneously work for passenger and freight flows, for instance, a mainline railway network. When perturbations happen and disorder the original plan, the provisional plan decides how to prioritize the interests of different clients. At the same time, some networks deploy one type of vehicles, such as the rail-bounded network; while the multimodal networks engage more than one mode of vehicles, including barge, rail and trucks. In the case of perturbations, the organizing agent considers which modes to use and how to align the chosen modes in time.

The last issue is the type of dispatching methods. The dispatching methods for rail-bound networks can be re-timing, re-ordering, re-routing and cancellations of the trains. The replanning decides what methods can be adopted when to apply these methods and how to integrate these methods. Even though the applied networks differ in the extent of their component vehicles and engaged clients, throughout the thesis, the overall topic is how to synchronize involved infrastructures, service providers and applied methods to complete the transport tasks to deal with the unexpected fluctuations. The following sub research questions explore in detail how to achieve the synchronization control in corresponding networks.

RQ 1: *What is the focus of the existing literature in rail-bound networks and what are the promising directions for further research?*

We focus on the academic literature that concerns real-time traffic management of railway networks. The purpose is to investigate what their studied traffic flows are, what the research objective is and what modelling approaches and methodologies are used.

According to Gleave. et al., 2015, European shippers often prefer road than rail even though the latter one is an environmentally friendly transport mode. With the increase of freight transportation demand, cargo trains can attract more freight flows if a higher service level can be achieved. To do so, the reliability of services, that is, on-time reliability should be improved. In that regard, we aim to review how the freight trains are currently treated in the existing studies. And related to that, how do replanning models deal with the balance of passenger and freight delays in an overarching network.

RQ 2: *How to improve the service level to passengers based on the previous studies and what can be the criteria to indicate the service level?*

Based on the literature review, it can be summarized that there are mainly two main control objectives of the previous studies, i.e., (1) minimizing the overall train delays, or (2) minimizing the overall passenger delays. The first objective considers only the negative impacts of the delays on the trains; while the second one evaluates the effect of delays to the overall passengers in the network. Thus the second objective can improve the service level to the passenger, however possibly cause extra delays to some trains. Still, it simply calculates the delay per individual passenger, thus ignoring the distinctions among different groups of passengers.

Many elements affect passengers' experiences during the delays such as the purpose of travelling, level of crowdedness in the carriage, broken connections and many more. For example the commute passengers, which are common for a metropolis metro network, experience the delays worse than the tourists passengers. Hence new methods should be introduced to analyze the level of discomforts of the delays to a different group of passengers. In this way, the overall passenger delays can be evolved to the overall passenger costs, which can further downplay the negative effects on passengers.

RQ 3: *How to integrate the freight flows with the passenger flows in the real-time traffic control while considering their specific characteristics and operating methods?*

To integrate the freight flows, the replanning has to think about (i) what the service objective of the freight trains is, and (ii) what the differences of freight trains from passenger trains are. Regarding the served clients of freight trains, the primary concern is the final delivery time at the destination. The delays can affect the delivery time and causing economic loss, depending on the amount and value of the freight. As to the differences from the passenger trains, the freights (goods) do not have the sensation of discomforts as explained in RQ 2. The shippers can let the service providers decide which vehicles to use and what route the cargo takes. Concerning the operation of cargo trains and the freights, the freight trains do not have to rigorously stick to the pre-designed time and routes in case of delays. As known from the literature study, the bottleneck areas are extremely difficult to handle in the replanning. The replanning can explore how to take advantage of this difference and release the capacity of preserved bottleneck area.

RQ 4: *How to model the cooperation among the service providers and the overall flexibility of synchromodality?*

Syncrhomodality can be interpreted as synchronized intermodality. This can be understood as the coordination among different service providers, as well as a flexibility of the service flows towards the shipment flows. Thus the first question of the replanning concerns how to achieve the coordination since some of them (trains) are already scheduled and some are not. This also involves how to deal with the operating costs when coordinating the different service modes, since the corresponding fixed costs and unit costs are considerably different.

The second question is how to synchronize the service flows with the shipment flows. The synchromodality can take reference from the passenger traffic management in the way how to decide rescheduling according to the passenger flows. At the same time, the replanning needs to take care of the differences in shipment flows, since freight has batches of containers that can be split and need to be re-merged.

1.4 RESEARCH APPROACH

Our methodology is based on transport optimization techniques used for mathematical modelling and analyzing of the transportation replanning operations.

The answer to acknowledge RQ 2 to RQ 4 is based on the integer linear programming to model the transport network. Level of detail of the formulation is chosen to fit on the studied network and the controlling object. For example, we used the mixed macro-microscopic level of modelling to describe the railway infrastructure and trains movement therein in Chapter 4. The macroscopic formulation describes the network from a higher level of aggregation, based on which we can explore a broader area. The microscopic formulation describes detailed, including the control signals, which secures a conflict-free solution. We build replanning models in the same way also for other networks.

For real-time traffic control, an instantaneous solution is necessary. We use commercial solvers to get quick solutions when they can be obtained instantly. In the cases where the calculation is too complex to be solved rapidly enough, we adopt algorithms to provide a near-optimal but fast solution.

The developed control methods are tested on practical networks in control and reducible way. Besides, we evaluate the output of the models and the solving methodologies using key performance indicators (KPIs) in the corresponding network. We also compare our replanning methods with benchmark methods from existing studies or theories to gain insight into their applications. In the following section, we present an outline of the remaining chapters of this thesis.

1.5 THESIS OUTLINE

The outline of this thesis is shown in Figure 1.1. Following this introductory chapter, Chapter 2 presents an overview of existing studies about real-time traffic management of the railway network to deal with disturbances. We study the literature according to the levels of modelling, the dispatching methods, and the controlling objective. Based on the system overview, we conclude the broad set of research challenges in this area, including the research reported in the thesis on the research direction and corresponding modelling approach for the following research.

Next in Chapter 3 we study the synchronization control of urban railway network (URN). We propose a transfer synchronization control model (TSCM) with the consideration of congestion cost caused by the over-saturated passenger flow. The movement of rolling stock and the status change caused by trains' arrivals or departures can be represented by a discrete event dynamic system (DEDS) adopting the max-plus algebra to solve this model. The test case shows that the 'wait or depart' decisions and the train's corresponding maximum waiting time change with different passenger flow density. It indicates that the decisions to determine where or when to hold trains can be adjusted according to the congestion cost to improve the service to passengers.

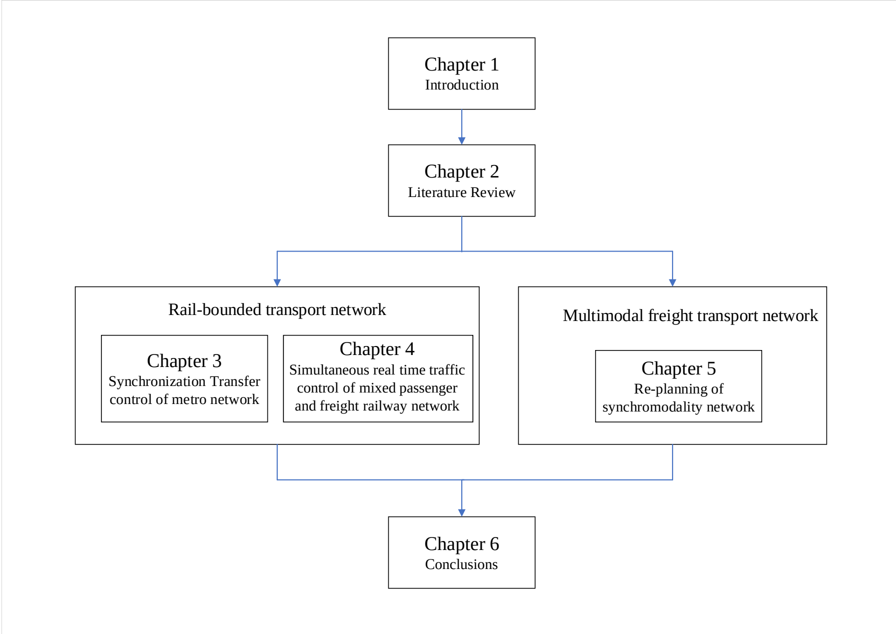

Figure 1.1: The simultaneous synchronization control of passengers and freight transport network to deal with perturbations.

Chapter 4 investigates a simultaneous rescheduling control for mixed passenger and freight traffic of the railway network. The simultaneous rescheduling problem is formulated via a mixed macro-microscopic modelling. The heterogeneous traffic is treated with their characteristics in the rescheduling process. The rerouting of the freight trains is studied as a solution, which can release the limited capacity of the bottleneck area. The mixed-integer programming is calculated using CPLEX solver. The results show an interesting gap between re-routed freight trains resolution and non-rerouted freight trains resolution in terms of the total train delays.

Chapter 5 introduces a re-planning model for hinterland freight transportation under the framework of synchromodality. The transportation replanning benefits from high operational flexibility and coordination. Flexibility is increased by allowing individual containers from single shipment to travel over different routes. Coordination gets improved by aligning departure times of service flows with the shipment flows. We implement these principles in a new transportation re-planning model for an independent, cross-modal Logistic Service Provider (LSP). We demonstrate the advantage of the model for the case of Rotterdam hinterland network. Besides, we highlight the benefits of flexibility in the load and time dimension over rigid transport operations. The results show that splitting and bundling of shipment gives a higher degree of freedom to enable better usage of large capacity services (LCS). Meanwhile, the rescheduling of transport modes at terminals synchronizes the service flows with the re-routed shipment flows, leading to mode shift from trucks to LCS and further utilization of assets.

Finally, in Chapter 6 we draw conclusions and outline avenues for future research.

Part I

RAIL-BOUND NETWORK TRAFFIC MANAGEMENT

This part concerns rail-bound network traffic management. We review traffic management of the rail-bound networks which handle disturbances. Based on that, we develop a delay management model for a metro network which deals with massive passenger volumes. Afterwards, we propose a real-time traffic management for a railway network with mixed passenger and freight flows.

2

A REVIEW OF REAL TIME RAILWAY TRAFFIC MANAGEMENT DURING DISTURBANCE

This chapter is reproduced from Qu, Corman, and Lodewijks, 2015:

W. Qu, F. Corman, G. Lodewijks, *A Review of Real Time Railway Traffic Management During Disturbances*, in Computational Logistics, Springer, pp. 658-672 (2015).

ABSTRACT

This paper gives an overview of real-time traffic management of the railway network in case of disturbances. After briefly introducing the problem of disturbance management and basic mathematical formulations, this paper overviews the existing literature according to the typologies of traffic and levels of detail in the infrastructure models used for railway traffic network representation. A precise placement is made based on the effect of management decisions towards the various stakeholders. The application of these models in real life railway systems is discussed based on the special constraints considered, the size of the railway network and the calculation time. Most railway disturbance management models are tested in an experiment setting at present, and if applied in practice they can be helpful to dispatchers to provide a higher quality service for all stakeholders involved.

2.1 INTRODUCTION

Railway network operates according to a pre-planned timetable, which specifies the route choice of the trains through the infrastructures and regulates the precise time slot of the trains' departure or arrival. Intercity trains or express trains stop mainly at big stations, regional trains or regular trains stop almost at every station. In the stage of timetable designing, in order to provide convenient to passengers, and coupling of express trains and regular trains are well considered at stations where passengers' transfers take place, not only at "hub" stations. The occurrence of delay and its propagation can be reduced by making delay-resistant timetables or improve the robustness of a railway system, see Liebchen and Stiller, 2009, and Dewilde, 2014.

However, there are always some unavoidable disturbances causing deviation from the original timetable, which calls for more effective real time trains management. The real time traffic management in case of disturbance (RTTM-disturbance) should make a decision whether to maintain or drop the pre-planned connections in the time-table. Besides, the management needs to deal with possible conflicts resulting from delayed trains require access to infrastructures pre-assigned to another train at the same time. It takes the current train time-space position and the traffic regulations as input, gives out a series of corresponding adjustments to the train schedule before execution, and aims at bringing the traffic status back to the normal timetable as soon as possible and limiting the economic loss caused.

Except for very few applications, in practice RTTM-disturbance is usually performed by dispatchers. Due to the complexity of the problem, dispatchers utilize some simplifying rules to implement their decisions and resolve conflicts accordingly. From the view of system effectiveness and efficiency, their decisions should be supported with appropriate tools because their immediate decisions may cause considerable train delay propagation in future interference. Many scholars have studied the RTTM-disturbance problem in many ways and from various perspectives over the years, to assist the decision makers to make more effective decisions, avoiding a suboptimal through put out. The application of those approaches in practices constitutes a scientifically challenging and promising perspective.

The existing researches use different approaches to address models and algorithms for RTTM-delay. A recent state-of-the-art paper about recovery models and algorithm for real-time railway rescheduling is Cacchiani et al., 2014. This review differs from the existing review articles in the following ways.

1. We concentrate on the RTTM in case of disturbance, and waive the questions concerning the RTTM-disruption. Delays can be categorised into disturbances and disruptions, however there is not a sharp distinction between disturbances and disruptions in term of time length. According to Cacchiani et al., 2014, disturbances are relative small perturbation to the railway system that can be handled by modifying the timetable, but without modifying the duties of rolling stock and crew. Disruptions are relatively large incidents, requiring both the timetable and the duties for rolling stock and crew to be modified. Railway operating companies are mostly faced with disturbances, instead of disruptions. For more information on railway disruption management, please refer to Jespersen-Groth et al., 2009.

2. The literature on RTTM-disturbance is reviewed from the perspective of the classes of traffic, i.e., distinguishing between passenger trains, freight trains, approaches considering unspecified traffic, or explicitly both at the same time, as shown in Figure 2.1. Here the unspecified traffic means the train types are not emphasized and distinguished in the models. Passenger and freight trains should be distinguished be-cause they differ a lot in served clients, body structures, organization of train operations, etc. Since they might share the same corridor, there is interplay in terms of infrastructure and time. It

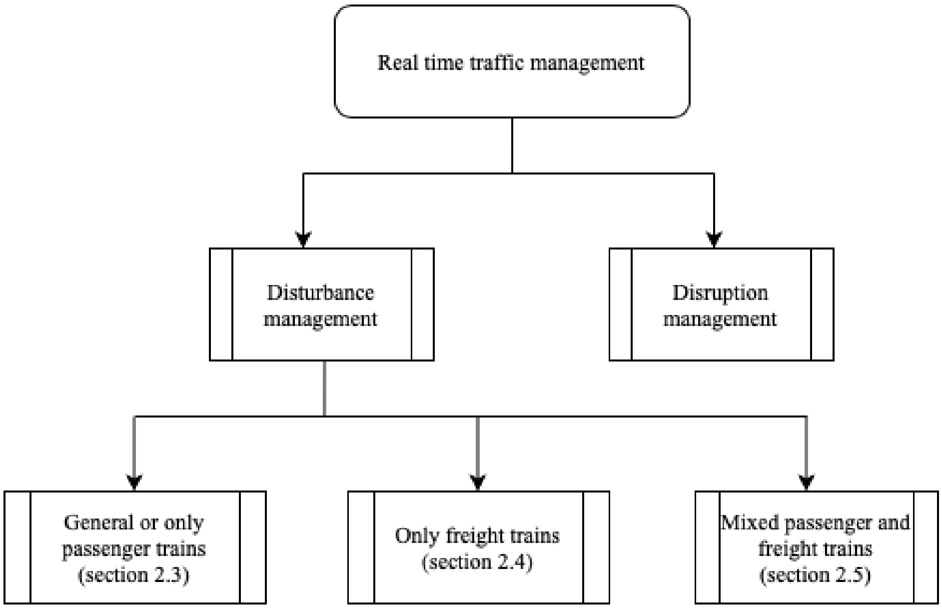

Figure 2.1: Basic categories of the RTTM according to the train types

is worthwhile studying how to integrate different types of train traffic, with respect to the diversities.

3. We discuss the dispatch decisions from the point view of different stakeholders. Early studies mainly aim at bringing down the overall train delays yet ignore the service quality perceived by passengers. In the last decades, scholars start to focusing on the passengers' interest, addressing the management and keeping of passenger connections, discomfort caused by congestion, value of time, etc. In addition, some RTTM models try to get a balance between the train delays and passenger delays.

2.2 BACKGROUND INFORMATION

The RTTM-disturbance strategy can be retiming, rerouting, reordering, and cancelling the connections, explained in Sec. 2.2.1. And Sec. 2.2.2 introduces the commonly used mathematical formulations.

2.2.1 *Operation strategy*

In case of disturbance, the traffic deviate from the original timetable and the timetable need to be up-dated to a new one, the so-called disposition timetable. The problem of finding a disposition timetable is also called re-scheduling, which may consist re-timing, rerouting, reordering or cancelling the passenger connections.

- **Re-timing**: Decision makers may adjust the time slots of one train entering or leaving one block section, arriving or departing from the stations.

- **Rerouting**: Decision makers may assign a new route from a set of feasible routes (inside one station or between stations) to a train, instead of executing the train's pre-planned route.
- **Reordering**: Decision makers may change the order of a pair of consecutive trains' access to the infrastructure.
- **Cancelling connections**: If the pre-planned passenger connection is dropped, there is no adjustment to the movement of the feeder trains. The decision of maintaining or dropping a connection (or wait-no-wait decisions) is called pure delay management.

2.2.2 *Mathematical Formulation*

The RTTM-disturbance model can be divided into microscopic model and macroscopic models, according to the different levels of detail in the infrastructure models used for railway traffic network representation. Macroscopic models consider the railway network at a relative high level, and have a more aggregated representation of some resources, i.e., stations are represented by nodes of a graph and tracks by arcs. Besides, signals are not considered. Most macroscopic models use *Event-Activity* mathematics to formulate the railway network.

While in microscopic models, the above aspects are considered in detail, and they include a lot more detail (e.g. a station is composed of a complex set-up of pieces of tracks separated by switches and signals). Most microscopic models use *Alternative Graph* mathematics to formulate the railway network.

2.3 RTTM-DISTURBANCE OF GENERAL OR ONLY PASSENGER TRAINS

Sec 2.3 discusses the RTTM-disturbance literature of general or only passenger trains, first microscopic and then macroscopic models. The basic structure is listed in Table 2.1 and Table 2.2. Some papers in this section do deal with the freight trains, however, they do not distinguish passenger trains (high speed or normal ones) and freight trains in the model. All trains are taken as the same in terms of value of time or travel time reliability. These models do not discuss the differences of passenger and freight transportation, such as the carrying vehicles, clients' perception and evaluation to the transportation service. The consequential differences, such set of priority in access to the infrastructures, are not discussed. Therefore, we classify them as models on general train flows.

2.3.1 *Microscopic RTTM-Disturbance Model of General or Only Passenger Trains*

Much of the literature at a microscopic level are based in the Alternative Graph model introduced in Section 2.2.2.

The delay management is a series of adjustment decisions, which can affect both the trains and passengers. If all connecting trains wait for the delayed feeder trains to provide passengers with smooth transfers, which however increases the overall train delays. On the other hand, if all connecting trains depart punctually to make an early arrival, passengers missed their connections and total passenger waiting time will be enormous. According to the stakeholders involved in the wait or no wait decisions, the existing literature are sorted according to the objective functions: trains orientation (i.e., aim at minimizing the total train delays), passenger orientation (i.e., aim at minimizing total passenger delays) or a trade-off between the two criteria.

Train Orientation. D'Ariano, Pacciarelli, and Pranzo, 2007 studies the train RTTM-disturbance on a regional railway network, by giving new conflict-free timetable of feasible arrival and departure times. It is modelled with an alternative graph formulation with *no-store constraints*. No-store constraints mean that a train, which has reached the end of a track segment, cannot enter the subsequent segment if the latter is occupied by another train. This prevents other trains from entering the former segment. A branch & bound (B&B) algorithm is adopted to speed up the computation time. The lower bound is achieve from Jackson pre-emptive schedule in Jackson (1955), and the upper bound as the best value obtained by three heuristics: "First Come First Served" dispatching rule, "First Leave First Served" dispatching rule and greedy algorithm AMCC described in Mascis and Pacciarelli, 2002. The objective function is to minimize the deviation between the disposition timetable and the original timetable at the last stop of the trip. Computational experiments based on the Schiphol area, a bottleneck of the Dutch railway, show that the truncated version of the algorithm provides proven optimal or near optimal solution within short time limits (120 seconds).

Table 2.1: Basic category of the microscopic RTTM-disturbance models according to train types.

Objective functions	Special constraints	Related Literature	Test network	Calculation time
Train orientation	train retiming	D'Ariano, Pacciarelli, and Pranzo, 2007	Schiphol, 2h (20km)	<120 s
	train retiming & rerouting	D'Ariano, Pacciarelli, and Pranzo, 2008	Utrecht Den Bosch	<30 s
		D'Ariano et al., 2008	Schiphol	<30 s
		Corman et al., 2010	Utrecht Den Bosch (1h)	<1 s
	track choices, extra-stops	Gély, Dessagne, and Lérin, 2006	Line between Tours and Bordeaux (8h)	-
	junctions areas	Rodriguez, 2007	Pierrefitte-Gonesse junction	<180 s
	congested bottlenecks	Caimi et al., 2012	Berne area (10 scenarios)	<1 min
	speed control	Albrecht, Binder, and Gassel, 2011	Utrecht to 's- Hertogenbosch	<-
Passenger orientation		Wegele and Schnieder, 2004	Deutsche Bahn AG (24 h)	-
	passenger rerouting	Corman et al., 2015	Utrecht and Den Bosch (2 h)	About 1m
A trade-off		Corman et al., 2012	Utrecht Central (1 h)	<30s

In D'Ariano, Pacciarelli, and Pranzo, 2007, trains are not allowed to change routes from the original timetable, while in practical, better solutions can be achieved by rerouting. D'Ariano, Pacciarelli, and Pranzo, 2008, D'Ariano et al., 2008 integrate trains' rerouting and extend it into a real time traffic management system ROMA (Railway traffic Optimization by Means of Alternative graphs). Still improvements are possible in term of computation time, especially when dealing with large-scale network.

D'Ariano et al., 2008 put out "flexible timetable" in comparison with "rigid timetable", by relaxing some timetabling constraints in the ROMA model. The authors construct a flexible timetable by replacing the scheduled arrival and departure times with maximum arrival and minimum departure times. Experimental results are performed on the congested area around Schiphol Amsterdam Airport with randomly generated disturbances. The results show that flexible timetables are promising. The best solution is always found within 30 s of computation time.

RTTM differs from the timetabling significantly in the required time, since that RTTM needs to bring the traffic status back to the normal timetable as soon as possible. On the basis of ROMA, Corman et al., 2010) develop a Tabu Search (TS) scheme to address the train conflict detection and resolution (CDR) problem. Tabu Search improves the calculation speed of the local research method by using memory structures called Tabu list which describe the visited solutions. The new TS algorithm can enhance significantly the performance of ROMA.

Gély, Dessagne, and Lérin, 2006 present a quite detailed mathematical model to describe the RTTM-disturbance problem. The model presents continuous variables (arrival or departure times of trains from each visited each node at a given order), track choices, and extra-stops. The objective is to minimize the total accumulated delay of trains. They use evolutionary algorithms and hybrid techniques to solve the problem. Test based to real-life instances of railway lines between Tours and Bordeaux of the French railway company between 15:00 and 23:00 shows the feasibility and the effectiveness of decision support systems.

High traffic density and heterogeneous traffic networks make the railway operations more sensitive to delays. There have been increasing studies on congestion related RTTM.

Small disruptions are amplified through the junction area. Rodriguez, 2007 points out that a disturbance which is at origin only a few seconds long can quickly lead to a delay of over 5 min. Thus, the researcher presents a constraint-programming model for rerouting and rescheduling of trains at a junction area. The model has been applied to a set of problem instances. Preliminary test results based on a real case study of traffic on the Pierrefitte-Gonesse node, North of Paris show that the model yields a significant improvement in performance within an acceptable computation time.

A special concern can be the congested bottleneck areas, where delays easily propagate from one train to another. Caimi et al., 2012 propose a dispatching assistant system in the form of a model predictive control framework for railway traffic management in bottleneck areas, aimed at maximizing customer satisfaction. In particular, they propose a closed-loop discrete-time control system, which suggests rescheduling trains according to solutions of a binary linear optimization model. The system is tested in collaboration with the Swiss Federal Railways, and successfully applied for an operational day at the central railway station area. The computation time is less than 1 minute.

Albrecht, Binder, and Gassel, 2011 introduce the concept of target points and target windows to distinguish and specify the management operations in stations dealing with different traffic intensity. Target points correspond to arrival and departure times for the purpose of maintaining passenger connections. When there is no need to consider connections (often in minor stations), target windows can be considered. Target windows impose an upper limit within which the planned time can be exceeded by a delay. The authors propose a two-level optimization approach: optimization of speed between consecutive target points, and optimization of running times between target windows, if target windows exist between the target points. Case study on parts of the Dutch and German railway networks shows that the method is able to improve timetable adherence, to save energy, and to improve throughput in the system-related bottlenecks.

Passenger Orientation. Most early papers aim to minimize the train delays or the number of cancelled trains. However, a growing number of papers begin to consider minimizing the negative effects of disturbances for passengers. Wegele and Schnieder, 2004 use genetic algorithm to achieve objective function of minimizing customers' annoyance, e.g., delays, change of platforms and missed connections. The method is tested with a real data from part of the German Railways. The evaluation runtimes in both cases lied at few minutes. Corman et al., 2015 integrate microscopic railway traffic rescheduling and passenger point of view into a single mixed integer linear programming. The objective function is to minimize the time spent by the passengers in the railway network. They use a fast iterative algorithms based on a decomposition of the problem and on the exact resolution of the sub-problems. Computational experiments based on a real-world Dutch railway network show that good quality solutions and calculation time within a limited computation time.

A Trade-Off. Transfer connections are relevant to the passenger satisfaction but do not affect the feasibility of railway operations, therefore one of the possible dispatching countermeasures to reduce delay propagation is to cancel some of scheduled connections. This operation reduces overall train delays but cause an extra delay to the passenger because of the missed connections. Train operating companies are clearly also interested in keeping as many connections as possible even in the presence of disturbed traffic condition.

As a series of coherent research work, Corman et al., 2012 combine the microscopic formulation with the minimization of passenger dissatisfaction in Ginkel and

Schöbel, 2007 (introduced in Sec 2.3.2). The authors consider the needs of different stakeholders (infrastructure company, train operating companies, passengers, etc.), and get a compromise solution between the minimization of train delays and missed connections. Test based on a complex and densely occupied Dutch railway network put out accurately the Pareto front in a limited computation time.

2.3.2 *Macroscopic Model of General or Only Passenger Trains*

In this section, we discuss RTTM-disturbance at a macroscopic level. Also the literature is sorted from the point view of train orientation or passenger orientation.

Train Orientation. Schöbel, 2009 provides a disturbance management model including the track capacity constraints. The author develops a branch-and-bound algorithm and several heuristic approaches to solve the problem. The macroscopic approach allows treating the most important capacity constraints using headway time constraint. The algorithmic approaches have been tested at a real-world data from the region of Harz, Germany with a calculation time within seconds.

Dollevoet et al., 2015 point out the limited capacity of the stations in disturbance management problem. They develop an integer programming (IP) formulation that takes into account the capacity within stations and allow rescheduling the platform track assignment. First, a fixed platform track assignment is used, and then the platform track assignment is improved systematically, using iterative heuristic.

Min and Wynter, 2011 propose a MIP formulation to solve the RTTM-disturbance problem in Seoul Railway, South Korea. They prove that when the railway network gets as large as a metropolitan area network, the computation times can be extensive and fluctuate significantly on the instances. They proposed a heuristic Fix and Regenerate Algorithm that exploits the structure of the problem iteratively. Tests based on the Seoul metropolitan railway network (passenger trains, subway trains, and freight trains) shows that the model provides a new timetable within 1 min.

Table 2.2: Basic category of the macroscopic RTTM-disturbance models according to train types.

Objective functions	Special constraints	Related Literature	Test network	Calculation time
Train orientation		Min and Wynter, 2011	Seoul metropolitan railway network	<1 min
		Dündar and Şahin, 2013	single track in Turkey (150 km)	not mentioned
	Track capacity	Schöbel, 2009	Deutsche Bahn	within seconds
	Station capacity	Dollevoet et al., 2015	Zwolle to Utrecht network (1 h)	<3 min
Passenger orientation		Schöbel, 2001	Rheinland-Pfalz and the Saarland	within seconds
	Passenger rerouting	Dollevoet et al., 2012	Zwolle to Utrecht network (1 hour)	<1 min
	Dynamic interacting	Kanai et al., 2011	Japanese railway (130 km)	<6 min
A trade-off		Ginkel and Schöbel, 2007	Verkehrsverbund Rhein-Neckar (30 m)	<1 min

Dündar and Şahin, 2013 study real-time management of railway traffic on a single-track railway line of about 150 km in Turkey. They develop a Genetic Algorithm (GA) to reschedule the trains on this line, and in particular to determine the meets and passes of trains in opposite directions. This model can be used for solving the smaller instances to optimality. However, this model cannot solve the larger instances.

Passenger Orientation. Schöbel, 2001 firstly using an IP model the delay (disturbance) management problem, deciding which connections should be maintained and which other connections can be dropped. The objective function is to minimize the total passenger delays at their destination stations.

In a following paper, Dollevoet et al., 2012 point out that, different from the typical assumption in classic DM models, passengers can choose another route instead oftaking their original planned route and waiting for one cycle time when delays occur. They adopt an IP formulation to manage the passenger re-routing and delay management at the same time. Computational experiments based on real-world data from part of Netherlands Railways show that significant improvements can be obtained by integrating the re-routing of passengers.

Kanai et al., 2011 attach high importance of passenger dynamic interaction and level of service to the RTTM-disturbance problem. A dynamic interaction means a phenomenon such that if a train is delayed for some time, more passengers than usual get on at the next station and the dwell time becomes longer than usual and the train is delayed further. In Japan, railway companies have to densely operate express trains and regular trains and transport a massive of passengers. The authors develop a delay management algorithm combining simulation and optimization. They adopt Tabu-search algorithm to finding good strategies for the management of train connections. The model is tested on data of the Japanese railway network with promising results, with an average computation time of about 6 min.

A Trade-Off Ginkel and Schöbel, 2007 present a bi-criteria model for the DM problem, aiming at a trade-off between the delay of the vehicles and the number of passengers who miss their connections. They present an IP formulation and a graph-theoretic approach based on discrete time/cost project network. The test based on south-west Germany show the applicability of the approach.

Table 2.1 and Table 2.2 show that when the main focus of RTTM is the feasibility of railway operations, the related papers take into account the constraints on the limited capacity of the railway network disregarding the passenger inconvenience. On the other hand, when the focus is on the minimization of passenger inconvenience, the signaling is often not taken into account.

2.4 DISTURBANCE MANAGEMENT OF FREIGHT TRAINS

At the moment, there are only limited literature studding the freight delay management. However, there is an environmental and European political vision for an increased use of the railway, especially for long-distance freight transportation. Thus, it is increasingly promising to investigate the freight transport management.

2.4.1 *Differences between Passenger Transport and Freight Transport*

Different Speed. passenger trains usually run at a higher speed than freight trains, thus have speed limited by construction. Passenger trains and freight trains are marshalling separately. In timetable designing, freight trains are inserted among the fixed schedule of passenger trains.

Different Value of Time, and Travel Time Reliability. In case of disturbance, the loss caused can be different according to value of time. For detailed information, please refer to the report by De Jong et al., 2007. Based on the value of time, researchers can get a set of priority order for different train flows, as discussed in Sec 2.6.

Different Number of Stops Alongside the Trips. Normally freight transport provides a point-to-point service to clients. However, there can be frequent stops along the lines. Clients are concerned with the total delivery time, instead of the stops or changing between different lines.

Different Operations. Passengers can accomplish the boarding/alighting themselves, on the platform. But for freight transport, the cargo handlings take place at the yards. Off-peak and night-time deliveries are widely considered in Freight transportation to avoid some of the traffic network congestion at peak hours.

2.4.2 *Disturbance Management of Freight Trains*

The railway freight transportation can be affected by the regional factors, such as the transportation demand, operations, policies, etc. The European railway freight transportation tends to have a schedule, while the US or Canadian Railways would more often wait to dispatch a train, scheduling only happens when sufficiently many cars have been accumulated.

Kraay and Harker, 1995 present a MIP model for the optimization of RTTM-disturbance over the entire freight network based on the American railway networks. The model aims to provide a link between strategic schedules (which might be decided every month) and line dispatching or computer assistant dispatching models (which need the scheduled arrival and departure times of each train at the ends of each lane at very frequent intervals). The model considers the current position of each train and the relative importance of the train. The model generates the time that each train should be at each major point in its itinerary. The model is tested on a portion of a major North American railroad to ascertain its efficiency and efficacy. These results were significantly better than the initial standard heuristic at that time in many of the problem instances.

Kuo, Miller-Hooks, and Mahmassani, 2010 point out that the demand of freight transportation in North America is not fixed. Instead, it is a function of the frequency of service. The researchers develop a train slot selection model for scheduling railway services. The freight scheduling methodology equilibrates between the performance of the freight transport system (given schedules, delays and user costs under different levels of demand) and the demand (given level of service provided). The objective function is to minimize both operating costs incurred by carriers and delays incurred by shippers, while ensure that the schedules and levels of demand are mutually consistent. They propose a column generation-based methodology for train slot selection to meet the frequency requirements. Its utility is illustrated through the development of weekly train and ferry timetables for the international freight services.

Oetting and Keck, 2015 put forward an evaluation approach to assess the effect of DM measures on transportation service quality based on European freight transportation. The so called monetarization analysis of delays is a customer-based cost-benefit-analysis for operations addressing changes infrastructure, operations and vehicle characteristics. Based on expert interview and a survey with 55 companies, the study gives an approach to integrate the effects of delays in a monetarized way, i.e., resulting cost arising from the effects of delays on production and logistics. It concludes in a non-linear time-and-process dependent valuation function for delays with assigned costs caused by delays to the customer.

2.5 DISTURBANCE MANAGEMENT OF PASSENGER AND FREIGHT TRAINS

Different from Section 2.3, literature in this section clearly distinguishes passenger and freight transportation in the model.

When passenger and freight trains use the same corridor, it creates a quite heterogeneous traffic flow. In case of delays, both passengers and freight transportation are affected. Passengers experience it as late arrival, broken connections, insufficient seating capacity, cancelled trains, etc. Also freight trains suffer from delays mainly in the form of rerouted trains and late arrivals. Also they compete for the same part of the train system at the same time. Decisions have to be made about which train can go first, and thus there arise the problem of priority orders among different traffic types.

Törnquist and Persson, 2007 propose a MIP model which takes into account both passenger and freight trains. The model in a geographically large and fine-grained railway network with highly interacting traffic, consider reordering and rerouting of trains. This model can formulate a highly complex setting such as a railway network composed of segments with a large number of tracks which can be unidirectional (only permitting one-way traffic) and bidirectional. There are two goals, to minimize the total traffic delays, and to minimize the cost function based on the train delays. The cost function allocates different delay cost per minute, different tolerance time windows and different fixed penalty cost and to different passenger and freight trains. The model is tested on the southern part of the Swedish railway network. Since the model contains a lot detail and deal with large-scale network, the computational time can be long. Still this model does not give more analysis of the interaction between passenger and freight trains when they can have some conflict with the occupation of the infrastructure resources.

A prompt decision is quite important in RTTM. Krasemann, 2012 improves Törnquist and Persson, 2007 using a heuristic greedy approach. In general the heuristic provides solutions that are good enough very fast, but the author holds that it could be further improved.

Acuna-Agost et al., 2011 also extend the model presented in Törnquist and Persson, 2007 . In particular, they use hard and soft fixing of integer variables with local-branching-type cuts. Tests based on instances of French railways show a best compromise can be obtained with the Iterative MIP based local search procedure. Good solutions (average gap $<1\%$) can be obtained within 5 min.

All the above literature concerns macroscopic models. The models distinguish passenger trains and freight trains using delay cost. However, in real practices, dispatchers can always meet such difficulties, some high-speed passenger trains should recover to the origin timetable as soon as possible. At this moment, slow freight trains are not in the scope of the emergency, and their recovery can be postponed to "some other time". Corman et al., 2011 propose an innovative microscopic model to cope with the passenger and freight train rescheduling, using a set of priority. At each step, the procedure focuses on the current priority class, preserving solution quality from the higher priority classes and neglecting the lower priority classes in the optimization of train orders and times. The iterative process is implemented from the highest one to the lowest one until a new network disposition timetable is finished. They use the alternative graph to formulate and a sophisticated B&B algorithm to solve the multi-class rescheduling problem. The test result based on Utrecht Central shows an interesting gap between single-class and multi-class rescheduling problems in terms of delay minimization.

Similar studies can be referring to Godwin, Gopalan, and Narendran, 2007. They describe a situation in Indian Railway system that freight trains are routed and scheduled in a railway network where passenger trains must adhere to a strict schedule. The freight train movements can be inserted at any time or as demands arise, and should not disrupt passenger trains. They adopt a binary (0-1) MIP and a hierarchical permutation heuristic. The computational experiences for solving various problems are based on real data.

2.6 CONCLUSIONS AND FUTURE RESEARCH

This paper gives an overview of the state-of-the-art in models and algorithms for real time traffic management in case of disturbance in railway network. In particular, we describe various approaches in terms of the train types, stakeholders, and constraints.

We draw the following conclusions:

1. Due to regional differences and other reasons, most research focus on passenger traffic management. Some American and Japanese researchers study railway freight transportation management. The European ongoing deregulation has created multiple competing actors such as different train operating companies for passenger and freight traffic, competing on the same tracks. Thereby infrastructure managers are to be neutral and cope with multiple conflicting requests and demands. The majority of previous research approaches does not account explicitly for this new situation. Furthermore, the complexity of the problem of having mixed traffic is often mentioned but less investigated. This is a clear direction of future research, about simultaneous management of railway traffic of different classes, including the diversity and interplay between passenger and freight trains.

2. There are two main tendencies of real time disturbance management models for railway traffic. Microscopic models consider all elements of the railway infrastructure for computing safe movements, thus are able to check feasibility and quality of the solutions from the viewpoint of operations managers. On the other hand, macroscopic models transfer the focus from minimizing train delays to improving the quality of service perceived by the passengers, i.e., less waiting time, more convenient transfers.

3. There are several challenges left open for research when dealing with optimization in the level of service, such as management of congestion on board or at stations, passenger behaviours, etc. Passengers' behaviour means their reactions when disturbance occur and lead to a change to their pre-planned path. Will they stick to their original plans or follow the real-time suggestion for alternative modes of connectivity? The analysis of recorded passenger flows might provide insights on how passenger flows react to unexpected events as investigated, for example, by Hurk et al., 2013.

4. Improvement in calculation time can be important, to give a quick response in the real time management. There is a trend to integrate more detail (such as track capacity, station capacity, passenger behaviour, interactions, etc.) in the model, thus the RTTM becomes increasingly complex. The development of algorithms for RTTM is currently still mainly an academic field, where the research is still far ahead of what has been implemented in practice. Some approaches are implemented in practice, despite being far from the academic results.

5. Although most of the presented models show promising results in the described experiments, most researches mainly focus on instances representing relatively small railway networks (see Table 2.1 and Table 2.2). It will be quite a challenge to get the calculation results for large-scale network in short time when bring these methods into real operations.
   Nevertheless, there are signals that practice has discovered the added value that can be provided by real-time management methods, based on the successes that have been achieved by the application of optimization methods in

the railway planning stage, see Caprara, Fischetti, and Toth, 2002. Despite the large amount of research that has been accomplished in the last decades, the regular application of the research results in practice will still need quite some time.

3

SYNCHRONIZATION CONTROL OF PASSENGER FLOWS IN AN URBAN RAILWAY NETWORK

This chapter reproduced from Qu et al., 2020:

Qu, W., Wu, J.J., Corman, F., Gao, Z.Y., *Synchronization control of passenger flows in an urban railway network*, to be submitted.

ABSTRACT

This paper focuses on transfer management in cases of delays involving metro networks with massive number of passengers. When delays occur and cause broken transfer connections, dispatchers should decide rapidly either to maintain or cancel a transfer. This decision also affects the trains movements and other passengers' travel time as a ripple effect of the original delay. We construct a binary integer programming model to provide decisions makers with support for the problem under consideration. The non-linear programming is a typical discrete dynamic system that is solved with max-plus algebra. The contribution of this study lies in its full consideration of the point view of passengers via the factor waiting costs. The effectiveness of the proposed model and solution approach are evaluated using a real-world data set. The test involving passenger flows in morning and evening peak hours shows that the decisions to wait or depart decisions change significantly depending on the direction of the passenger flows. A comparison with a reference model shows that the decisions from the proposed model decisions more reasonable about how long the train in question should wait.

3.1 INTRODUCTION

Rail bound transportation networks play an important role in urban transport systems. Their advantages mainly lie in their high transport speed, large capacity, and low emissions. However, the rail bound transportation cannot provide a point-to-point service. Passengers usually need to make several transfers during one trip, which means they face additional transfers waiting time and sometimes interchanging walks, making it clear that it is important to provide smooth transfer connections, because they improve the service levels.

There are two main stages at which the transfer connections at a given infrastructure capacity can be enhanced, namely (i) scheduling planning and (ii) rescheduling control. In the former stage, the timetable planners put effort to design convenient transfer connections according to the passenger flows and their corresponding travel routes. Equally important in practice unavoidable disturbances during operations cause deviations and broken connections, in which case the original timetable can not be adhered to and the latter stage of rescheduling control comes into play. To manage the consequences of the broken transfer connections, real-time decisions have to be made.

Rail rescheduling is a complex and systematic process in which several factors play a role. Early rescheduling studies were aimed mainly at recovering to the original timetable as quickly as possible (Mascis and Pacciarelli, 2002). The main objective was to minimize the overall delays for all the trains in the network, without taking into account passenger delays. Later researchers focus more on the passenger-orientated rescheduling. For example, Schöbel, 2001 studies the wait or depart decisions with the aim of minimizing waiting times for passengers. The passengers behaviour in the transfer connections is studied in (Dollevoet et al., 2012).

However, despite the growing attention to the effect delays have on passenger waiting times, their impact of passenger flows to the real-time transfer connection control has yet to be examined. It is a relatively new phenomenon of the massive passenger flows in metropolitan cities. Most passengers commute in the morning or evening peak hours, or visit major cultural or sports activities within certain time windows. At the moment, the effect of passenger flows to a large extent is not yet modelled in to the rescheduling phase.

This study examines the real-time transfer connection control decisions in different passenger flows scenarios. We refer to this control as *transfer synchronization control*. We build an integer programming model to decide whether to maintain or cancel a certain existing transfer connection based on the current flows. To that end, we augment a standard transportation model by adding a realistic waiting cost component. This allows us to evaluate the effect of congestion to *passengers' perceptions* during the transfers. The natural way is to aim at reducing the overall *passenger waiting cost*, which is in contrast to the classical approach of merely reducing the overall waiting time. In other words, our model introduces a human component into the standard passenger-oriented transportation models. Based on the case study involving the Beijing urban railway network, this model suggests very different wait or depart decisions for the same transfer connections in the morning and evening peak hours. That is because the passenger flows for this transfer pair vary significantly in these two time windows. A comparison with a reference model that does not consider the discomfort caused by congestion shows that the proposed model proposes longer waiting times when passengers are in massive flow. The research results of this study can be used to help decision-makers to deal with delays for a heavily burdened urban railway system.

The reminder of this paper is organized as follows. Firstly, a review of previous studies is demonstrated in Section 3.2. Section 3.3 explains the formulation of train movements and passenger transfer connections for an urban rail network. The model especially proposes the waiting cost, which integrate in-carriage congestion and platform waiting coefficient. This makes the model more suitable for a metropolitan metro network. Section 4 briefly describes the max-plus algebra for the solving methodology. The discussion and analysis is given with an actual Beijing metro network in the Section 3.5. Finally, section 3.6 looks at avenues for future research.

3.2 LITERATURE REVIEW

In this section, we provide an overview of real-time traffic control and the related transfer connections. We focus particularly on (i) general real-time traffic control of rail bounded transport and (ii) transfer synchronization control dealing with large volume passenger flows. For an overview of passenger flows in the metro networks, readers can consult Goulet-Langlois, Koutsopoulos, and Zhao, 2016 and Zhu, Koutsopoulos, and Wilson, 2017.

Table 3.1 lists the studies discussed in this section. We compare the various studies in terms of their planning phase, control operations, control objective, and solution strategies.

Table 3.1: literature of studies dealing with massive passenger flows

Planning phase	Literature	Objective	Main operations	Solution approach
Strategy	Currie, 2010	Reduce peak hours congestion	finance strategy	
	Yang and Tang, 2018	Reduce peak hours congestion	fare reward scheme	
Scheduling	Wong et al., 2008	Min. passenger transfer waiting time	time adjustment	heuristic
	Niu, Zhou, and Gao, 2015	Min. passenger waiting time	timetable adjustment	optimization solver
	Niu and Zhou, 2013	Construct congestion-sensitive timetables	timetable adjustments	genetic algorithm
	Jiang, Cacchiani, and Toth, 2017	Schedule more trains to relief congestion	timetable adjustment	heuristic algorithm
	Shang et al., 2018	Min. passenger waiting cost	stop-skipping	Lagrange relaxation
Rescheduling	Mascis and Pacciarelli, 2002	Min. train delays	re-timing,re-ordering	heuristic algorithm
	Törnquist and Persson, 2007	Min. train delays	re-timing, re-routing	commercial solver
	Schöbel, 2001	Min. passenger delays	wait-depart decisions	relaxation approximation
	Schöbel, 2009	Min. passenger delays	re-ordering, wait-depart decisions	
	Zhu and Goverde, 2019	Min. disruption impacts	flexible stopping and turning	optimization software
	Gély, Dessagne, and Lérin, 2006	Min. train delays	re-ordering and local rerouting	hybrid algorithm
	Dollevoet et al., 2012	Min. passenger waiting time	wait-depart decisions passenger rerouting	commercial solver
	Dollevoet et al., 2015	Min. passenger waiting time	wait-depart decisions platform track re-assignment	iteration
	Goverde, 1998	Min. passenger delay	wait depart decisions	max-plus algebra
	Goverde, 2010	Min. trains delay	re-timing	graph algorithm

3.2.1 *General real-time traffic control of rail bounded transport*

Cacchiani et al., 2014 defines the railway rescheduling in case of (a) disturbance(s) as a range of operations designed to reschedule the timetable, which include the following:

- *re-timing*: modifying the times for the trains to enter the stations or block sections, e.g., Mascis and Pacciarelli, 2002
- *re-routing*: selecting a new route from a set of feasible routes for the trains inside or between stations, e.g., Törnquist and Persson, 2007
- *re-ordering*: changing the order of a pair of consecutive trains on a track, e.g., Schöbel, 2009
- *wait or depart decisions*: maintaining or dropping passenger connections, e.g., Schöbel, 2001

Another recent operation, which was not included in in Cacchiani et al., 2014 is that of *stop-skipping* (to skip certain stops to keep some passengers at their stations and reroute their journeys). We note here that that is used mainly for the flexible scheduling or rescheduling of metro system (Niu, Zhou, and Gao, 2015), or for a general railway having to deal with longer delays (Zhu and Goverde, 2019).

The decisions which operations to adopt can be determined by the control objective. The control objective can roughly speaking be divided into two categories: (i) train-oriented objectives and (ii) passenger-oriented objectives. The intention of the former is to minimize the caused delay propagation, as well as to return the railway system to the default timetable as soon as possible, or to minimize the number of cancelled trains. Mascis and Pacciarelli, 2002 is the seminal work on the re-timing problem. The occupation of the infrastructure by the trains is formulated as no-store constraints, which means that a job cannot leave a machine until the subsequent machine becomes available again. In this way the author formulates the re-timing problem so as to minimize the total train delays. Gély, Dessagne, and Lérin, 2006 uses the re-ordering and local re-routing (i.e. track reassignment within the station) operations. The aim is here to minimize the deviations of the actual schedule from the original plan and the costs due to violating the regularity (i.e. varying headways between consecutive departures). These studies are the early explorations about the delays management, which lays foundation for the research of the second categories.

The second category, i.e. passenger-oriented objectives, is a trend that has emerged in recent decades. Researchers have started to focus more on improving the service level delivered to the passengers. In addition to the above-mentioned operations, the wait or depart operations at transfer stations receive greater attention. Schöbel, 2001 defines the *delay management problem* as making wait-depart decisions based on a perturbed timetable for all vehicles in the network. In this way the sum of all delays for all the passengers involved is minimized. Schöbel, 2009 and Dollevoet et al., 2015 expand the *delay management problem* with the capacity of open track segment and the capacity of stations, based on the study of Schöbel, 2001. Furthermore, Dollevoet et al., 2012 and Corman, 2020 investigate passenger rerouting behaviour and its effect on rescheduling. These studies look at to the service as perceived by the passengers and keep improving that service as the public transportation develops.

Remark that the transfer connection studies are also referred to as *transfer synchronizations control* in the rescheduling (Goverde, 1998) and scheduling phase (Wong et al., 2008, Goverde, 2007). The synchronization control emphasize the concatenation of feeder and connecting trains.

3.2.2 *Transfer synchronization control dealing with large volume passenger flows*

The large volume of passenger flows is an emerging challenge, along with the rise of mega-cities. Here, we discuss how that massive volume flow is handled by the urban rail networks, with a particular emphasis on metro networks.

There are a few fare pricing schemes aiming to shift a part of the passengers to travel off-peak hours. Currie, 2010 describes the "Early Bird" scheme used in Melbourne's rail transport. The policy was that passengers could enjoy a free travel if they finish their trips before 7 AM in the morning. The Dutch Railways (Nederlandse Spoorwegen) also use a price policy to shift the travel time. Passengers can get a price discount when avoiding peak hours (from 6:30 AM to 9:00 AM, and from 4:00 PM to 6:30 PM on workdays). In this regard, several attempts have been proposed to take this policy into account while constructing a railway model. One of these is the work by Yang and Tang, 2018, who propose a controlled fare-reward scheme to manage the departure time choice of commuters in a rail transit bottleneck. However, although the price policy can be effective to some passengers, those who commute will in generally still want to do so in peak-hours. As such the over-loading of peak hours reduces but is still substantial.

Researchers intend to design ad-hoc timetables to encounter with the large volume passenger flows in highly congested railway lines or for congestion hours. The work by Niu and Zhou, 2013 is one of the early studies to acknowledge the large volume flows and limited vehicle capacity. Niu and Zhou, 2013 consider that not all passengers can be board the first train that arrives after they get to the platform. To model that, they propose a dynamic timetable to deal with the over-saturated situation. Jiang, Cacchiani, and Toth, 2017 study the problem of scheduling passenger trains on a highly congested double-track railway line. The work aims at adding additional trains to a given train schedule to try and solve the large volume passenger problem. The schedule also integrates adding/skipping stops by considering accelerating or decelerating times. Shang et al., 2018 also propose the stop-skipping solution in the scheduling phase to deal with the platform congestion problem for a metro network. An stop-skipping is intended to keep some passengers at stations and reroute their journeys in the time dimension based on the available capacity of each train. We note here that each of the above-mentioned studies focuses on the timetable phase. Note that the idea of flexible timetabling therein can provide insight into the real-time control phase.

As yet, few studies have examined large volume passenger flows in the real-time transfer control phase, particularly in terms of the transfer connection problem. To remedy that state of affairs, we propose a transfer synchronization control model (TSCM) with adequate consideration of the passenger flow. The wait or depart decisions at the peak hours vary substantial compared to those at the normal hours.

3.3 THE SCHEDULED TRAINS URBAN RAILWAY NETWORK

In this section, we describe our railway synchronization control model, which incorporates the passenger flows. First, we provide a brief problem description, followed by the basic assumptions. Then, we present the traditional model of the movement of the trains in case of delays. Finally, we add factor passenger flows to the traditional model.

3.3.1 *Problem description*

We consider a metro network in a metropolis city with a given schedule. The network is made up of stations S and tracks in between. Usually, there are several metro lines L. Each line $l \in L$ contains a set of stations S^l. Any two successive trains from the same line to the same direction have a headway h. The headway serves to keep a safe distance between two trains. The value of h is determined by the speed of the trains and the distance between the corresponding stations. The different lines may or may not share the same headway. The period of the metro network T equals the least common multiple of the headway of all lines. In other words, the state of the network repeats after time T.

This study adopts the synchronization model proposed by Goverde, 1998. We note here that the description of time is based on the stations and the periods involved. For example, a train departures from station i in period k at time t. The traditional train movement model of our choice is described in greater detail in section 3.3.3. This model is based on events from a certain line at the successive stations.

The given timetable ϕ arranges the time when trains occupy the stations and tracks. The timetable reserves running time supplements and transfer buffer times to prevent and recover train delays. For a running activity $ji \in \mathcal{A}_{run}$, the scheduled process time is indicated as a^o_{ji}. It is the superposition of a_{ji} (the minimum running time) and s_{ji} (running time supplement). The same applies to the transfer activities $\mathcal{A}_{trans}$ and dwell activities $\mathcal{A}_{dwl}$.

The metro transports passengers from their origins to their destinations. We denote the number of passengers who depart from station i as n_i. In the same way, n_{ji} and t_i denote the number of passengers who travel from station i to station j, and that of who transfer at station i, correspondingly.

Suppose that an initial delay $d_i(k)$ occurs at the event i at the period k. In this study, we also consider the situation in which more than one initial delay occurs at the same time. There are two decisions have to be made when that happens. First, the disposition schedule should be determined. This variable is denoted as $x_i(k)$, which refers to the real-time departure time from station i. The second decision is whether to maintain or to drop a pre-designed transfer connection. This variable is denoted as $z_{ji}(k)$. In this study, we refer to this process as real-time transfer control. Our goal is to minimize the total delay costs for all the passengers involved.

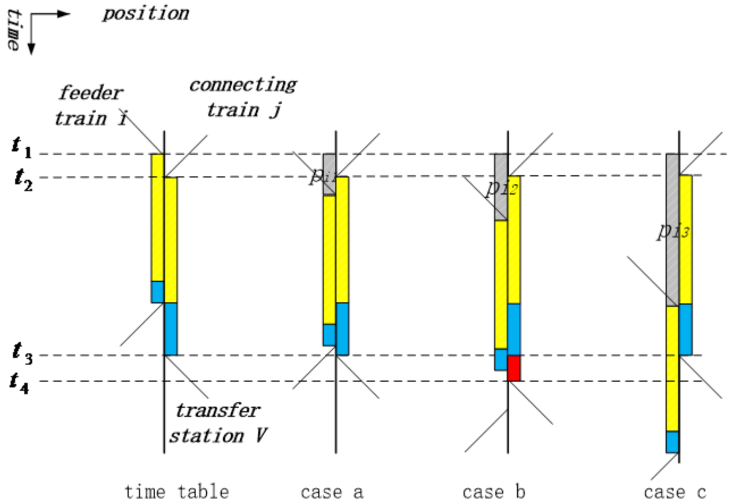

Figure 3.1: Arrivals delay and synchronization controls: The feeder train and its connecting train at the transfer station V. This figure is originally from Goverde, 1998

To maintain or drop a transfer connection is a trade-off problem in the real-time traffic management. Figure 3.1 describes three situations involving delays and the corresponding decisions. In case a, the slight delay of the feeder train d_{i1} (indicated in grey) is shorter than the buffer time. The connecting train can departure at t_3 as scheduled and causes no delays. In case b, the delay d_{i2} is a bit longer and cannot be eliminated by the buffer time. The operator holds the connecting train until t_4 to maintain the current transfer connection. This process is called *transfer synchronization control*, the additional time $t_4 - t_3$ is called *synchronization control time* s_j, Goverde, 1998. In this situation, the transfer passengers involved will not have to wait a long time for the next connecting train. As a result, other passengers at the next stations will wait a bit longer. In case c, the feeder train is significantly delayed and the operators cancel the connection. A long synchronization control time s_j is undesirable, because it might cause waiting time disequilibrium between the current transfer passengers and other passengers involved in the network. This will also cause long-term perturbations to the whole network.

As mentioned that this work is based on Goverde, 1998, we further develop the synchronization model by adding the influence of passenger flows. The passenger flows in Beijing Metro network are described in section 3.5.1.

3.3.2 *Basic assumptions*

To facilitate the statement, the following basic assumptions are given in this paper.

ASSUMPTION 1. Only the initial departure delay(s) is (are) considered. The subsequent delays are caused by the initial delay(s) through propagation in the network. In case another wave of initial delay(s) occur(s), we restart the real-time control.

ASSUMPTION 2. Trains will not depart before their scheduled times.

ASSUMPTION 3. We assume that passengers who missed a connection can aboard the next coming train. This differs from the scheduling phase, which assumes limited capacity of the trains to ensure a proper frequency. In the real-time control phase, we assume that the trains can take all the passengers who are waiting on the platform. Additionally, we assume the passengers will board on the first train arriving after they reach the platform.

ASSUMPTION 4. We assume that the scheduled running time has a positive running time supplement. When initial delay occurs (or initial delays occur), the train will accelerate to mitigate the delays.

3.3.3 *Trains movement in case of delays*

The trains movement in case of delays is affected by three kinds of constrains: (i) the timetable constraints, (ii) the preceding events, these include the running, dwelling and transfer activities, and (iii) the departure headway for safety reason.

E.q. (3.3.1) describes the basic regulations of 'no early departure', which applies to all departure events i during the entire period $k \in K$. It means that the real departure time $x_i(k)$ can not be earlier than the designated timetable departure time $\phi_i(k)$. $x_i(k)$ is also affected by the initial delay $d_i(k)$ if there exists one. The scheduled departure time $\phi_i(k)$ can be derived from the departure time of the first period $\phi_i(0)$, plus the period in between (k), as shown in e.q. (3.3.2).

$$x_i(k) \geq \phi_i(k) + d_i(k), \tag{3.3.1}$$

$$\phi_i(k) = \phi_i(0) + kT. \tag{3.3.2}$$

All departure and arrival events can occur only after their preceding processes have been completed. Explicitly, e.q (3.3.3) regulates that an arrival event can only take place (at time x_i) after the physical train has departed (at x_j) from its previous station and finished the running process (which lasts for a_{ji}). Here $u_{ji} \in \mathbb{N}_0 := (\mathbb{N} \cup \{0\})$ is the period shift, i.e. it denotes whether an activity falls within one period or takes place in more than one. The period shift is calculated via e.q. (3.3.4) Goverde, 2007 in which $\lfloor \cdot \rfloor$ denotes the standard floor operator. E.q. (3.3.5) ensures that a departure event can only happen (at x_i) after the physical train has arrived (at x_j) and the dwell time (which lasts for a_{ji}) has been fully executed for the passengers to board and leave the train. If the departure is involved in a transfer connection A_{trans}, a decision regarding the maintenance ($z_{ji} = 1$) or cancellation ($z_{ji} = 0$) has to be made. This is described in e.q. (3.3.6).

$$x_i(k) \geq a_{ji} + x_j(k - u_{ji}) \qquad \forall ji \in A_{\text{run}} \tag{3.3.3}$$

$$u_{ji} = \left\lfloor (a_{ji}^0 + \phi_j^0 - \phi_i^0)/T \right\rfloor \qquad \forall ji \in (A_{\text{run}} \cup A_{\text{dwl}} \cup A_{\text{trans}}) \tag{3.3.4}$$

$$x_i(k) \geq a_{ji} + x_j(k - u_{ji}) \qquad \forall ji \in A_{\text{dwl}} \tag{3.3.5}$$

$$x_i(k) \geq a_{ji} + \mathcal{M}(z_{ji} - 1)(k - u_{ji}) \qquad \forall ji \in A_{\text{trans}} \tag{3.3.6}$$

For the transfer pair $ji \in A_{trans}$, the waiting time of the passengers depend on: (i) the wait or depart decision z_{ji} and (ii) the rescheduled time of the relevant trains. In case of maintained connection $z_{ji} = 1$, the transfer waiting time is $x_i - x_j$; otherwise the passengers have to wait for the next connecting train i', which is assumed to be the following train after i in the same direction. In this situation the waiting time is $x_{i'} - x_j$. In total the waiting time can be calculated as $z_{ji}(x_i - x_j) + (1 - z_{ji})(x_{i'} - x_j)$.

Minimum departure headway and arrival headway are required to ensure a safe distance between two successive trains from departing and arriving respectively, i.e., $ji \in A_{\text{succ}}$. A departure headway is compulsory to isolate two consecutive departure events form the same station in the same direction. The evaluation of the departure headway for a given train should be the maximum of transverse time of safety distance for the following train. The same applies to the arrival headway.

$$x_i(k) \geq a_{ji} + x_j(k - u_{ji}), \quad \forall ji \in A_{\text{succ}}. \tag{3.3.7}$$

3.3.4 *Waiting Time*

The goal of this transfer synchronization model is to minimize the total passenger delay costs $\mathcal{C}$. This waiting cost $\mathcal{C}$ is based on the total passenger waiting time $\mathcal{T}$, although it takes into account the passenger's perception both on the platform and inside the train.

First, we calculate the overall waiting time $\mathcal{T}$. It is the superposition of the individual waiting time of all passengers at all stations from the initial delay(s) ($k = 0$) to the end of the control ($k = K$) (see Goverde, 1998). We consider three types of passengers: (i) originating passengers O_p for those who begin their trips at departure events j, (ii) through passengers R_p of the dwelling activities $\mathcal{A}_{dwl}$ and (iii) the transfer passengers F_p of the transfer activities $\mathcal{A}_{trans}$. The waiting time $\mathcal{T}$ is therefore composed of three components, each associated with one type of passengers: the waiting time of originating passenger $\mathcal{T}_j^O$, the waiting time of through passengers $\mathcal{T}_{ji}^R$ and those of transferring passengers $\mathcal{T}_{ji}^F$.

$$\mathcal{T}^O = \sum_{k=0}^{K} \sum_{n_j \in O_p} \mathcal{T}_j^O(k) = \sum_{k=0}^{K} \sum_{n_j \in O_p} n_j(k) \cdot [x_j(k) - \phi_j(k)] \tag{3.3.8}$$

$$\mathcal{T}^R = \sum_{k=0}^{K} \sum_{n_{ji} \in R_p} \mathcal{T}_{ji}^R(k) = \sum_{k=0}^{K} \sum_{n_{ji} \in R_p} n_{ji}(k) \cdot [x_i(k + u_{ji}) - x_j(k)] \tag{3.3.9}$$

$$\begin{aligned} \mathcal{T}^F = & \sum_{k=0}^{K} \sum_{n_{ji} \in F_p} \mathcal{T}_{ji}^F(k) = \sum_{k=0}^{K} \sum_{n_{ji} \in F_p} n_{ji}(k) \cdot \\ & [z_{ji}(k) \cdot (x_i(k + u_{ji}) - x_j(k)) + (1 - z_{ji}(k)) \cdot (x_{i'}(k + u_{ji'}) - x_j(k))] \end{aligned} \tag{3.3.10}$$

The overall passenger waiting time $\mathcal{T}$ is thus calculated as $\mathcal{T} = \mathcal{T}^O + \mathcal{T}^R + \mathcal{T}^F$.

3.3.5 *Waiting Cost*

Based on $\mathcal{T}$, we can calculate the overall passenger waiting cost $\mathcal{C}$, which considers the passengers discomfort caused by congestion. This is based on the fact that passengers consider the crowded carriages of a metro network unattractive compared to the privacy of their own cars, especially in case of delays. In this condition, the waiting times spent in the morning peak hours or on a congested holiday should be valued differently to those spent in an spacious train in the off peak hours. The weight of a given waiting time is determined by the discomforts caused by the congestion.

The weighted coefficient can be divided into two categories: the coefficient in-carriage α and the one on platform β. The former coefficient is calculated as e.q. (3.3.11). It is based on the number of on-board passengers $n_{i,\text{board}}$ of train i and the number of designed seats $n_{i,seat}$ therein (Wu et al., 2014). $n_{i,\text{congest}}$ is the margin of passengers number that passengers a train i can hold without discomfort. In case of $n_{i,\text{board}} > n_{i,\text{congest}}$ there will be discomfort caused by congestion. The latter coefficient, the platform congestion coefficient β, is evaluated according to Goeverden, 1990. As such, we take $\beta = 3$ for the waiting times outside of the train. We note here, recent studies Polinder et al., 2020 and Polinder, 2020 also distinguish passenger waiting time spent in the train (named as in-train time therein) and on the platform (named as adaption time), to evaluate the timetable from the passengers perspective.

$$\alpha(n_{\text{i,board}}) = \begin{cases} \dfrac{n_{i,\text{board}} - n_{i,\text{seat}}}{n_{i,\text{seat}}} & n_{i,\text{board}} \leq n_{i,\text{congest}} \\ \dfrac{n_{i,\text{congest}} - n_{i,\text{seat}}}{n_{i,\text{seat}}} + 2\dfrac{n_{i,\text{board}} - n_{i,\text{congest}}}{n_{i,\text{seat}}} & n_{i,\text{board}} > n_{i,\text{congest}} \end{cases} \tag{3.3.11}$$

Last but not least we evaluate the platform congestion coefficient β according to Goeverden, 1990. As such, we take $\beta = 3$ for the waiting times outside of the train.

The waiting cost is thus calculated as the sum of (i) weighted in-carriage waiting time plus (ii) weighted platform waiting time. We compute the waiting cost according the three groups of passengers.

For the originating passengers O_p, their departure time are assumed as the timetable departure time $\phi_i(k)$ of the train they plan to take $i(k)$. We remind that for a departure event $i(k)$, it can have a preceding arrival event $j(k - u_{ji})$ at the same station with a period shift of u_{ji}. The passengers can only get on board after the train's arrival $(x_j(k - u_{ji}(k))$. They wait afterwards in the carriage until the train departs at x_i. We have their waiting time on platform and in the carriages as following.

$$\mathcal{T}^O_{i,plat} = \sum_{k=0}^{K} \sum_{n_j \in O_p} n_{ji}(k) \cdot [\max\left(x_j(k - u_{ji}), \phi_i(k)\right) - \phi_i(k)] \tag{3.3.12}$$

$$\mathcal{T}^O_{i,carr} = \sum_{k=0}^{K} \sum_{n_j \in O_p} n_{ji}(k) \cdot [x_i(k) - \max\left(x_j(k - u_{ji}), \phi_i(k)\right)] \tag{3.3.13}$$

The through passengers spend their waiting time inside the carriages.

$$\mathcal{T}^R_{ji,plat} = 0 \tag{3.3.14}$$

$$\mathcal{T}^R_{ji,carr} = \sum_{k=0}^{K} \sum_{j \in F_p} n_{ji}(k) \cdot [x_i(k + u_{ji}) - x_j(k)] \tag{3.3.15}$$

The transfer passengers get off from the feeder train at $x_j(k)$, and walks towards the connecting train. If the pre-designed transfer connection is maintained ($z_{ji}(k) = 1$), they can take the train $i(k+u_{ji})$; otherwise ($z_{ji}(k) = 0$) they have to wait for the next connecting train $i'(k+u_{ji'})$. In either situation, passengers can only get on board after the connecting train has arrived. We denote the preceding arrival event of the connecting train $i(k+u_{ji})$ as $l(k')$. Similarly the preceding arrival event of the connecting train $i'(k+u_{ji'})$ is denote as $l'(k'')$. w_{ji} is the transfer walk time for the transfer pairs ji.

$$\begin{aligned}\mathcal{T}^{F}_{ji,plat} = \sum_{k=0}^{K}\sum_{j\in F_p} n_{ji}(k)\cdot[z_{ji}(k)\cdot[max\left(x_l(k'),(x_j(k)+w_{ji})\right)-x_j(k)] \\ +(1-z_{ji}(k))\cdot[\max\left(x_{l'}(k''),(x_j(k)+w_{ji})\right)-x_j(k)]] \qquad (3.3.16)\end{aligned}$$

$$\begin{aligned}\mathcal{T}^{F}_{ji,carr} = \sum_{k=0}^{K}\sum_{j\in F_p} n_{ji}(k)\cdot[z_{ji}(j)\cdot[x_i(k+u_{ji})-\max\left(x_l(k'),(x_j(k)+w_{ji})\right)] \\ +(1-z_{ji}(k))\cdot[x_{i'}(k+u_{ji'})-\max\left(x_{l'}(k''),(x_j(k)+w_{ji})\right)]] \qquad (3.3.17)\end{aligned}$$

The overall passenger waiting cost $\mathcal{C}$ is thus calculated as $\mathcal{C} = \alpha\mathcal{T}^{O}_{j,carr} + \beta\mathcal{T}^{O}_{j,plat} + \alpha\mathcal{T}^{R}_{ji,carr} + \alpha\mathcal{T}^{F}_{j,caar} + \beta\mathcal{T}^{F}_{j,plat}$.

3.3.6 *The objective function*

The aim of this real-time transfer synchronization model is to minimize the total passenger waiting cost (weighted total passenger waiting time). We compute the total waiting costs based on the trains movements constraints and the passengers flows.

$$\text{Objective function:} \qquad \min\mathcal{C} = \alpha\mathcal{T}_{\text{Carr}} + \beta\mathcal{T}_{\text{plat}}. \qquad (3.3.18)$$

3.4 DISCRETE EVENT DYNAMIC SYSTEM (DEDS) AND MAX-PLUS ALGEBRA FORMULATION

3.4.1 *Discrete Event Dynamic System (DEDs)*

The above described model includes the departure and arrival events. They occur at discrete instants of time. Meanwhile the activity consists of the running, dwelling and exchange activity, all of which last for some time. The events correspond to the beginning or the end of the activities. For example the running activity starts with the departure and ends at the arrival event.

E.q.(3.3.1)-(3.3.7) can be summarized as e.q. (3.4.1) using the DEDS methodology. Here n is the total number of all preceding events for event $i(k)$. As a reminder, $x_i(k)$ denotes the rescheduled time for the departure or arrival event $i(k)$.

$$x_i(k) = \max(\max_j(a_{ji} + z_{ji}(k)\cdot x_j(k-u_{ji})), d_i(k)), i = 1,...,n. \qquad (3.4.1)$$

The trains movement of the metro network is a typical discrete event dynamic system (DEDS), which is defined as a dynamic, asynchronous system, where the state transitions are initiated by events that occur in 3.4.1. At a certain moment, all those on-going events are named as *the state of the system*. It can be analogous to a screenshot of the system at a given moment. The state of the system repeats every period for the metro network when there are no delays.

From (3.4.1), we can see that events occur at discrete instants of time and initiate the state transitions. Here the max-operation converts the discrete event system into a set of non-linear equations. We adopt the so-called max-plus algebra to analyse and solve the non-linear formulation. As such, it can be effectively represented by a discrete event dynamic system using max-plus system algebra, see Braker and Olsder, 1993, Heidergott and Vries, 2001, Braker and Olsder, 1993 and Goverde, 2007, Goverde, 2010.

3.4.2 *Brief introduction to max-plus algebra*

Max-plus algebra shares a large number of similarities with conventional algebra. However it extends the set of real numbers $\mathbb{R}$ with an element $\varepsilon := -\infty$. Thus the applied field is enlarged as $\mathbb{R}_\varepsilon := \mathbb{R} \cup \{\varepsilon\}$. $(\mathbb{R}_\varepsilon, \oplus, \otimes)$ defines a max-plus algebra structure in which $\oplus, \otimes$ denote the maximization and addition operations, respectively. These are defined for integer inputs a and b as

$$\begin{aligned} a \oplus b &= \max(a, b) \\ a \otimes b &= a + b. \end{aligned} \tag{3.4.2}$$

As such, we refer to $\oplus$ as max-plus algebra addition and $\otimes$ as max-plus algebra multiplication. We have $a \oplus \varepsilon = \varepsilon \oplus a = a$, thus $\varepsilon = -\infty$ is the zero element for the addition operation in max-plus algebra. Similarly, $a \otimes 0 = 0 \otimes a = a$, hence $e = 0$ is the neutral element of multiplication operation.

The extension of the max-plus algebra operators to matrices is a follows. Let $A, B \in \mathbb{R}_\varepsilon^{m\times n}$ and $C \in \mathbb{R}_\varepsilon^{n\times p}$ be matrices sized $m \times n$ and $n \times p$ correspondingly. The matrix addition and multiplication operations in max-plus algebra are defined as:

$$\begin{aligned} (A \oplus B)_{ji} &= a_{ji} \oplus b_{ji} = \max(a_{ji}, b_{ji}) \\ (A \otimes B)_{ji} &= \bigoplus_{k=1}^{n} (a_{jk} \otimes b_{ki}) = \max_{k=1,\dots,n} (a_{jk} + b_{ki}). \end{aligned} \tag{3.4.3}$$

The associated neutral elements for addition and multiplication are $\mathcal{E}_{ji} = \varepsilon$ and $E_{ji} = e$ for all j, i, respectively. The above notions of addition and multiplication form an algebraic structure called an idempotent semi-ring (or dioid). The operations satisfy a number of algebraic properties (associativity, distributively and neutral elements), but in comparison to a field a semi-ring has no additive or multiplicative inverses. Instead, addition in max-plus algebra is idempotent, i.e. $A \oplus A = A$, which is a powerful property that compensates the lack of inverses. For more detailed descriptions we refer the reader to Goverde, 2007.

3.4.3 *Max-plus Formulation of the System*

The formulation of the trains movement in max-plus algebra is as follows. The event time vector when all events in the synchronization control occur is $x(k) = (x_1(k), ..., x_n(k))^T$ and the vector of timetable time $\phi(k) = (\phi_1(k), ..., \phi_n(k))^T$.

We introduce the *process time matrix* A which is composed as $(A)_{ji} = a_{ji}$ where we recall that a_{ji} is the technical minimum process time. Subsequently, we also introduce the maximum period shift as $v := \max_{ji}(u_{ji})$, in which u_{ji} is defined in e.q. (3.3.4). v is referred to as *the order of the system*. Following that, the A is split according to the corresponding period shift as $A = A_0 + A_1 + ... + A_v$. Here the subscript of A_u corresponds to the period shift u_{ji}. The remaining elements in $A_0, A_1, ..., A_v$ are defined as the zero element of addition: $a_{ji} = \varepsilon$. The rationale behind ε follows from the observation that (i) these two events have no precedence constraints ($(A)_{ji} = \varepsilon$) or (ii) they have precedence constraints but with another value of the period shift u_{ji}. For example, if the period shift for activity a_{ji} is u, the evaluation of the split is

$$(A_l)_{ji} = \begin{cases} a_{ji} & \text{if } l = u \\ \varepsilon & \text{if } l \neq u \end{cases}. \tag{3.4.4}$$

The system of recursive linear E.q. (3.4.1)

$$x(k) = \left(\oplus_{l=0}^{p} A_l \otimes x(k-l)\right) \oplus (\phi(0) \otimes T^{\otimes k}). \tag{3.4.5}$$

Note that (3.4.5) defines a DEDS. Here the second part is the timetable event time produced by:

$$\phi(0) \otimes T^{\otimes k} = \phi(k-1) \otimes T = \phi(k). \tag{3.4.6}$$

It can be derived that a power in a max-plus algebra denotes n times repeated multiplication. The $l = 0$ component in (3.4.5) describes the constrains within one same period shift as

$$x(k) = \max(A_0(k) \otimes x(k)) \oplus (\phi(0) \times T^{\otimes k}). \tag{3.4.7}$$

Note that $x(k)$ exists on both sides of the equation indicating an implicit system. However, if A_0 is strictly lower triangular then the system is recursive Goverde, 2007. So under this assumption, the successive event time vectors $A, x(k), k \in \mathbb{N}$ can be computed recursively by (3.4.5) for a given initial timetable $d(0)$ and initial conditions.

3.5 CASE STUDY

In this section we test the application of the transfer synchronization model in a real metro network. First we describe the setup of the network. Following that we describe the configuration with the events and activities. Subsequently, we present the wait/depart decisions at given initial delays and passenger flows. We analyse the proposed model by studying the wait/depart decisions at different periods of the day. Finally, we compare it with a typical classical model based on waiting time without congestion coefficient.

3.5.1 *Network description*

Events, activities and minimum headways

The core part of Beijing metro network (Figure 3.2), Lines 1 and 2, is chosen to test the synchronization model described above. The main motivation for considering this part of metro network lies in the fact that it covers the city centre. The volume and direction of passenger flows here are strongly determined by what time of the day it is. Usually a massive number of commuters travel from the outside into the central business district (CBD) during morning peak hours. Figure 3.3 depicts the passengers passing through station FuXingMen throughout an entire day. Moreover, the metro handles large passenger volumes on holidays, such as Spring Festival and the national holidays. Additionally, some stations in the trading centre area (e.g., Xidan station), tourist area (e.g., Yonghe Lama Temple Station), and stations where several transportation modes connect (Beijing Central Station) tend to handle larger passenger flows compared than other stations.

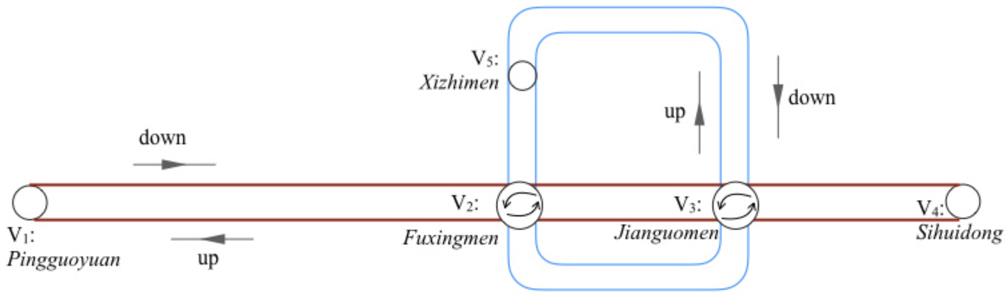

Figure 3.2: The studied network, Beijing metro network Line 1 and Line 2

During rush hours, most trains in these stations arrive with some delays around 0 to 5 min. Passenger volumes in these stations are so big that it is inadvisable to base the synchronization control solely on a classic model. The classic model neglects the congestion that causes inconvenience to passengers.

3.5.2 *Events and activities*

We start by describing the departure and arrival events of the network under study. The practical headways of Line 1 and Line 2 are correspondingly 5 and 6 minutes, respectively. Thus, the period for this network is $T = 30$ minutes. Both lines run in two directions, up and down. A total rolling stock of 40 trains is allocated to these two lines. The scheduled event time slots are shown in Table 3.2. For the convenience of notation, the starting time slot is denoted as zero.

The activities in the metro network include running, dwelling/turning, and transfer activities. The technical minimum process time a^0_{ji} and the buffer time $s_{ji} = a_{ji} - a^0_{ji}$ for these activities are shown in Table 3.3. The transfer activities take place at stations v_2 and v_3. The transfer pairs and the corresponding a^0_{ji} and z^0_{ji} are

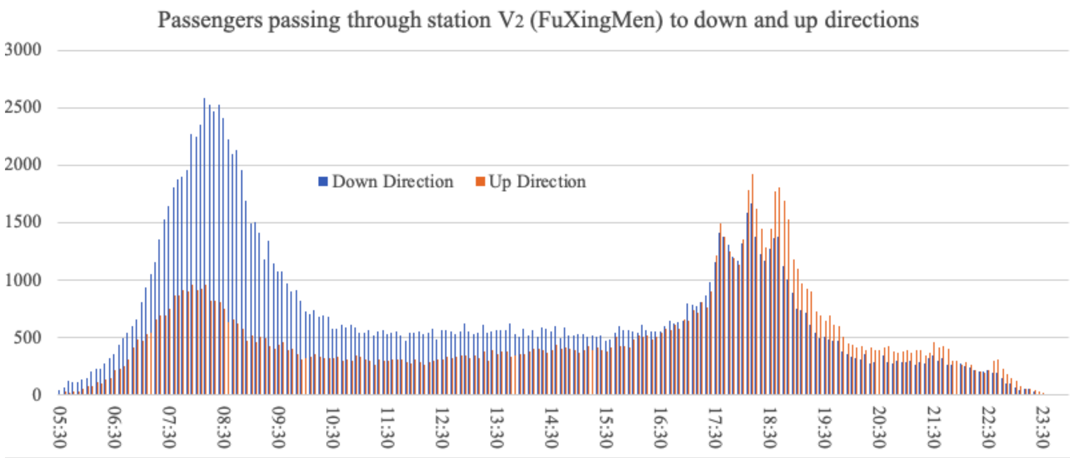

Figure 3.3: Passengers passing through station FuXingMen (the numbers are sampled at 5 minutes granularity)

Table 3.2: The events and the timetable for the described network

Line	Departures	D*	S_i	time (min)	Arrivals	D*	S_i	time(min)
	$\phi_1,...,\phi_5$	d	V_1	1,7,13,19,25	$\phi_{67},...,\phi_{71}$	d	V_2	2,8,14,20,26
	$\phi_6,...,d10$	d	V_2	3,9,15,21,27	$\phi_{72},...,\phi_{76}$	d	V_3	3,9,15,21,27
L1	$\phi_{11},...,\phi_{15}$	d	V_3	4,10,16,22,28	$\phi_{77},...,\phi_{81}$	d	V_4	5,11,17,23,29
	$\phi_{16},...,\phi_{20}$	u	V_4	1,7,13,19,25	$\phi_{82},...,\phi_{86}$	u	V_3	2,8,14,20,26
	$\phi_{21},...,\phi_{25}$	u	V_3	3,9,15,21,27	$\phi_{87},...,\phi_{91}$	u	V_2	3,9,15,21,27
	$\phi_{26},...,\phi_{30}$	u	V_2	4,10,16,22,28	$\phi_{92},...,\phi_{96}$	u	V_1	5,11,17,23,29
	$\phi_{31},...,\phi_{36}$	d	V_5	0,5,10,15,20,25	$\phi_{97},...,\phi_{102}$	d	V_3	1,6,11,16,21,26
L2	$\phi_{37},...,\phi_{42}$	d	V_3	2,7,12,17,22,27	$\phi_{103},...,\phi_{108}$	d	V_2	1,6,11,16,21,26
	$\phi_{43},...,\phi_{48}$	d	V_2	2,7,12,17,22,27	$\phi_{109},...,\phi_{114}$	d	V_5	3,8,13,18,23,28
	$\phi_{49},...,\phi_{54}$	u	V_5	0,5,10,15,20,25	$\phi_{115},...,\phi_{120}$	u	V_2	1,6,11,16,21,26
	$\phi_{55},...,\phi_{60}$	u	V_2	2,7,12,17,22,27	$\phi_{121},...,\phi_{126}$	u	V_3	1,6,11,16,21,26
	$\phi_{61},...,\phi_{66}$	u	V_3	2,7,12,17,22,27	$\phi_{127},...,\phi_{132}$	u	V_5	3,8,13,18,23,28

D*: stands for Direction

listed in Tables 3.4 and 3.5. The departure frequency affects these transfer activities. Namely, the frequency of trains in line 1 (per 6 minutes) is lower than that of Line 2 (per 5 minutes). One connecting train from Line 1 has two feeder trains from Line 2, as shown in Figure 3.5. During one period T, there departs a train on Line 2 without any feeder train, as shown in Figure 3.4.

The last group of time constraints concerns the minimum headway. In this network, the minimum headway of Line 1 and Line 2 are both 2 minutes. According to Han et al., 2015, it takes 6.5 minutes for the transfer walking from Line 1 to Line 2 at station v_2 (FuXingMen Station), and 2.5 minutes for the other direction. It takes 3 minutes for the transfer walking from Line 1 to Line 2 at station v_3 (JianGuoMen Station), and 1.5 minutes for the other direction.

Table 3.3: The scheduled process time a_{ji}, technical minimum process time a^0_{ji}, running slack time and buffer time s_{ji} (in minutes)

Line	Running	Direction	a^0_{ji}	a_{ji}	s_{ji}	Dwelling	a^0_{ji}	a_{ji}	s_{ji}
L1	(v_1, v_2)	d	31	30	1	v_2	1	1	0
	(v_2, v_3)	d	12	11	1	v_3	1	1	0
	(v_3, v_4)	d	13	12	1	v_4	2	2	0
	(v_4, v_3)	u	13	12	1	v_3	1	1	0
	(v_3, v_2)	u	12	11	1	v_2	1	1	0
	(v_2, v_1)	u	31	30	1	v_1	2	2	0
L2	(v_5, v_3)	d	16	15	1	v_3	1	1	0
	(v_3, v_2)	d	14	13	1	v_2	1	1	0
	(v_2, v_5)	d	6	5	1	v_2	2	2	0
	(v_5, v_2)	u	6	5	1	v_5	1	1	0
	(v_2, v_3)	u	14	13	1	v_3	1	1	0
	(v_3, v_5)	u	16	15	1	v_5	2	2	0

Table 3.4: The transfer activities at station V_2 (in minutes)

Transfer direction	$j(k1)$–$i(k2)$	a^0_{ji}	a_{ji}	Transfer direction	$j(k1)$–$i(k2)$	a^0_{ji}	a_{ji}
L1d-L2d	67(0)-45(0)	10	6.5	L1u-L2d	87(0)-45(0)	9	6.5
	68(0)-46(0)	9	6.5		88(0)-46(0)	8	6.5
	69(0)-47(0)	8	6.5		89(0)-47(0)	7	6.5
	70(0)-48(0)	7	6.5		90(0)-43(1)	11	6.5
	71(0)-44(1)	11	6.5		91(0)-44(1)	10	6.5
L1d-L2u	67(0)-57(0)	10	6.5	L1u-L2u	87(0)-57(0)	9	6.5
	68(0)-58(0)	9	6.5		88(0)-58(0)	8	6.5
	69(0)-59(0)	8	6.5		89(0)-59(0)	7	6.5
	70(0)-60(0)	7	6.5		90(0)-55(1)	11	6.5
	71(0)-56(1)	11	6.5		91(0)-56(1)	10	6.5
L2d-L1d	103(0)-7(0)	8	2.5	L2u-L1d	115(0)-7(0)	8	2.5
	104(0)-7(0)	3	2.5		116(0)- 7(0)	3	2.5
	105(0)-8(0)	4	2.5		117(0)-8(0)	4	2.5
	106(0)-9(0)	5	2.5		118(0)-9(0)	5	2.5
	107(0)-10(0)	6	2.5		119(0)-10(0)	6	2.5
	108(0)-6(1)	7	2.5		120(0)-6(1)	7	2.5
L2d-L1u	103(0)- 26(0)	3	2.5	L2u-L1u	115(0)-26(0)	3	2.5
	104(0)-27(0)	4	2.5		116(0)-27(0)	4	2.5
	105(0)-28(0)	5	2.5		117(0)-28(0)	5	2.5
	106(0)-29(0)	6	2.5		118(0)-29(0)	6	2.5
	107(0)-30(0)	7	2.5		119(0)-30(0)	7	2.5
	108(0)-26(1)	8	2.5		120(0)-26(1)	8	2.5

Table 3.5: The transfer activity at station V_3 (in minutes)

Transfer direction	$j(k1)$–$i(k2)$	a^0_{ji}	a_{ji}	Transfer direction	$j(k1)$–$i(k2)$	a^0_{ji}	a_{ji}
	72(0)-38(0)	4	3		82(0)-38(0)	5	3
	73(0)-39(0)	3	3		83(0)-39(0)	4	3
L1d-L2d	74(0)-41(0)	7	3	L1u-L2d	84(0)-40(0)	3	3
	75(0)-42(0)	6	3		85(0)-42(0)	7	3
	76(0)-37(1)	5	3		86(0)-37(1)	6	3
	72(0)-62(0)	4	3		82(0)-62(0)	5	3
	73(0)-63(0)	3	3		83(0)-63(0)	4	3
L1d-L2u	74(0)-65(0)	7	3	L1u-L2u	84(0)-64(0)	3	3
	75(0)-66(0)	6	3		85(0)-66(0)	7	3
	76(0)-61(1)	5	3		86(0)-61(1)	6	3
	97(0)-11(0)	3	1.5		121(0)-11(0)	3	1.5
	98(0)-12(0)	4	1.5		122(0)-12(0)	4	1.5
L2d-L1d	99(0)-13(0)	5	1.5	L2u-L1d	123(0)-13(0)	5	1.5
	100(0)-14(0)	6	1.5		124(0)-14(0)	6	1.5
	101(0)-15(0)	7	1.5		125(0)-15(0)	7	1.5
	102(0)-15(0)	2	1.5		126(0)-15(0)	2	1.5
	97(0)-21(0)	2	1.5		121(0)-21(0)	2	1.5
	98(0)-22(0)	3	1.5	.	122(0)-22(0)	3	1.5
L2d-L1u	99(0)-23(0)	4	1.5	L2u-L1u	123(0)-23(0)	4	1.5
	100(0)-24(0)	5	1.5		124(0)-24(0)	5	1.5
	101(0)-25(0)	6	1.5		125(0)-25(0)	6	1.5
	102(0)-21(1)	7	1.5		126(0)-21(1)	7	1.5

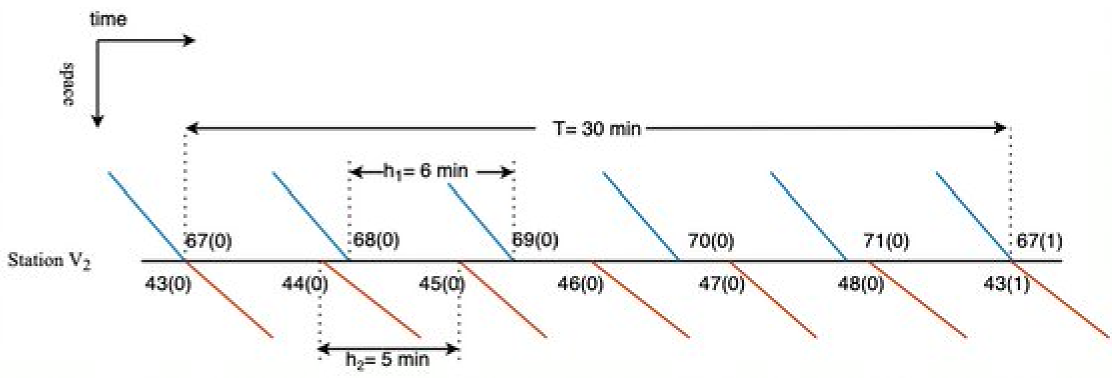

Figure 3.4: Transfers from line 1 to line 2 within one period T

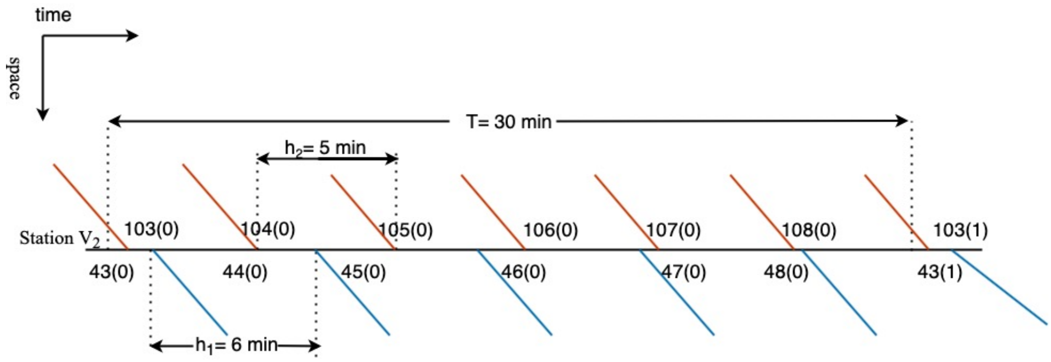

Figure 3.5: Transfer from line 2 to line 1 within one period T

Passenger flows

The passenger flow varies significantly during one day in this example. The data in question is collected on February 20, 2014, a normal workday (Thursday). Note that the metro network uses *ad-hoc* timetables in the weekends, during holidays, and with some cultural or sport events. In these situations the proposed model is also suitable to generate quick decisions in case of delays.

We now specify the other passenger flow parameters. The service trains in Beijing metro network are made up of 6 carriages, and each carriage has 46 seats inside. Passengers get the discomfort when the number of standing passengers is almost 3 times the number of the people sitting. As such, the parameters in e.q. (3.3.11) are $n_{i,\text{seat}} = 276$ and $n_{i,\text{congest}} = 1104$ for all trains i.

3.5.3 *Test with passenger flows in the morning and evening peak hours*

We first compare two different decisions in case of delays, namely the 'always depart' decision and the 'always wait' decision. The aim is to check the delay propagation and passengers choice at given decisions. We take the initial delay of arrival event $70(0)$ as $d_{70}(0) = 3$ minutes. The movement of the trains under 'never always wait' and 'always no-wait' decisions are shown as Figure 3.6 and 3.7, respectively.

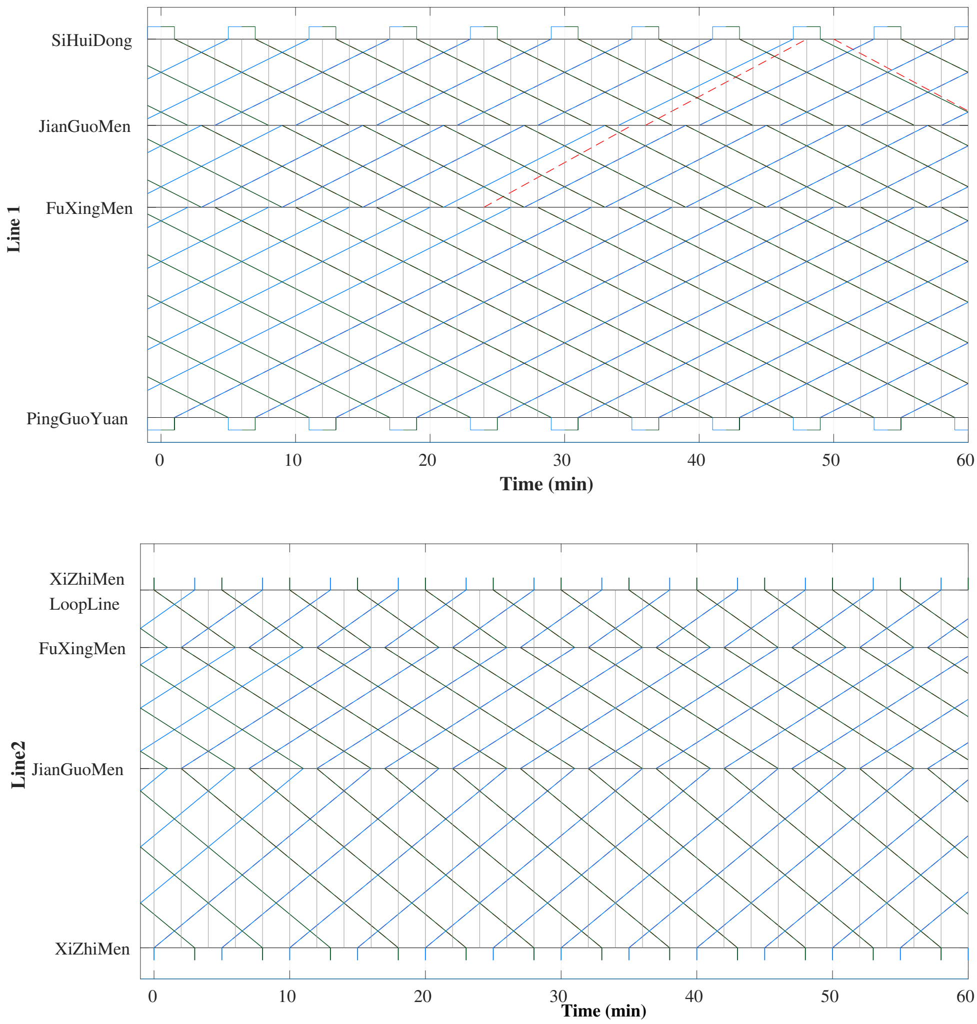

Figure 3.6: Delay propagation under the 'never-wait' policy. The blue and green full lines represent the trains movement designed by the timetable. The colors are used to distinguish up (blue) and down (green) directions. The red dashed line stands for the train movement in case of delays.

We observe from Figure 3.6 that the delay propagates within the trains on Line 1. Line 2 is not affected due to the 'never wait' decisions. Thus transfer passengers from Line 1 at station FuXingMen and JianGuoMen will miss their planned connections.

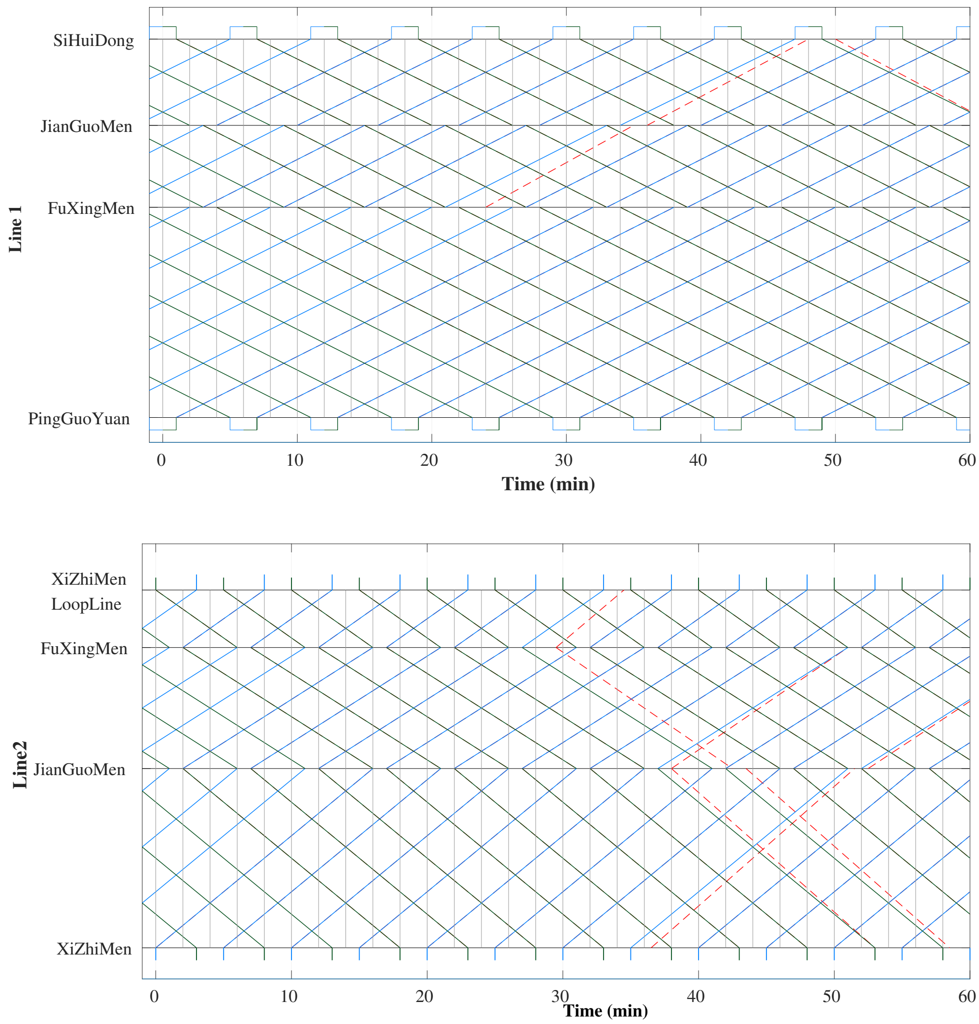

Figure 3.7: Delay propagation under the 'wait' policy. The blue and green full lines represent the trains movement designed by the timetable. The colors are used to distinguish up (blue) and down (green) directions. The red dashed line stands for the train movement in case of delays.

In Figure 3.7, it can be seen that trains from Line 2 also face delays, which are propagated from the maintained transfer connections. In this situation, the overall train delays increase. However, the passengers are guaranteed with smooth transfer connections.

From the comparison of Figure 3.6 and 3.7, we learn that the 'never wait' decision can effectively stop the delay propagation among the trains and allow the system to recover to the origin timetable status. As a consequence, it also causes many passengers to miss their transfer connections. The wait or depart decisions can be made according to the main objective which is to minimize overall train delays or passenger delays.

Solution strategy

We briefly explain the solution strategy. First we calculate the longest possible recovery time $\bar{\Delta}$ for given delays d_i, and the compute the wait-depart decisions z_{ji} and the rescheduled operation time x_i in this recovery time. For the first step, the longest possible recovery time windows $\bar{\Delta}$ for occurred delay(s) are calculated under the 'always wait decisions', i.e., under the assumption $\forall z = 1$. We recall here that recovery windows is the duration of periods since the emergence of initial delays till the vanish of the propagated delays. Secondly, the synchronization control problem in this recovery time window is solved using yalmip (Lofberg, 2004) toolbox in matlab. The optimal wait-depart decisions z, the number of affected passengers and the average waiting cost of all passengers will also be updated in the second step.

Test with passengers in morning and evening peak hours

We use the transfer pair $70(0)$-$48(0)$ at station V_2 as the observed object. A variety of initial delay combinations are given to test the model. The delay of feeder train $d_{70}(0)$ and delay of connecting train $d_{48}(0)$ both range from 0 to 5 min. The passenger flows are sampled from the morning peak hours and the evening peak hours. The calculations are performed using Matlab (version R2019) on a 2,3 GHz Intel Core i5 laptop. The solution for each scenario is available within a few seconds.

Table 3.6 gives the maximum possible recovery time windows $\bar{\Delta}$ in each delays scenario, i.e., the recovery time windows in case of always wait policy.

Table 3.6: The maximum possible recovery time windows in case of all delays combinations

		$d_{70(0)}$										
		0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5
$d_{47(0)}$	0	1	2	2	3	3	3	3	4	4	4	4
	0.5	2	2	2	3	3	3	3	4	4	4	4
	1	2	2	2	3	3	3	3	4	4	4	4
	1.5	3	3	3	3	3	3	3	4	4	4	4
	2	3	3	3	3	3	3	3	4	4	4	4
	2.5	3	3	3	3	3	3	3	4	4	4	4
	3	3	3	3	3	3	3	3	4	4	4	4
	3.5	4	4	4	4	4	4	4	4	4	4	4
	4	4	4	4	4	4	4	4	4	4	4	4
	4.5	4	4	4	4	4	4	4	4	4	4	4
	5	4	4	4	4	4	4	4	4	4	4	4

The control results to minimize the overall passenger costs with passenger flows and evening peak hours are displayed in the Figures 3.8-3.11.

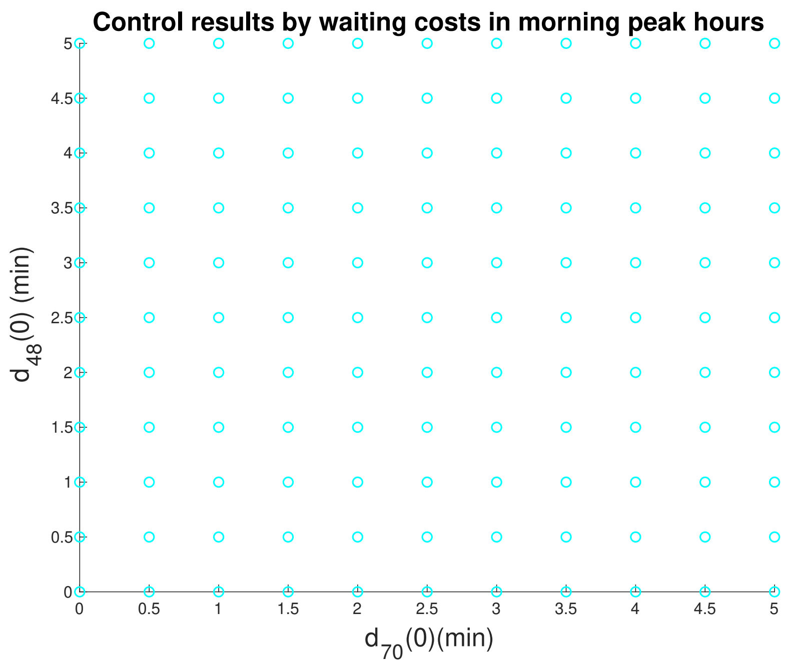

Figure 3.8: One example of control decisions in morning peak hours to minimize overall waiting costs. The blue circles represent waiting decisions for this transfer pair in the current period.

Figure 3.8 illustrates the control results with the passenger flows in the morning peak hours. The horizontal and vertical axes represent $d_{70}(0)$ and $d_{48}(0)$ (in minutes), respectively.

In the morning peak hours, the model gives the wait decisions $z_{70(0),48(0)} = 1$ for all tested initial delays. The underlying rationale is that the number of connecting passenger, from the rural area into the city centre, is relatively high in the morning peak hours. The congestion coefficient in e.q. (3.3.11) synergies with this large number. This is also the reason of the low percentage of depart decisions in the morning peak hours.

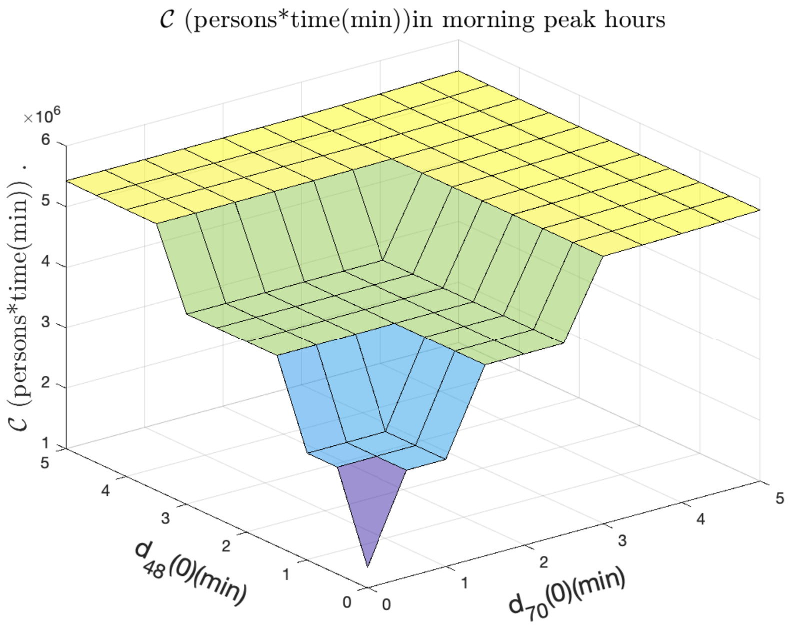

Figure 3.9: Overall waiting cost in morning peak hours

Next, in Figure 3.9 we show the overall waiting cost $\mathcal{C}$ that correspond to the dispatch decisions of Figure 3.8. Here two horizontal axis and vertical axis represent $d_{70}(0)$, $d_{48}(0)$ (min) and the value of $\mathcal{C}$ (passenger · min) during the longest possible recovery time window. We learn that the overall passenger waiting cost typically increases when the delay of the connecting train increases.

The passenger flows in the evening peak hours are sampled to test the wait or depart decisions. The configuration uses the same transfer pairs and the same initial delay combinations. The control results are showed in Figure 3.10 and 3.11.

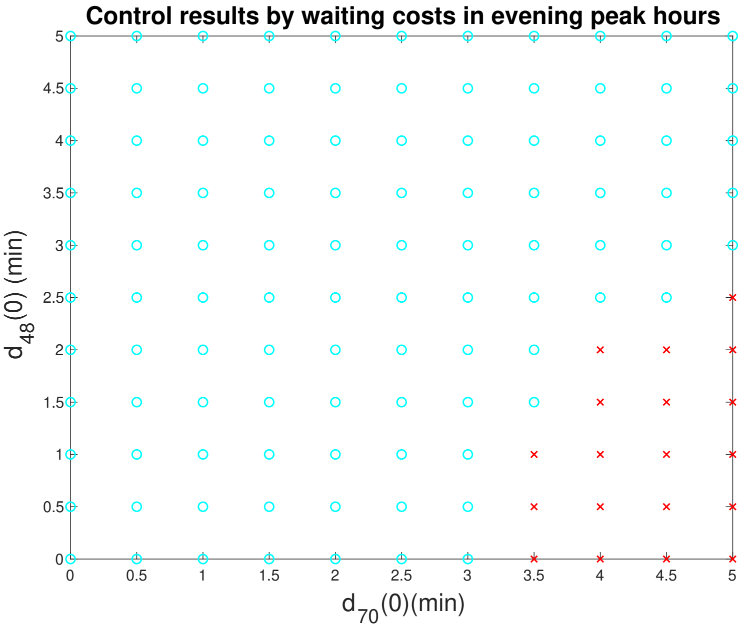

Figure 3.10: One example of control decisions in evening peak hours to minimize overall waiting costs. The blue circles represent waiting decisions for this transfer pair in the current period; the red crosses represent no wait decisions.

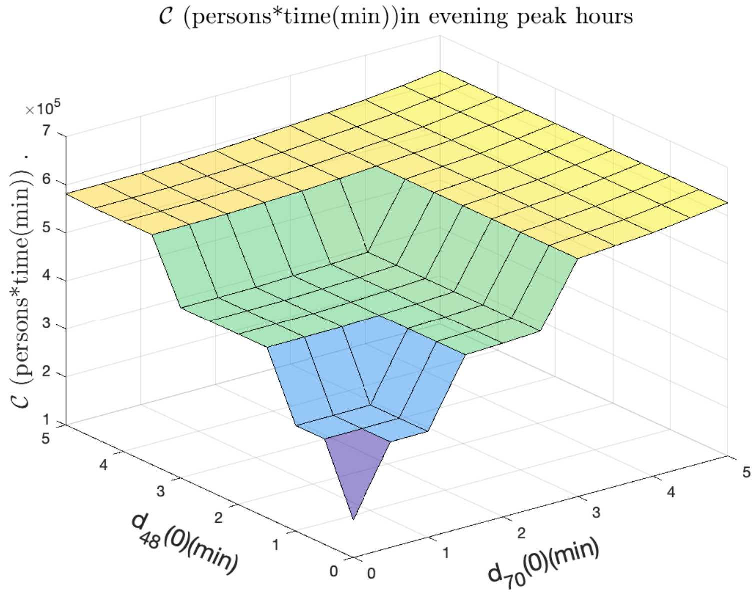

Figure 3.11: Overall waiting cost in evening peak hours

Figures 3.10 shows that some decisions are made to cancel the transfer connections in the evening peak hours. The reason for those cancellations is (i) the number of passengers going towards the city center via the feeder train 70(0) are much less compared those in the morning peak hours,as shown in Figure 3.3; and (ii) the congestion coefficient α augments the waiting time in the connecting train $48(0)$. In evening peak hours, it is less crowed and the model suggests less wait to avoid spreading the delays to longer distributions.

To summarize, the decisions are the result of a comprehensive evaluation of the delay propagation, the volume and direction of the passenger flows. In the morning the connecting train can be hold as long as 5 minutes extra after the scheduled departure time to let the transfer passengers get on board. This will cause additional delays, propagated by the following running actives, minimum departure headway and transfer connections. However, the maintained transfers reduce the waiting cost of passengers from the feeder train. This is a significant volume during the morning peak hours. In the evening hours, on the other hand, some of the connections are cancelled. There are fewer passengers from the feeder train compared to those on the connecting train. Thus, waiting time is relatively short, which can prevent severe delay propagation and additional waiting costs for other passengers.

3.5.4 *Total passenger waiting time*

We compare our proposed model (3.3.18) to reference model (3.5.1). The reference model is a typical transfer synchronization model and aims at minimizing the overall passenger waiting time $\mathcal{T}$. This does not include the in-carriage congestion e.q. (3.3.11) and platform waiting coefficient. The reference model uses the same trains movement as described in e.q. (3.3.1)-e.q. (3.3.10).

$$\text{Objective function:} \qquad \min \mathcal{T} = \mathcal{T}^{\mathrm{O}} + \mathcal{T}^{\mathrm{R}} + \mathcal{T}^{\mathrm{F}}. \tag{3.5.1}$$

The basic settings, i.e., transfer pairs, initial delays and passenger flows, are unchanged. The corresponding control results to minimize the overall waiting time are shown in Figures 3.12-3.15.

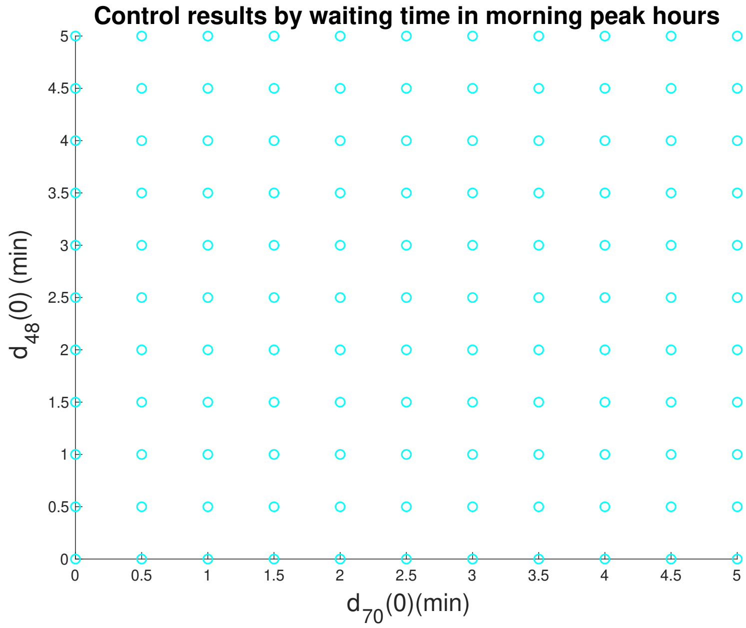

Figure 3.12: One example of control decisions in morning peak hours to minimize overall waiting time. The blue circles represent waiting decisions for this transfer pair in the current period.

Figure 3.12 shows that the transfer connection pair $70(0) - 48(0)$ is maintained in all initial delay combinations. Even not considering discomfort caused by congestion, the wait decisions are made in the classical reference model due to the massive flow of passengers towards the city center.

Figure 3.13 illustrates the corresponding overall waiting time of the whole network during the longest maximum recovery time window $\bar{\Delta}$.

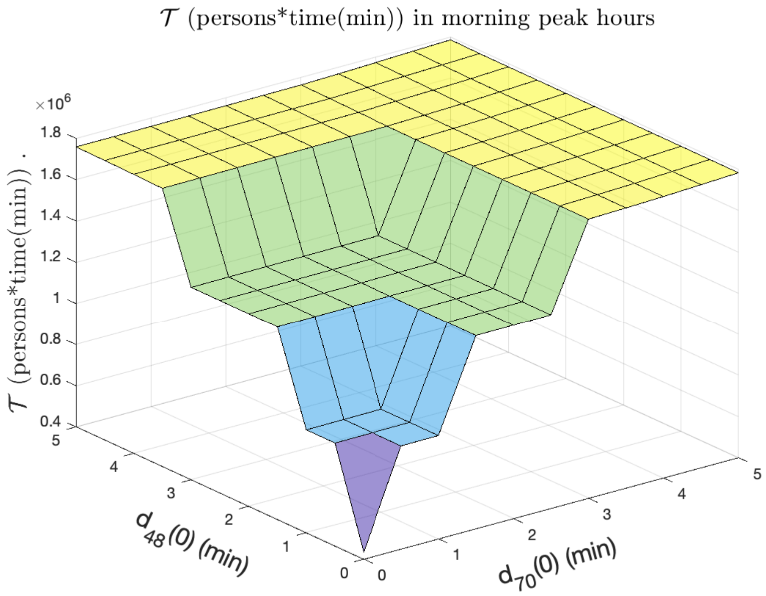

Figure 3.13: The overall waiting time in morning peak hours

Figures 3.10 and 3.14 show the different control decisions in the proposed model (3.3.18) and the reference model (3.5.1) in the evening peak hours.

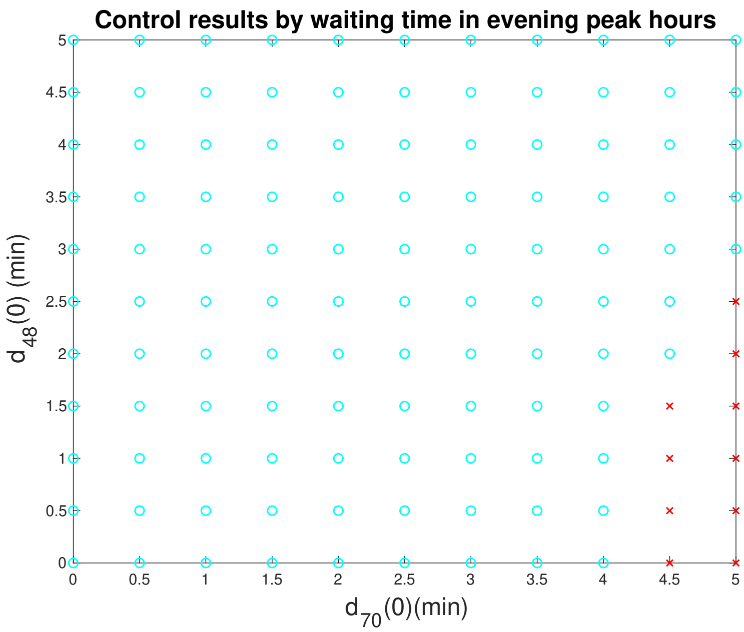

Figure 3.14: Overall waiting time in morning peak hours

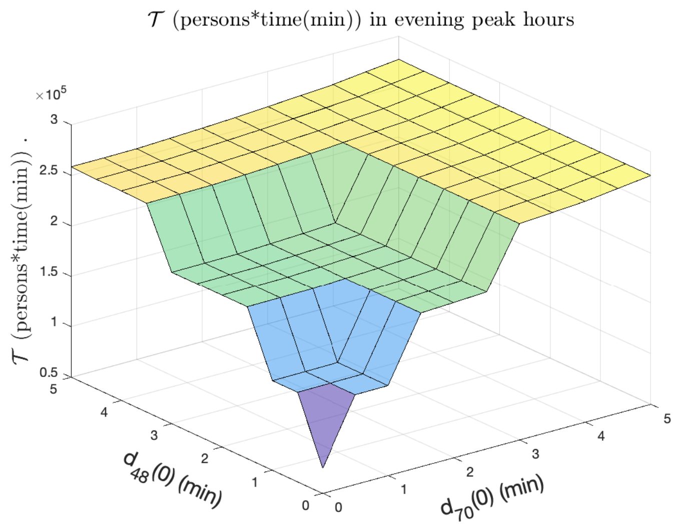

Figure 3.15: Overall waiting time in evening peak hours

We observe that the connecting trains are held longer. This implies a relatively large volume of passengers from train $48(0)$ have to wait for a relatively small group of passengers from train $70(0)$. These decisions are less suitable for a metro network of a major city struggling to handle the congestion problems.

3.6 CONCLUSIONS

Urban traffic congestion will continue to grow along with the number of citizens, which means that management designed to deal with travel congestion should be optimized.

In this paper, a transfer synchronization model is developed for a metropolitan metro network having to deals with enormous numbers of passengers. The key feature of the model is the introduction of a wait-depart decisions for transfer connections based on the so-called waiting cost instead of the traditional approach based on waiting time. In this way the passenger experience becomes an essential component of the model. Apart from the ability to reschedule trains, the model makes it possible to maintain or drop a transfer connection. The wait and depart decisions have to do with the decisions: (i) *whether* to hold the connecting train and (ii) *how long* the connecting trains can wait for a specific delay of the feeder train. We have tested the model on the core part of the Beijing metro network using various time periods. The decision of *whether* to wait or not is significantly different in the morning and evening peak hours. The comparison with a traditional waiting-time based model reveals that the model proposed in this paper provides more reasonable decisions involving *how long* the connecting train should wait for a specific delay of the feeder train.

Based on the results of this study, there are several avenues for future research. To begin with, an avenue for further study would involve testing the performance of the proposed methodology on more complicated networks, for example the whole Beijing metro network. Secondly it is practical to integrate passengers' reaction to perturbations with the rescheduling. This is based on the fact that passengers have their own intuitive choices when the conditions change. Besides, metro organizer can also pose intervention, for example stop skipping, which distributes the passenger flows and thus avoids congestion. This integration can be achieved via an iterative formulation.

4

REAL-TIME RAILWAY TRAFFIC CONTROL WITH MIXED PASSENGER AND FREIGHT TRAFFIC

This chapter is reproduced from Qu, 2020:

W. Qu, Y. Maknoon, F. Corman, *real-time railway traffic control with mixed passenger and freight traffic*, to be submitted.

ABSTRACT

This study presents a simultaneous rescheduling model for mixed passenger and freight railway traffic with capacity constraints. The heterogeneous traffic flows are regulated based on the service demand from the clients in case of delays. We propose the rerouting of the freight trains, which does not hinder the delivery of the freight and can release the limited capacity in the delayed region. To provide a faster resolution, a set of alternative routes is generated in advance for each origin-destination pair. The simultaneous rescheduling problem is formulated via the event-activity network at a macroscopic level. The capacity constraints are introduced to ensure a conflict-free resolution. The proposed model is evaluated on a crucial part of the Dutch railway network and the results are compared to non-flexible reference models which do not allow the stop-skipping and rerouting of the freight trains. The test results show that, while maintaining the service levels to passengers, the flexible replanning model offers a way to reduce the freight delays.

4.1 INTRODUCTION

Increasing demand for freight transportation and the mode shift from trucks to trains calls for an improvement of the service level of freight transportation. In 2017, road was the leading mode of freight transport at within the EU (51.5%), followed by maritime transport (32.4%) and rail transport (11.6%) (EuroStat, 2017). It is essential to improve the service quality of railway freight transportation to attract more freight to railways. This also has to take into account the already congested situation involving passengers trains, which share the existing railway lines with freight trains in practice.

During the scheduling process, the timetabling of passenger trains starts a long time ahead of the real-time operations, whereas the freight trains to a large extent are scheduled more *just-in-time* (Cacchiani et al., 2014). In the implementation of a timetable, there are inevitably things that disrupt the existing schedule, breaking the pre-designed access of the trains to the stations. A real-time traffic management plan is needed to deal with such disruptions and move the passengers and freight to their destinations.

There are several considerations that needed to be taken into account when drawing a rescheduling plan. The main one is to ensure conflict-free movements of the trains, i.e., to operate within the capacity of the stations and open track segments. Other objectives can be (i) to minimize the deviations from the original timetable and restore to the timetable status as soon as possible (D'Ariano et al., 2008), (ii) to minimize the delay propagation (Goverde, 2010), and (iii) to provide unhindered transfer connections and minimize passengers waiting times (Schöbel, 2009, Dollevoet et al., 2015, Corman et al., 2011).

However, in most existing rescheduling studies, freight trains are not taken into consideration, and those that do, treat them as passenger trains with a low weight coefficient (Törnquist, 2012) or a low priority (Corman et al., 2011). So even mentioning freight trains explicitly and trying to integrate them is a step forward. Ranking trains by their order of importance makes it possible to recover the high-priority trains first, at the expense of trains with lower priorities(mainly the freight trains). Moreover, these methods used in studies so far have overlooked the differences between passenger trains and freight trains, including the particular (distinctive) requirements during the real-time traffic management and the operating characteristics. Those can be used to improve the the real-time traffic management.

Using the different characteristics of the mixed passenger and freight traffic leads to one operational question: What are the differences between operating passenger trains and freight trains? The requirements of the transportation clients in questions provide the main clue. Passengers perceive the train journey as a service. Thus for the passenger trains, the goal should be to minimize the deviation from the original timetable and to provide effortless transfer connections. In the case of freight trains, on the other hands, the clients care most about the final delivery time. This means that the precise routing and waiting time at the intermediate stations matter less. There is a greater flexibility when it comes to rescheduling freight trains when it comes to rescheduling. For example, the routing choice and the dwelling time at the intermediate stations can be modified.

Another question is how to formulate the movement of trains in the network. The railway rescheduling models can be divided between into macroscopic and microscopic models according how detailed the formulations are. Macroscopic models look at a railway network at a relative high level of aggregation, thus can be used for large-size network. The microscopic models are much more detailed, making them suitable for the local dispatching. These two formulations will be discussed in the literature review section. In this study, we adopt a macroscopic formulation with capacity constraints. There are two reasons for adopting the macroscopic formulation with the capacity constraints: (i) the chosen network and time horizon should be large enough to explore the rerouting of the freight trains; (ii) the capacity constraints of the open track segments and the stations are added to ensure the safe movement of all the trains.

The contributions of this study include the following: (1) we present a mathematical model integrating the passenger and freight transport, while maintain their distinguishing characteristics. (2) we introduce the rerouting and stop-skipping of freight trains. This can release the existing limited capacity of the congestion area to provide access to passenger trains. At the same time, the service level for the freight transports is maintained. (3) We construct a macroscopic formulations to explore management of a broader network and longer time window and to improve the feasibility of the operations at the local-network level (capacity constraints).

The remaining part of this paper is organized as follows. Section 4.2 presents a literature involving railway rescheduling models and the methods they use. Section 4.3 provides the problem description. Section 4.4 presents the mathematical formulation of the problem with the event activity network. In section 4.5 we provide the methodology and in section 4.6 we report and analyse the results based on the Dutch railway network. Section 4.7 provides conclusions and avenues for further research.

4.2 LITERATURE REVIEW

This section contains a literature review involving real-time railway traffic management (rtRTM) problems. First, we study rtRTM problems of generic rail networks (not distinguishing between passenger and freight trains) or passenger-only networks. The focus lies on the problems under examination and the corresponding solutions. Secondly, we look at the limited body of literature regarding a mix of passenger and freight rail transport flows. We discuss the regulation of the heterogeneous traffic flows via the differentiation of passenger and freight trains. The studies in the review are listed in Table 4.1 based on the level of formulation detail and their objective functions, incorporating further information as the corresponding special constraints, the planning horizon and tested area.

Table 4.1: Basic category of the Macroscopic models

formulation	objective functions	Literature	Special constraints	planning horizon	test area
Macroscopic	trains delays	Dündar and Şahin, 2013			single track in Turkey (150 km)
		Schöbel, 2009	track capacity		Deutsche Bahn
	passenger delays	Schöbel, 2001			RRheinland-Pfalz and the Saarland
		Dollevoet et al., 2015	station capacity	1h	Zwolle to Utrecht network
		Dollevoet et al., 2015	passenger rerouting	1h	Zwolle to Utrecht network
		Kanai et al., 2011	dynamic interacting		Japanese railway (130 km)
		Min et al., 2011			Seoul metropolitan railway network
	mixed	Ginkel and Schöbel, 2007			Verkehrsverbund Rhein-Neckar (30 m)
		Törnquist and Persson, 2007			
Microscopic	passenger delays	D'Ariano, Pacciarelli, and Pranzo, 2007	train re-timing	2h	Schiphol (20km)
		D'Ariano et al., 2008	train re-timing and rerouting		Utrecht Den Bosch
		Corman et al., 2010	train re-timing and rerouting	1h	Utrecht Den Bosch
		Gély, Dessagne, and Lérin, 2006	track choices, extra-stops	8h	Line between Tours and Bordeaux
		Albrecht, Binder, and Gassel, 2013	speed control		Utrecht to 's-Hertogenbosch
		Rodriguez, 2007	junctions areas		Pierrefitte-Gonesse junction
		Caimi et al., 2012	congested bottlenecks		area of Berne (10 scenarios)
	train delays	Wegele and Schnieder, 2004		24h	Deutsche Bahn AG
		Corman et al., 2015	passenger rerouting	2 h	Utrecht and Den Bosch
	Mixed	Corman et al., 2012		1h	Utrecht Central

Passenger: minimizing total passenger delays; trains: minimizing trains delays; mixed: minimizing mixed passenger and freight delays;

4.2.1 *Real-time traffic management of generic trains*

In this section, we review the rtRTM problem in the context of generic trains, i.e., which do not distinguish between passenger and freight trains, or look only at passenger trains. The focus is the traffic control approaches, the main objective of the rescheduling, and the exact problem involved. Different formulations and the corresponding solutions are used according to the aims of the system under study.

In the early stage of rtRTM studies, the main task was to minimize the overall trains delays so the focus was on getting the trains move again as soon as possible. Due to the management approach is mainly re-timing, i.e., modifying the times for the trains to enter the stations or block sections. These studies are named as re-timing problem, see Törnquist and Persson, 2007, Törnquist, 2012, Dündar and Şahin, 2013, etc. Besides the re-timing, succeeding re-ordering (changing the order of a pair of consecutive trains on a track) and local re-routing (selecting another platform track for the trains inside stations) approaches discussed. This stream of the study is named as rescheduling problem (D'Ariano et al., 2008) or real-time train dispatching problem (Lamorgese and Mannino, 2015). Researchers also give attention to minimizing the passenger delays via providing smooth transfer connections, which is also called the delay management problem (Schöbel, 2001, 2007, Dollevoet et al., 2015, 2012).

The formulations of the rtRTM models can be divided into the categories macroscopic and microscopic, according to the level of detail. Macroscopic models describe the railway network at a relative high level and contain a more aggregated representation of certain resources. Stations are represented by nodes, and the tracks are represented by arcs. An example of a macroscopic model is event activity network (Schöbel, 2001). The details of block sections and signals are not considered. Meanwhile in microscopic models, the above aspects are considered in detail. Concretely, a station is composed of a complex set-up of pieces of tracks separated by switches and signals. The alternative graph is an intensively used microscopic model that is used in many studies (see D'Ariano et al., 2008).

Macroscopic and microscopic models target on different planning scales and serve different replanning purposes. Macroscopic models focus mainly on the application of a large-scale network. The application can be the delay propagation of initial delays over the network (Goverde, 2010), specifying the passengers' transfer connections (Schöbel, 2001), the influence of passenger flows (Schöbel, 2009) and the way passengers respond to the delays (Dollevoet et al., 2015), etc. The cover of large-scale network and long time windows is achieved from the high level of aggregations, however compromise the details in the description of trains' movement.

Microscopic models, on the other hand, are used mainly for regional (local) networks, including major stations with a complex topology of interlocking areas (Caimi et al., 2012), or multiple small stations with a simple topology and with possible points of conflict between major stations such as junctions ((Rodriguez, 2007)), level crossings etc. They are mainly used to deal with safety-related issues, such as speed control (Albrecht, Binder, and Gassel, 2013), setting local routes for trains within the stations (D'Ariano et al., 2008, Corman et al., 2010), predicting and solving conflicts at a local level and control processes that take place on the part of infrastructure under their supervision. According to Pellegrini, Marllière, and

Rodriguez, 2014, traffic management performance can be different depending on the granularity of the representation of the infrastructure. Route-lock sectional-release control methods provide an alternative to route-lock route-release interlocking, and they require longer calculation times.

There is a trend with regard to macroscopic models to include capacity constraints to make the models more feasible in real application. Schöbel, 2009 integrates the open track segment constraints into the macroscopic delay management model. Dollevoet et al., 2015 proposed a rescheduling model dealing with the platform track capacity, and platform tracks reassignment. In the following researches, Lamorgese and Mannino, 2015 divide the train dispatching problem into a master problem (re-timing and re-ordering of the trains associated with the line) and the slave sub-problems (the platform tracks arrangements within the stations). The slave problems communicate with the master problem through suitable feasibility cuts in the variables of the master problem.

4.2.2 *Mixed passenger and freight traffic management during disturbances*

Although there is an abundance of rtRTM studies, the number of studies focusing on freight flows is limited (Qu, Corman, and Lodewijks, 2015). To the best of our knowledge, there are a number of studies mentioning freight trains, (Törnquist and Persson, 2007, Törnquist, 2012 , Acuna-Agost et al., 2011, Corman et al., 2011), which integrate the heterogeneous traffic via the objective function (i) the *Value of Time* (VoT) and (ii) the *set of priority*.

The VoT allocates a different cost functions (unit delay cost, extra penalty for severely delays) to passenger and freight trains (see Törnquist and Persson, 2007). Usually, the theoretical VoT for freight is significantly smaller than it is for passengers. Furthermore, Törnquist and Persson, 2007 accumulate the passenger delays at each stop, but take into account the freight delay only at the destination stop (or the last stop of the planning horizon or area). The freight trains are often assigned a lower priority than passenger trains. Törnquist, 2012 and Acuna-Agost et al., 2011 adopted the VoT to handle the mixed traffic flows in a rescheduling problem, and further improve the solving heuristics.

Corman et al., 2011 propose using a set of priorities to cope with mixed traffic flow rescheduling problems. They treat the delayed trains hierarchically, according to the type of the trains. At each step, the procedure focuses on the current priority class, preserving solution quality from the higher priority classes and neglecting lower priority classes in the optimization of train orders and times. The iterative process is implemented from the highest one to the lowest priority, until a new network disposition timetable is finished. Using this approach, high-priority trains, for example high-speed trains, have exclusive access to the infrastructure and will be the first to return to the original timetable. As a result, slow freight trains are usually excluded from the scope of an emergency, and their recovery can be postponed.

Similar approaches using a set of priorities are used in the scheduling problem from a longer planning horizon, e.g. Godwin, Gopalan, and Narendran, 2007 and Cacchiani, Caprara, and Toth, 2010. These papers describe a situation where the passenger trains are scheduled first, and then the freight trains use the residual capacity. The freight trains cannot cause any change to the schedule of the passenger trains. Another recent related contribution is the work of Bešinović, Widarno, and Goverde, 2020 which studies adjusting freight train paths to infrastructure possessions.

Based on the studies discussed above, we propose a model for dealing with mixed passenger and freight trains on the basis of their operational properties. Our model provides the following benefits with regard to train scheduling: (1) freight trains can change between passing through a station or stopping at the station; (2) freight trains can change their routes and do not have to stick to the pre-planned stops; (3) different arrival times, departure times, arrival and departure headways are used to distinguish the passenger trains and freight trains, since they have different braking or accelerating longitudinal dynamics; (4) our model includes the open track segment capacity (OTS-Capacity), overtaking constraints due to the limited capacity of platform tracks.

Our model adopts the Event-Activity Network Formulation (macroscopic) with capacity constraints, for the following reasons: (1) freight transportation usually covers vast geographical areas and long time windows. Thus the event-activity networks are chosen as the basic framework of the formulation. (2) because it is important to distinguish the network capacity when handling passenger and freight traffic, respectively. Thus capacity constraints are integrated into the model, to make sure it provides feasible and safe solutions. The macroscopic approaches in this work emphasize the capacity constraints to regulate the uncertainty caused by the flexibility of the freight trains. The rerouting and stop-skipping of the freight trains are within the capacity of the open track segments and stations.

4.3 PROBLEM DESCRIPTION AND ASSUMPTIONS

This section describes the rtRTM problem for mixed passenger and freight traffic flows, followed by the main assumptions of this study. The network and movement of the heterogeneous traffic are formulated using the *event activity network* with capacity constraints. The real-time railway traffic management starts with the initial delay and ends with the reduction of all subsequent delays; we terminate before reduction when the propagation exceeds a certain time window. The aim in this study is to minimize the overall passenger and freight delay costs.

4.3.1 *Railway train network*

Consider a railway network engaged with passenger and freight trains. Let S denotes the set of stations (nodes), with instances $s \in S$. We denote $K(s)$ as the set of platform tracks of station s. An open track segment (OTS) $q = (s, u)$ is the directed track from $s \in S$ to $u \in S$. Q is set of the open track segments. For this macroscopic formulation with capacity constraints, we do not further describe the blocks and signalling system. Schöbel, 2007 defined the blocks (also known as block sections) as the track between two signals) and the signalling system.

The passenger trains P and freight trains F co-exist in the network. Each train is represented by a sequential number i. For each train $i \in (P \cup F)$, the pre-designed path in the original timetable is denoted as ρ_i. For a freight train $i \in F$, $\alpha(i)$ and $\beta(i)$

stand for the first stop and the last stop along its trip within the replanning horizon and replanning region.

We use the *event and activity network* $\mathcal{N} = (\mathcal{E}, \mathcal{A})$ as a basis to formulate the movement of the trains in the network. A departure or arrival event indicates a train i starting from or arriving at station s, as denoted by $e_i^{s,-}$ or $e_i^{s,+}$, respectively. The set of *events* $\mathcal{E} = (\mathcal{E}_{dep} \cup \mathcal{E}_{arv})$ corresponds to the departure and arrival events of the trains. Recall that a train has a pre-designed timetable if no delays or disturbances occur to the system. We define ϕ as the time associated with an event and denote this as $\phi_i^s := \phi(e_i^s)$ where e_i^s is a departure or arrival event.

An activity is the procedure between two successive events. A running activity $a \in \mathcal{A}_{run}$ connects one train's departure $e_i^{s,-}$ and the following arrival at the next station $e_i^{u,+}$. A dwell activity $\mathcal{A}_{dwl}$ associates a train's arrival $e_i^{u,+}$ and the precedence departure $e_i^{u,-}$ from the same station. A transfer activity $\mathcal{A}_{trans}$ is the passengers' movement between one train's arrival $e_i^{s,+}$ and another train's departure $e_j^{s,-}$ at the same station. The set of *activity* is $\mathcal{A} = (\mathcal{A}_{run} \cup \mathcal{A}_{dwl} \cup \mathcal{A}_{trans})$. For an activity $a \in \mathcal{A}$, the duration time is denoted as L_a.

When the primary delay(s) d_i^s unexpectedly happen(s) to the network, the movement of the trains should be rearranged. The capacity constraints, are applied to ensure a conflict-free movement. The OTS capacity is regulated using the departure headway $h_{i,j}^{s,-}$ and arrival headway $h_{i,j}^{s,+}$. It is the minimum time to separate the train pair i and j using the same segment $q = (s, u) \in Q$. In this work we use the headway time from PETER(Goverde and Odijk, 2002). Schöbel, 2007 described the generic requirements about departure/arrival headway. $h_{i,j}^{s,+}$ and $h_{i,j}^{s,-}$ in that work are evaluated according to the train speeds and the length of the block (the track between two signals), i.e.,

$$h := \max \{\text{transverse time in b } : \text{ b is the block on OTS q}\}.$$

Passengers and freight trains are integrated and distinguished in the recovery process. Both passenger and freight trains are rescheduled in the *event activity network*. Meanwhile, the rerouting of freight trains is integrated as a solution. Moreover, the freight trains can either stay at or pass through the stations on their chosen path, depending on the circumstances. For the rerouting, the set of alternative possible paths $\mathcal{R}(i)$ for train i is generated in advance via the k shortest methods (Yen, 1971) to save the computational time. Here, an alternative path $r(i) \in \mathcal{R}(i)$ goes from $\alpha(i)$ to $\beta(i)$ in the network. A binary parameter $b^{s,r}$ describes whether a station s belongs to an alternative route r or not. Then $b^{s,r} \cdot b^{u,r}$ denotes whether an OTS $q = (s, u)$ belongs to alternative route r. After making the rerouting choice, another decision should be made regarding whether the train will pass through or dwell at the station. In the latter case, the train will lose extra time due to the braking and acceleration process. We incorporate this effect in our model via an arrival time $L_i^{s,+}$ and departure time $L_i^{s,-}$, as shown in Fig 4.1.

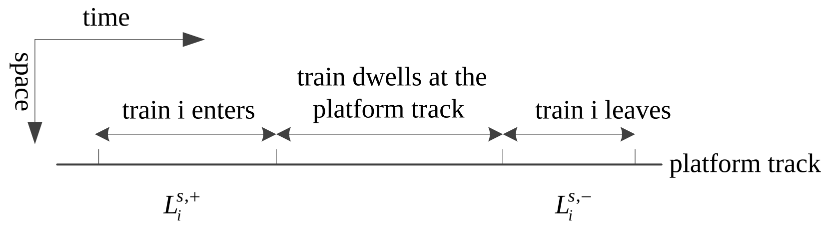

Figure 4.1: A freight train's dwelling time at a station.

4.3.2 *Decision Variables and main assumptions*

The decision variable $x_i^{s,+} := x(e_i^{s,+})$ is the reschedule departure time for event $e_i^{s,+}$. In the same manner, we define the $x_i^{s,-}$ for the rescheduled arrival time. We use $(y_{i,j}^s)$ to decide the order of two departure events $e_i^{s,-}$ and $e_j^{s,-}$ to the same OTS q. $y_{i,j}^q = 1$ means that $e_i^{s,-}$ is arranged before $e_j^{s,-}$; otherwise $y_{i,j}^q = 0$.

The sets of binary variables $z = (z_i^k)$ and $\zeta = (\zeta_{i,j}^k)$ determine the platform track assignment and the corresponding occupancy sequence. $z_i^k = 1$ denotes that the train i is assigned to the platform track k, otherwise $z_i^k = 0$. In case $z_i^k = 1$ and $z_j^k = 1$, the occupancy sequence should be ordered. $\zeta_{i,j}^k = 0$ arranges train i before train j, otherwise $\zeta_{i,j}^k = 1$.

For passenger trains $i, j \in P$, $\xi = (\xi_{i,j}^s)$ denotes the wait or depart decisions for the pre-designed transfer connection at station s. If $\xi_{i,j}^s = 1$ the connecting train j will be hold for the passengers from the delayed feeder train i. For a freight train $i \in F$, there are two decisions to be made: (i) the route choice and (ii) the passing through or dwelling at the station. If the route $r \in \mathcal{R}$ is chosen, then the binary $v_i^r = 1$; otherwise, $v_i^r = 0$. Concerning the second decision, we introduce variable w_i^s to indicate whether the freight train $i \in F$ passes through ($w_i^s = 1$) or dwells at ($w_i^s = 1$) station $s \in r$.

Assumptions

Our model is constructed based on the following assumptions:

- The delays in this study are those up to 30 minutes. In case of disruptions longer over than 30 minutes, the extra operation strategies involve rescheduling of rolling stock and the crew duties, which as such falls beyond the scope of this study.
- For the passenger trains, the itinerary and dwelling decisions are kept the same as in the original plan. Flexible stopping strategies for passenger trains are not considered here. This is to maintain service levels as high as possible in case of delays.

- The model does not deal with the freight operations in transit. It is assumed that all freight trains pass through the stations, or simply dwell at stations without any operations. A future extension of the model can involve integrating simple operations such as disassembling or assembling of the trains.

4.4 THE COORDINATED MODEL

In this section, we illustrate the formulation of the real-time traffic management of mixed passenger and freight trains with the corresponding capacity constraints. The presentation starts with the traditional *event activity network* formulation of the passenger traffic without capacity constraints (section 4.4.1). Section 4.4.2 integrates freight traffic into the model. Next, we augment the framework with capacity constraint in section 4.4.3.

4.4.1 *Event-Activity formulation of passenger traffic without capacity constraints*

Constraints (4.4.1)-(4.4.4) make up a traditional event-activity formulation for the delay management problem without capacity constraints (Schöbel, 2001). Constraint (4.4.1) ensures that passenger trains cannot departure or arrive earlier than the pre-scheduled time ϕ_i^s plus the initial delay d_i^s if it exists. Constraint (4.4.2) describes a dwelling activity. Then (5.4.11) describes a running activity from station u to the successive station v. Constraint (4.4.4) makes the wait or depart decisions for the transfer connection from feeder train i to its connecting train j at station s. To sum up, these constraints (4.4.1)-(4.4.4) describe how the initial delay(s) propagate(s) within the network. For each train i the consecutive delay d_i^s can be calculated as $x_i^s - \phi_i^s$.

$$x_i^s \geq \phi_i^s + d_i^s, \qquad \forall i \in P,\ \forall s \in S,\ \forall e_i^s \in (\mathcal{E}_{dep} \cup \mathcal{E}_{arv}) \quad (4.4.1)$$

$$x_i^{s,-} - d_i^{s,-} - x_i^{s,+} \geq L_a, \qquad \forall i \in P,\ a = (e_i^{s,+}, e_i^{s,-}) \in \mathcal{A}_{dwl} \quad (4.4.2)$$

$$x_i^{u,+} - d_i^{u,+} - x_i^{s,-} \geq L_a, \qquad \forall i \in P,\ \forall s, u \in S,\ a = (e_i^{s,-}, e_i^{u,+}) \in \mathcal{A}_{run} \quad (4.4.3)$$

$$x_j^{s,-} - x_i^{s,+} \geq L_{trans} - \mathcal{M} \cdot \xi_{ij}^s, \qquad \forall i, j \in P,\ \forall s \in S,\ a = (e_i^{s,+}, e_j^{s,-}) \in \mathcal{A}_{trans} \quad (4.4.4)$$

4.4.2 *Mixed traffic rescheduling without capacity control*

At this point, we integrate freight trains into the event activity network. Note that freight trains are operated differently than the passenger trains in reality and that is taken into account in this study. We introduce three groups of constraints, namely (I) constraints associated with early departure relaxation, (II) constraints associated with rerouting and (III) constraints associated with flexibility in dwelling or passing through:

I Early departures or arrivals are acceptable except for the first station $\alpha(i)$, as described as constraint (4.4.5).

II In case of delays, freight trains can switch to another route. Constraint (4.4.6) regulates the rerouting of $i \in F$, i.e., one and only one route will be chosen from

$\mathcal{R}(i)$. For a chosen route ($v_i^r = 1$), the running time of train i on OTS $q = (s, u)$ is calculated as constraint (4.4.7). If q is not included in the chosen route, $v_{s,u}^r = 0$ and constraint (4.4.7) is automatically satisfied. Similarly, constraint (4.4.8) indicates that the departure time is not connected with the arrival time for those unused stations.

III Following up the rerouting decisions, we have to decide whether a freight train should pass through or dwell at a station. Constraint (4.4.9) makes sure that a dwelling activity can take place only if the station is included in the chosen route. Constraint (4.4.10) describes the departure time $x_i^{s,-}$, depending on whether the train passes through ($w_i^s = 0$) or stops at ($w_i^s = 1$) the station. In the latter situation ($w_i^s = 1$), the departure $x_i^{s,-}$ should take place after the superposition of (i) the real arrival time ($d_i^{s,-} + x_i^{s,+}$), (ii) the extra entering time $L_i^{v,+}$ and (iii) the leaving time $L_i^{v,-}$. In the other case ($w_i^s = 0$), the real departure time $x_i^{s,-}$ will be equal to the real arrival time ($d_i^{s,-} + x_i^{s,+}$).

$$x_i^{\alpha(i)} \geq \phi_i^{\alpha(i)} + d_i^{\alpha(i)}, \qquad \forall i \in F,\ \forall e_i^{\alpha(i)} \in (\mathcal{E}_{dep} \cup \mathcal{E}_{arv}) \tag{4.4.5}$$

$$\sum_{r \in \mathcal{R}(i)} v_i^r = 1, \qquad \forall i \in F \tag{4.4.6}$$

$$x_i^{u,+} - d_i^{u,+} - x_i^{s,-} + \mathcal{M} \cdot (1 - \sum_{r \in \mathcal{R}(i)} v_i^r \cdot b^{(s,u),r}) \geq L_a,$$
$$\forall i \in F,\ \forall a = (e_i^{s,-}, e_i^{u,+}) \in \mathcal{A}_{run}, \tag{4.4.7}$$

$$x_i^{s,-} - d_i^{s,-} - x_i^{s,+} + \mathcal{M} \cdot (1 - \sum_{r \in \mathcal{R}(i)} v_i^r \cdot b^{s,r}) \geq L_a,\ \forall i \in F,\ s \in S, \tag{4.4.8}$$

$$w_i^s \leq \sum_{r \in \mathcal{R}(i)} v_i^r \cdot b^{s,r}, \qquad \forall i \in F,\ \forall s \in r(i) \tag{4.4.9}$$

$$x_i^{s,-} - d_i^{s,-} - x_i^{s,+} \geq w_i^s \cdot (L_i^{s,-} + L_i^{s,+}), \qquad \forall i \in F,\ \forall s \in r(i) \tag{4.4.10}$$

4.4.3 *Mixed traffic control with the capacity constraints*

This section formulates the capacity constraints of the OTS to ensure a conflict-free movement of the traffic therein.

Constraint (4.4.11) arranges the sequence of any two possible departure/arrival events to/from the same OTS. There are two situations, namely (i) if both i and j will run on the same OTS q, one and only one of them will go first, i.e., either $y_{i,j}^q = 1$ or $y_{j,i}^q = 1$; and (ii) otherwise, if any of i or j does not use the described OTS q, then (4.4.11) will evaluate $y_{i,j}^s$ and $y_{j,i}$ as zero. This will void the OTS capacity between these two trains via constraints (4.4.12)-(4.4.13). Note that the route choice is also applied to passenger trains. For $i \in P$, the $\mathcal{R}(i)$ contains only one route, which is the pre-designed one. Constraints (4.4.12) and (4.4.13) add headway constraints to the departure and arrival events if relevant trains run on the same OTS.

$$y_{i,j}^q + y_{j,i}^q = \sum_{r \in \mathcal{R}(i)} (v_i^r \cdot b^{q,r}) \cdot \sum_{r \in \mathcal{R}(j)} (v_j^r \cdot b^{q,r}), \qquad \forall i, j \in P \cup F,\ \forall s \in S \tag{4.4.11}$$

$$x_j^{s,-} - x_i^{s,-} + \mathcal{M} \cdot (1 - y_{i,j}^q) \geq h_{i,j}^{s,-}, \qquad \forall i, j \in P \cup F,\ \forall s \in S,$$

$$\forall e_i^{s,-}, e_j^{s,-} \in \mathcal{E}_{dep} \tag{4.4.12}$$

$$x_j^{s,+} - x_i^{s,+} + \mathcal{M} \cdot (1 - y_{i,j}^q) \geq h_{i,j}^{s,+}, \qquad \forall i,j \in P \cup F,\ \forall s \in S,\ \forall e_i^{s,+}, e_j^{s,+} \in \mathcal{E}_{arv} \tag{4.4.13}$$

The variable $y_{i,j}^q$ coordinates both the departure order and subsequent arrival order. In this study, overtaking in this study only takes place at the stations with more than two platform tracks or at certain junctions. The re-ordering can take place at the station, i.e., $y_{i,j}^{q'}$ can be different than $y_{i,j}^q$, here q' is the precedence track after q for both trains. In this study, the departure/ arrival headway are taken from PETER (Goverde and Odijk, 2002). Besides, the no-overtaking constrains are also constructed for those stations with no more than two platform tracks. The set of those stations can be denoted as $\bar{S}$.

$$x_i^{s,-} - x_j^{s,-} \begin{cases} \geq 0 & \text{if } x_i^{s,+} - x_j^{s,+} \geq 0 \\ \leq 0 & \text{if } x_i^{s,+} - x_j^{s,+} \leq 0 \end{cases}, \qquad \forall i,j \in P \cup F,\ \forall s \in \bar{S} \tag{4.4.14}$$

4.4.4 *Objective function*

The optimization model is formulated as a mixed-integer programming problem (MIP). The objective is to minimize the total delay costs.

The set of parameter $c = (c_i^s, c_a)$ equals delay costs according to the number of passengers or value of the freight involved. For a passenger train $i \in P$, c_i^s denotes the number of passengers who arrive at the destination at event $e_i^{s,+}$. For a transfer activity $a \in \mathcal{A}_{trans}$, the number of passengers using this transfer activity is assumed to be known from history data c_a. In case of cancelled connection, passengers have to wait for the next connecting train, denoted as t_a. It is the interval of the missed connecting train i and the next connecting train i'. Thus, we have the total passenger delays as,

$$\mathcal{D}(P) = \left(\sum_{i \in P} \sum_{e_i^s \in \mathcal{E}_{arv}} c_i^s \cdot (x_i^s - \phi_i^s) + \sum_{a \in \mathcal{A}_{trans}} z_a \cdot c_a \cdot t_a \right) \tag{4.4.15}$$

For a freight train $i \in F$, c_i^s converts the freight delay costs into passenger delay costs, according to the quantity and value of carried goods being transported. Since the freight train can be rerouted or depart earlier, thus it is possible that the freight train can arrive earlier than scheduled at the fixed destination. It is allowed in this work but taken as off-target delivery and penalised according to the freight amount, freight value and the length of off-target time.

$$\mathcal{D}(F) = \sum_{i \in F} c_i \cdot \left| x_i^{\beta(i)} - \phi_i^{\beta(i)} \right| \tag{4.4.16}$$

In total we have the overall integrated passenger and freight delay costs as below.

$$(\mathrm{RTM-MPF}) \qquad \min \qquad \mathrm{f(x,y,z,v)} = \mathcal{D}(\mathrm{P}) + \mathcal{D}(\mathrm{F}) \tag{4.4.17}$$

The real-time traffic management with mixed passenger and freight trains (RTM-MPF) optimization problem now consists of the objective function (5.4.18) and the constraints (4.4.1)-(4.4.13).

4.5 METHODOLOGY

The optimization procedure results in the following decisions: the rescheduling of train operating time (x), the re-routing choice of freight trains (v) together with the pass through or stop decisions (w), and the wait or depart decisions (z) for passenger transfer connections. For this real-time traffic management problem, these decisions are expected within minutes. In this section, we describe our solution strategy for the optimization problem RTM-MPF.

4.5.1 *Generate the set of possible routes for freight trains in advance*

There are two ways to achieve the freight rerouting. The first one is to explore the situation from station to station. It starts from the station where train is located or moves towards (denoted as $\alpha(i)$) and ends at the fixed destination (denoted as $\beta(i)$). The second way is to select one route from a set of possible routes. As indicated before, we adopt the second solution to reduce the calculation time.

The k shortest path method (Yen, 1971) is applied to generate the set of possible paths from $\alpha(i)$ to $\beta(i)$. After the automatic generation of the loopless path sets, certain low quality routes are excludes to reduce the size of the problem. To that end, we discard a path if:

- it detours more than 80% of the pre-designed path.
- its net running time exceeds 180% of that of the pre-designed route.

4.5.2 *Customized value of $\mathcal{M}$*

The big-$\mathcal{M}$ is used in the constraints (4.4.4), (4.4.6), (4.4.7), (4.4.12) and (4.4.13). These constraints all regulate the service operation time. We evaluate

$$\mathcal{M} := \max\{x_i^{s,+} \mid \forall i \in P \cup F,\ \forall s \in S\} + \sum_{\forall i \in P \cup F} d_i^s, \tag{4.5.1}$$

i.e., the sum of the latest arrival time and all the initial delays.

4.5.3 *The problem size*

We examine the size of the problem and initially check the calculation time. A test case is made based on a complex and densely used railway network of the Netherlands, as shown in Figure 4.2. The network contains 30 stations, deploying 373 passenger trains and 14 freight trains within a 2-hour time windows.

For a better understanding of the problem size, we use two reference models hereby. These two reference models are described in Appendix A. For a brief depiction, Reference model 1 deals with the freight trains as though they were passenger trains. The freight trains cannot switch from passing through or dwelling at a station. Reference model 2 assigns freight trains with the flexibility to select between passing through or dwelling at the stations. Note that this flexibility is regulated only if (i) the station has spare platform tracks, and (ii) there are no freight operations involved or the time for freight operations is shorter than the actual dwell time.

The code of the optimization of the three models is implemented in CPLEX. All the tests are performed on an Intel Core 2 GHz laptop. The number of events and activities involved are listed in Table 4.2. The corresponding variable and calculation time for a test run are recorded in Table 4.3. Hereby the test run is the situation without any disturbances.

Table 4.2: Size of the Network

Case	num($\mathcal{E}_{dep}+\mathcal{E}_{arv}$) [1]		num($\mathcal{A}_{run}$, $\mathcal{A}_{dwl}$, $\mathcal{A}_{trans}$, $\mathcal{A}_{head}$) [2]		
	$\mathcal{E}(P)$ [3]	$\mathcal{E}(F)$ [3]	$\mathcal{A}_{run,dwl}+\mathcal{A}_{trans}(P)^3$	$\mathcal{A}_{run,dwl}(F)^3$	$\mathcal{A}_{head}(PP,PF,FF)^4$
RTM-FasP	663	70	953+20	126	3640,394, 48
FSta	663	70	953 +20	126	3640, 410, 70
RTM-MPF	663	388	953+20	762	3640,3453 , 2020

Note: num($\mathcal{E}_{dep}+\mathcal{E}_{arv}$)[1]: total number of departure events ($\mathcal{E}_{dep}$) and arrival events($\mathcal{E}_{arv}$);

num($\mathcal{A}_{run}$,$\mathcal{A}_{dwl}$, $\mathcal{A}_{trans}$, $\mathcal{A}_{head}$) [2]: number of running /dwelling /transfer/headway activities;

(P) (F) [3]: the events or activities take place to the passenger trains(P) or freight trains;

(PP, PF, FF)[4]: the simultaneous departure or arrival activities between two passenger trains/two freight trains/passenger train and a freight train.

The simultaneous departure or arrival activities refer to timetable times within 30 min.

Table 4.3: Size of the computation

Case	size of the calculation		calculation time (second)	
	constraints	variables	root before processing	root+ branch & cut
reference 1	15083	5740	0.08	0.08
reference 2	15652	6310	0.62	0.62
RTM-MPF	36960	33046	27.01	452.88

As can be seen from Table 4.2, the events $\mathcal{E}_P$ and activities A_P for the passenger trains remain the same in all three models. However, the $\mathcal{E}_F$ A_F increases as the level of flexibility growing. The corresponding variable and calculation time for a test run are recorded in Table 4.3.

4.6 COMPUTATIONAL TESTS AND ANALYSIS

A computational study is performed to evaluate the advantage of allowing greater flexibility to the freight trains in real-time traffic management. First, we describe the network and cases used in this study. Afterwards, we show that the proposed RTM-MPF model leads to better solutions compared to the two reference models. Finally, we conclude with a discussion of our findings.

4.6.1 *Test Network and Data Set*

The RTM-MPF model and the two reference models are tested in a real world application from a crucial section of the Dutch railway networks. We display the network in Figure 4.2. The network consists of 30 major stations, and some minor stations and junctions. The minor station and junctions are left out to make the figure easier to read. The names of the station are denoted with their official abbreviations. Their full names are listed in Table 4.4.

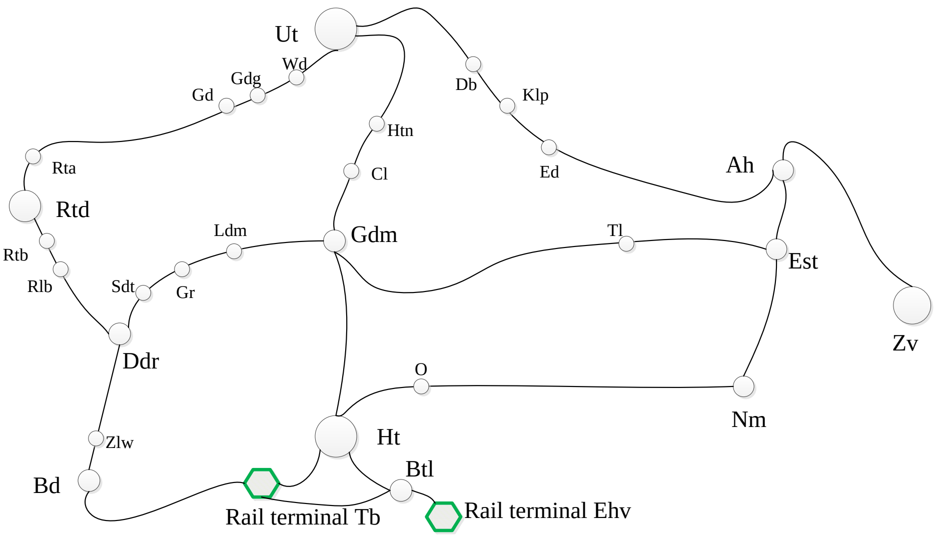

Figure 4.2: An illustration of the network. Minor stations and junctions are not visualized for the sake of clarity.

In this network, stations *Gr*, *Klp*, *Ldm*, *Rta*, *Sdt* and *Tl* have no more than two platform tracks. Thus the extra dwelling of freight trains (Eq. (4.4.2)) and overtaking (Eq. (4.4.14) can not take place in these stations. Meanwhile, there is a junction between *Gdm* and *Ht* that allows for overtaking.

The information regarding the departure and arrival time is taken from PETER (Goverde and Odijk, 2002). The passenger numbers c_i^s (arriving at station s via train i) and c_a (transfer connection a) are generated according to the passenger data (see

Table 4.4: Stations' official abbreviations and the full names

Abbr.	Stations	Abbr.	Stations	Abbr.	Stations
Ah	Arnhem	Bd	Breda	Btl	Boxtel
Cl	Culemborg	Db	Driebergen-Zeist	Ddr	Dordrecht
Ed	Ede-Wageningen	Ehv	Eindhoven	Est	Elst
Gd	Gouda	Gdg	Goverwelle	Gdm	Geldermalsen
Gr	Gorinchem	Ht	's-Hertogenbosch	Htn	Houten
Klp	De Klomp	Ldm	Leerdam	Nm	Nijmegen
O	Oss	Rlb	Lombardijen	Rta	Alexander
Rtb	Rotterdam Blaak	Rtd	Rotterdam Centraal	Sdt	Sliedrecht
Tb	Tilburg	Tl	Tiel	Ut	Utrecht Centraal
Wd	Woerden	Zlw	Lage Zwaluwe	Zv	Zevenaar

TreinreizigerNL, 2019) and the shortest path assumption. The coefficients c_i used for the freight trains are evaluated as 50. A two-hour time window is chosen to observe the real-time traffic management. Note that, some train routes are not limited to this network. We do not consider the movements outside the test network. The delay of the last stop of its route is taken as the final delay of that train.

The usage of the lines is checked to gain a basic understanding of the utilization of the lines. The most frequently used line is the *Ut–Btl* line. Figures 4.3 and 4.4 depict the trains running downwards and upwards on this line. The blue lines represent the movement of passenger trains, and the orange ones stand for that of freight trains. As can be seen, the frequency can reach 10 times per hour. This line also contains the three important transfer stations *Ut*, *Gdm* and *Ht*.

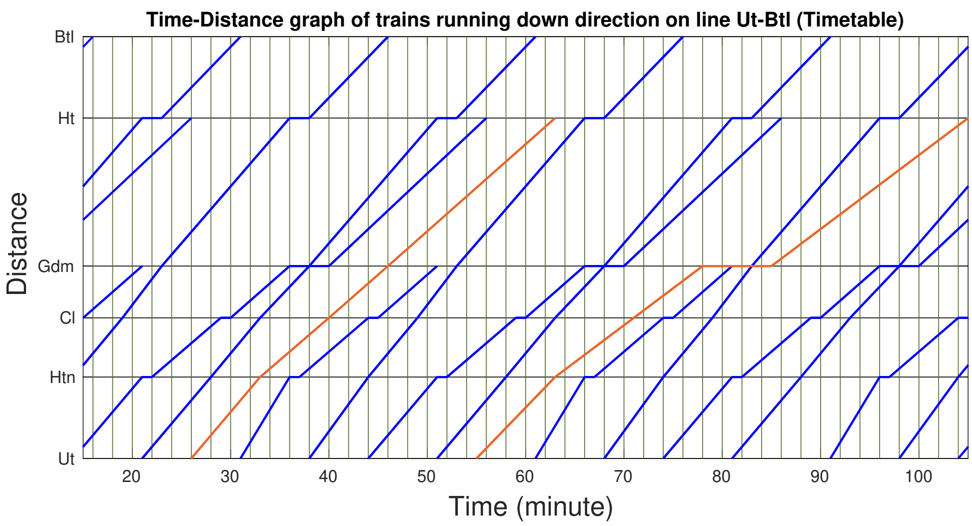

Figure 4.3: The timetable designed movement of trains on the line Ut-Btl (downwards).

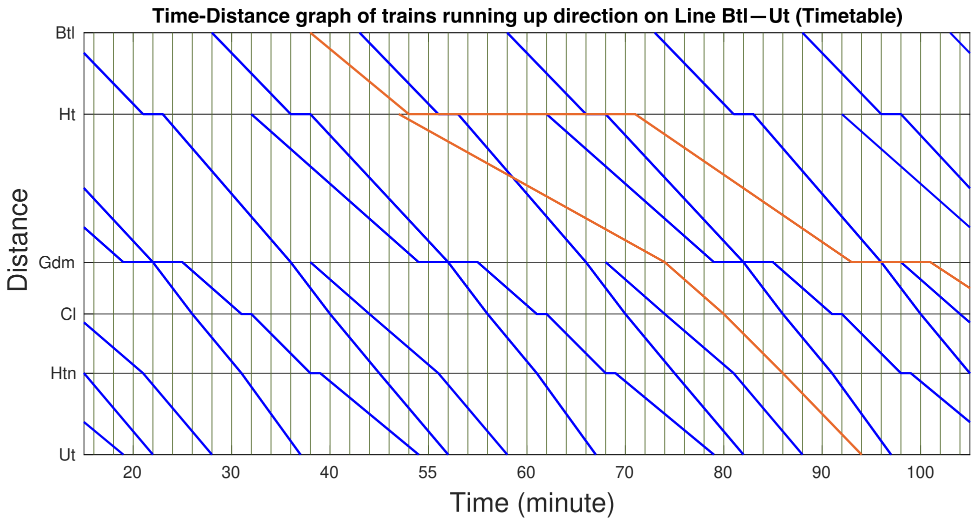

Figure 4.4: The timetable designed movement of trains on the line Ut-Btl (upwards).

The line with the lowest usage frequency is the *Ddr-Est* line, which is only used twice per hour in each direction. Part of this line is single track, and the stations on this line do not allow overtaking and extra dwelling.

4.6.2 *Test of the Model*

In this subsection, we present the solution of the proposed RTM-MPF model and compare it to the two reference models. We consider initial delays, which are departure or/and arrival delays of maximum 15 minutes. We limit the discussion to the following three most illustrative cases: (i) Early departure of the freight train (ii) Skipping certain stops (dwelling at the station) of the freight train, and (iii) Rerouting of freight trains. We also compare the total passenger delays and total freight delays in each case. We describe cases (i) and (ii) in section 4.6.2.1 and case (iii) in section 4.6.2.2. All the solutions here are optimal results from the proposed model and reference models.

4.6.2.1 *Early departure or the stop-skipping of the freight train*

The first contribution of the model lies in the flexibility reserved for the freight trains at the stations. It allows early departure of the freight trains except for at the first stop, see constraint (4.4.9). Note the early departure operations are based on the condition that there are no freight operations involved.

Freight train *F477* moves from Station *Zv* to *Ddr*. During this trip, it has access to the open track segment *Nm-O* after the passenger train *P550*, with a 3-minute headway. Consider the situation that *P550* is faced with a 5-minute departure delay from the station *Nm*. As a result train *F477* is also delayed due to the headway constraint. There are two decisions to be made. First, the decision concerning which train should go first (variable $y^{Nm-O}_{P550,F477}$). Following that, the departure times of the two trains (variable x^{Nm}_{P550} and x^{Nm}_{F477}) have to be rescheduled. The proposed RTM-MPF model changes the departure order from $y^{Nm-O}_{P550,F477} = 1$ into $y^{Nm-O}_{P550,F477} = 0$.

This means that the freight train departs 1 minute earlier than the scheduled time. The replanning solution is shown in Figure 4.5.

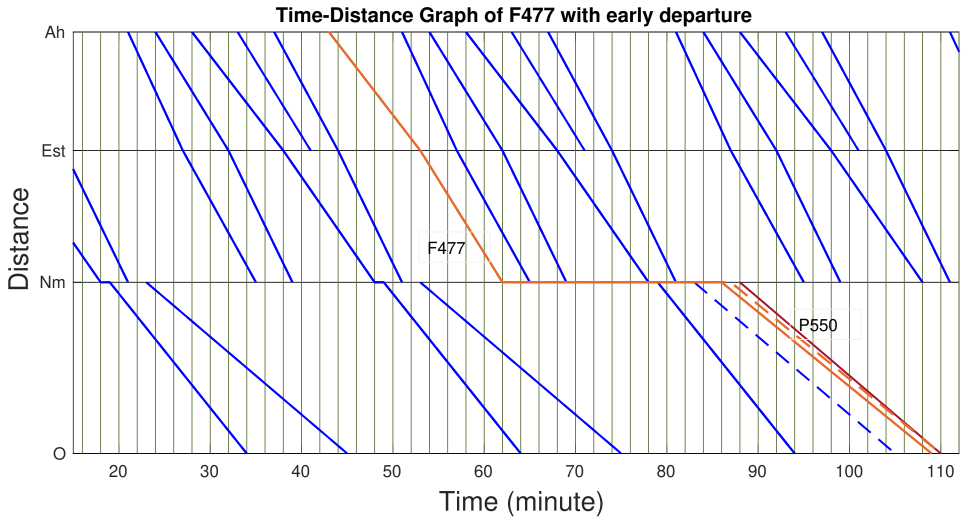

Figure 4.5: RTM-MPF. The movement of train F477 with early departure.

In Figure 4.5, the red line stands for the movement of the delayed train. The dashed lines in blue and orange colours represent the designed movement of passenger trains and freight trains respectively. The re-planned trains movement are depicted in full lines. The usage of the lines are the same in the following time distance figures.

Using the reference model RTM-FasP which does not allow early departure of the freight train leads to a delay of *F477*. This solution is shown in Figure 4.6.

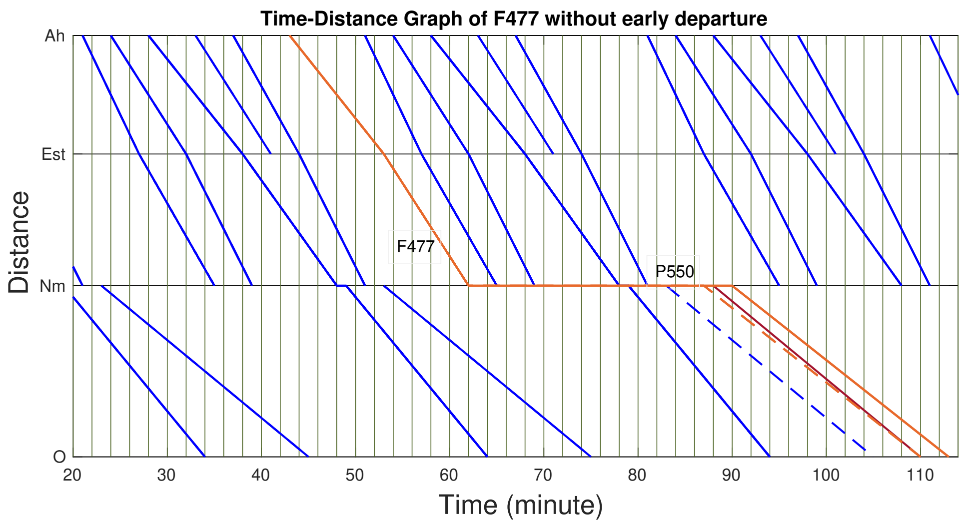

Figure 4.6: RTM-FasP. The movement of train F477 without early departure.

Stop-skipping is another form of flexibility granted to the freight trains at stations. A freight train *F408* and a passenger train *P180* use the open track segment *Ut-Htn*. Given a 4-minute departure delay of *P180*, the solutions of the proposed RTM-MPF model and the reference model 1 are given in Figure 4.7 and 4.8. It can be observed that the freight train *F408* in the RTM-MPF model skips the stop at station *Gdm*. Thus it avoids another delay at the next station *Gdm*.

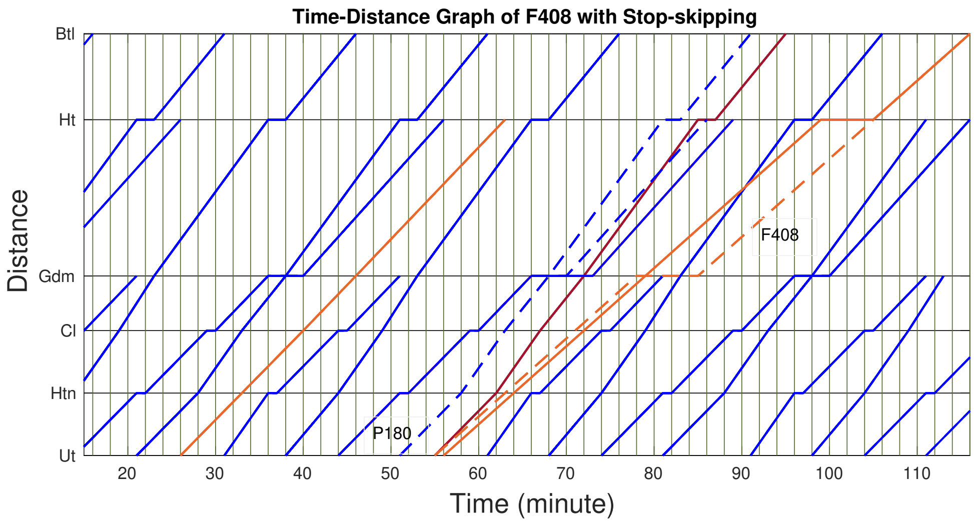

Figure 4.7: RTM-MPF. The movement of freight train F408 with stop skipping.

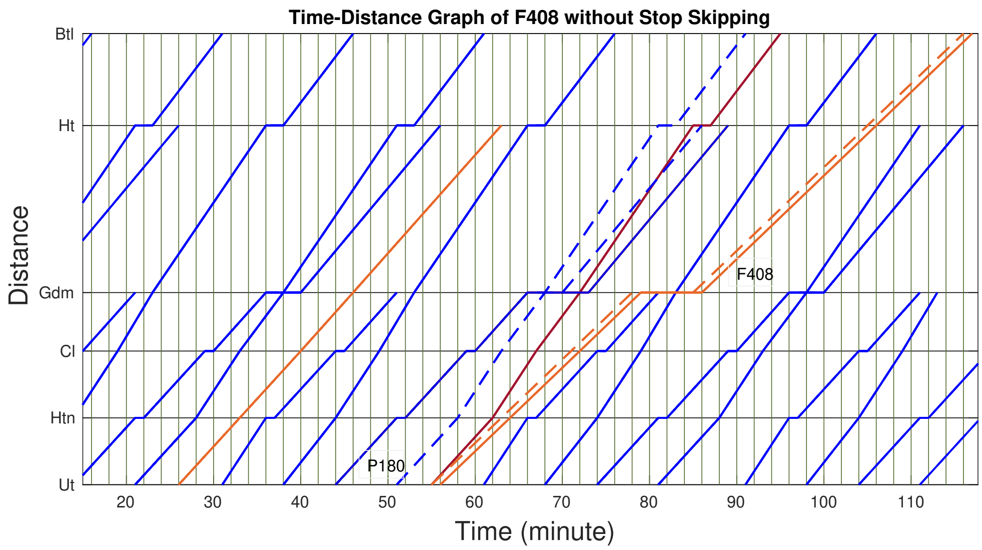

Figure 4.8: RTM-FasP. The movement of freight train F408 without stop skipping.

4.6.2.2 *Rerouting of freight trains*

Another contribution of our model is the rerouting of the freight trains in case of delays. In this network, the up and down directions of line *Ut-Btl* are both busy, as shown in Figures 4.3 and 4.4. The headways between two trains using the same segments subsequently are comparatively short. If delay occurs on this line, it will result in heavy congestion via the unavoidable propagation. We propose the rerouting of freight trains in this case to save the limited capacity of the busy line.

Hereby we provide an example of the rerouting of the freight train. Part of train *F434*'s route is on the line *Ut-Btl*. When there are two delays, which are $d_{P210}^{Tb+} = 3(min)$ and $d_{P697}^{Gdm-} = 4(min)$, the freight train *F434* will be delayed twice if it stays on the original route. The proposed RTM-MPF model reroutes, which takes 5 minutes more. This solution is shown in Figure 4.9.

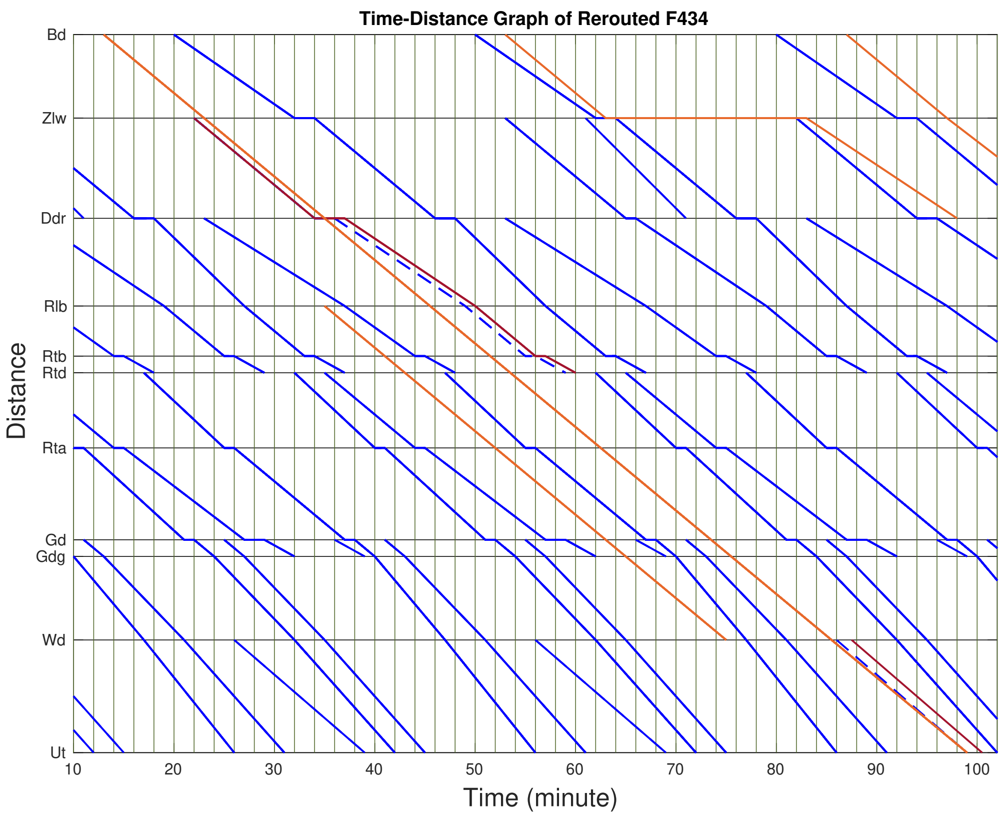

Figure 4.9: RTM-MPF. The movement of freight train F434 with rerouting.

The rerouting of the freight train *F434* takes place because of the following reasons: (i) the original route *Bd-Tb-Ht-Gdm-Cl-Htn-Ut* is partly located on a busy line on which delays occur; (ii) the net running time of the alternative route *Bd-Zlw-Ddr-Rlb-Rtb-Rtd-Rta-Gd-Gdg-Wd-Ut* takes 5 minutes more (this additional running time is less than 10% of the original running time and therefore justified); (iii) the passenger trains running on the alternative route are not carrying many passengers (if this takes place during the peak hours, when the passenger trains are carrying many more passengers, the solution may be different).

Meanwhile, *F434* will end with a longer delay if it stays on its original route, as shown in Figure 4.10.

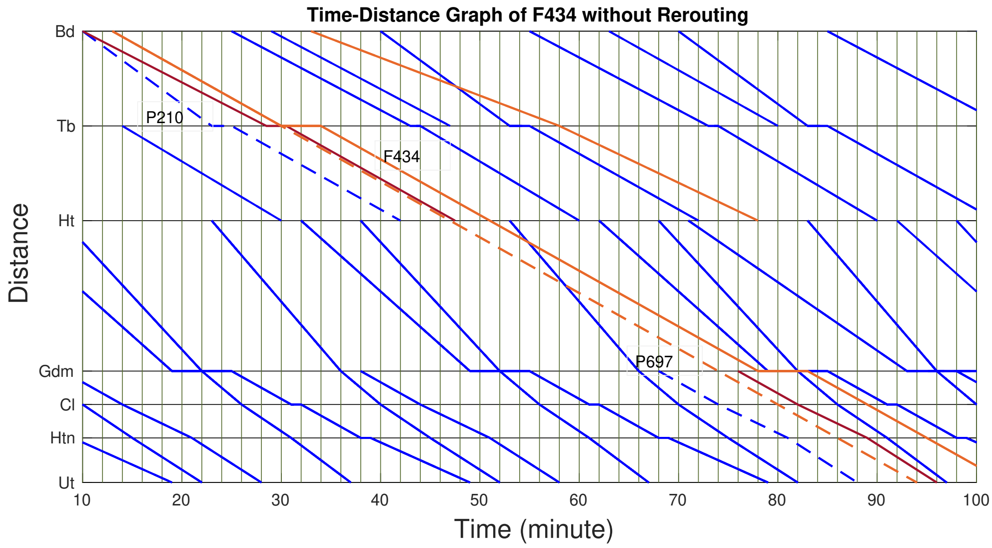

Figure 4.10: RTM-FasP. The movement of freight train F434 without rerouting.

4.6.3 *Discussion*

Table 4.5 sums the delays of the above-mentioned cases using the three different models. The delays are listed as passenger-train delays, total passenger delays, the freight-train delays, and total freight delays.

Table 4.5: Solution comparison, the train delays, the total passenger delays and the total freight delays

Case	solution	Passenger delay**		Freight delay**	
		train	$\mathcal{D}(P)$	train	$\mathcal{D}(F)$
Case 1	RTM-MPF & RTM-FSta	5	700	$\lvert -1 \rvert$*	$\lvert -50 \rvert$*
	RTM-FasP	5	700	3	150
Case 2	RTM-MPF & RTM-FSta	27	2480	0	0
	RTM-FasP	27	2480	1	50
Case 3	RTM-MPF	67.5	5875	5	250
	RTM-FasP & RTM-FSta	63	5795	8	400

*An early arrival of the freight train. RTM-MPF treats early arrivals as off target.

**: in minutes.

Table 4.5 shows that the freight-train delays and total freight delays are all shorter in the solution with the proposed RTM-MPF model. Passenger train delays and total passenger delays are longer in the third case. With the elementary examples outlined above, we demonstrate how the RTM-MPF model introduces greater flexibility to the freight trains. We note, however, that the trade-off between the passenger trains and freight trains needs to be explored in greater detail. The proportion of the freight trains in the data we use is relatively low (14 freight trains out of a total of 387 trains). With the current growth of freight transportation, the authors anticipate that a flexible replanning strategy, such as the RTM-MPF model, will become more useful.

We close with some remarks about the computational time. For the current network and time window in our example, it takes about 10 minutes to solve the RTM-MPF model. It is possible to use a benders decomposition algorithm to obtain a solution more quickly. The main idea is of this algorithm is to decompose the problem into a Master Problem (MP) and a Sub-problem (SP). First, we recall Table 4.2 and observe that if the route choice of the freight trains (v_i) is known, then the number of headway constraints involving the freight trains can be reduced significantly. As such, we suggest formulating an MP that provides a solution for the variables regarding (i) the rescheduling of trains operational time (x), (ii) the rerouting choice of freight trains (v) together with the pass through or stop decisions (w) and (iii) the wait or depart decisions (z) for passenger transfer connection. Afterwards, a SP may be formulated which handles the priority order of track usage (y) via the suggested solution of the MP. This SP provides feasible solutions. See IBM, 2020 for more information about the Benders Decomposition Algorithm via CPLEX solvers.

4.7 CONCLUSION

In this paper we discussed a real-time traffic management model for mixed passenger and freight trains (RTM-MPF). The proposed RTM-MFP model provides the replanning scheme with the least overall passenger and freight delays. The key concept is to combine the rigidity of the passenger train timetable with a more flexible freight train schedule. The underlying rationale is to maintain the entire promised passenger service while aiming only for minimum delay only at the final stop as far as the freight trains are concerned.

The proposed RTM-MPF model takes into account the specific characteristic of the heterogeneous traffic flows. Specifically, the contributions of the RTM-MPF model lie in (i) the rerouting of the freight trains, and (ii) the possibility of early departure or stop-skipping of the freight trains, and (iii) the capacity constraints to ensure a conflict-free movement of the trains at a macroscopic level. We have tested the replanning model on a crucial part of the Dutch railway network. Our results demonstrate that making the freight train schedules more flexible maintains passenger train service level intact and improves the punctuality of the freight delivery.

There are a number of possibilities for further extensions, of which we name a few here. First, we suggest to applying the replanning model to larger networks or/and longer time windows, to benefit more from freight train flexibility. Freight trains usually cover longer distance, and thus more running time. This means that the proposed replanning gains can then be more substantial. We suggest adopting a benders decomposition algorithm to reduce the computational time involved. Additionally, the trade-off between the two different traffic flows can be examined in further detail, to study the effect of freight train flexibility on the passenger trains. A Monte Carlo test can be used to gain a better understanding of the proposed solutions of the model. Last but not least, it can also be an interesting research direction to reduce the unnecessary stops of the freight trains in the transitional plan.

The freight trains usually carry heavier loads than passenger trains and thus the energy consumption of braking and restarting can be more expensive. For example we refer the reader to Corman et al., 2009 in which the authors build a detailed optimization model adjusting the dwelling time at stations in order to avoid extra stops and speed profile modifications in open corridors.

APPENDIX

4.A REFERENCE RTM-FASP MODEL

The RTM-FasP treats with the freight trains as though they were passenger trains, ignoring their operating characteristics. It does not allow the freight trains to depart early or be re-routed. It is formulated as (4.4.4), (4.4.12)-(4.4.14) and the following constraints.

$$\text{(RTM} - \text{FasP)} \qquad \min \qquad f(x,z) = \mathcal{D}(P) + \mathcal{D}(F)$$
$$= \left(\sum_{i \in P} \sum_{e_i^s \in \mathcal{E}_{arv}} c_i^s \cdot (x_i^s - \phi_i^s) + \sum_{a \in \mathcal{A}_{trans}} z_a \cdot c_a \cdot t_a \right) + \sum_{i \in F} c_i \cdot \left(x_i^{\beta(i)} - \phi_i^{\beta(i)} \right) \tag{4.A.1}$$

$$x_i^s \geq \phi_i^s + d_i^s, \qquad \forall i \in P \cup F,\ \forall s \in S,\ \forall e_i^s \in (\mathcal{E}_{dep} \cup \mathcal{E}_{arv}) \tag{4.A.2}$$
$$x_i^{s,-} - d_i^{s,-} - x_i^{s,+} \geq L_a, \qquad \forall i \in P \cup F,\ a = (e_i^{s,+}, e_i^{s,-}) \in \mathcal{A}_{dwl} \tag{4.A.3}$$
$$x_i^{u,+} - d_i^{u,+} - x_i^{s,-} \geq L_a, \qquad \forall i \in P \cup F,\ \forall s,u \in S,\ a = (e_i^{s,-}, e_i^{u,+}) \in \mathcal{A}_{run} \tag{4.A.4}$$
$$y_{i,j}^q + y_{j,i}^q = 1, \qquad \forall i,j \in P \cup F,\ \forall s \in S \tag{4.A.5}$$

Freight trains can be holder longer at certain stations, or assigned additional stops at some stations in case of delays.

4.B REFERENCE RTM-STA MODEL

The RTM-FSta deals with the flexibility of freight trains flexibly within the stations. It allows a freight train $i \in F$ to departs earlier than its scheduled departure time, except for the first stop $\alpha(i)$. Meanwhile a freight train can also skip a stop if there are no operations involved. The RTM-PSta is formulated as (4.4.1)-(4.4.5), (4.4.12)-(5.4.18) and the following constraints.

$$\text{(RTM} - \text{FSta)} \qquad \min \qquad f(x,z,w) = \mathcal{D}(P) + \mathcal{D}(F)$$
$$= \left(\sum_{i \in P} \sum_{e_i^s \in \mathcal{E}_{arv}} c_i^s \cdot (x_i^s - \phi_i^s) + \sum_{a \in \mathcal{A}_{trans}} z_a \cdot c_a \cdot t_a \right) + \sum_{i \in F} c_i \cdot \left| x_i^{\beta(i)} - \phi_i^{\beta(i)} \right| \tag{4.B.1}$$

$$x_i^{u,+} - d_i^{u,+} - x_i^{s,-} \geq L_a, \qquad \forall i \in F,\ a = (e_i^{s,-}, e_i^{u,+}) \in \mathcal{A}_{run}, \tag{4.B.2}$$

$$\begin{cases} x_i^{s,-} - d_i^{s,-} - x_i^{s,+} \geq L_i^{s,-} + L_i^{s,+} & \text{if } w_i^s = 1 \\ x_i^{s,-} - d_i^{s,-} - x_i^{s,+} = 0, \quad \text{if } w_i^s = 0 \end{cases}$$

$$\forall i \in F,\ s \in S,\ a = (x_i^{s,+}, x_i^{s,-}) \in \mathcal{A}_{DwlTh} \tag{4.B.3}$$

$$y_{i,j}^q + y_{j,i}^q = 1, \qquad \forall i, j \in P \cup F,\ \forall s \in S \tag{4.B.4}$$

Part II

SYNCHROMODALITY

In this part we present the synchromodality framework and present a synchromodality replanning model for hinterland freight transportation.

5

HINTERLAND FREIGHT TRANSPORTATION REPLANNING MODEL UNDER THE FRAMEWORK OF SYNCHROMODALITY

This chapter is reproduced from Qu et al., 2019:

Wenhua Qu, Jafar Rezaei, Yousef Maknoon, Lóránt Tavasszy, *Hinterland freight transportation replanning model under the framework of synchromodality*, Transportation Research Part E: Logistics and Transportation Review 131, 308-328 (2019).

ABSTRACT

Hinterland freight transportation is managed according to a pre-designed schedule. In daily operations, unexpected uncertainties occur and cause deviation from the original plan. Thus replanning is needed to deal with the perturbations and complete the transportation tasks. This paper proposes a mixed-integer programming model to re-plan hinterland freight transportation, based on the framework of synchromodality. It is a holistic resolution of shipment flow rerouting, consequence transshipment organization in the intermediate terminals, and corresponding service rescheduling. The replanning benefits from a high operational flexibility and coordination via a split of shipment and aligning the departure time of service flows with the shipment flows.

5.1 INTRODUCTION

Hinterland freight transportation is carried out in accordance with an elaborated tactical plan. A service Network Design (SND) problem lies at the core of the tactical planning process, addressing issues like the selection and scheduling of services, specification of terminal operations and routing of shipment flows (Crainic, 2000). These plans are designed on the basis of predicted freight volume. Stochastic and robust planning are studied extensively to prevent or limit the influence of prediction bias.

At the operational level, the tactical schedules are implemented on a local scale with a shorter time window. During this process, the transportation network is vulnerable to uncertainties and variations. Those not only rise from the external environment (unexpected additional or cancelled freight demands, weather caused hazards, delays of the shipment release), but can also internal from system fluctuations (congestion, breakdown and human-caused malpractices). The stochastic or robust planning plan is unable to handle all unexpected scenarios in everyday practice. It is still on open filed to regulate uncertainties at the operational level.

Synchromodal transportation is an upcoming new solution succeeding intermodal transportation towards a more flexible and cooperated freight transportation (Tavasszy et al., 2010 at operational level to deal with uncertainties. Under the framework of synchromodality, shippers accept a mode-free booking and only determine the price and quality requirements (Behdani et al., 2016). The involved service providers collaborate under the centralized supervision of one Logistic Service Provider (LSP) (see also SteadieSeifi et al., 2014, Li, Negenborn, and De Schutter, 2017 and Rivera and Mes, 2017). This implies that it is possible to perform real-time switching among different modes (Behdani et al., 2016).

The centralized LSP holds the promise of generating specific replanning solutions to the specific uncertainties encountered in each individual case. The replanning questions come in: (1) Which part of the original SND plan can be re-planned? (2) To what degree can the re-planning solutions deviate from the SND plan? As mentioned above, the SND mainly consists of three components: the shipment route choices, the terminal operations, and the service schedules. With regard to the first component, it is important to deliver shipments to their destinations on time. Hence, complete flexibility can be given to the shipment route replanning. This is under the agreement of mode-free booking and real-time switching. The thorough re-routing of shipments unavoidably involves the replanning of terminals operations and services schedules. Those can be integrated in the replanning and benefit from the cooperation of all synchromodality operators.

This replanning model further carries out two consecutive synchronization tasks: to derive the corresponding transshipment flows at the terminals, and to synchronize the service rescheduling with the shipment rerouting. A dilemma emerges for to the LSP while rescheduling the LCS. The higher degree of flexibility given, the more flows can be diverted from truck services to LCS. However, that will also generate higher costs as a result of deviating from the SND plan (paying for cancellations or postponing services) and retrospective re-positions. The trade-off between flexibility and deviations is crucial for the rescheduling of LCS. This work explores how LSP can synchronize these intertwined tasks for hinterland freight transportation replanning.

This study includes the following contributions to existing research. We construct a model that regulates the shipment flow re-routing and services re-scheduling. It can be used for a hinterland freight transportation network under the framework of synchromodality. The shipment rerouting is given a higher degrees of freedom, involving the split and bundling of shipments. Consecutively, the transshipment flows at the intermediate terminals are extracted from the shipment flow rerouting. Transshipment is an important link in the overall process, albeit one that is sometimes neglected and over-simplified in some studies. Last but not least, moderate flexibility is given to the operating times of services, to be synchronized with the shipment flows. Alignment is introduced to LCS, i.e., the barge and rail services are given a buffer time that transcends the scheduled time. Thus shipment flows *en route* can be shifted from truck services towards LCS.

The model can be used for daily planning at the operational level. It can be embedded as the 'real-time layer' of the SYNCHRO-NET service platform (see Figure 2 of Giusti et al., 2019). In this paper, the model is tested in the Rotterdam hinterland transportation network, in cases involving late release of shipments, latency of LCS, volume fluctuation of shipments, and a variety of mixed perturbations. The model can provide a replanning resolution within seconds. In addition, we compare this model to other rigid replanning methods. The results show that the proposed synchromodality replanning model can save more overall operating costs, improves the modal split of barge and rail services, and improves LCS utilization. Finally, the model is tested with theoretical instances, with different network sizes, shipment amount and network typologies, providing that the model is able to solve small-size problems efficiently.

The remainder of this paper is organized as follows. The next section provides a review of relevant studies those incorporate uncertainties in transportation planning/replanning. The problem is formally described in Section 3, and presented mathematically in Section 4. The mathematical model describes the synchronization scheme of shipment re-routing, transshipment flows at the intermediate terminals and service rescheduling. The replanning algorithm is calibrated and described in Section 5. Section 6 tests the proposed model in the Rotterdam hinterland network with a variety of perturbations, with the solutions being provided tailored to the specific situations. The model is further analyzed in Section 7 in three aspects: (1) critical path and re-scheduling flexibility; (2) comparison with other two rigid replanning methods; (3) theoretical instances to check the solving difficulties. Finally, in Section 8 we present our conclusions and discuss possible avenues for future research directions.

5.2 LITERATURE REVIEW

In this section, we take a look at the studies which deal with unexpected fluctuations in freight transportation planning. The studies discussed here can be divided on the basis of the uncertainties (also called perturbations) involved.

Uncertainties in freight transportation can rise in exogenous and endogenous form. Exogenous uncertainties mainly come from the shipments, in the form of volume fluctuations, release date deviations, and the corresponding mutual impact. Endogenous uncertainties are usually time-dependent, such as longer travel times due to congestion, late arrivals of services, and late releases of empty containers. Both exogenous and endogenous uncertainty studies can be divided further on the basis of the planning horizon, i.e., the tactical, operational, and mixed tactical-operational level. In practice, the planning is sometimes crossed, resulting in mixed tactical-operational level problems. Table 5.2.1 provides an overview of the corresponding literature.

Table 5.2.1: Literature of studies dealing with uncertainty.

	literature	Uncertainty	Planning level	Modes	Routes	Shipment split	Transshipment	Solutions
Exogenous	Lium, T., and S., 2009	shipment	t	H	p	✓	-	outsourcing
	Hoff et al., 2010	shipment	t	H	p	✓	-	outsourcing
	Bai et al., 2014	shipment	t	H	p	✓	-	outsourcing and services' itinerary re-planning
	Meng, Wang, and Wang, 2012	shipment	t	B	a	-	-	changing ship size
	Rivera and Mes, 2017	shipment	t-o	B,R,T	p	-	-	picking up date re-planning
Endogenous or mixed	Andersen and Christiansen, 2009	transit time	t	R	p	-	-	-
	Clark and Watling, 2005	transit time	t-o	T		-	-	-
	Escudero et al., 2013	transit time	t-o	T		-	-	shipment flow re-assigning
	Goel, 2010	transit time	t-o	R,T	p	-	✓	shipment flow reassigning
	Li, Negenborn, and De Schutter, 2015	mixed	t	B,R,T	a	-	-	shipment flow reassigning
	Riessen et al., 2015	mixed	t-o	B,R,T	p	✓	✓	shipment flow reassigning

Note: Modes: B: Barge, R:Rail, T: Truck. H: homogeneous, not mentioned which one;
Planning level: t: tactical level planning, t-o: mixed tactical and operational level;
Routes: p: path level; a: arc level;

The emphasis of this review is on the types of uncertainty and the corresponding solution methodologies. As such, some of the studies discussed here concerns single mode transportation or supply chain logistics. For a more detailed review of multimodal freight transportation, see SteadieSeifi et al., 2014.

5.2.1 *Exogenous uncertainty*

Shipment stochasticity is the most studied group of exogenous uncertainty. It is mainly addressed from the tactical planning. Usually the shipments are assumed to have a known probability distribution. And the prevention or mitigation plan is modelled and solved via two-stage approach. The first stage is formulated as a deterministic network flow planning (NFP) or service network design (SND) problem. And the uncertainties are integrated in the following stage via *stochastic programming* or *robust optimization*.

Lium, T., and S., 2009 studied shipment volume stochasticity (including volume, and release date) in the SND problem. The continuous distributed stochastic volume is approximated by scenario generations, and handled by outsourcing the unexpected shipment to other service providers. Based on the work of Lium, T., and S., 2009, Hoff et al., 2010 developed a variable neighbourhood search-based approach to solve a large-scale SND problem with shipment volume stochasticity. Bai et al., 2014 extend the work by extra solution of itinerary re-planning, i.e. adjusting the service itinerary of services. Meng, Wang, and Wang, 2012 propose to handle the seasonal shipment uncertainty by deploying different amounts and sizes of ships for a liner ship fleet management. We note that each of the above mentioned works considers homogeneous services within the network.

As an alternative to stochastic programming, researchers also handle uncertainty in SND by robust optimization. Ukkusuri, Mathew, and Waller, 2007 propose a robust SND model in which the service network is designed to be less sensitive and higher resilient to shipment uncertainty. Atamtürk and Zhang, 2007 describe a two-stage robust optimization strategy by deferring a subset of the shipment decisions to the second stage, i.e. taking the decision after the realization of the uncertain demand. Including the option to decline unexpected shipments is proposed by List et al., 2003. If an accepted shipment is out of the service capacity, then the model allows the service provider to choose a different fleet size. We emphasize that also the discussed robust optimization literature deals with homogeneous modes of services.

The stochastic programming (Lium, T., and S., 2009, Hoff et al., 2010 and Bai et al., 2014) is a considerable step forward toward incorporating the effect of uncertainty into stochastic programming. However, it should be pointed out that outsourcing the shipment flows done there is not a good recourse to address unexpected shipments. The service provider loses business to competitors and resultant operational costs are expensive. On the other hand, the solution of robust optimization handles the uncertainty by providing extra service capacity or buffer time at critical nodes and links, (see Ukkusuri, Mathew, and Waller, 2007). It appears that the solution obtained via robust optimization needs higher investment. Both stochastic programming and robust optimization result in a deterministic solution which obviously is not capable of handling all possible uncertainties occurring in the future.

Rivera and Mes, 2017 is one of the early studies solving the shipment uncertainty under the framework of synchromodality. They assume the LSP has the probabilistic knowledge of the shipment and will gradually known the probabilistic over time. The problem is settled using a look-ahead planning, i.e., deciding to directly pick-up or choose to postpone the shipments. Since Rivera and Mes, 2017 concerns the day to day planning, some details are neglected. For example, the model does not deal with the operating schedules of services, modal shares, capacity utilization rate, and related transshipment connections.

5.2.2 *Endogenous uncertainty*

Travel time is the most studied endogenous uncertainties. The uncertainty is usually affected by congestion, weather condition or other time-dependent factors. Thus it is mainly addressed at operational level or mixed tactical-operational level.

Andersen and Christiansen, 2009 include travel time uncertainty in the SND problem of a railway network. They directly convert the consequence of this time uncertainty into unit flow cost, and quantify the impact of uncertainty to service quality. Clark and Watling, 2005 propose a way to estimate the probability distribution of the transportation time for a road (truck) network. These aforementioned works focus on quantifying the travel time uncertainty.No discussion are provided about the resolution approach.

Goel, 2010 quantifies the influence of travel time visibility for a road and rail network. The shipment routing is adjusted according to the travel time uncertainty. He concludes that the on-time delivery performance can be improved by increasing the visibility of the travel time uncertainty. Escudero et al., 2013 suggest a dynamic approach to solve the transit time uncertainty for a daily drayage problem. To handle the changed condition, the model reassigns the shipments on condition of real-time knowledge of the trucks' location.

The above literature exclusively deals with one sort of uncertainty. However, in real practice, the network can experience multiply uncertainties. Li, Negenborn, and De Schutter, 2015 integrates shipments uncertainty and highway travel time uncertainties into NFP problem. They calculate the truck travel time based on the density of truck flows on the highway. The following NFP planning tackles the uncertainty under a cooperative mechanism among different service providers. We note that the due time requirements for shipments are not provided in this work.

At the same time, delays (high frequency low impact) or disruption (low frequency high impact) are inevitable to the transportation network. There are very few works investigating how to handle delays or disruptions and get the network recovered in the domain of multi-modal freight transportation. Riessen et al., 2015 presents one of the very few disputations management strategies for an intermodal network. The strategy is to perform the dynamic SND planning using the updated information of available services and remaining shipments tasks. No new approaches, such as rerouting, re-assignment or itinerary replanning, are introduced in replanning stage. Properly modeling flexible management which can handle the deviations from the original plan at the operational level is still missing.

To sum up, most of the existing studies from the tactical level are mainly dealing with shipment uncertainty for a single mode network. The main target is to prevent and mitigate the effect of uncertainties in advance. The Dynamic SND, at mixed tactical-operational level, can capture some time-dependant uncertainty. Those are a step forward from the deterministic programming. However, how to do a re-planning to handle any unexpected uncertainty for a multimodal network is still an open field. Meanwhile, there are pervasive studies for re-planning in public (passenger) transport system and (passenger or mixed passenger and freight) railway system, just to name some, Binder, Maknoon, and Bierlaire, 2017, Corman et al., 2017. Those studies provide useful ideas for the freight transportation network to handle uncertainty at the operation level, such as the rescheduling of the services, itinerary or local-itinerary replanning. However there is an important difference between passenger transportation and freight transportation: freight flows are under full control. Thus, one significant solution can be obtained from the flow assignment replanning. Synchromodality adopts a centralized LSP organization, and allows mode-free booking, timely switching among services. Thus synchromodality places a corner stone for the flexible management to address any kind of uncertainty at the operational level.

Based on the aforementioned literature study and the development of synchromodality, we propose a model for freight transportation at the operational level. It can be used for a daily planning, or act as a 'shuttle/transaction plan' for the time window after disturbances and before the recovery to the regular plan. It takes the tactical planning as input, and formulate the flow assignment re-planning and re-scheduling of services. Note here we do not consider the situation of major disruptions, it is assumed that the terminals and roads are capable to continue the services.

5.3 PROBLEM DESCRIPTION AND ASSUMPTION

Consider a synchromodal network engaged in transporting shipments (or freights) from many origins to many destinations, using various service modes. Our model implements the re-planning of involved operators to cope with uncertainties at the operational level.

Let N denote the set of nodes of the network, standing for terminals which can be ports, railway stations or truck hubs. Mark each service by a sequence number $v \in V$. The arc (also named as link or leg) $a \in A = (i, j, v)$ indicates a service v from terminal $i \in N$ to another one $j \in N$. Thus the physical transportation network can be represented as a hyper graph $G(N, A)$.

A service v is described using an attributes tuple $(m_v, u_v, f_v, \psi_v, p_v)$. The first entry m_v denotes the service mode, which is either a barge, rail, or truck; u_v is the capacity of the service; f_v represents the fixed cost as long as the service is used. ψ_v refers to cancellation costs, applicable to the situation where a pre-planned barge or rail service v is not used in the replanning; p_v is the itinerary of the service, which is pre-specified from the SND plan.

An arc a is regulated with the tuple $[\pi_a^{\mathrm{dep}}, \pi_a^{\mathrm{arv}}, k_a, c_a]$. Remind that for the rail and barge services, tactical planning pre-specifies the itinerary, frequencies, departure

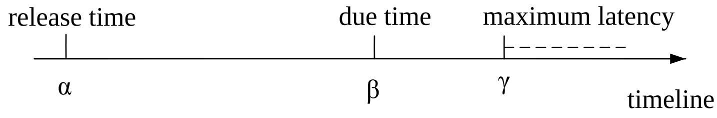

Figure 5.3.1: Release time, and delivery window of one shipment

times and corresponding arrival times. π_a^{dep} and π_a^{arv} represent the pre-plan the departure time and arrival time. k_a is the required traverse time and the variable cost of using this arc. c_a is the unit cost of the container transportation, which will be used in combination with the load volume.

Here we distinguish the arc $a = (i, j, v)$ and the service v. There can be parallel arcs between two terminals if that region contains more than one service. Meanwhile, the service's itinerary p_v can cover more than two terminals, i.e., $||p_v|| \geq 2$. The order of an element $i \in p_v$ is marked as $q_{i,l}$; The itinerary and frequency (or fleet size) of LCS remain the same as the original plan in this study. However, an additional buffer time ϕ_a^{dep} is introduced to reschedule the departure time of services. Consequently, the real departure time of LCS are regulated between the range $[\pi_a^{dep}, (\pi_a^{dep} + \phi_a^{dep})]$.

In a similar way, we characterize each shipment $s \in S$ by $(n^s, d^s, \alpha^s, \beta^s, \gamma^s, [\rho^{s,-}, \rho^{s,+}])$. Here s is a load of containers, with the volume denoted as n^s, to be transported from the same origin node o^s to the same destination node d^s. Each s has a release time α^s at the origin and expected due time β^s at the destination terminal, as shown in Fig. 5.3.1. Both early and late arrivals are allowed, where the early arrival is charged via the storage costs $\rho^{s,-}$, while a delay penalty $\rho^{s,+}$ is incorporated due to a decrease of service quality. The arrival time of each commodity can not exceed the pre-set maximum latency, noted as γ^s.

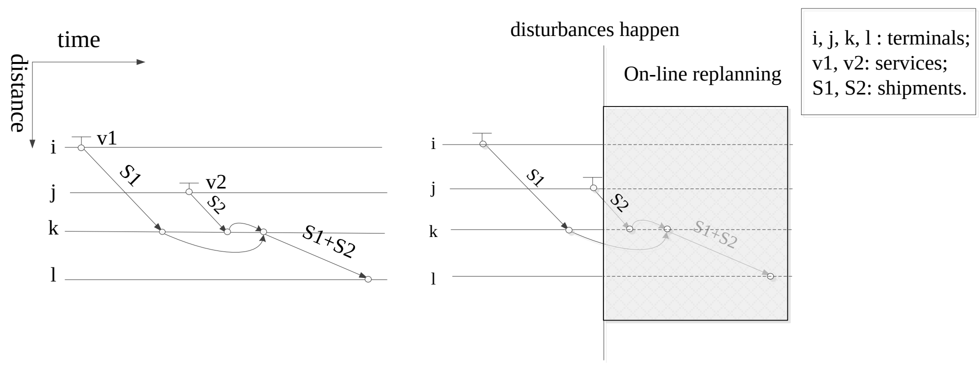

Figure 5.3.2: Initial Service schedules and Online planning after disturbances.

In daily operation, various sources of disturbances inevitably cause perturbations to the pre-specified transport operations. We define the state of a system as disturbed if: any shipment s is released later than α^s at the origin, any pre-specified service v is broken down, any v which carries shipment(s) experiences a late departure or arrival, any arc a is known to be in congestion or blockage and will lead to longer transit time, or a combination of some of those situations.

Possible influences can be: broken connection of the shipments on the way, violation of the due time, a large increase of in operational costs, etc. In this situation, LSP can use our model for short-term online replanning, to deal with the occurred disturbance(s) and complete the transportation of the shipments. As shown in Fig. 5.3.2, the replanning will be conducted directly after the delays, using the updated information of services and the shipments. Moreover, the system goes back to normal once the aforementioned situation has been solved and the transportation network can operate as originally planned.

Remind that under the framework of synchromodality, all service providers work together under the centralized organization of LSP. Perboli et al., 2017 studied the collaboration of the stakeholders from the business perspectives. Our model is build based on the assumption of this collaboration. For the synchromodal replanning, we only consider shipments those have already been accepted. This means no new shipments will be accepted during the replanning. Rather than accepting new shipments and generating new revenue, it is more rational to finish existing shipments and to avoid the penalty of delayed or failed delivery. Finally, the updated information about the real-time locations of services is assumed to be available. Based on the full deterministic information, the resultant model will be optimized for one single stage. This open-loop control avoids the issues involving the consistency of solutions across different stages, which is the inherent problem of closed-loop control (multi-stage planning).

We assume that cancelling certain pre-planned services is feasible with cancellation costs when the services are quite sufficient for the shipments. Meanwhile adding extra services or changing pre-planned routes lie beyond the scope of this study.

5.4 MATHEMATICAL FORMULATION

The aim of this section is to construct an integrated and detailed model of the shipment flow assignment replanning and real-time deployment of the service fleet. We model the online replanning in three parts. (1) the shipment flow re-planning for a capacitated network denoting the transportation of shipments, at the arc level; (2) The service rescheduling at the arc level, mainly involving the service rescheduling and service usage. (3)The shipment flow re-planning and service rescheduling are connected through synchronization at the terminal, providing a dynamic solution within a day (Zhang and Pel, 2016, Corman, Viti, and Negenborn, 2017).

This model is expressed in terms of the decision variables with x_a^s and y_a^{dep}. x_a^s denotes the number of containers from shipment s routed to arc $a = (i, j, v)$; and $x_a^s \in \mathbb{N}^0 = \{0, 1, 2, ...\}$. y_a^{dep} denotes the rescheduled departure time via arc $a = (i, j, v)$.

5.4.1 *Shipment flow conservation*

Using the decision variable x_a^s, we regulate the consistence of the shipments from their origins to destinations throughout the network. Here we define $\delta_i^+ =$

$\{(i,j,v) \mid j \in N, j \neq i, v \in V\}$ as the set of arcs starting from node i. Similarly, $\delta_i^- = \{(j,i,v) \mid j \in N, j \neq i, v \in V\}$ denotes the set of arcs arriving at node i.

Constraint (5.4.1) indicates that the difference value of incident volume and outgoing volume of shipment s at a certain node i should equal the required volume. To be more explicit, the required volume equals $-n^s$, n^s, and 0, at the shipment's origin terminal o^s, destination terminal d^s, and other transshipment nodes $i \in N \setminus \{o^s, d^s\}$.

$$\sum_{a \in \delta_i^+} x_a^s - \sum_{a' \in \delta_i^-} x_{a'}^s = \begin{cases} -n^s, & \forall s \in S, \text{ if } i = o^s \\ 0, & \forall s \in S, \text{ if } i \in N \setminus \{o^s, d^s\}\ ; \\ n^s, & \forall s \in S, \text{ if } i = d^s \end{cases} \tag{5.4.1}$$

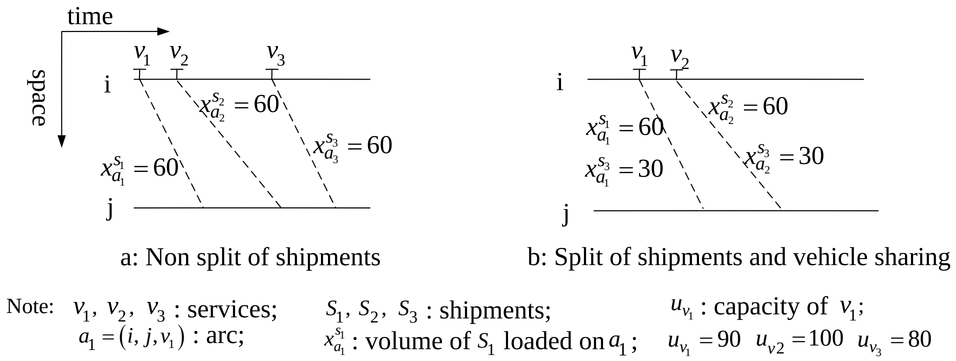

Figure 5.4.1: A sample of shipment split and service sharing.

Constraint (5.4.1) makes it possible to split one shipment into more than one group of containers. It makes the most of the capacity of LCS. This shifts shipment flows to LCS and leads to a reduction in the transportation costs. Fig. 5.4.1 provides a simple example of the comparison of a *non-split* and a *split* of shipment. In the latter case, shipment s_3 is split among different services, using the remainder of those services' unused capacity.

Furthermore, constraint (5.4.2) makes sure the total load to arc $a = (i,j,v)$ from all shipments cannot exceed the capacity of the service v. To avoid any confusion, we want to make it clear again at this point that service v is pre-designed from node i to node j. Here the services' itineraries are consistent with the original SND plan.

$$\sum_{s \in S} x_a^s \leq u_v, \quad \forall a = (i,j,v) \in A; \tag{5.4.2}$$

To synchronize the shipment and service flows, the LSP needs to track the occupancy of arc a with the shipment of s. Thus an auxiliary binary variable b_a^s is introduced as an occupancy indicator. Meanwhile we adopt a sufficiently big number M to create the following constraints. In constraint (5.4.3), the M forces the $b_a^s = 1$ if the arc a is engaged by shipment s.

$$x_a^s \leq \mathcal{M}_{load} \cdot b_a^s, \quad \forall s \in S,\ \forall a = (i,j,v) \in A; \tag{5.4.3}$$

The LSP also has to monitor whether or not the pre-arranged service v is used or not in the re-planning. To that end, another auxiliary binary variables z_v is used for this purpose. Constraint (5.4.4) regulates $z_v = 1$ if the service v is used via any arc along its itinerary p_v.

$$\sum_{s\in S}\sum_{a\in p_v} x_a^s \le \mathcal{M}_{load}\cdot z_v, \quad \forall v \in V; \tag{5.4.4}$$

5.4.2 *Transshipment*

This section extracts the transshipment at intermediate terminals as a result of the rerouting of the shipment flows. It is necessary to trace the transshipment regarding both the load and time dimension, since they generate costs and will further affect services rescheduling.

There are two kinds of transshipment: (1) transshipment involving the shipment flows. We model the transshipment with variable $n^s_{a,a',i}$, representing the number of containers belonging to shipment s from arc a to arc a' at a transshipment terminal $i \in N \setminus \{o^s, d^s\}$. At this point, we have to mention that there is no transshipment if the container uses the same service vehicle. (2) transshipment involving service vehicles. The binary variable $e_{a,a',i}$ is used to indicate whether services from arc a and arc a' need to be synchronized from the time dimension.

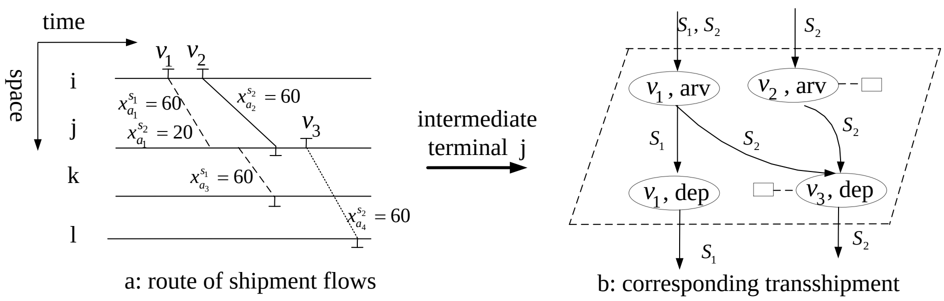

Figure 5.4.2: A schematic illustration for transshipment

For the sake of simplicity, we define the arc before and after the transshipment as the feeder arc and the connecting arc respectively. Based on the pre-specified itinerary p_v, we record $F(v)$ and $L(v)$ as the first and last node of the itinerary. If node $i \in (p_v \setminus \{F(v), L(V)\})$, then the service v transverses node i via incident arc $\overleftarrow{(i,v)}$ and outgoing arc $\overrightarrow{(i,v)}$. For example, in Fig. 5.4.2, $\overleftarrow{(j,v_1)} = (i,j,v_1)$, $\overrightarrow{(j,v_1)} = (j,k,v_1)$, $L(v_2) = j$, $F(v_3) = j$. If the volume from shipment s should be unloaded from feeder arc a at node i, this volume is denoted by $\Delta^{s,-}_{a,i}$. Similarly, $\Delta^{s,+}_{a,i}$ refers to the volume from shipment s which will be loaded to connecting arc a.

The derivation procedure of tracking $n^s_{a,a',i}$ (the transshipped volume of shipment s) from $n^s_{a,a',i}$ (the re-routed shipment flow s) is described as following:

1) There exists a transshipment for shipment s at terminal i ($n^s_{a,a',i} > 0$) if and only if $\Delta^{s,-}_{a,i} > 0$ and $\Delta^{s,-}_{a,i} > 0$.

2) Constraints (5.4.5) and (5.4.6) describe the possible volume fluctuation (increase or reduction) of shipment s load on service v at node $i \in p_v$. The first line of Constraint (5.4.5) describe the situation in which the service v transverses i, meanwhile the second line describes when i is the last stop of service v.

$$\Delta_{a,i}^{s,-} \geq \begin{cases} (x_a^s - x_{a''}^s) & \text{if } L(v) \neq i, \text{here } a'' = \overrightarrow{(i,v)}, \\ x_a^s & \text{if } L(v) = i, \end{cases}$$
$$\forall s \in S, \forall i \in N \setminus \{o^s, d^s\}, \forall a = (i,j,v) \in \delta_i^-; \qquad (5.4.5)$$

$$\Delta_{a',i}^{s,+} \geq \begin{cases} (x_{a'}^s - x_{a'''}^s), & \text{if } F(v') \neq i, \text{here } a''' = \overleftarrow{(i,v')}, \\ x_{a'}^s, & \text{if } F(v') = i, \end{cases}$$
$$\forall s \in S, \forall i \in N \setminus \{o^s, d^s\}, \forall a' = (i,j,v) \in \delta_i^+; \qquad (5.4.6)$$

3) The value of $n_{a,a',i}^s$ can be determined as a minimum value between $\Delta_{a,i}^{s,-}$ and $\Delta_{a',i}^{s,+}$, which is presented as constraint (5.4.7). If $\Delta_{a,i}^{s,-} \geq \Delta_{a',i}^{s,+}$, then volume $\Delta_{a,i}^{s,-}$ are shifted to more than one services. Otherwise, the increased flow $\Delta_{a,a',i}^{s,+}$ are from more than one services.

$$n_{a,a',i}^s = \min\left(\Delta_{a,i}^{s,-}, \Delta_{a',i}^{s,+}\right), \forall s \in S, \forall a \in \delta_i^-, \forall a' \in \delta_i^+,$$
$$\forall i \in N \setminus \{o^s, d^s\}, v' \neq v; \qquad (5.4.7)$$

4) Finally, the binary variable $e_{a,a',i}$ takes value one if there is a transshipment between arc a and arc a', which is presented in constraint (5.4.8).

$$\sum_{s \in S} n_{a,a',i}^s \leq \mathcal{M}_{\text{load}} \cdot e_{a,a',i}, \qquad \forall i \in N \setminus \{o^s, d^s\}, \forall a \in \delta_i^-,$$
$$\forall a' \in \delta_i^+, v \neq v' \qquad (5.4.8)$$

5.4.3 *Service re-scheduling to synchronize shipment flows and service flows*

Constraints (5.4.9a)-(5.4.13) discipline the rescheduling of services in the network. The actual departure time of service y_a^{dep} should be synchronized with the re-routed shipment flow x_a^s and the transshipment connection $e_{a',a,i}$. Meanwhile, the operating time of a barge or rail service should be based on the original timetable. Since it is impractical for these services to deviate too much from the pre-planned timetable. Note here the truck services are more flexible in the departure time.

The traverse time of an arc is k_a, making it possible to obtain the actual arrival time y_a^{dep} from the departure time y_a^{dep}. Parameters $h_{a,i}^{s,+}$ and $h_{a,i}^{s,-}$ are used to denotes the time needed to load or unload shipment s to or from an arc a at node i. Parameter $r_{a,a',i}$ stands for the duration for transshipment from arc a to a' at terminal i. Note in this work, we assume two or more loading activities can be executed pair-wisely. This also applies to the unloading and transshipment activities. We define ϕ_a^{dep} as a buffer time to regulate the maximum extra time it can be held. Fig. 5.4.3 illustrates the logic behind this buffer time.

Constraints (5.4.9a) and (5.4.9b) adjust y_a^{dep}, i.e., the actual departure time of a service. It is decided by the pre-designed departure time π_a^{dep} and the extra buffer

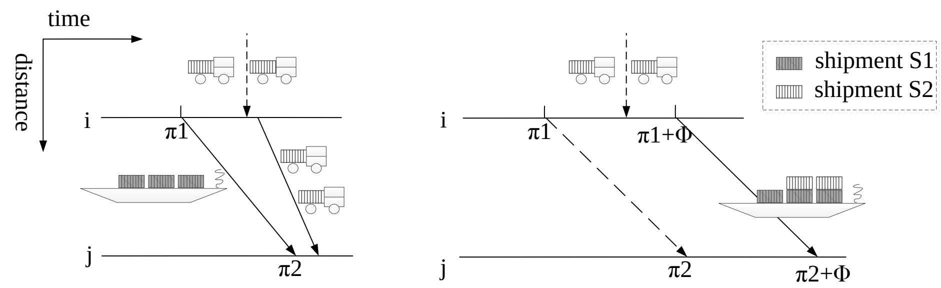

Figure 5.4.3: Alignment of services departure time to synchronize the shipment flows.

time ϕ_a^{dep}. The evaluation of the buffer time depends on the service $a = (i, j, v)$. More precisely, it depends on the service mode m_v, the terminals i and j. (5.4.9a) reschedules barge and truck services, π_a^{dep} is the maximum length of holding time. For the rail services, we assume step-wise buffer time as shown in (5.4.9b). The actual departure time of a rail service is quite difficult to predict due to the complexity of railway operations. Thus we assume it is possible within every time unit ϕ_a^{dep}, for example 30min or 60 min. The maximum holding time can be $k_{max} \cdot \phi_a^{dep}$.

$$y_a^{dep} \leq \pi_a^{dep} + \phi_a^{dep}, \qquad \forall a = (i, j, v) \in \delta_i^+,\ m_v \in \{barge, truck\}, \forall i \in N; \tag{5.4.9a}$$

$$y_a^{dep} = \pi_a^{dep} + k \cdot \phi_a^{dep}, \qquad \forall a = (i, j, v) \in \delta_i^+,\ m_v = rail, k \in \{0, 1, 2..k_{max}\},\ \forall i \in N; \tag{5.4.9b}$$

In addition, at the origin terminal of shipment s (i.e., o^s), the departure time of service a must be after the release time, which is regulated in constraint (5.4.10). Recall from constraint (5.4.3), $b_a^s = 1$ if flow from shipment s is rerouted to arc a.

$$y_a^{dep} + \mathcal{M}_{\text{time}} \cdot (1 - b_a^s) \geq \alpha^s + h_{a,i}^{s,+}, \quad \forall s \in S, \forall a = (i, j, v) \in \delta_i^+, i = o^s; \tag{5.4.10}$$

Constraint (5.4.11) derives the auxiliary decision variable y_a^{arv} from the running activity. The arrive time (at the end of the arc) equals the departure time (at the beginning of the arc) plus the transverse time in the case if $b_l = 1$, which means the service is used from constraint (5.4.4).

$$y_a^{arv} - y_a^{dep} + \mathcal{M}_{\text{time}} \cdot (1 - b_a) \geq k_a, \qquad \forall a = (i, j, v) \in A; \tag{5.4.11}$$

Finally, constraints (5.4.12) and (5.4.13) regulate how long services dwells at terminal i. The departure time should be later than its anterior arrival time, plus transshipment time, if that exists.

$$y_{a'}^{dep} - y_a^{arv} \geq \mathcal{M}_{\text{time}} \cdot (1 - z_v), \qquad \forall i \in N, \forall a' = (j, i, v) \in \delta_i^+, \forall a = (i, j', v) \in \delta_i^-; \tag{5.4.12}$$

$$y_{a'}^{dep} - y_a^{arv} \geq h_{a,i}^{s,-} + r_{a,a',i} + h_{a,i}^{s,+} - \mathcal{M}_{\text{time}} \cdot (1 - e_{a,a',i}), \forall i \in N, \forall a' = (j, i, v) \in \delta_i^+, \forall a = (i, j', v) \in \delta_i^-; \tag{5.4.13}$$

5.4.4 *Shipment due time constraints*

Constraints (5.4.14a)-(5.4.17) are used to make sure that all the shipments arrive within the expected time window. Shipment s can arrive ahead of or after the expected due time γ^s. Since the splitting of shipment is built into constraint (5.4.1), one shipment s can arrive at the destination $j = d^s$ in the shape of embranchments via any possible arc $a = \delta_i^+$. From (5.4.3), we can obtain the distribution of the shipment flows to all these possible arcs. If $b_a^s = 1$ ((part of) the shipment is assigned to the arc a) the arrive time equals the arrive time of the arc y_a^{arv} plus the unloading time. On the other hand, if $b_a^s = 0$, then delivery time is not related to the arrival event.

Taking into account the aforementioned description, variable t_a^s is used to represent the final delivery time of containers from shipment s delivered via $a \in \delta_j^+$ at the destination terminal $j = d^s$, which is determined by constraints (5.4.14a) and (5.4.14b).

$$t_a^s \geq y_a^{\mathrm{arv}} + h_{a,i}^{s,-} - \mathcal{M}_{\mathrm{time}} \cdot (1 - b_a^s), \quad \forall s \in S, \forall a = (i,j,v) \in \delta_j^-, j = d^s; \tag{5.4.14a}$$

$$t_a^s \leq y_a^{\mathrm{arv}} + h_{a,i}^{s,-} + \mathcal{M}_{\mathrm{time}} \cdot (1 - b_a^s), \quad \forall s \in S, \forall a = (i,j,v) \in \delta_j^-, j = d^s; \tag{5.4.14b}$$

Constraint (5.4.15) makes sure that the delivery time of each split shipment t_a^s should be before γ^s, which is the latest delivery time.

$$t_a^s \leq \gamma^s, \qquad \forall s \in S, \forall a = (i,j,v) \in \delta_j^+, j = d^s; \tag{5.4.15}$$

We use $[w_a^{s,-}, w_a^{s,+}]$ to denote the corresponding earliness or latency. Constraints (5.4.16) and (5.4.17) exacts the deviation of t_a^s (the shipment arrival time) from β^s (the expected due time). Note that for shipment t_a^S only one of earliness or delay can occur.

$$w_a^{s,-} \geq (\beta^s - t_a^s) - \mathcal{M}_{\mathrm{time}} \cdot (1 - b_a^s), \quad \forall s \in S, \forall a = (i,j,v) \in \delta_j^-, j = d^s, \tag{5.4.16}$$

$$w_a^{s,+} \geq (t_a^s - \beta^s) - \mathcal{M}_{\mathrm{time}} \cdot (1 - b_a^s), \quad \forall s \in S, \forall a = (i,j,v) \in \delta_j^-, j = d^s. \tag{5.4.17}$$

5.4.5 *Objective function*

The optimization model is formulated as a mixed-integer programming problem (MIP). The objective is to minimize the total operating cost based on on-time delivery expectations.

We briefly introduce related parameters for the operating costs. f_v is the fixed cost of using the service v, η_v is the cancellation fee if the pr-planned LCS are not used, c_a is the variable cost of routing a unit of shipment on arc a. In addition we define g_i^s as the unit transshipment cost for shipment s at terminal i. The early arrival penalty $\rho^{s,-}$ is generated from storage cost at the destination, and the late penalty $\rho^{s,+}$ is made from the contract with the shipper, depending on the value and urgency of the goods.

The total operating costs are expressed as $\mathcal{C}$ in (5.4.18), which includes: the fixed costs of used services $\mathcal{C}^{\mathrm{f}}$, the flow transportation costs $\mathcal{C}^{\mathrm{c}}$, the transshipment costs $\mathcal{C}^{\mathrm{t}}$, and the penalty costs due to early delivery $\mathcal{C}^{\mathrm{e}}$ or late delivery $\mathcal{C}^{\mathrm{l}}$ at the destinations.

The fixed costs in (5.4.19) include the total fixed costs of the service being used from all transportation modes, and the cancellation fees for the pre-specified services. Constraint (5.4.20) describes the total variable transportation costs, which increase linearly with the transported volume x_a^s. The transshipment costs are calculated by constraint (5.4.21). The penalty of off-target time delivery at the destination $i = d^s$ consists of early arrival penalty (5.4.22) and late arrival penalty (5.4.23). The early and late arrival penalty synchronize the shipments, which occurs when several shipments are waiting to be transported at the same terminal while service capacity is insufficient. Thus, the loading or the holding of the shipments is balanced according to the delivery time.

$$\text{Min} \quad \mathcal{C}^{\mathrm{f}} + \mathcal{C}^{\mathrm{c}} + \mathcal{C}^{\mathrm{t}} + \mathcal{C}^{\mathrm{e}} + \mathcal{C}^{\mathrm{l}}, \tag{5.4.18}$$

where,

$$\mathcal{C}^{\mathrm{f}} = \sum_{v \in V} (z_v \cdot f_v) + \sum_{v \in V} (1 - z_v) \cdot \psi_v, \tag{5.4.19}$$

$$\mathcal{C}^{\mathrm{c}} = \sum_{s \in S} \sum_{a \in A} (x_a^s \cdot c_a), \tag{5.4.20}$$

$$\mathcal{C}^{\mathrm{t}} = \sum_{s \in S} \sum_{i \in N \setminus \{o^s, d^s\}} \sum_{a, a' \in A} \left(n_{a,a',i}^s \cdot g_{a,a',i}^s \right), \tag{5.4.21}$$

$$\mathcal{C}^{\mathrm{e}} = \sum_{s \in S} \sum_{a \in \delta_j^-, j = d^s} \left(x_a^s \cdot \rho^{s,-} \cdot w_a^{s,-} \right), \tag{5.4.22}$$

$$\mathcal{C}^{\mathrm{l}} = \sum_{s \in S} \sum_{a \in \delta_j^-, j = d^s} \left(x_a^s \cdot \rho^{s,+} \cdot w_a^{s,+} \right). \tag{5.4.23}$$

5.4.6 *Linearization of objective function*

The mathematical model is formulated in linear constraints (5.4.1)-(5.4.17), with a non-linear term objective function (5.4.22) and (5.4.23). To linearize the objective function, we assume that storage costs or delay penalty are charged per TEU and per time interval (for example per 10 minutes interval). This assumption is consistent with current practice in synchromodal transportation.

Take (5.4.22) as demonstration, the maximum possible value of auxiliary variable $w_a^{s,-}$ can be estimated in advance. It can be calculated as the due time minus the earliest arrival time of all services at the destination terminal. Let $\dot{\pi}^{s,arv}$ denote the earliest possible arrival time, $\dot{\pi}^{s,arv} := min(\pi_a^{arv})$, where $\forall a \in \delta_j, j = d^s$). The maximum earliness duration can be equally divided into $\mathcal{K}$ time slots, each of which equals $(\beta^s - \dot{\pi}^{s,arv})/\mathcal{K}$, where β^s is the due time of shipment s. Another auxiliary variable $\theta_a^{s,\mathfrak{k}}$ will be used, if $\theta_a^{s,\mathfrak{k}} = 1$, denoting the earliness $w_a^{s,-}$ eventually is true in the $\mathfrak{k}$-th time slot, otherwise $\theta_a^{s,\mathfrak{k}} = 0$. Thus constraint (5.4.22) can be expressed as following:

$$\mathcal{C}^{\mathrm{e}} = \sum_{s \in S} \sum_{a \in \delta_j, j = d^s} \sum_{\mathfrak{k}=1}^{\mathcal{K}} \left(x_a^s \cdot (\rho^{s,-}/\mathcal{K}) \cdot \theta_a^{s,\mathfrak{k}} \right) \tag{5.4.24}$$

$$r_a^s + \mathcal{M}_{time} \cdot (1 - \theta_a^{s,\mathfrak{k}}) \geq (\beta^s - \dot{\pi}^{s,arv})/\mathcal{K}, \quad \forall s \in S, \forall a = (i,j,v) \in \delta_j, j = d^s, \forall \mathfrak{k} \in 1, ..., \mathcal{K} \tag{5.4.25}$$

The same process is applied to deal with (5.4.23), resulting in (5.4.26) and (5.4.27):

$$\mathcal{C}^l = \sum_{s \in S} \sum_{a \in \delta_j, j = d^s} \sum_{\eta=1}^{\mathcal{H}} (x_a^s \cdot (\rho^{s,+}/\mathcal{H}) \cdot \lambda_a^{s,\eta}) \tag{5.4.26}$$

$$w_a^s + \mathcal{M}_{time} \cdot (1 - \lambda_a^{s,\eta}) \geq (\gamma^s - \beta^s)/\mathcal{H} \cdot \eta, \quad \forall s \in S, \forall a = (i,j,v) \in \delta_j, j = d^s, \forall \eta \in 1, ..., \mathcal{H} \tag{5.4.27}$$

As such, the model is reformulated with objective (5.4.18) (deposed into (5.4.19)-(5.4.21), (5.4.24)-(5.4.27)), and is constrained by (5.4.1)-(5.4.16). The model will be referred to as the STP (synchromodality transportation planning) model for short.

5.5 RE-PLANNING ALGORITHM

The general steps of the replanning procedure are included in Algorithm 1:

Algorithm 1 Re-planning Algorithm

1: **Input**: Set of unexpected events, current status of the network.
2: **Output**: Re-planning solution, the integrated shipment flow rerouting, transshipment and service rescheduling.
3: **procedure** PROCEDURE RE-PLAN WHEN (AN) UNEXPECTED EVENT(S) OCCUR(S)
4:     **Step 1**: Initialize the network.
5:     **Step 2**: Apply the pre-processing:
6:     **Step 3**: Solve the STP model.

5.5.1 *Initialize the network*

The initialization includes inputting the information of the terminals, services, arcs and shipments. The status of the network should also be recorded, including the on-going transportation operations. The STP model can be used to re-plan the operations directly after an unforeseen situation occurs, which may include any fluctuations or deviation from the original plan.

5.5.2 *Applying the Pre-processing*

In the following section, we describe the pre-processing steps designed to improve the computational performance of the algorithm. The pre-processing including (1) tightening the re-planning time window, (2) fixing (part of) variables, and (3) reducing of the number of transshipment variables.

Tightening the re-planning time window

The replanning starts immediately after the unexpected event occurs, which is denoted as $\mathcal{T}^{start}$. The replanning ends at $\mathcal{T}^{end} = \max\{\gamma^s | \quad \forall s \in S\}$. Thus the replanning window is set as $[\mathcal{T}^{start}, \mathcal{T}^{end}]$.

Fixing part of variables

All the decisions those have been made before the unexpected event occurs are not affected. This means that any variables corresponding to those decisions are fixed. Meanwhile, the arcs in which the services are broken down will be eliminated from the network.

Reducing of the number of transshipment variables

The number of transshipment variables is the largest number among all the variables, because of the uncertainties in three dimensions: the feeder arcs, the connecting arcs, and the rerouting of shipment flow. Consider the possible transshipment connection from arc $a \in \delta_i^-$ to arc $a' \in \delta_i^+$. We can extract from this arc the earliest possible arrival time $\tilde{t}_a^{arv}$ and the latest possible departure time $\tilde{t}_{a'}^{dep}$. Then $n_{a,a',i}^s = 0$, $e_{a,a',i} = 0$ in the case of $\tilde{t}_a^{arv} \geq \tilde{t}_{a'}^{dep}$. Thus, the problem-solving efficiency can be improved significantly if the number of the transshipment variables can be reduced.

5.5.3 *Solving the STP model*

After initializing and pre-processing, the STP model can be solved using the CPLEX solver. A minimum cost replanning proposal will be generated, which is an integrated shipment flow replanning, terminal transshipment, and the service rescheduling. Here need to consider the value of big-$\mathcal{M}$ to be given to the solver.

Customized value of $\mathcal{M}$

The big-$\mathcal{M}$ is used in the constraints. The evaluation of $\mathcal{M}$ can be divided into two categories according to the applied situation. $\mathcal{M}_{load}$ is used in (5.4.3), (5.4.4), (5.4.8). We evaluate $\mathcal{M}_{load} := \max\{u_v) \mid \forall v \in V\}$. The second category $\mathcal{M}_{time}$ is used in (5.4.10)-(5.4.14b), (5.4.16), (5.4.17), (5.4.25) and (5.4.27) to regulate the services operation time. We evaluate $\mathcal{M}_{time} := \mathcal{T}^{end}$.

5.6 COMPUTATIONAL EXPERIMENTS

In this section, we demonstrate the applicability of the model in a real-word transportation case. The optimization model is coded and solved in CPLEX 12.8.0 on a computer with an Intel Core i5-4690 3.2 GHz processor and 8GB of RAM. All the instances are solved within the time limit within seconds.

The computational experiments are performed in two parts. Firstly, we present a description of the Rotterdam hinterland network and its daily transportation tasks. A base case is provided as an original plan. Secondly, we test the STP model on different initial perturbations. These include includes exogenous perturbations from shipments (late release and volume fluctuations), endogenous disturbances from services (delays), and some mixed events. The computational experiments show that the STP model has the flexibility to deal with perturbations via the rerouting of shipment flows and the service rescheduling.

5.6.1 *Network description*

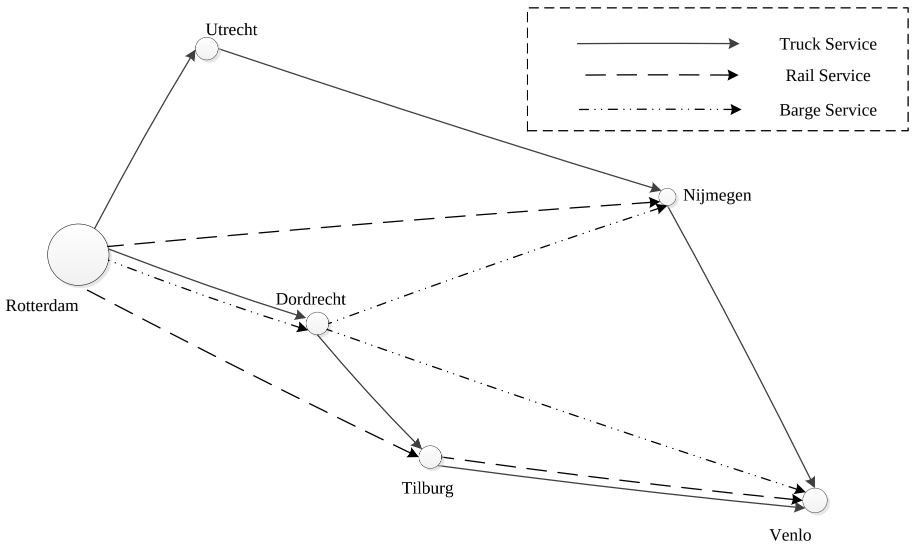

Figure 5.6.1: The Rotterdam hinterland transportation network.

To test the STP model, we construct our case based on the network presented by Guo et al., 2018 and Guo et al., 2020, as shown in Fig. 5.6.1. There are six terminals in the studied network, i.e., Port of Rotterdam (PoR in short), Utrecht, Nijmegen, Dordrecht, Tilburg and Venlo. All shipments are transported from PoR to the other five terminals. In this network, three different transportation modes (i.e., barge, rail and truck services) are used in combination. For instance, the two arcs from PoR to Dordrecht represent a barge service and truck services.

A summary of the parameters used in our case study is presented in Table 5.6.1. The fixed costs (€/ service) of the service are based on the typical monetary costs of services per mode. The transportation time is calculated as the sum of the loading time, the net travelling time and the unloading time. Besides, we set the loading and unloading time of LCS and truck services to one hour and a half hours respectively. We consider one hour for transshipment time (Li, Negenborn, and De Schutter, 2017). Transshipment cost is set as 23.89 (€/TEU) (Li, Negenborn, and De Schutter, 2017). Note that we also allow transshipment between services of the same mode (e.g. transshipment from one barge service to another is possible).

During the re-planning, we assume the cancellation of LCS will generate half of the fixed cost, i.e., $\psi_l := 0.5 \cdot f_v$ in constraint (5.4.19). The values of the buffer time in constraints (5.4.9a) and (5.4.9b) are set to $\phi_a^{dep} := 3$ (hours) and $k_{max} \cdot \phi_a^{dep} := 1$

(hour). The customized evaluations of ϕ_a^{dep} should be decided by the local LSP according to the specific situations, which is briefly discussed in Section 5.7.1.

Table 5.6.1: Data set of the services in the Rotterdam hinterland

Arc (Dep S, Arv S, Service)	Mode	Capacity (TEU)	Travel Time (hour)	Fixed Cost (€/service)	Variable Cost* (€/TEU)
PoR, Dordrecht, v0001	Barge	120	2	60	2.45
PoR, Tilburg, v0002	Rail	100	2	30	30.16
PoR, Nijmegen, v0003	Rail	60	2	30	30.16
Dordrecht, Nijmegen,v0004	Barge	120	5	60	4.29
Dordrecht, Venlo, v0005	Barge	120	9	60	6.73
Tilburg, Venlo, v0006	Rail	100	1	30	22.16
PoR, Utrecht, v0007-v0506	Truck	1	1	15	61.96
PoR, Dordrecht, v0507-v1006	Truck	1	0.5	15	30.98
Utrecht, Nijmegen,v1007-v1506	Truck	1	1	15	61.96
Dordrecht, Tilburg,v1507-v2006	Truck	1	1	15	30.98
Tilburg, Venlo, v2007-v2506	Truck	1	1	15	30.98
Nijmegen, Venlo, v2507-v3006	Truck	1	1	15	30.98

Variable Cost * is from Guo et al., 2018

5.6.2 *Base case*

We put five shipments into the network, as described in Table 5.6.2. Each shipment s is characterized by its origin terminal (o^s), destination terminal (d^s), container volume n^s, release time (α^s), due time (β^s), latest delivery time (γ^s), and penalties for early ($\rho^{s,-}$) or late delivery ($\rho^{s,+}$).

Table 5.6.2: Transportation shipments from port of Rotterdam to 5 hinterland terminals

s	o^s	d^s	n^s **(TEU)**	α^s **(h)**	β^s **(h)**	γ^s **(h)**	$\rho^{s,-}$ **(€/TEU/h)**	$\rho^{s,+}$ **(€/TEU/h)**
S1	PoR	Utrecht	50	7	18	24	0.5	1.5
S2	PoR	Dordrecht	50	7	18	24	0.5	1.5
S3	PoR	Tilburg	50	7	18	24	0.5	1.5
S4	PoR	Nijmegen	100	7	18	24	0.5	1.5
S5	PoR	Venlo	100	7	18	24	0.5	1.5

Table 5.6.3 shows the pre-planned solution if there are no disturbances. We use this solution as a base to verify how the STP model deals with the disturbances.

For the base case, the total cost $\mathcal{C}$ (€) is 15961, in which C^{rmf} = 1320, C^c = 10476, C^t = 3345, C^e=205, and C^l = 615. These costs are described in the objective function (5.4.18)-(5.4.23). For the solution, 10.42% of the shipment is transported by truck services. While, the shares of rail and barge services are 38.64% and 50.94%, respectively. The modal split is calculated as the ratio of a freight mileage (TEU $\times$ km) of a certain mode against the overall used freight mileage. Meanwhile, the utilization rates of barge and rail services are 49.27% and 85.67%. Note that the utilization rate (%) is calculated as the ratio of (i) the used freight mileage (TEU $\times$ km) and (ii) the product of the full capacity of the service and the pre-planned running distance.

Table 5.6.3: The base case: the original shipment assignment and service schedules

Arc	Mode	S1	S2	S3	S4	S5	Start Time	End Time
PoR, Dordrecht, v0001	Barge		30		40	50	7	11
PoR, Tilburg, v0002	Rail			50		50	10	14
PoR, Nijmegen, v0003	Rail				60		14	18
Dordrecht, Nijmegen, v0004	Barge				40		15	22
Dordrecht, Venlo, v0005	Barge					50	12	23
Tilburg,Venlo, v0006	Rail					50	15	18
PoR, Utrecht,v0007-v0506	Truck	50					16	18
PoR, Dordrecht,v0507-v0526	Truck		20				16.5	18

Note: S1-S5: Shipment 1-Shipment 5. DepS/ArvS: Departure/Arrival Station

Start T: time to start loading, the service departures afterwards, which takes $h^s_{a,i}$

End T: time when unloading finished, the service arrives before, with interval $h^s_{a,i}$

5.6.3 *Key performance indicators (KPI)*

We use 11 Key Performance Indicators (KPIs) to test the performance of the STP model. The KPIs are categorized into three groups indicating (1) the flexibility of the solution, (2) the LSP performance or (3) the service level. There are two KPIs in the first group: the amount of the re-routed shipment flows (TEU) and the rescheduled time of all LCS (vehicles $\times$ delayed hours). The second group of KPIs includes the variation rate of fixed service cost $\pm\mathcal{C}^{f}$ (where $\pm$ denotes the variation of the costs), the variation rate of variable cost $\pm\mathcal{C}^{c}$, the variation rate of transshipment cost $\pm\mathcal{C}^{t}$, the variation of the modal split rate (%) of LCS, the variation of utilization rate (%) of LCS, the number of cancelled LCS, and the number of additional truck services. The final group consists of the early delivery (TEU $\times$ h) and late delivery (TEU $\times$ h) at the destinations.

5.6.4 *Delays, STP solutions and corresponding KPIs*

In this section, we discuss the results of the STP replanning in the cases where occur the delay of shipment release or LCS. The results are listed in Tables 5.6.4 and 5.6.5, respectively. In these tables, the first two columns represent the incident case and the third column demonstrates the associated delay. For instance, the first row indicates the replanning solutions if shipment $s2$ encounters a delay of 0.5 hour. The STP model tailors different replanning solutions according to the duration of the delay. If shipments are slightly delayed, no adjustments are needed. That is because of the time gap between the departure services and the expected release time of the shipments. However, when there are longer delays, the STP model initially resorts to service rescheduling. The LCS vehicles are held longer than the original timetable to process the intake of the delayed shipments. When the delays get even longer, the STP model combines the rerouting of shipments and the service rescheduling.

Table 5.6.4: Late release of shipments and corresponding KPIs of the STP solutions

Shipment	Delays	Severity	Shipment re-routing (TEU)	LCS Rescheduling (vehicle × h)	$\pm\mathcal{C}^f$ (%)	$\pm\mathcal{C}^c$ (%)	$\pm\mathcal{C}^t$ (%)	Cancelled LCS	Extra Trucks	± LCS modal split (%)	± LCS utilization (%)	± Early Delivery (TEU × h)	± Delayed delivery (TEU × h)
	S2	0.5h	0	0	0	0	0	0	0	0	0	0	0
		1h	0	1	0	0	0	0	0	0	0	-10	50
		1.5h	60	0	34.09	-4.31	0	0	30	-2.1	-4.7	-210	150
	S3	$\leq$ 3h	0	0	0	0	0	0	0	0	0	0	0
		3.5h	0	1	0	0	0	0	0	0	0	-25	25
		4.0h	0	2	0	0	0	0	0	0	0	-50	50
	S4	0.5h	0	1	0	0	0	0	0	0	0	-15	25
		1h	0	2	0	0	0	0	80	-11.48	-8.2	-30	50
		1.5h	80	0	22.73	-2.88	0	0	20	0	0	140	100
	S5	0.5h	0	1	0	0	0	0	0	0	0	-15	25
		1h	0	2	0	0	0	0	0	0	0	-30	50
		1.5h	80	1.5	34.09	8.17	0	0	30	-4.8	-2.5	-10	0

Table 5.6.5: Latency of the LCS and corresponding KPIs of the STP solutions

	Delays	Severity	Shipment re-routing (TEU)	LCS Rescheduling (vehicle × h)	$\pm\mathcal{C}^f$ (%)	$\pm\mathcal{C}^c$ (%)	$\pm\mathcal{C}^t$ (%)	Cancelled LCS	Extra Trucks	± LCS modal split (%)	± LCS utilization (%)	± Early Delivery (TEU × h)	± Delayed delivery (TEU × h)
Services	**v0001**	0.5h	0	0	0	0	0.5	0	0	0	0	-15	25
		1h	0	1	0	0	0	0	0	0	0	-30	50
		1.5h	60	2	34.09	8.17	0	0	30	-4.8	-2.5	-10	0
	v0002	0.5h	70	0.5	0	0	0	0	0	0	0	-25	25
		1h	0	1	0	0	0	0	0	0	0	-50	50
		1.5h	100	1	55.68	-7.19	0	1	50	-7.7	-3.4	-335	250
	v0003	0.5h	0	0	0	0	0	0	0	0	0	0	30
		1h	0	0	0	0	0	0	0	0	0	0	60
		1.5h	0	0	0	-2.88	0	0	20	0	0	140	100
	v0004	0.5h	0	0	0	0	0	0	0	0	0	0	90
		1h	0	0	0	0	0	0	0	0	0	0	40
		1.5h	0	0	0	0	0	0	0	0	0	0	60
	v0005	0.5h	0	0.5	0	0	0	0	0	0	0	-15	25
		1h	0	1	0	0	0	0	0	0	0	-30	50
		1.5h	80	3	154.55	113.62	135.70	1	50	-7.3	-5.1	320	560
	v0006	0.5h	0	0.5	0	0	0	0	0	0	0	-25	25
		1h	0	1	0	0	0	0	0	0	0	-50	50
		1.5h	0	1	0	0	0	0	0	0	0	-50	75

The STP model can also manage service delays, as shown in Table 5.6.5. Rerouting of shipment flows is used as the primal solution. In addition, the rescheduling of other LCS is also used to solve severe delays.

Sometimes the delay propagates from one service to another one from the transshipment connection. It needs to be pointed out that there are two cases in which a barge or a rail service is cancelled. In the first case where rail service $v0002$ is delayed for 1.5 hour, while the STP does not assign any shipment flows to rail service $v0006$. This is because of the buffer time $\phi_a^{dep} := 1$ hour for the arc $Dordrecht, Venlo, v0006$ in constraint (5.4.9b), which precludes transshipment from service $v0002$ to service $v0006$.

5.6.5 *Shipment fluctuations, STP solutions and corresponding KPIs*

In this section, we test the STP model with volume fluctuations of shipments. Tables 5.6.6-5.6.7 presents the input fluctuations and corresponding KPIs. The first column illustrates to from which shipment the volume fluctuation rises from, and the second column indicates the severity of the fluctuation.

Table 5.6.6: Volume Fluctuations and corresponding KPIs

	Shipment	Severity	Flow Re-routing (TEU)	LCS Re-scheduling (vehicle $\times$h)	$\pm\mathcal{C}^f$ (%)	$\pm\mathcal{C}^c$ (%)	$\pm\mathcal{C}^t$ (%)	Cancelled LCS	Extra Trucks	$\pm$ LCS modal split (%)	$\pm$ LCS utilization (%)	$\pm$ Early Delivery (TEU $\times$ h)	$\pm$ Delayed delivery (TEU $\times$ h)
Volume increase	**S2**	+10%	0	0	5.68	1.48	0	0	5	-0.7	0	0	0
		+20%	0	0	11.36	2.96	0	0	10	-1.4	0	0	0
		+30%	0	0	17.05	4.44	0	0	15	-2.1	0	0	0
	S3	+10%	10	0	5.68	0.72	0	0	5	-0.6	-0.6	-42.5	47.5
		+20%	20	0	11.36	1.44	0	0	10	-1.2	-1.2	-30	50
		+30%	30	0	17.05	2.16	0	0	15	-1.8	-1.8	45	75
	S4	+10%	20	0	11.36	3.37	7.14	0	10	-1.2	1.4	-70	40
		+20%	40	0	22.73	6.73	14.28	0	20	-2.3	2.8	-140	80
		+30%	60	0	34.09	10.10	21.43	0	30	-3.3	4.3	-210	120
	S5	+10%	20	0	11.36	3.60	7.14	0	10	-1.0	2.4	-70	50
		+20%	40	1	22.73	7.20	14.28	0	20	-2.0	4.9	-165	125
		+30%	60	0	24.09	10.80	21.43	0	30	-2.8	7.4	-210	150

Table 5.6.7: Volume Fluctuations and corresponding KPIs

	Shipment	**Severity**	**Flow Re-routing (TEU)**	**LCS Re-scheduling (vehicle $\times$h)**	$\pm\mathcal{C}^f$ **(%)**	$\pm\mathcal{C}^c$ **(%)**	$\pm\mathcal{C}^t$ **(%)**	**Cancelled LCS**	**Extra Trucks**	**$\pm$ LCS modal split (%)**	**$\pm$ LCS utilization (%)**	**$\pm$ Early Delivery (TEU $\times$ h)**	**$\pm$ Delayed delivery (TEU $\times$ h)**
Volume decrease	**S2**	-10%	0	0	-5.68	-1.48	0	0	-5	0.7	0	0	0
		-20%	0	0	-11.36	-2.96	0	0	-10	1.5	0	0	0
		-30%	0	0	-17.05	-4.44	0	0	-15	2.3	0	0	0
	S3	-10%	10	0	-5.68	-0.72	0	0	-5	0.6	-0.6	15	-25
		-20%	20	0	-11.36	-1.44	0	0	-10	1.3	-1.2	30	-50
		-30%	30	0	-17.05	-2.16	0	0	-15	2.0	-1.9	45	-75
	S4	-10%	10	0	-11.36	-3.37	-7.14	0	-10	1.3	-1.4	70	-40
		-20%	20	0	-22.73	-6.73	-14.28	0	-20	2.7	-2.9	140	-80
		-30%	30	0	-22.73	-11.54	-21.43	0	-20	2.4	-5	140	-70
	S5	-10%	10	0	-11.36	-3.60	-7.14	0	-10	1.1	-2.5	70	-50
		-20%	20	1	-22.73	-7.20	-14. 28	0	-20	2.4	-4.9	115	-75
		-30%	20	0	-22.73	-12.24	-21.43	0	-20	2.0	-8.1	140	-100

The solutions mainly use the re-routing of the shipment flows. That is because the perturbations are caused by volume fluctuations. In the rerouting module, the increased volume is mostly handled by the LCS. This is the result of the lower fixed cost f_v and variable cost c_l of LCS compare to those of truck services. Meanwhile, the re-scheduling of LCS is updated when necessary to extend the possibility of taking in more flows. However, additional truck services are still used when there is no available capacity from the LCS.

When the volume of shipments decreases, the STP model primarily cancels the usage of trucks. This also leads to a higher modal split of LCS in these particular cases.

5.6.6 *Mixed events, STP solutions, and corresponding KPIs*

To further examine the solution strategy and its performance in greater detail, we introduce simultaneous events to the network and the STP model.

Table 5.6.8: Mixed events, solutions and the corresponding KPIs

Perturbation 1	Perturbation 2	Flow Re-routing (TEU)	LCS Re-scheduling (vehicle $\times$h)	$\pm \mathcal{C}^f$ (%)	$\pm \mathcal{C}^c$ (%)	$\pm \mathcal{C}^t$ (%)	Cancelled LCS	Extra Trucks	$\pm$ LCS modal split (%)	$\pm$ LCS utilization (%)	$\pm$ Early Delivery (TEU $\times$ h)	$\pm$ Delayed delivery (TEU $\times$ h)
S2+2h	**S3+2h**	60	0	34.09	-4.32	0	0	30	-4.7	-2.1	210	150
S2+2h	**S5+2h**	70	3	34.09	8.17	0	0	30	-4.8	-2.5	-10	0
S3+2h	**S4+2h**	150	3	54.54	13.62	35.71	1	50	-7.4	-5.2	-90	150
S4+2h	**S5+2h**	70	3	34.09	8.17	0	0	30	-4.8	-2.5	-10	0
v0001+1h	**v0002+1h**	0	2	0	0	0	0	0	0	0	-80	100
v0001+1h	**v0004+1h**	100	1	22.73	-2.88	0	0	20	-3.14	-1.37	90	210
v0001+1h	**v0005+1h**	70	2	34.09	8.17	0	0	30	-4.76	-2.47	-10	50
v0002+1h	**v0006+1h**	70	3	34.09	8.17	0	0	30	-4.76	-2.47	-60	50

To summarize, we obtain the following attributes of the STP models from the cases in Table 5.6.4 - Table 5.6.8. The STP model tailors different re-planning solutions to the specific encountered disturbance each time. Adjustments in the time dimension (service rescheduling) are made to deal with the delays; adjustments in the space dimension (rerouting of shipment flows) are made to deal with volume fluctuations. If the disturbances are comparatively severe, the STP resorts to a combination of the rescheduling and rerouting.

The operating costs increase when dealing with the late release of shipments, service delays and increased shipment volume. The costs can increase sharply in the case of cancelling a LCS. The cancellation is due to the delay propagate from one service to another. The resulting delay substantially violates the planned operating time. Finally, some shipments are delivered later than the expected due time, which is unavoidable in managing the disturbances.

5.7 Computational Analysis

Based on the computational experiments, we also analyze it looking at three other aspects. Firstly, we check the critical path of the example network to see if it is possible to improve the service rescheduling flexibility any further. Secondly, the proposed model is compared to other rigid re-planning methods to access the value of shipment split and flexible service scheduling. Finally, the model is tested based on more theoretical instances to explore the difficulty of the model resolution.

5.7.1 *Critical path and the re-scheduling flexibility*

Refinements can be made to tailor the value of ϕ_a^{dep} in constraint (5.4.9a). This parameter regulates the limit of extra waiting time before a departure event. If ϕ_a^{dep} can be individually tailored, the rescheduling flexibility of model can be reinforced. To illustrate this more clearly, we take the delivery of shipment $S5$ as an example.

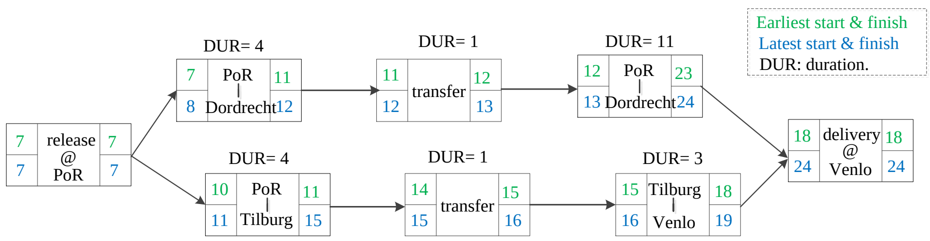

Figure 5.7.1: The Critical Path for shipment 5 to use Large Capacity Services

We visualize the paths of $S5$ in which only barge or rail services are used (see Fig.5.7.1). It is constructed based on the critical path method. The upper route $PoR, Dordrecht, v0001$ - $Dordrecht, Venlo, v0005$ (route 1 in short) has an elasticity of one hour. Here, it is determined by the maximum due time (γ^{S5}). On the other hand, the lower route $PoR, Tilburg, v0002$ - $Tilburg, Venlo, v0006$ (route 2 in short) also has one hour elasticity. Here it is determined by ϕ_a^{dep} for the arc $a = PoR, Venlo, v0006$, which limits the possibility of additional waiting and transshipment. Whereas as shown in Table 5.6.5 the connecting train $v0006$ is cancelled when service $v0002$ is 1.5 hours delayed.

The are three reasons for extending this buffer time. (1) The constraint of the latest finish time (γ^{S5}) can still be satisfied; (2) We assume rail service $v0006$ or the locomotive has no other immediate tasks. (3) The resulting transshipment costs and delay penalty can be compensated by switching from trucks into rail services.

The rescheduling of railways in the multi-modal transportation

The rescheduling of the rail services in practice is quite complex in the real application. The railway is an important component of the multi-modal network. However, most intermodal or synchromodal literature avoid talking about the exact operation time of rail services. For example, Rivera and Mes, 2017 model it as a day to day service in synchromodal transportation. This kind of assumption makes the work cannot be used for flexible replanning in real-time.

One of the few studies to deal with exact freight trains departure time in the multi-modal freight network was Corman, Viti, and Negenborn, 2017. Another study include the terminal freight train operation was Hu et al., 2018. Both studies regulate the departure time by using a pre-planned time plus a buffer time. Here we use the same logic as Corman, Viti, and Negenborn, 2017 and Hu et al., 2018. We additionally refined it as a step-wise buffer time according to constraint (9b).

The departure times of rail services are more limited for safety reasons. If the freight rail services run on dedicated rail tracks (like the Betuweroute line from Rotterdam to Germany), the departure times should meet a headway separation constraint. If the rail cargoes share the same railway track with passenger trains, usually priority is given to passenger trains, Corman et al., 2011. Generally speaking, the rescheduling solution of railways is expected in minutes Crainic, Hewitt, and Rei, 2014.

With the developed information & communication and technology (ICT) and the synchromodal platform, the railway operators can keep up with the expected flexibility. Secondly if the rail service is held within a buffer time, it will take accumulated shipment *en route*; otherwise the train will be cancelled and the shipment flow goes to truck services. From the optimization aspects, this study illustrates what the railway company can do (re-scheduling) in the multimodal transportation, and how the railway company can benefit (more shipment flows and improved modal split).

5.7.2 *The value of shipment split and flexible scheduling*

In this section, we compare the solution obtained by the STP model against two reference models. In the first case, we consider a case splitting of the shipment is not conducted (N-STP). For the second case, we consider a case where the network has a rigid schedule (R-STP).

To keep the comparison detailed and at the same time keep this section concise, we use one particular case in which LSP is informed in advance that the release time shipments $S4$ and $S5$ will be delayed by two hours. In other words, $\alpha^{S4} = 7 \rightarrow 9$, $\alpha^{S5} = 7 \rightarrow 9$. The reference N-STP model and R-STP model will be discussed in the following subsections. Table 5.7.1 summaries the KPIs obtained by solving the three models. For the detailed solutions, see Tables 5.A.1-5.A.3 in Appendix 5.A.

Table 5.7.1: Results of the solutions by STP model, N-STP model, and R-STP model

KPI	STP	N-STP	R-STP	Improvement of STP over N-STP	Improvement of STP over R-STP
Total Service Cost (€)	17261.8	20043	19078	16.11 %	10.52%
Fixed Cost (€)	1770	3240	2370	83.05%	33.89%
Variable Cost(€)	11332.2	10600	12473.4	-6.46%	10.07%
Transshipment Cost (€)	3344.6	4778	3344.6	42.85%	0
Earliness Penalty (€)	200	75	275	-62.5%	37.5%
Delay Penalty (€)	615	1350	615	119.51%	0
Modal split Barge (%)	34.13	57.98	29.08	-23.85	5.05
Modal split Rail (%)	49.36	11.76	48.29	37.6	1.07
Modal split Truck (%)	16.51	30.26	22.63	-13.75	-6.12
Barge services utilization (%)	44.93	83.33	39.13	-38.4	5.8
Rail services utilization (%)	85.67	22.29	85.67	63.38	0
LCS utilization (%)	62.50	57.00	59.20	5.5	3.3

The STP resolutions save 16.11% and 10.52% of the operational costs than the N-STP resolutions and the R-STP resolutions. These saving costs can be tracked to the fixed and variable costs of the services, since the STP model uses fewer truck services. Meanwhile, the STP resolutions are less punctual compared to two reference models, which means that there are some additional early and late delivery penalties. That is due to the less usage of truck services. Furthermore, the results indicate that by splitting the shipment, the LSP can increase the share of barge and rail services. Moreover, the utilization rates of the available barge and rail services from the solution of the STP model are higher than those of the two reference models. It is due to time & space-related flexibility that the STP model shows a superior performance.

We have to mention that for each replanning, the solutions and the gap between the two models depending on the network, the configuration of services and the events.

The reference N-STP model

The reference N-STP model does not allow for shipments to be split. The entire volume n^s can only be assigned to services from the same mode in each case. Each shipment is operated as one indivisible unit. One binary variable χ_a^s is used to replace the decision variables x_a^s in the STP model. If $\chi_a^s = 1$ then shipment s is assigned to arc a with the entire volume n^s. For a detailed description of the N-STP model, we refer to Appendix 5.B. Note that, in the N-STP model, there the flexible schedules of barges and rail services are maintained for all the services.

The split of shipments gives more freedom to the flow assignment, and further promotes the loading rate of LCS via vehicle sharing. We note that this is analogous to the feature "message fragmentation and reassembly" in the Multipurpose Internet Mail Extension (Media Types), where a message is first divided into several parts and then transparently reassembled back into the original message at the recipient's end. This allows an e-mail client to improve performance when using a slow connection (because the parts are sent in parallel) as well as to circumvent size restrictions. Similarly, splitting of shipments improves the use of barge and rail services.

The reference R-STP model

The R-STP model operates the LCS according to the original schedule when dealing with perturbations. The barge and rail services depart at pre-scheduled times, while truck services maintain a flexible departure time. Thus, the flexible departure constraint (5.4.9a) of STP model is not part of the R-STP model. The reference R-STP model can be written as the object (5.4.18), constraints (5.4.1)-(5.4.10), and constraints(5.4.11)-(5.4.25). Note that, in this R-STP model, it is possible to split shipments. The corresponding solutions are shown in Table 5.A.3.

We depict the time distance graphs of STP solutions and R-STP solutions in Fig. 5.7.2 and Fig. 5.7.3. Reader can clearly see the differences and benefits of synchronizing the service flows (barge service $v0001$) with the shipment flows (delayed shipment $S4$).

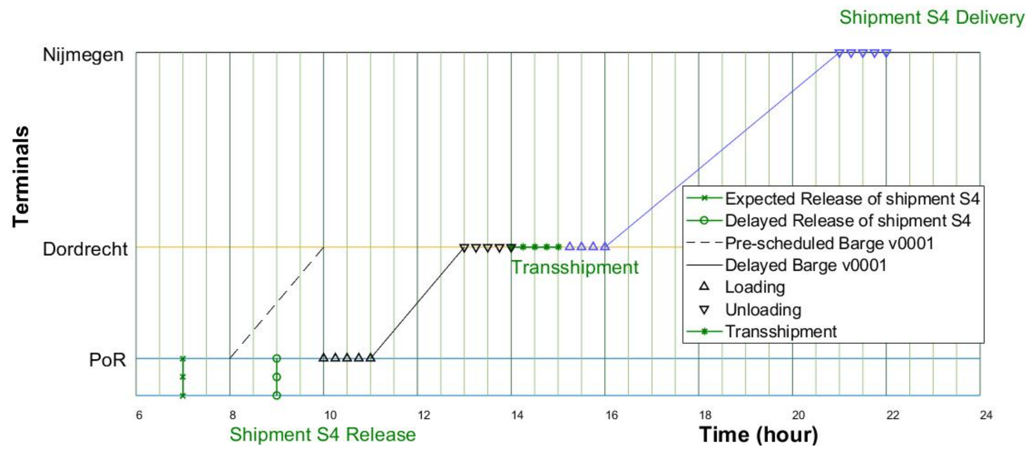

Figure 5.7.2: The time distance graph of transportation of delayed shipment $S4$ by STP model

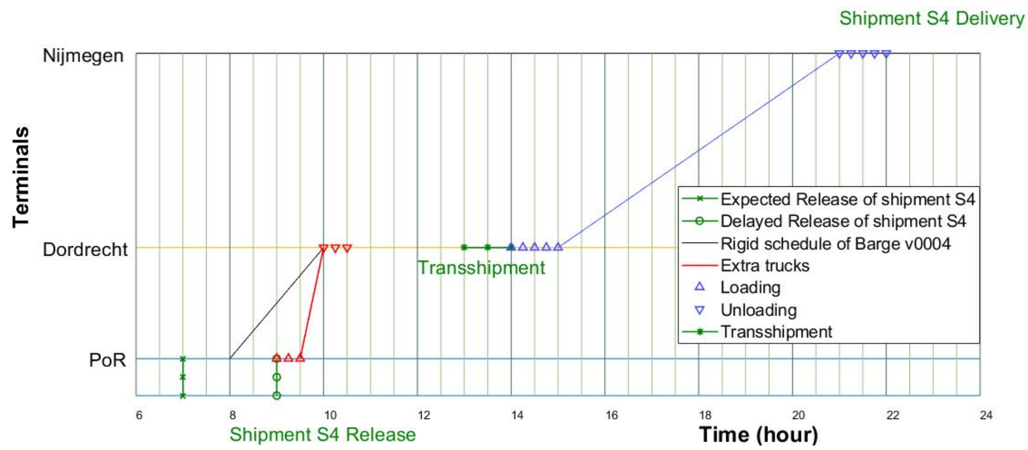

Figure 5.7.3: The time distance graph of transportation of delayed Shipment $S4$ by R-STP model

5.7.3 *More computational tests on theoretical instances*

To check the difficulties in solving the model, we first test the impact of different network sizes and shipment numbers. The indicators for the solution difficulties are indicated by the number of variables, the number of constraints, and the calculation time. The results are summarized in Table 5.7.2. *The supporting data set is shared on the open data platform 4TU.Research Data. (Qu et al., accessed 24 September 2019).*

The first part of Table 10 shows that the network size has a comparatively small impact on the problem size. Meanwhile, the number of shipments has a significant effect on the problem size. This is because the model allows shipment split, which further increases the possible transshipment and the service rescheduling. The calculation is within seconds (root+ branch& cut, CPU time). We indicate ticks here to illustrate the comparisons more clearly.

The above-mentioned networks are the same topology of the Rotterdam hinterland network. To further test the application of the model, we use different network typologies, including the line, ring, star, tree, and fully connected network typologies. The problem sizes are shown in Table 5.7.3.

Table 5.7.2: Theoretical instances of different network size and number of shipments to test the difficulties of solving the model

	Num of nodes	Num. of arcs (LCS+Trucks)	Num. of shipments	Num. of constraints	Num. of variables (binary, integer, other)	Calculation time* (ticks)
Network	6	6 + 600	5	7135	16412 (1722, 1106, 732)	1125.37
	7	8 + 600	5	8968	20121 (2091, 1238, 880)	203.56
	8	10+ 600	5	7558	17075 (1723, 1292, 961)	259.68
	9	12+800	5	10162	22473 (2249, 1827, 1357)	207.58
	10	14+800	5	8558	19508 (1892, 1922, 1352)	210.48
Shipments	10	14 +800	**5**	8558	19508 (1892, 1922, 1352)	210.48
	10	14 + 800	**6**	10063	23084 (2212, 2311, 1634)	261.15
	10	14 + 800	**7**	11534	26592 (2532, 2695, 1888)	367.62
	10	14 + 800	**8**	12282	28524 (2662, 3091, 2174)	566.58
	10	14 + 800	**9**	13100	30466 (2792, 3497, 2410)	442.82
	10	14 + 800	**10**	14623	34078 (3112, 3870, 2711)	774.91
	10	14 + 800	**15**	23466	53918 (4902, 5777, 3942)	882.48

Calculation time: the CPU time, include root + branch &cut

Table 5.7.3: Theoretical instances of different network typology to test the difficulties of solving the model

Network topology	Num. of Nodes	Num. of Arcs (LCS+Truck)	Num. of shipments	Num. of constraints	Num. of variables (binary, integer, other)	Calculation time (ticks)
line	6	13 + 500	5	8152	4819 (1898, 1631, 1290)	2904.14
star	6	12 + 500	5	8139	5016 (1884, 1755, 1377)	255.76
ring	6	14 + 500	5	8478	5138 (1964, 1763, 1411)	1602.03
tree	6	13 + 500	5	7353	3727 (1613, 1157,957)	665.13
fully connected	6	12 + 500	5	6620	2320 (1547, 450, 323)	104.06

The test shows that the STP model can be applied to all these network typologies. As shown in Table 5.7.3, the difficulties of solving the problem is not affected by the topology in question. In each case, the calculation time is also within seconds, which is compatible with the claim that the synchromodality can be used in the local network for real-time replanning. Replanning for larger networks can be divided into several sub-networks replanning problems and solved in rolling horizon fashion.

5.8 CONCLUSIONS AND FUTURE RESEARCH

In this paper, we propose a synchromodality transportation replanning (STP) model for a hinterland freight transportation network. To this end we present a mixed-integer linear programming model, determining the minimum cost solution by combining decisions integrating shipment rerouting and service rescheduling. We demonstrated the performance of the approach in the case of the Rotterdam hinterland transportation network. The STP model provides specific resolutions according to the specific perturbations those are encountered in each case. In addition, we demonstrate the benefit of synchromodality (i.e., flexibility operations in modal split and service rescheduling) compared to rigid planning and non-split shipment planning.

In this study, all the terminals involved have the ability to implement the new resolutions. However, in some cases overloaded terminals could cause delays and delay propagation. In future research, it would be interesting to include the effects of delays and congestion into the model.

Another direction can be to extend this daily-based replanning model into a weekly or monthly plan. However, this can not simply be done in a copy and duplicate way. To minimize the off-target penalty, the model can delay the shippers whose shipments are relatively cheaper. As time goes on, it will cause frequent delays for these shippers and a possibility of losing these customers. Concerning this, the long term replanning can build a customer′ equity mechanism, which balances the delays among different shippers. Moreover, the weekly or monthly planning can also be combined with the supply chain management of the shippers, if provided with storage capacity.

APPENDIX

5.A ORIGINAL SOLUTIONS AND RE-PLANNING SOLUTIONS

Shipments $S4$ and $S5$ are released two hours later than expected time. Table 5.A.1, 5.A.2, and Table 5.A.3 correspondingly give the solutions of the STP model, N-STP model, and R-STP model.

Table 5.A.1: STP Solution: flow assignment and the timetable

Arc	Mode	S1	S2	S3	S4	S5	Start Time	End Time
PoR, Dordrecht, v0001	Barge		50		40		10	14
PoR, Tilburg, v0002	Rail			50		50	10	14
PoR, Nijmegen, v0003	Rail				60		14	18
Dordrecht, Nijmegen, v0004	Barge				40		15	22
Dordrecht, Venlo, v0005	Barge					50	12	23
Tilburg,Venlo, v0006	Rail					50	15	18
PoR, Utrecht,v0007-v0506	Truck	50					16	18

Table 5.A.2: N-STP solution: flow assignment and the adjusted timetable

Arc	Mode	S1	S2	S3	S4	S5	Start Time	End Time
PoR, Dordrecht, v0001	Barge				100		10	14
PoR, Tilburg, v0002	Rail			50			11	15
PoR, Nijmegen, v0003	Rail						not used	not used
Dordrecht, Nijmegen,v0004	Barge				100		15	22
Dordrecht, Venlo, v0005	Barge					100	12	23
Tilburg,Venlo, v0006	Rail						not used	not used
PoR, Utrecht,v0007-v0506	Truck	50					16	18
PoR,Dordrecht,v0507-v0516	Truck		50				16.5	18
PoR,Dordrecht,v0517-v0606	Truck					100	9	10.5

5.B REFERENCE NON-SPLIT SHIPMENT TRANSPORTATION PLANNINGS

The reference non split shipment transportation planning can be formulated with (5.4.4), (5.4.7), (5.4.8), (5.4.9a)-(5.4.13), (5.4.15), (5.4.18)-(5.4.19), (5.4.21), (5.4.24)-(5.4.25) and the following sub-objective and constraints:

Table 5.A.3: R-STP solution: flow assignment and the adjusted timetable

Arc	Mode	S1	S2	S3	S4	S5	Start Time	End Time
PoR, Dordrecht, v0001	Barge		50				7	11
PoR, Tilburg, v0002	Rail			50		50	10	14
PoR, Nijmegen, v0003	Rail				60		14	18
Dordrecht, Nijmegen,v0004	Barge				40		15	22
Dordrecht, Venlo, v0005	Barge					50	12	23
Tilburg,Venlo, v0006	Rail					50	15	18
PoR, Utrecht,v0007-v0506	Truck	50					16	18
PoR,Dordrecht,v0507-v0546	Truck				40		9	10.5
PoR,Dordrecht,v0547-v0596	Truck					50	9	10.5

$$\mathcal{C}^l = \sum_{a \in A} (\chi_a^s \cdot n^s \cdot k_a \cdot c_l) \tag{5.B.1}$$

$$\sum_{a \in \delta_i^+} \chi_a^s - \sum_{a' \in \delta_i^-} \chi_{a'}^s = \begin{cases} -1, & \forall s \in S, i = o^s \\ 0, & \forall s \in S, i \in N \setminus \{o^s, d^s\} \ ; \\ 1, & \forall s \in S, i = d^s \end{cases} \tag{5.B.2}$$

$$\sum_{a \in A} \chi_a^s \leq 1, \qquad \forall s \in S; \tag{5.B.3}$$

$$\sum_{s \in S} \chi_a^s \cdot \ n^s \leq u_v, \quad \forall a = (i,j,v) \in A; \tag{5.B.4}$$

$$\Delta_{a,i}^{s,-} = \begin{cases} \max\left((\chi_a^s - \chi_{a_2'}^s) \cdot n^s, 0\right) & \text{if } p_v(|p_v|) \neq i \\ \max\left((\chi_a^s - 0) \cdot n^s, 0\right) & \text{if } p_v(|p_v|) = i \end{cases}, \quad \forall s \in S, \forall i \in N \setminus \{o^s, d^s\}, \forall a \in \delta_i^+; \tag{5.B.5}$$

$$\Delta_{a',i}^{s,+} = \begin{cases} \max\left((\chi_{a'}^s - \chi_{a_1'}^s) \cdot n^s, 0\right) & \text{if } p_v(|p_v|) \neq i \\ \max\left((\chi_{a'}^s \cdot n^s - 0), 0\right) & \text{if } p_v(|p_v|) = i \end{cases}, \quad \forall s \in S, \forall i \in N \setminus \{o^s, d^s\}, \forall a' \in \delta_i^-; \tag{5.B.6}$$

$$t_a^s \geq y_a^{\mathrm{arv}} - \mathcal{M} \cdot (1 - \chi_a^s), \qquad \forall s \in S, \ \forall a = (i,j,v) \in \delta_j^+, j = d^s; \tag{5.B.7a}$$

$$t_a^s \leq y_a^{\mathrm{arv}} + \mathcal{M} \cdot (1 - \chi_a^s), \qquad \forall s \in S, \ \forall a = (i,j,v) \in \delta_j^+, j = d^s; \tag{5.B.7b}$$

$$w_a^{s,-} = \max\left((\beta^s - t_a^s - \mathcal{M} \cdot (1 - \chi_a^s), 0\right), \qquad \forall s \in S, \ \forall a = (i,j,v) \in \delta_j^+, j = d^s; \tag{5.B.8}$$

$$w_a^{s,+} = \max\left((t_a^s - \beta^s - \mathcal{M} \cdot (1 - \chi_a^s), 0\right), \qquad \forall s \in S, \ \forall a = (i,j,v) \in \delta_j^+, j = d^s. \tag{5.B.9}$$

Part III

WRAPPING UP

In this part we draw conclusions and outline avenues for future research.

6

CONCLUSIONS AND RECOMMENDATIONS

This thesis aims to explore how can the transport systems manage the perturbations at the operation level. The solutions in this work emphasize the synchronization of the involved flows to recover promptly from perturbations and implement the agreed transport tasks. The work first reviews the existing literature concerning the real-time traffic control of rail-bound networks, especially on the studied problems and the corresponding solution methods (Chapter 2). Following that, a study has been conducted to investigate the transfer synchronization control of the metro network which serves massive passenger flows (Chapter 3). Next, the passenger traffic is integrated with the freight traffic in the real-time traffic control of the mainline railway network (Chapter 4). Finally, a consecutive study addresses freight transportation replanning in the framework of the synchromodality which deploys different transport modes (Chapter 5).

6.1 RESULTS AND MAIN FINDINGS

In this section, we present the results and main findings based on the Research Questions from the introduction (Chapter 1).

General RQ: *How can multi-class transport networks best respond to flow perturbations with real-time flow management?*

To deal with perturbations, the network can synchronize (i) the involved service flows including different transport modes, and (ii) the involved client flows including the heterogeneous passenger and freight flows. The operation methods of the first group can be re-timing (Chapter 3, 4 and 5), re-ordering (Chapter 4) and re-routing (Chapter 4). Generally speaking, the synchronization among vehicle flows is operators-oriented. It mainly impacts the operating costs, the safe movement of the vehicles and the recovery time. Second, The operation methods of the second group can be flow re-assignment (Chapter 5) and re-timing (Chapter 4 and Chapter 5). The synchronization of the vehicle flows with the client flows is customer-oriented. It influences the service level, such as the passenger transfer connections or the punctuality rate of the final delivery.

The synchronization control approaches can be designed in a way that it complies with the applied networks and the control aims. The synchronization control of the metro network in this study is customer-oriented. The 'wait' decisions can provide smooth transfer connections to passengers, however it leads to more delay propagation and longer recovery time (see Figure 3.6 and Figure 3.7). On the other hand, the synchromodality replanning model is operator-oriented, thus the control aim is calibrated to minimize the overall operating costs. Therefore the usage of the Large Capacity Services causes higher early or late delivery, but it reduces the operating costs (See Table 5.10).

The level of feasible flexibility also influences the synchronization control. For the replanning of the mainline railways in Chapter 4, the shipments are bond with the freight trains. The case is based on the assumption that the agreement with the shippers does not allow switching their shipments to other vehicles. However, the shipments can be shifted to other modes of vehicles in the framework of the synchromodality. The centralized organization of synchromodality (Chapter 5) further improved flexibility during the synchronization control via the shipment split and vehicle sharing.

RQ 1: *What is the focus of the existing literature in rail-bound networks and what are the promising directions for further research?*

In Chapter 2 we summarize the literature on real-time traffic control of the rail-bound networks. Most studies focus on passenger trains and ignore freight trains. Few studies include freight trains but treat them as lower prioritized passenger trains. This ignores the special characteristic and operating methods of freight trains. Meanwhile, the volume of freight transport keeps growing and gaining more attention on resilience optimization of the logistics. Moreover, in real practice freight traffic coexists with passenger traffic. Based on that we study the simultaneous traffic control of heterogeneous traffic, exploring the diversity and interplay of the two different client flows.

To return to the five points raised in the end of Chapter 2, we note that all of these are linked to the demand growth or saturated usage of the network capacity. In this thesis we have looked into several of the aspects raised, in particular we note the following. Firstly, we mention that the congestion due to a massive passenger volume is studied in Chapter 3. Some of the challenges resulting from mixed passenger and freight flows are studied in Chapter 4. As an extension, vehicles of different modalities are coordinated in Chapter 5.

Besides, the combination of macro- and microscopic formulations appears in Chapter 4. There are two main variants of the mathematical formulations, i.e., the macroscopic and microscopic (Cacchiani et al., 2014, Qu, Corman, and Lodewijks, 2015). The classification is mainly made based on the level of aggregation. The first group mainly describes the stations and tracks as nodes and arcs. It can be used for the control of large-scale networks. Meanwhile, the second group further takes the block sections and signalling into account. Thus these models can check the feasibility and quality of the solutions from the viewpoint of operation managers. The research aims to explore the interaction of passenger and freight traffic. Freight trains usually cover a longer distance and running time, which requires a broader network and a longer observe time. Thus the model in Chapter 4 is positioned as the macroscopic level of formulations and solutions.

RQ2: *How to improve the service level to passengers based on the previous studies and what can be the criteria to indicate the service level?*

Real-time traffic control of a metro network concerns two decisions, (i) the rescheduling of trains and (ii) the wait or depart decisions for pre-planned transfer connections. The first group of decisions is operators-oriented. The second group of decisions directly influences whether the passengers can make the transfer connections or have to wait extra for the next connecting train. Thus to optimize the wait or depart decisions are crucial for the level of services. Besides, the volume of passenger flows varies significantly during the day. The congestion problem during the morning and peak hours is still a challenging problem to solve. Thus to integrate the passenger flow is essential to improve the service level.

Chapter 3 develops a transfer synchronization model for metropolitan metro networks that deal with massive passenger flows. The model aims to improve the level of services to passengers. The model introduces a rescheduling strategy based on the so-called waiting cost instead of the traditional waiting time approach. In this way, the passenger experience becomes an essential component of the model. The wait and depart decisions concern (i) *whether* to hold the connecting train and (ii) *how long* the connecting trains can wait for a specific delay of the feeder train. We have tested the model on the core part of the Beijing metro network using various periods. The decision of *whether* to wait or not is significantly different in the morning and evening peak hours. The comparison with a classical waiting-time based model reveals that the newly proposed model provides different decisions of *how long* the connecting train should wait for a specific delay of the feeder train. The waiting time is relatively long when the volume of transfer passengers is high. Therefore it saves passenger extra waiting times and loss of comfort caused by congestion.

We conclude in this study that the shorter passenger delays can be achieved via longer train delays and longer recovery time windows. The large volume group of passengers is given higher priority than the smaller group. This biased decision is based on the assumption that the metro networks of big cities transport a massive volume of passengers into or away from the business area in peak hours. Besides, the transport organizer is willing to reduce the congestion-caused discomfort in the disturbance management.

RQ 3: *How to integrate the freight flows with the passenger flows in the real-time traffic control while considering their specific characteristics and operating methods?*

The increasing freight transportation demand and the trend of mode shift from truck to trains ask for improving the service level of freight transportation and new re-arranging methods. It is essential to improve the service quality of railway freight transportation to attract more freight to railways. This also has to take into account the already congested situation of the passenger trains, that share the railway lines with freight trains in practice.

The specific characteristics and operation methods lie in the requirements of the transportation clients. The passengers perceive the whole trip as a service. Thus for the passenger trains, the goal should be to minimize the deviation from the original timetable and provide effortless transfer connections, while for the freight trains, the clients care mainly about the final delivery time. The exact route and waiting time at the intermediate stations are less important.

We present a mathematical model integrating the passenger and freight traffic in Chapter 4, while clearly distinguishing their characteristics. More flexibility can be given to the freight trains in a re-planning model. For example, the routing choice and the dwelling time at the intermediate stations can be modified. We test the model on a core part of the Dutch railway network provides the following conclusions. Giving more flexibility to the freight trains, both passenger train delays and freight train delays can be reduced. It can also leads to higher passenger train delays, then there is a trade-off between the heterogeneous traffic. If the open track segments and extra station capacity are released in case of delays, the freight trains can depart earlier and arrive earlier than planned in the timetable. This implies that the freight train delays reduce. Moreover, we found that both passenger traffic and freight traffic can benefit from rerouting of freight trains

This study indicates that freight trains can travel a longer distance but arrive earlier in time when delays occur. This is the under the condition that freight trains are treated as lower-valued or lower-prioritised passenger trains. Besides, the re-planning of the freight trains can benefit from a longer time windows or broader area.

RQ 4: *How to model the cooperation among the service providers and the overall flexibility of synchromodality?*

Hinterland freight transportation is managed according to a pre-designed schedule. In daily operations, unexpected perturbations occur and cause deviations from the original plan. This asks for replanning in order to deal with the perturbations and complete the transportation tasks. Synchromodal transportation is an upcoming new solution strategy succeeding intermodal transportation that provides a more flexible and cooperated freight transportation at the operational level with the purpose of dealing with uncertainties. Under the framework of synchromodality, shippers accept a mode-free booking and only determine the price and quality requirements. The involved service providers collaborate under the centralized supervision of one Logistic Service Provider (LSP). This implies that it is possible to perform real-time switching among different modes. The centralized LSP holds the promise of generating specific replanning solutions to the specific uncertainties encountered in each individual case.

In Chapter 5, we proposed a synchromodality transportation replanning model for a hinterland freight transportation network. To this end we present a mixed-integer linear programming model, determining the minimum cost solution. More concretely, the replanning benefits from high operational flexibility and coordination. Flexibility is increased by allowing individual containers from a single shipment to travel by different routes. Coordination is improved by aligning departure times of service flows with the shipment flows.

We demonstrate the advantage of the model for the Rotterdam hinterland network. The network is exposed to different perturbations including the late release of shipments, delay of vehicles, and volume fluctuations. Key performance indicators are used to evaluate the flexibility of the solutions, and the level of service. The results show that our model can synchronize from the space and time dimension to tackle various perturbations. From the time dimension, the alignment of vehicle departure time can effectively solve the late release of shipments and vehicle delays. From the space dimension, the shipment rerouting can tackle the volume fluctuations. The comparisons with reference models show that vehicle alignment, shipment split and vehicle sharing can improve the modal split and utility rate of Large Capacity Services. This further brings down the total service cost since the fixed cost and variable costs of large capacity services are cheaper than those of trucks.

We conclude from this study that the coordination among the vehicles of different modes can react flexibly to perturbations. This leads to solutions with lower operating costs, less off-target deliveries and higher modality shares of large capacity vehicles. This coordination and flexibility can not be achieved without the synchromodality framework: the shippers give up mode choices, and the carriers allow the free switch of shipment flows.

6.2 APPLICATION RECOMMENDATIONS

All the models presented in this work seek to build replanning model for transport networks to deal with perturbations. However, it is impossible to build an all-inclusive application that can be adapted to a specific environment.

The rescheduling models in Chapter 3 can be used for the metro dispatcher in particular to those in metropolis cities to deal with delays. It aims to adjust the wait depart time of the metro trains according the passenger flows to provide smooth transfer connections and to reduce the congestion in carriages. That is based on the situation that the metro transits a massive volume of passengers towards or away from the central business districts in the morning and evening peak hours. This model can provide off-line decision tables generated by history passenger flows. Meanwhile, it can also be used for some big cultural or sports events to provide online decision support to real-time traffic control. The adjustment in the operating times can be informed to the passenger in advance to avoid confusion and non-preparation. This can be done via phone apps and the media inside metro stations.

The real-time traffic management model in Chapter 4 is aimed at providing decision support to mainline railways dispathers. It can be used to balance different stakeholders represented by the heterogeneous traffic flows. The transport organizer can coordinate the heterogeneous traffic by considering their characteristics and operating methods, instead of using a set of priorities and postponing the lower prioritized ones.

The application of this simultaneous traffic management involves the transportation organizer and the service operators. For the dutch railway network, ProRail is the railway infrastructure manager, who allocates the capacity to the passenger trains operators (such as NS Reizigers, Arriva, Connexxion) and the freight operators (KNV, Koninklijk Nederlands Vervoer). From the ProRails' side, it takes effort to set capacity for the new route of the freight trains in the perturbation management. There are extra concerns about how to comply the freight trains with paths which have a higher average speed. From the KNV's side, it is economically beneficial only if estimated gains of avoiding late delivery penalties can cover the extra costs of traveling longer distances.

The synchromodality model of Chapter 5 finds its use as a decision support tool for a Logistic Service Provider to re-organize the freight transportation. It is a holistic resolution of shipment flow rerouting, consequence transshipment organization in the intermediate terminals, and corresponding service rescheduling. The case study of Rotterdam hinterland network shows that the usage of Large Capacity Service can reduce the transport costs and cause comparatively smaller penalty for off-time delivery. That is to say, the model shifts more shipment flows to barge or railways to deal with the perturbations. However, it can cause a higher number of early or late deliveries.

The synchromodality framework assumes an ideal centralized organization situation, where the shippers give up their own preferences, and the logistic service providers (carriers) agree with the free switch of shipment from one carrier to another. This assumption of the coordination comes with the promise of a later avenue compensation to those carriers who lose their contract(s). In real practice, independent carriers compete with each other and bargain with the shippers in the logistics market, which is a typical discrete optimization problem. To change the minds of shippers and independent carriers, the policy makers need to illustrate clearly the benefits of coordinated perturbation organization. Moreover, because of the above-mentioned competition, the carriers, the shippers and the terminals generally do not share their data regarding orders, costs and the trading price. Moreover, there may occur errors in the data of the individual carriers or the data may be incomplete. The confidentiality increases the difficulties of the synchronization modelling and analysis of the benefits.

Except for the transformation from distributed optimization to a centralized optimization in the perturbation management, there are other obstacles as well. One of them is that the flexible operations will bring subsequent uncertainties, and high level of flexibility brings more uncertainties. We mention some source of uncertainties in the following. The proposed rescheduling and rerouting of the trains in Chapter 3 and Chapter 4 aim at providing smooth transfers to passengers and reducing late deliveries of the freight goods. Those imply additional uncertainties of platform tracks re-assignment, the rolling stock allocation and recruitment management. There are studies dealing with those problems but none all-inclusive study. The proposed shipment re-assignment and vehicles rescheduling provides flexible reactions to the perturbations. The shipment flow re-assignment calls for extra storage of the earlier arrived shipment split and the combination with the later arrived split. The vehicle rescheduling introduces another uncertainty, namely the working time constraints of the vehicle drivers. Besides, the shipment flow re-assignment and vehicle rescheduling requires extra modifications to the transshipment in the intermediate terminals. Thus the terminals have to re-plan the cranes and vehicles in a short time. There are also other concerns such as re-positioning of empty containers, empty runs of the service vehicles and passenger rerouting choice.

The author does not suggest to solve these consequence uncertainties at once. The calculation becomes more challenging when more uncertainties are integrated. Using heuristic algorithms can get faster solutions and is beneficial to explore more complicated problems. However, those consequence uncertainties can also be solved in the next stage re-planning, or in a rolling horizon manner. The proposed models can be adopted as the first stage plan, concerning on implementing the transit tasks. The contribution of this work is that it explores the interaction and synchronization of the involved flows in different applied networks. In practice, the transport planer can check the proposed model for reference to react to the perturbations and adopt certain operating measures based on the applied environment. The presented case studies and analysis provide insights for corresponding organizers about balancing different stakeholders and at the same time, taking advantage of the operating characteristics of the service flows.

6.3 FUTURE RESEARCH

This section provides possible future research directions and extensions. Each separate chapter concludes the future research in the end for specific parts of the thesis, which will not be repeated here. Here we add several other venues for possible extensions at an overarching level.

A first suggestion is to use a dynamic modelling approach such as *model predictive control*. The models proposed in this work are snap-shot solutions to deal with the current disturbances. However, we note that the external circumstances outside the applied network keep changing. To take this into account the network should be dynamically updated. The dynamic modelling needs to keep using real-time information. Another possible research direction concerns the local optimization of a sub-network and its interaction with the global optimization problem. For example, optimizing the internal operations of a terminal under the synchromodality framework. The optimization model in Chapter 5 is operator-oriented and aims to provide a holistic solution for the logistic service provider. A single terminal is regarded as one body and assigned with tasks within its capacity. Still the terminal contains its components and can be further optimized with its operation. Similarly, the rescheduled and rerouted trains require the platform tracks re-assignments and usage time re-arrangements. For large stations, such as Utrecht station who has 16 platform tracks and is busy with receiving and sending out trains, is an complex sub-network and needs to be well re-planned in the perturbations. With enhanced capacities, the sub-networks can improve the performance of the global network.

For the synchromodality, it is important to take into account shippers' level of willingness to delegate control over their shipment to LSPs as it is key for a successful implementation of synchromodality. In that regard, it would be interesting to see how shippers who are not willing to delegate their control could be motivated to do so; what mechanisms we could use to make synchromodality more attractive to them, see Khakdaman, Rezaei, and Tavasszy, 2020.

Beyond the synchromodality framework, the carriers from different modalities compete with each other in the logistics market, especially those of similar carrier types and aim at the same target of orders. An agent-based modelling paradigm can be used to simulate the relationship between carriers, and the connection between the market and the carriers. The policy makers or 'logistics orchestrator' can study the trading strategy and investigate how to formulate the coordination among the carriers. As mentioned in Chapter 1, the freight demand keeps growing and therefore a rising perturbation scenario can be that the shipment volume of a committed contract turns out higher than expected. The carriers can have a capacity shortage but still are not willing to transfer part of the contract to their competitors. One possibility is that carriers can pay competitors to rent vehicles for a short period. In this regard, it will be useful to build a platform for temporary capacity lease among the carriers, with time constraints regarding to the due time of the shipments. The sharing of assets and capacity is also the idea of the physical internet(Ballot, Montreuil, and Zacharia, 2021).

In general we recommend looking into more efficient solution strategies to allow the application of client-orientated optimizations models to real-world problems.

BIBLIOGRAPHY

Acuna-Agost, R., Michelon, P., Feillet, D., and Gueye, S. (2011). “A MIP-based local search method for the railway rescheduling problem.” *Networks* 57, pp. 69–86.

Albrecht, T., Binder, A., and Gassel, C. (2011). “An overview on real-time speed control in rail-bound public transportation systems.” *Proceedings of the 2nd International Conference on Models and Technologies for Intelligent Transportation Systems*. Leuven Belgium, pp. 1–4.

Albrecht, T., Binder, A., and Gassel, C. (2013). “Applications of real-time speed control in rail-bound public transportation systems.” *IET Intelligent Transport Systems* 7, pp. 305–314.

Andersen, J. and Christiansen, M. (2009). “Designing new European rail freight services.” *Journal of the Operational Research Society* 60, pp. 348–360.

Atamtürk, A. and Zhang, M. (2007). “Two-Stage Robust Network Flow and Design Under Demand Uncertainty.” *Operations Research* 55, pp. 662–673.

Bai, R., Wallace, S.W., Li, J., and Chong, A.Y.L. (2014). “Stochastic service network design with rerouting.” *Transportation Research Part B: Methodological* 60, pp. 50–65.

Ballot, E., Montreuil, B., and Zacharia, Z.G. (2021). *Physical Internet: First results and next challenges*.

Behdani, B., Fan, Y., Wiegmans, B., and Zuidwijk, R. (2016). “Multimodal Schedule Design for Synchromodal Freight Transport Systems.” *European Journal of Transport and Infrastructure Research* 16, pp. 424–444.

Bešinović, N., Widarno, B., and Goverde, Rob M.P. (2020). “Adjusting freight train paths to infrastructure possessions.” *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, pp. 1–6.

Binder, S., Maknoon, Y., and Bierlaire, M. (2017). “The multi-objective railway timetable rescheduling problem.” *Transportation Research Part C: Emerging Technologies* 78, pp. 78–94.

Braker, J.G. and Olsder, G.J. (1993). “The power algorithm in max algebra.” *Linear Algebra and its Applications* 182, pp. 67–89.

Cacchiani, V., Caprara, A., and Toth, P. (2010). “Scheduling extra freight trains on railway networks.” *Transportation Research Part B: Methodological* 44, pp. 215–231.

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

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

Caprara, A., Fischetti, M., and Toth, P. (2002). "Modeling and Solving the Train Timetabling Problem." *Operations Research* 50, pp. 851–861.

Clark, S. and Watling, D. (2005). "Modelling network travel time reliability under stochastic demand." *Transportation Research Part B: Methodological* 39, pp. 119–140.

Corman, F. (2020). "Interactions and Equilibrium Between Rescheduling Train Traffic and Routing Passengers in Microscopic Delay Management: A Game Theoretical Study." *Transportation Science* 54, pp. 785–822.

Corman, F., D'Ariano, A., D.M., Alessio, Pacciarelli, D., and Sama, M. (2017). "Integrating train scheduling and delay management in real-time railway traffic control." *Transportation Research Part E: Logistics and Transportation Review* 105, pp. 213–239.

Corman, F., D'Ariano, A., Hansen, I.A., and Pacciarelli, D. (2011). "Optimal multi-class rescheduling of railway traffic." *Journal of Rail Transport Planning Management* 1. Robust Modelling of Capacity, Delays and Rescheduling in Regional Networks, pp. 14–24.

Corman, F., D'Ariano, A., Pacciarelli, D., and Pranzo, M. (2012). "Bi-objective conflict detection and resolution in railway traffic management." *Transportation Research Part C: Emerging Technologies* 20. Special issue on Optimization in Public Transport+ISTT2011, pp. 79–94.

Corman, F., D'Ariano, A., Pacciarelli, D., and Pranzo, M. (2009). "Evaluation of green wave policy in real-time railway traffic management." *Transportation Research Part C: Emerging Technologies* 17.6, pp. 607–616.

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

Corman, F., Pacciarelli, D., D'Ariano, A., and Samá, M. (2015). "Rescheduling Railway Traffic Taking into Account Minimization of Passengers' Discomfort." *Computational Logistics*. Ed. by F. Corman, S. Voß, and R.R. Negenborn. Cham: Springer International Publishing, pp. 602–616.

Corman, F., Viti, F., and Negenborn, R.R. (2017). "Equilibrium models in multimodal container transport systems." *Flexible Services and Manufacturing Journal* 29, pp. 125–153.

Crainic, T.G. (2000). "Service Network design in freight transportation." *European Journal of Operation Research* 122 (2), pp. 272–288.

Crainic, T.G., Hewitt, M., and Rei, W. (2014). “Scenario grouping in a progressive hedging-based meta-heuristic for stochastic network design.” *Computers & Operations Research* 43, pp. 90–99.

Currie, G. (2010). “Quick and Effective Solution to Rail Overcrowding: Free Early Bird Ticket Experience in Melbourne, Australia.” *Transportation Research Record* 2146, pp. 35–42.

D’Ariano, A., Corman, F., Pacciarelli, D., and Pranzo, M. (2008). “Reordering and Local Rerouting Strategies to Manage Train Traffic in Real Time.” *Transportation Science* 42, pp. 405–419.

D’Ariano, A., Pacciarelli, D., and Pranzo, M. (2007). “A branch and bound algorithm for scheduling trains in a railway network.” *European Journal of Operational Research* 183, pp. 643–657.

D’Ariano, A., Pacciarelli, D., and Pranzo, M. (2008). “Assessment of flexible timetables in real-time traffic management of a railway bottleneck.” *Transportation Research Part C: Emerging Technologies* 16, pp. 232–245.

De Jong, G., Tseng, Y., Kouwenhoven, M., Verhoef, E., and Bates, J. (2007). “The value of travel time and travel time reliability.” *The Netherlands Ministry of Transport*.

Dewilde, T. (2014). “Improving the robustness of a railway system in large and complex station areas.” PhD thesis. Ku Leuven.

Dollevoet, T., Huisman, D., Kroon, L., Schmidt, M., and Schöbel, A. (2015). “Delay management including capacities of stations.” *Transportation Science* 49, pp. 185–203.

Dollevoet, T., Huisman, D., Schmidt, M., and Schöbel, A. (2012). “Delay management with rerouting of passengers.” *Transportation Science* 46, pp. 74–89.

Dündar, S. and Şahin, İ. (2013). “Train re-scheduling with genetic algorithms and artificial neural networks for single-track railways.” *Transportation Research Part C: Emerging Technologies* 27. Selected papers from the Seventh Triennial Symposium on Transportation Analysis (TRISTAN VII), pp. 1–15.

Escudero, A., Muñuzuri, J., Guadix, J., and Arango, C. (2013). “Dynamic approach to solve the daily drayage problem with transit time uncertainty.” *Computers in Industry* 64. Decision Support for Intermodal Transport, pp. 165–175.

EuroStat (2017). *Database: Freight transport statistics*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Freight_transport_statistics. [Online; accessed 27 August 2019].

Gély, L., Dessagne, G., and Lérin, C. (2006). “Modelling train re-scheduling with optimization and operational research techniques: results and applications at SNCF.” *World Congress on Railway Research*. Vol. 29.

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

Giusti, R., Iorfida, C., Li Y., and Manerba D., Musso, S., Perboli, G., Tadei, R., and Yuan, S. (2019). "Sustainable and de-stressed international supply-chains through the SYNCHRO-NET approach." *Sustainability* 11, p. 1083.

Gleave., S. D., Dionori, F., Casullo L., Ellis S., Ranghetti, D., Bablinski, K., Vollath, C., and Soutra, C. (2015). *FREIGHT ON ROAD: WHY EU SHIPPERS PREFER TRUCK TO TRAIN*. https://www.europarl.europa.eu/RegData/etudes/STUD/2015/540338/IPOL_STU(2015)540338_EN.pdf. [Online; accessed 20 December 2020].

Godwin, T., Gopalan, R., and Narendran, T.T. (2007). "A heuristic for routing and scheduling freight trains in a passenger rail network." *International Journal of Logistics Systems and Management* 3, pp. 101–133.

Goel, A. (2010). "The value of in-transit visibility for supply chains with multiple modes of transport." *International Journal of Logistics Research and Applications* 13, pp. 475–492.

Goeverden C.D., Behr F.H. Schoemaker T.J.H. Wiggenraad P.B.L. van (1990). *Onderzoek weging tijdelementen (Deelrapport 5) (in Dutch)*. http://resolver.tudelft.nl/uuid:0c7999c8-c96e-4a4b-9e2c-da920e2e8bc9.

Goulet-Langlois, G., Koutsopoulos, H. N., and Zhao, J. (2016). "Inferring patterns in the multi-week activity sequences of public transport users." *Transportation Research Part C: Emerging Technologies* 64, pp. 1–16. ISSN: 0968-090X.

Goverde, R.M.P. (1998). "Synchronization control of scheduled train services to minimize passenger waiting times." *Transport, Infrastructure and Logistics. Proceedings 4th TRAIL Congress 1998*. TRAIL.

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

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

Goverde, Rob and Odijk, M.A. (Jan. 2002). "Performance evaluation of network timetables using PETER." *Computers in Railways* 8, pp. 731–740.

Guo, W., Atasoy, B., Beelaerts, W., and Rudy Negenborn, R. (2020). "A dynamic shipment matching problem in hinterland synchromodal transportation." *Decision Support Systems* 134, p. 113289. ISSN: 0167-9236.

Guo, W., Blokland W., B. van, Lodewijks, G., and Negenborn, R.R. (2018). "Multi-Commodity Multi-Service Matching Design for Container Transportation Systems."

Han, B., Zhou, W., Li, D., and Yin, H. (2015). "Dynamic Schedule-Based Assignment Model for Urban Rail Transit Network with Capacity Constraints." *The Scientific World Journal*.

Heidergott, B. and Vries, de. R. (2001). "Towards a (max,+) control theory for public transportation networks." *Discrete Event Dynamic Systems* 11, pp. 371–398.

Hoff, A., Lium, A. G., Løkketangen, A., and Crainic, T.G. (2010). "A metaheuristic for stochastic service network design." *Journal of Heuristics* 16, pp. 653–679.

Hu, Q., Corman, F., Wiegmans, Bart., and Lodewijks, G. (2018). "A tabu search algorithm to solve the integrated planning of container on an inter-terminal network connected with a hinterland rail network." *Transportation Research Part C: Emerging Technologies* 91, pp. 15–36.

Hurk, E. van der, Kroon, L.G., Maróti, G., Bouman, P., and Vervest, P.H.M. (2013). "Network reduction and dynamic forecasting of passenger flows for disruption management." *Proceedings of the 5th International Conference on Railway Operations Modelling and Analysis (RailCopenhagen2013)*. Denmark, pp. 13–15.

IBM (2020). *CPLEX implements Benders algorithm*. https://www.ibm.com/support/knowledgecenter/SS9UKU_12.10.0/com.ibm.cplex.zos.help/CPLEX/UsrMan/topics/discr_optim/benders/introBenders.html. [Online; accessed 10 December 2020].

Jespersen-Groth, J., Potthoff, D., Clausen, J., Huisman, D., Kroon, L., Maróti, G., and Nielsen, M.N. (2009). "Disruption management in passenger railway transportation." *Robust and online large-scale optimization*. Springer, pp. 399–421.

Jiang, F., Cacchiani, V., and Toth, P. (2017). "Train timetabling by skip-stop planning in highly congested lines." *Transportation Research Part B: Methodological* 104, pp. 149 –174.

Kanai, S., Shiina, K., Harada, S., and Tomii, N. (2011). "An optimal delay management algorithm from passengers' viewpoints considering the whole railway network." *Journal of Rail Transport Planning Management* 1, pp. 25–37.

Khakdaman, M., Rezaei, J., and Tavasszy, L.A. (2020). "Shippers' willingness to delegate modal control in freight transportation." *Transportation Research Part E: Logistics and Transportation Review* 141, p. 102027. ISSN: 1366-5545.

Kraay, D.R. and Harker, P.T. (1995). "Real-time scheduling of freight railroads." *Transportation Research Part B: Methodological* 29, pp. 213–229.

Krasemann, J.T. (2012). "Design of an effective algorithm for fast response to the re-scheduling of railway traffic during disturbances." *Transportation Research Part C: Emerging Technologies* 20, pp. 62–78.

Kuo, A., Miller-Hooks, E., and Mahmassani, H.S. (2010). "Freight train scheduling with elastic demand." *Transportation Research Part E: Logistics and Transportation Review* 46, pp. 1057–1070.

Lamorgese, L. and Mannino, C. (2015). "An Exact Decomposition Approach for the Real-Time Train Dispatching Problem." *Operations Research* 63, pp. 48–64.

Li, L., Negenborn, R.R., and De Schutter, B. (2015). "Intermodal freight transport planning– A receding horizon control approach." *Transportation Research Part C: Emerging Technologies* 60, pp. 77–95.

Li, L., Negenborn, R.R., and De Schutter, B. (2017). "Distributed model predictive control for cooperative synchromodal freight transport." *Transportation Research Part E: Logistics and Transportation Review* 105, pp. 240–260.

Liebchen, C. and Stiller, S. (2009). "Delay resistant timetabling." *Public Transport* 1, pp. 55–72.

List, G.F., Wood, B., Nozick, L.K., Turnquist, M.A., Jones, D.A., Kjeldgaard, E.A., and Lawton, C.R. (2003). "Robust optimization for fleet planning under uncertainty." *Transportation Research Part E: Logistics and Transportation Review* 39, pp. 209–227.

Lium, A. G., T., Crainic G., and S., Wallace W. (2009). "A Study of Demand Stochasticity in Service Network Design." *Transportation Science* 43, pp. 144–157.

Lofberg, J. (2004). "YALMIP : a toolbox for modeling and optimization in MATLAB." *2004 IEEE International Conference on Robotics and Automation (IEEE Cat. No.04CH37508)*, pp. 284–289.

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

Meng, Q., Wang, T., and Wang, S (2012). "Short-term liner ship fleet planning with container transshipment and uncertain container shipment demand." *European Journal of Operational Research* 223, pp. 96–105.

Min, W. and Wynter, L. (2011). "Real-time road traffic prediction with spatio-temporal correlations." *Transportation Research Part C: Emerging Technologies* 19, pp. 606–616.

Min, Y., Park, M.J., Hong, S.P., and Hong, S.H. (2011). "An appraisal of a column-generation-based algorithm for centralized train-conflict resolution on a metropolitan railway network." *Transportation Research Part B: Methodological* 45, pp. 409–429.

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

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

Oetting, A. and Keck, A.K. (2015). "Monetarization of delay valuation for freight." *6th International Conference on Railway Operations Modelling and Analysis-Rail Tokyo 2015*.

Pellegrini, P., Marllière, G., and Rodriguez, J. (2014). "Optimal train routing and scheduling for managing traffic perturbations in complex junctions." *Transportation Research Part B: Methodological* 59, pp. 58–80.

Perboli, G., Musso, S., Rosano, M., Tadei, R., and Godel, M. (2017). "Synchro-Modality and Slow Steaming: New Business Perspectives in Freight Transportation." *Sustainability* 9, p. 1843.

Polinder, G. J. (2020). "New Models and Applications for Railway Timetabling." PhD thesis. Erasmus University Rotterdam.

Polinder, G. J., Cacchiani, V., Schmidt, M., and Huisman, D. (2020). "An Iterative Heuristic for Passenger-Centric Train Timetabling with Integrated Adaption Times." Available at SSRN: `https://ssrn.com/abstract=3629576`.

Qu W., Maknoon Y. Corman F. (2020). "Real time railway traffic control with mixed passenger and freight traffic." *to be submitted*.

Qu, W., Corman, F., and Lodewijks, G. (2015). "A review of real time railway traffic management during disturbances." *International Conference on Computational Logistics*, pp. 658–672.

Qu, W., Rezaei, J., Maknoon, Y., and Tavasszy, L. (2019). "Hinterland freight transportation replanning model under the framework of synchromodality." *Transportation Research Part E: Logistics and Transportation Review* 131, pp. 308–328.

Qu, W., Rezaei, J., Maknoon, Y., and Tavasszy, L.A. (accessed 24 September 2019). "Theoretical test for the work hinterland freight transportation replanning model under the framework of synchromodality." *Online; 4TU.Centre for Research Data. Dataset*.

Qu, W., Wu, J.J., Corman, F., and Gao, Z.Y. (2020). "Synchronization control of passenger flows in an urban railway network." *to be submitted*.

Riessen, B. van, Negenborn, R. R., Lodewijks, G., and Dekker, R. (2015). "Impact and relevance of transit disturbances on planning in intermodal container networks using disturbance cost analysis." *Maritime Economics & Logistics* 17, pp. 440–463.

Rivera, A.E.P. and Mes, M. (2017). "Anticipatory Scheduling of Freight in a Synchromodal Transportation Network." *Transportation Research Part E: Logistics and Transportation Review* 105, pp. 176–194.

Robenek, T., Azadeh, S. S., Maknoon, Y., Lapparent, M. de, and Bierlaire, M. (2018). "Train timetable design under elastic passenger demand." *Transportation Research Part B: Methodological* 111, pp. 19–38.

Rodriguez, J. (2007). "A constraint programming model for real-time train scheduling at junctions." *Transportation Research Part B: Methodological* 41, pp. 231–245.

Schöbel, A. (2001). "A model for the delay management problem based on mixed-integer-programming." *Electronic notes in theoretical computer science* 50, pp. 1–10.

Schöbel, A. (2007). "Integer programming approaches for solving the delay management problem." *Algorithmic methods for raileeay optimization*. Springer, pp. 145–170.

Schöbel, A. (2009). "Capacity constraints in delay management." *Public Transport* 1, pp. 135–154.

Shang, P., Li, R., Liu, Z., Yang, L., and Wang, Y. (2018). "Equity-oriented skip-stopping schedule optimization in an oversaturated urban rail transit network." *Transportation Research Part C: Emerging Technologies* 89, pp. 321 –343.

SteadieSeifi, M., Dellaert, N.P., Nuijten, W., Woensel, T. van, and Raoufi, R. (2014). "Multimodal freight transportation planning: A literature review." *European Journal of Operational Research* 233, pp. 1–15.

Tavasszy, L.A., Lugt, van der, G.R., Janssen, and Hagdorn, E. (2010). *Verkenning Synchromodaal Transportsysteem*. Delft, the Netherlands.

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

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

TreinreizigerNL (2019). *Aantal in- en uitstappers per station bij NS 2013-2018*. https://www.treinreiziger.nl/aantal-in-en-uitstappers-per-station-2013-2018/. [Online; accessed 10 December 2020].

Ukkusuri, S.V., Mathew, T.V., and Waller, S.T. (2007). "Robust Transportation Network Design Under Demand Uncertainty." *Computer-Aided Civil and Infrastructure Engineering* 22, pp. 6–18.

Wegele, S. and Schnieder, E. (2004). "Dispatching of train operations using genetic algorithms." *Proceedings of the 9th International Conference on Computer-Aided Scheduling of Public Transport (CASPT), San Diego, USA*, p. 11.

Wong, R.C.W., Yuen, T.W.Y., Fung, K.W., and Leung, J.M.Y. (2008). "Optimizing timetable synchronization for rail mass transit." *Transportation Science* 42, pp. 57–69.

Wu, J., Sun, H., Gao, Z., Han, L., and Si, B. (2014). "Risk-based stochastic equilibrium assignment model in augmented urban railway network." *Journal of Advanced Transportation* 48, pp. 332–347.

Yang, H. and Tang, Y. (2018). "Managing rail transit peak-hour congestion with a fare-reward scheme." *Transportation Research Part B: Methodological* 110, pp. 122–136.

Yen, J.Y. (1971). "Finding the k Shortest Loopless Paths in a Network." *Management Science* 17, pp. 712–716.

Zhang, M. and Pel, A.J. (2016). "Synchromodal hinterland freight transport: Model study for the port of Rotterdam." *Journal of Transport Geography* 52, pp. 1–10.

Zhu, Y. and Goverde, R.M.P. (2019). "Railway timetable rescheduling with flexible stopping and flexible short-turning during disruptions." *Transportation Research Part B: Methodological* 123, pp. 149–181.

Zhu, Y., Koutsopoulos, H., and Wilson, N. (2017). "Inferring left behind passengers in congested metro systems from automated data." *Transportation Research Part C: Emerging Technologies* 94.

SUMMARY

Transport networks are operated under a pre-designed plan to carry out the transportation tasks. Perturbations can inevitably arise exogenously or endogenously, and cause disturbances to the transit of passengers or the delivery of freights. A transitional plan is needed to deal with the perturbations and accomplish the transport tasks. The transitional plan is referred to as *perturbations management* in this study.

The first contribution of this work is that we construct the re-planning models for multi-class networks. The perturbation management for the single-class transport network, for instance, a passenger railway network is well studied, however, it is still missing for multi-class transport networks. In real applications, passenger flows and freight flow sometimes co-exist in one system. The volume of freight transport keeps increasing. On the other hand, there are usually multi-mode services associated to finish the transit of one object. The current research is centred around this point. We study perturbation management for multi-class networks, which involve the mixed passenger and freight railway networks and deploy various modes of service vehicles.

There are several important research questions regarding the management of heterogeneous traffic flows in case of delays. The first challenge is to get an overview of the focus of the existing literature in rail bound networks and find out what are the promising directions for further research. Second, concerning the modeling part, the challenge is to come up with a replanning strategy and determine the control objective. Another task is to find out how to manage the different flows involved, either for the served clients or for the serving vehicles. Last concern is to explore how to achieve flexibility for the integrated network.

The second contribution of this work is that the replanning models arrange the heterogeneous traffic according to their characteristics. When delays occur to a mixed passenger and freight railway network, the replanning gives more flexibility to the freight trains. The underlying reason is that the freight clients, i.e., the shippers care mainly about the finally delivery time instead of the taken route or the waiting time at the intermediate stations. While in the synchromodality replanning, the three modes of service are scheduled according to their applied sub-networks. In the multimodal transport network, the large capacity vehicles are aligned with the freight flows.

Another contribution of this thesis is that the models achieve the synchronization of the involved flows. This is vital to perturbation management, even if the applied networks are different. We pursue the synchronization of flows from both time and space dimensions. We adjust the scheduling of the metro vehicles in agreement with

the massive passenger flows (Chapter 3). We align the scheduling of three modes of service vehicles according to the delayed preceding vehicles or shipment release (Chapter 5). Meanwhile, the passenger transfer management (Chapter 3 and Chapter 4) and the freight flows re-assignment (Chapter 4 and Chapter 5) is integrated with the re-scheduling. The synergy of the rescheduling and re-assignment can optimally deploy the affected capacity to accomplish the transport tasks and minimize the operation costs.

The models are tested on a real application and are analyzed via corresponding key performance indicators to check the benefits of the synchronization of the flows. Based on these tests of synchronization control of the Beijing metro network (Chapter 3), we conclude that the wait decisions in this study can cause longer train delays but can decrease effectively the overall passenger delay costs. In the simultaneous traffic control for a Dutch railway network (Chapter 4), it is verified that the flexibly organized freight trains can decrease the freight delays, while not increasing the passenger delays. The replanning of the Rotterdam hinterland network shows that the shipment split and vehicle departure time rescheduling can decrease the operational costs and improve the utilization rate of large capacity vehicles.

Based on these studies, we recommend that the replanning should aim to maintain the service level even with the occurrence of the perturbations. Transportation is a service to the customers, either passenger or freight clients. This is reflected in the congestion management of the metro network (Chapter 3), the smooth transfer connections in the rail bound network (Chapter 3 and Chapter 4), and the on-time delivery of the freight at the destination stations or terminals in the multimodal network (Chapter 4 and Chapter 5). The service is managed via synchronizing the vehicle flows with the client flows. Explicitly, the departure times of metro vehicles are significantly advised to be synchronized with the massive passenger flows in a metropolis city.

The service level can be affected by the previously made agreements with the clients. The timetable is designed for the passengers according to which they arrange their trips. The passenger trains are therefore not rescheduled to depart earlier than the timetable. Meanwhile, the freight trains are not restricted with this constraint. For freight transport, there two possible situations, the shippers do not allow mode-shift, or they do not pose this restriction. In the former situation, the operations can be the rerouting of the service vehicles. In the latter situation, i.e. which is the principle of synchromodality, the freight themselves can be rerouted to other modes of vehicles. This creates extra flexibility and thus can lead to a higher punctuality rate.

Except for the service level, the operators also have other concerns. Examples include ensuring the safe movement of the vehicles, reducing the length of the perturbation recovery windows, and decreasing the operating costs. These are all operator-oriented and are achieved via the synchronization of all the vehicle flows. In the rail bounded network, the headway constraints are very important to guarantee conflict-free movements. In case there is an agreement about the authority of the organizer to the heterogeneous traffic, then the organizer has to balance the interest of the different stakeholders. In the synchromodality transport network, we allocate the shipments to different carriers according to the current situation.

SAMENVATTING

Transportnetwerken worden geëxploiteerd volgens een vooraf ontworpen plan om de transporttaken uit te voeren. Verstoringen kunnen onvermijdelijk exogeen of endogeen optreden en veroorzaken vertragingen in de doorreis van passagiers of de levering van vracht. Er is een overgangsplan nodig om de verstoringen het hoofd te bieden en de transporttaken te volbrengen. Het transitieplan wordt in deze studie *storingsmanagement* genoemd.

De eerste bijdrage van dit werk is de constructie van herplanningsmodellen voor netwerken met meerdere klassen. Het verstoringsbeheer voor het vervoersnet van één klasse, bijvoorbeeld een spoorwegnet voor reizigers, is goed bestudeerd, maar ontbreekt nog voor vervoersnetwerken met meerdere klassen. In praktische toepassingen bestaan passagiersstromen en vrachtstromen soms naast elkaar in één systeem. Het volume van vrachtvervoer blijft toenemen. Aan de andere kant zijn er meestal multi-mode services gekoppeld om de doorvoer van één object te voltooien. Het huidige onderzoek concentreert zich rond dit punt. We bestuderen storingsmanagement voor netwerken met meerdere klassen, waarbij de gemengde passagiers- en goederenspoorwegnetwerken betrokken zijn en waarbij verschillende vormen van dienstvoertuigen worden ingezet.

Er zijn een aantal belangrijke onderzoeksvragen met betrekking tot het beheer van heterogene verkeersstromen bij vertragingen. De eerste uitdaging is om een overzicht te krijgen van de focus van de bestaande literatuur in spoorgebonden netwerken en te achterhalen wat de veelbelovende richtingen zijn voor verder onderzoek. Ten tweede, wat betreft het modelleringsgedeelte, is het de uitdaging om een herplanningstrategie te bedenken en het controledoel te bepalen. Een andere taak is om erachter te komen hoe de verschillende betrokken stromen kunnen worden beheerd, hetzij voor de bediende klanten, hetzij voor de bedienende voertuigen. De laatste zorg is om te onderzoeken hoe flexibiliteit voor het geïntegreerde netwerk kan worden bereikt.

De tweede bijdrage van dit werk is dat de voorsgestelde herplanningsmodellen het heterogene verkeer rangschikken op basis van hun kenmerken. Wanneer er vertragingen optreden in een gemengd reizigers- en goederenspoorwegnet, geeft de herplanning meer flexibiliteit aan de goederentreinen. De onderliggende reden is dat de vrachtklanten de uiteindelijke levertijd belangrijk vinden en minder geven om de afgelegde route of de wachttijd op de tussenstations. In de herplanning via synchromodaliteit worden de drie servicemodi gepland volgens hun toegepaste subnetwerken. In het multimodale transportnetwerk worden de voertuigen met grote capaciteit afgestemd op de goederenstromen.

Een andere bijdrage van dit proefschrift is dat de modellen de synchronisatie van de betrokken stromen bewerkstelligen. Dit is essentieel voor storingsmanagement, ook al zijn de toegepaste netwerken verschillend. We streven naar de synchronisatie van stromen uit zowel tijd- als ruimtedimensies. We passen de planning van de metrovoertuigen aan in overeenstemming met de enorme passagiersstromen (Hoofdstuk 3). We stemmen de planning van drie modi van servicevoertuigen af op de vertraagde voorafgaande voertuigen of vrijgave van zendingen (Hoofdstuk 5). Ondertussen is het beheer van de passagiersoverdracht (Hoofdstuk 3 en Hoofdstuk 4) en de herverdeling van de goederenstromen (Hoofdstuk 4 en Hoofdstuk 5) is geïntegreerd met de herplanning. De synergie van de herschikking en hertoewijzing kan de getroffen capaciteit optimaal inzetten om de transporttaken uit te voeren en de operationele kosten te minimaliseren.

De modellen worden getest op een echte applicatie en worden geanalyseerd via overeenkomstige key performance indicators om de voordelen van de synchronisatie van de stromen te controleren. Op basis van deze tests van synchronisatiecontrole van het metronetwerk van Peking (Hoofdstuk 3), concluderen we dat de wachtbeslissingen in deze studie langere treinvertragingen kunnen veroorzaken, maar de totale vertragingskosten voor passagiers effectief kunnen verlagen. Bij de gelijktijdige verkeersleiding voor een Nederlands spoorwegnet (Hoofdstuk 4) wordt geconstateerd dat de flexibel georganiseerde goederentreinen de vrachtvertragingen kunnen verminderen, maar de reizigersvertragingen niet kunnen vergroten. De herplanning van het Rotterdamse achterlandnetwerk laat zien dat de verschuiving van zendingen en het herschikken van de vertrektijd van voertuigen de operationele kosten kan verlagen en de bezettingsgraad van voertuigen met een grote capaciteit kan verbeteren.

Op basis van deze onderzoeken bevelen we aan dat bij de herplanning wordt gestreefd naar het handhaven van het serviceniveau, zelfs wanneer de verstoringen optreden. Transport is een dienst aan de klanten, zowel passagiers- als vrachtklanten. Dit komt tot uiting in het congestiebeheer van het metronetwerk (Hoofdstuk 3), de vlotte overstapverbindingen in het spoorgebonden netwerk (Hoofdstuk 3 en Hoofdstuk 4), en de tijdige levering van de vracht op de bestemmingsstations of -terminals in het multimodale netwerk (Hoofdstuk 4 en Hoofdstuk 5). De service wordt beheerd door de voertuigstromen te synchroniseren met de klantstromen. Het wordt expliciet aangeraden om de vertrektijden van metrovoertuigen te synchroniseren met de enorme passagiersstromen in een metropoolstad.

Het serviceniveau kan worden beïnvloed door de eerder gemaakte afspraken met de klanten. De dienstregeling is ontworpen voor de passagiers op basis waarvan ze hun reizen organiseren. De reizigerstreinen worden daarom niet eerder ingepland dan de dienstregeling. Ondertussen zijn de goederentreinen niet beperkt door deze beperking. Voor goederenvervoer zijn er twee situaties mogelijk: de verladers staan mode-shift niet toe, of ze leggen deze beperking niet op. In de eerste situatie kunnen de operaties het omleiden van de servicewagens zijn. In de laatste situatie, d.w.z. het principe van synchromodaliteit, kan de vracht zelf worden omgeleid naar andere voertuigmodi. Dit zorgt voor extra flexibiliteit en kan dus leiden tot een hoger punctualiteitstarief.

Behalve het serviceniveau hebben de operators ook andere zorgen. Voorbeelden zijn onder meer het waarborgen van de veilige verplaatsing van de voertuigen, het verkorten van de lengte van de storingsherstelvensters en het verlagen van de bedrijfskosten. Deze zijn allemaal operatorgericht en worden bereikt door de synchronisatie van alle voertuigstromen. In het spoorgebonden netwerk zijn de doorvoerbeperkingen erg belangrijk om conflictvrij verkeer te garanderen. Indien er overeenstemming is over de bevoegdheid van de organisator voor het heterogene verkeer, dan moet de organisator de belangen van de verschillende belanghebbenden afwegen. In het transportnetwerk synchromodaliteit verdelen we de zendingen aan verschillende vervoerders volgens de huidige situatie.

PUBLICATIONS

ARTICLES

(1). W. Qu, J. Rezaei, Y. Maknoon, L. Tavasszy, *Hinterland freight transportation replanning model under the framework of synchromodality*, Transportation Research Part E: Logistics and Transportation Review 131, 308-328 (2019).

(2). W. Qu, J.J., Wu., F. Corman, *Transfer Synchronization Control with Over-saturated Flow on UrbanRailway Network*, to be submitted

(3). W. Qu, Y. Maknoon, F. Corman, *Real time railway traffic control with mixed passenger and freight traffic*, to be submitted

(4). W. Qu, F. Corman, G. Lodewijks, *The Real Time Traffic Control with Mixed Passenger and Freight Flows*, 7th International Conference on Railway Operations Modelling and Analysis (2017).

(5). W. Qu, F. Corman, and G. Lodewijks, *A review of real time railway traffic management during disturbances*, in Computational Logistics, Springer, pp. 658-672 (2015).

SELECTED TALKS

(1). W. Qu, J. Rezaei, L. Tavasszy, *Shippers Satisfaction in Perturbation Management of Hyperconnected Logisitic Service Network*, 8th International Physical Internet Conference, Online, June 14-16, 2021

(2). W. Qu, J. Rezaei, L. Tavasszy, *Hinterland freight transportation replanning model under the framework of synchromodality*, 6th International Physical Internet Conference, London, UK, July 9-11, 2019

(3). W. Qu, J. Rezaei, L. Tavasszy, Siga2 2018 Conference 'Maritime and Ports': The Port and Maritime Sector: Key Developments and Challenges - University of Antwerp, Antwerp, Belgium

(4). W. Qu, J. Rezaei, L. Tavasszy, EURO Mini Conference on "Advances in Freight Transportation and Logistics". Padova (Italy), March 7-9, 2018

ACKNOWLEDGMENTS

I would like to thank all the people who helped me during the past (many) years. Among all these I would like to mention a number of people who contributed to make this thesis possible.

First and foremost, I would like to express my tremendous gratitude to Prof. Lóránt Tavasszy. Your support and guidance are crucial for me to finish my research work and manage those perturbations happened in this process. You keep informing our team about the new developments, other researchers and new chances in the field. I really appreciate your trust and patience, especially during some difficult moments in the past years.

I want to express my sincere thanks to my copromotor Dr. Jafar Rezaei, not only because of the opportunity that you gave to me to continue my research at Delft University of Technology, but also for your supervision during the past years. You sharp mind, smart ideas and efficient working patterns provide valuable guidance to me. Our weekly meetings make me keep going forward.

I owe my great appreciation to Prof. Francesco Corman. Thank you very much for your supervision in the first years when I came to the Netherlands. You gave full support and tolerance in the beginning of my Ph.D. I can still recall those moments how patiently you listened to me, communicated with me, wrote comments on my drafts and even spent hours checking my programs. Even you moved to ETH Zurich, I felt that your help stayed with me in Delft. I truly appreciate the time you spent to help me. Thank you for the time and effort that you put into this thesis.

Dr. Yousef Maknoon has had a great impact on my research career. I remember those days in which we went through our draft paper rounds after rounds. Nowadays I always check my own writing using your criteria; following your words 'the writing should be *crystal clear*'. I cherish those difficult but motivating questions and moments I received during our meetings.

I would also like to thank my committee members, Prof. Eric Ballot, Prof. Rob Zuidwijk, Prof. Rudy Negenborn, and Prof. Rob Goverde, for their insightful, constructive and detailed comments and questions. Thank you Prof. Rob Goverde, for your time and energy on this thesis and thank you for your help ever since my Go/No-Go meeting.

I would like to thank my colleagues from T&L (and before MT&T) for generating a warm atmosphere. Fanchao, Ioanna, Baiba, Patrick, Masoud, Xiao (Liang), Qu, Shijie, Jialun, Xiao (Lin), Xiaojie, Peijuan, Fei, Pengling, Jie, Yazhou, Bing and Hong, thank you for nice time together (and for some checking my drafts and being the audience for the rehearsals). Conchita, thank you for your help especially during the production process of the hard copy thesis. I would also like to thank my friends for the being around. Nikos and Semina, thank you for those visits, talks and games during the lockdown. Victoria, Sebastian, Susie and Servais thank you for being my friends and encouraging me for so many years.

Marco, thank you for your continuous love and support. Thank you for being there and providing encouragement and taking care of me during those difficult moments in my life. Also thank you for proof reading some parts of the dissertation and designing the cover of this thesis. You and your family provide many joy and happy moments, which make me less homesick in the Netherlands. I am also proud to be there at your defence and happy to have you at mine.

Lastly, my deepest thanks go to my parents and brother. My parents are not brain workers, but you made the smartest decision of having me, and paying the price for violating the one child policy. Thank you Baba and Mama for always putting my brother and I first. With all your unconditional love and care, I will keep going.

Wenhua Qu
Delft, December 2020

CURRICULUM VITAE

Wenhua Qu

April 6, 1990	Born in Zhaoyuan, China.

Education

	Beijing Jiaotong University
2007-2011	Bachelor of Science in Traffic Engineering
20011-2014	Master of Science in Systems Engineering
	Delft University of Technology
2014-2016	Researcher at the Faculty of Mechanical, Maritime and Materials Engineering
2017-2020	Ph.D. Researcher at the Faculty of Technology, Policy and Management
2021-	Postdoctoral Researcher at the Faculty of Aerospace Engineering