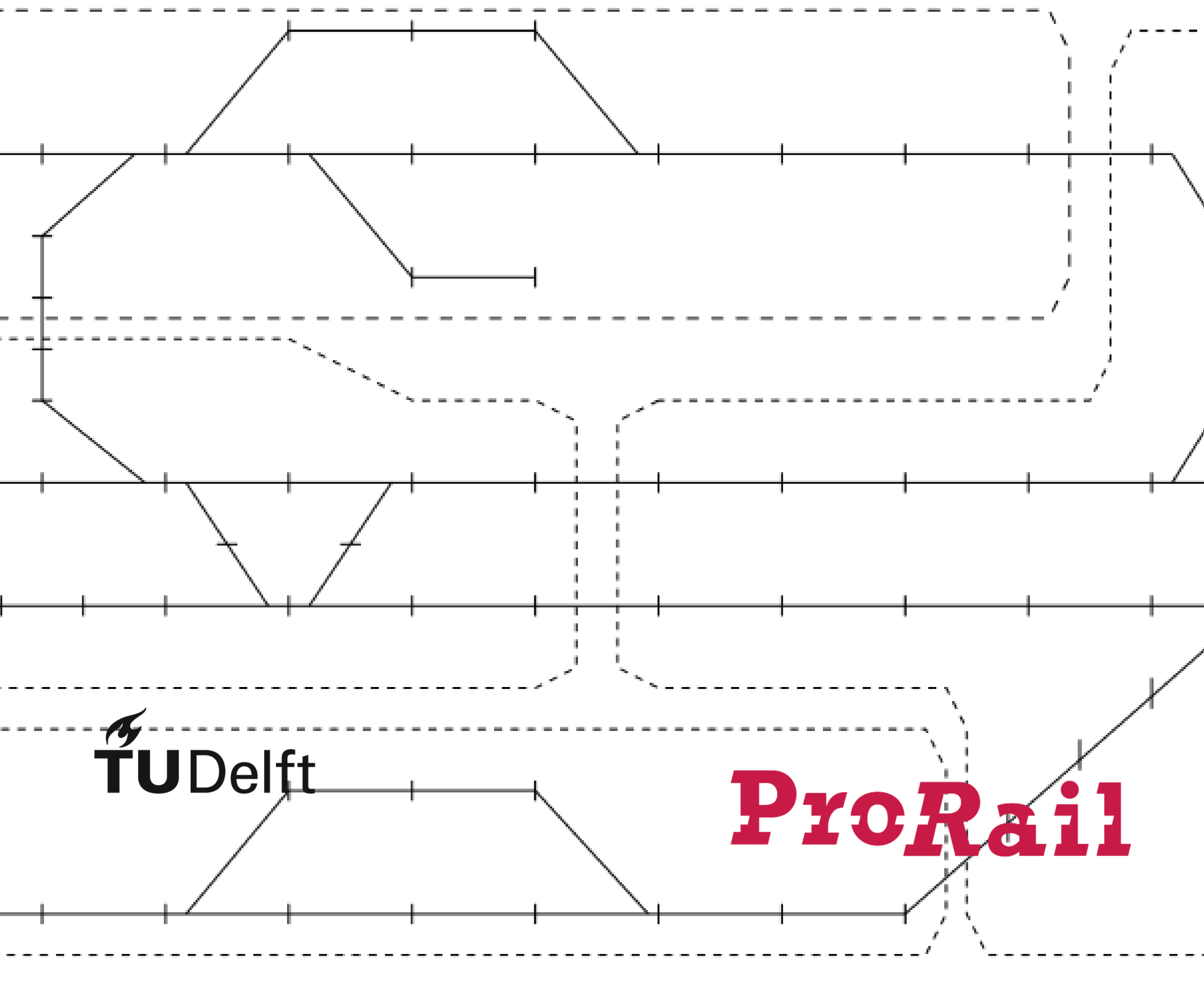
G.A.M. Hornung

Hourly Traffic Density Based Decomposition in Non-Centralised Railway Traffic Management

Master Thesis

2023.MME.8766



TU Delft

ProRail

Hourly Traffic Density Based Decomposition in Non-Centralised Railway Traffic Conflict Resolution

By

G.A.M. Hornung

in partial fulfilment of the requirements for the degree of

Master of Science
in Mechanical Engineering

at the Department Maritime and Transport Technology of Faculty Mechanical, Maritime and Materials Engineering of Delft University of Technology
to be defended publicly on Tuesday 21 February, 2023 at 13:00

Cover Photo [56]

Student number:	4443535	
MSc track:	Multi-Machine Engineering	
Report number:	2023.MME.8766	
Thesis committee:	Dr. V. Reppa,	TU Delft, chair & supervisor
	Dr. ir. E. Quaglietta	TU Delft, supervisor
	D. Middelkoop	ProRail, company supervisor
	Prof. dr. R.M.P Goverde	TU Delft, committee member
	Dr. B. Atasoy	TU Delft, committee member
Date:	6 February, 2023	

Preface

Dear reader,

In a few weeks I will have to say goodbye to being a student. Something that has been a large part of my identity for many years. Every single one of these years I have worked hard, but I have to say, this project is my personal academic magnum opus. This thesis has taken me quite some will power to finish and I am proud to have the 100 page document that lies before you to show for it.

But of course, there was more than only hard work. I decided to do my graduation project at ProRail. Two years ago, I had followed a course that had gotten me interested in the railway system and the immense logistical puzzle that it involves. The past year I learnt even more about all the aspects that are involved in getting the Dutch population from one place to another by the train. This knowledge fueled my own curiosity, but also allowed me to advise my Dad on making his model railway more realistic. The people at ProRail brightened my internship. Were it by interesting in-depth discussions or just pleasant small talk. Besides that it was also an honour to work in the beautiful and historic office, de Inktpot. The historic tour made the building seem almost magical. A few weeks ago, when I saw a bat flying through the hallway, of which thousands are estimated to be living in the building in the winter, my Inktpot experience was complete.

I have many people to thank for helping this thesis come about. Thank you Vasso, for being my supervisor since the start of my literature assignment over a year ago. You have always given me good feedback, helped me stay on track and supported me to reach my full potential. Thank you Egidio, for all your railway expertise, discussions and explanations and a special thanks for enriching my life with the expression "You've got to cook the meatball first, before you put it in the sauce". Thank you Dick, for showing me the ropes at ProRail, explaining all the ways the Dutch railway system functions and keeping me in touch with the practical applicability of the research. And of course thank you Rob and Bilge for being part of my thesis committee. Thank you Mathilde, for helping me find this project in the first place and taking the time to provide me with advice and documents even though you had already left your thesis behind you some time ago. Thank you to all the people at ProRail for providing me with your knowledge and making me feel welcome, especially Edith, Emdzad, Jelle and Dennis.
Believe it or not, there was also a life outside of my thesis. Sometimes I was in need of advice, other times in need of distraction and others just good companionship. Thank you Mom and Dad, for hearing my stories every week over the phone and for making the choice to cancel your great sabbatical to see your daughter graduate. Thank you Jasper, for supporting me every step of the way. You are the one who has heard the ups and downs every day of the week and has tirelessly believed in me and empowered me. Thank you Sophie, for being a good friend through times when I sometimes did not have the space to be a good friend to you. Thank you to Fiona, Emma, Myrthe, Loes, Britt, Kris and Koen for good conversations, distractions and kind words.

Lastly, I have thought of many train jokes while writing this thesis. I have permitted myself to leave one in there as an Easter egg and I hope it gives you a little laugh between all the serious matters.

Happy reading,

G.A.M. Hornung
6 February 2023

Executive Summary

The Dutch railway network is a dense network with 7000 kilometers of railway tracks spread over only 42 000 square kilometers [1, 2]. Additionally, the railways are responsible for 1.4 million passengers every working day [1]. These factors make both the infrastructure and timetable of the Dutch railway network extremely intricate. On top of that, the amount of train trips is expected to grow with 40 % by 2040 [3] and with that the complexity of managing the railway network will increase even more. In the case of conflict in the timetable, action must be taken to resolve this conflict in order for the train operations to continue. In current Dutch practice this task resides with the train dispatchers. However it can be difficult for the dispatcher to oversee the situation when multiple trains get involved and taken actions will have consequences on yet another selection of trains. This increasingly complex task has resulted in a grown interest in so-called *Traffic Management Systems (TMS)*, which are an intelligent systems that use a conflict resolution algorithm to find a solution to a conflict. A TMS can be used to support the train dispatcher.

The development of Traffic Management Systems has been a research subject for many years. Initially these approaches were mainly *centralised*, meaning that one single decision system, or *agent*, will find a solution to the whole region. For a large or complex situation the railway traffic management problem can be challenging and time consuming, if done by one single agent [4–6]. That is why research has shown increased interest in *non-centralised* approaches. In these approaches there are multiple agents that are each responsible for their own subsystem of the region, for example a subregion. Using a non-centralised approach has several advantages: increasing the flexibility within the finding of a solution, easing the introduction of additional factors to the problem and reducing computation time. A challenge is that the subsystems need to be coordinated such that all the sub solutions can form an overall feasible or even optimised solution. One of the aspects of a non-centralised railway traffic management system is the *decomposition*: how the system is divided into subsystems. Only a few publications have focused on the decomposition approaches, e.g. Lamorgese and Mannino (2015) [7] and Kersbergen et al. (2016) [8]. Within these papers the two main components of the railway system, the infrastructure and the train traffic, are not equally taken into account. Moreover, no research could be found that studies what the individual impact of a decomposition approach can be on the performance of the non-centralised conflict resolution.
Non-centralised approaches have also been regarded in practice. In the current state of practice, the network is divided into regions to form three levels. The national control center views the entire Dutch network, 12 traffic control centers are responsible for their own region and those regions are subdivided into 4 to 12 subregions, or dispatcher posts. The state of practice of how a region is subdivided into subregions is essentially also a decomposition. This state of practice decomposition has come about historically and organically and has not yet been scientifically founded.

Following from these challenges, the objective of this research is to develop a decomposition approach in non-centralised railway traffic management that incorporates both aspects of the railway system, the infrastructure and the train traffic. Additionally this research aims to provide insight into what the impact of a decomposition approach can be on the non-centralised conflict resolution.This leads to the research question of this thesis:

What is the impact of an hourly traffic density based decomposition approach on non-centralised railway traffic management?

This thesis develops a decomposition that results in sub areas, which takes the form of an optimisation model that assigns parts of the infrastructure to certain subregions, drawing inspiration from the geographical decomposition presented in Luan et al. (2020) [9]. The aim of developing a decomposition approach is to equally divide the *hourly traffic density* among the agents, to improve computation, performance and flexibility. The main criterion for the decomposition is defined as the hourly traffic density in trains per track per hour.

This combines both the infrastructure and the train traffic travelling over it. Also, it describes the ratio between supply and demand: it describes to what extent the supply of infrastructure can satisfy the demand of the train transport operators. It is calculated by dividing the traffic flow in the area by the amount of tracks the area has. The other criterion in the decomposition is the subregion crossings, which dictates that the amount of trains crossing over a subregion border in their route should be minimal. Contrary to previously presented microscopic decomposition models, this approach will take a macroscopic approach. As infrastructural elements, this model defines 'dienstregelpunten' (timetabling points). These are points in the infrastructure that are recognised to need scheduling actions by network administrator ProRail. Other inputs to the model include: the routes of the trains within the region; the frequency of trains travelling the aforementioned routes and how many subregions the region should be divided into.

The aspects that make up the non-centralised railway traffic management model are the coordination that ensures communication between the agents and the conflict resolution happening inside the agent. For the conflict resolution approach, this research uses the Alternative Graph formulation as presented in D'Ariano et al. (2007) [10]. This method models the conflict resolution using graph theory. The nodes represent operations of a certain train at a certain time and the arcs represent the succession of events. The model also contains alternative arcs that show the precedence between two trains. The arcs come in alternative arc pairs and in the solution only one of those arcs can be chosen. By changing the precedence of trains and/or the time of certain operations, the Alternative Graph finds a solution to the conflict problem. This solution will be chosen based on how the maximum secondary delay can be minimised. As a coordination method, it is chosen to use boundary constraints. These are coupling constraints that couple the subregions by constraining a certain characteristic of a train crossing a subregion border to ensure continuity. This research proposes a delay transition constraint, which entails that the delay of a train is coordinated between agents to be the same at both ends of a subregion border. To introduce delay-driven communication a coordinator is added as an additional entity. The coordinator is placed in a higher layer than the conflict resolution agents, which allows it to oversee the global conflict resolution task. It has three main tasks: (1) Communicating initial delay to the relevant agent(s); (2) Assigning the conflict resolution task to the agents and (3) Prompting communication between agents, or not.

The hourly traffic density based decomposition approach is validated by applying it to a case study within the Dutch network. The proposed decomposition is compared to the baseline decomposition: the train dispatcher subregions currently used at ProRail. This comparison is made both on the level of the decompositions as the impact they have on the conflict resolution problem. The test case is chosen as the network between Roosendaal and Lage Zwaluwe, and Den Bosch and Liempde. For the timetable a 15 minute interval is used of a basic timetable for Tuesday in 2022 by ProRail and 17 delay scenarios are constructed by delaying a train at the entrypoints of the network. By analysing these measurements in a series of graphs several conclusions are drawn. The trend in the computation time showed that the proposed decomposition allows up to *20 % faster computation* of the conflict resolution problem. This suggests that the proposed decomposition is able to more equally distribute the decisional complexity of the conflict resolution problem over the subregions than the baseline decomposition. In the interactions between agents no real distinction could be found between using the proposed decomposition and the baseline decomposition. The relative total output delay shows that the proposed decomposition allows the conflict resolution to *mitigate on average 5 % more delay.* The maximum and minimum delays shows that *the difference between minimum and maximum is 21 % smaller on average and up to 90 % smaller* when using the proposed decomposition. Because the maximum values are identical for both decompositions, this shows the proposed decomposition allows the conflict resolution to mitigate the smaller delays. Additionally, the proposed decomposition is able to allow the conflict resolution to *reduce the number of delayed trains with 17 %.*

From this it can be concluded that the proposed decomposition can distribute decisional complexity evenly across subregions and thereby reduce the decisional complexity and workload of a single subregion, providing more effective and time-efficient rescheduling solutions. This shows great promise for further research and the state of practice. The decomposition approach could be used by ProRail to reflect on the decomposition that is currently used. For further research, the reduction of decisional complexity means that additional constraints can be added to expand the viewed problem to consider for example local transport or passenger satisfaction. The reduction of delay means that general performance of non-centralised approaches could be improved by paying more attention to the decomposition. There are three main recommendations for future research. Firstly, it would be interesting to include the amount of subregions the region is to be divided into in the optimisation, as to determine the optimal number of subregions for that specific region. Secondly, it is recommended that future research looks into expansions and variations of decomposition criteria to explore different ways that the complexity of the conflict resolution problem can be distributed over subregions. Lastly, it is advised that in future research the decomposition is tested with more advanced forms of the non-centralised railway traffic management approach.

Contents

1. Introduction

Within the Dutch railway network, a carefully established timetable is constructed to ensure operations are as smooth and safe as possible. Due to variations in traffic, unforeseen circumstances and uncertainties, a train can build up a delay. A delay can be of any source and size, for example it can be a single minute as a result of a busy platform during rush hour or an hour as a result of a sudden infrastructure defect. Any delay challenges the margins incorporated in the timetable and if a delay exceeds its specific margin then this may cause a **conflict** [11]. It may occur that two trains aim to be at the same place at the same time. Such a conflict imposes an impasse on the train operations: the conflict must be resolved in order for train operation to resume. In current practice, the resolving of the conflict, also known as **conflict resolution**, is done by **train dispatchers** [12], as shown in Figure 1.1. These train dispatchers take the relevant action or actions to mitigate a conflict, for example a train can be asked to wait for another train to pass first. A delay in the complex timetable and infrastructure of the Dutch railroad network, can lead to complex puzzles for the train dispatchers to solve. A solution to the puzzle is found based on the train dispatcher's own insights, their knowledge from experience and a set of preset action plans that contain the recommended actions for specific delay situations [12]. It can be imagined that it is a demanding task to keep the Dutch train operation running 24/7. This task is under the responsibility of the company ProRail. They are responsible for the maintenance, renewal, expansion and safety of the railroad network and one of their core activities is the managing of the railway traffic in the Netherlands.

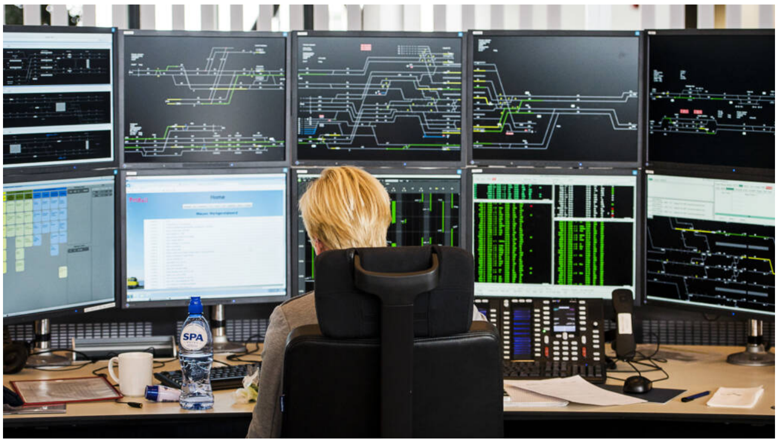


Figure 1.1: A ProRail train dispatcher [13]

The Dutch railway network is a dense network with 7000 kilometers of railway tracks spread over only 42 000 square kilometers [1, 2], see Figure 1.2. Additionally, the railway transport is responsible for 1.4 million train trips every working day [1]. These factors make both the infrastructure and timetable of the Dutch railway network extremely intricate. On top of that, the amount of train trips is expected to grow 40 % by 2040 [3]. Figure 1.3 shows the expected growth of trains trips on the respective parts of the Dutch network by means of the blue bars. With such an increase in the amount of train trips, the complexity of managing the railway network will increase even more.

Despite the preset action plans available to the train dispatcher, in the case of conflict which action(s) will lead to the least delay and passenger discomfort can be become a large puzzle. It can be difficult to oversee the situation when multiple trains get involved and taken actions will have consequences on yet another selection of trains. Besides that it is a difficult and demanding task to manage these conflicts, a train dispatcher will not always be able to take all factors into account when finding a solution, which means the solution will not always be the most effective or efficient one. In these cases a **Traffic Management System (TMS)** can support the train dispatcher. A Traffic Management System is an intelligent system that is able to oversee the traffic and the network and can provide multiple and/or optimised solutions in the case of conflicts [11]. This can serve as a decision support system to the train dispatcher by providing them with possible solutions to a conflict. The train dispatcher will then be able to combine these solutions with their expert knowledge to come to a computer-aided decision.

Figure 1.2: Train services in a part of the Dutch railway network [14]

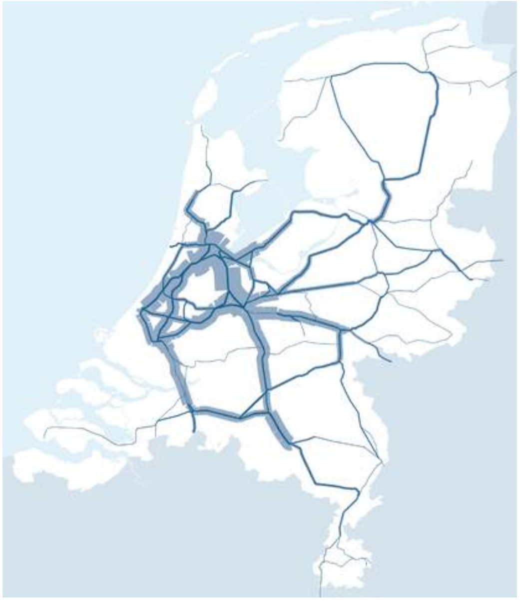


Figure 1.3: Relative expected growth in train trips per route [3]

The development of Traffic Management Systems has been a research subject for many years. Initially these approaches were mainly **centralised**, meaning that one single decison system will find a solution to the whole network. With this approach, finding a decision for a large or complex situation can be challenging and time consuming, especially because the railway traffic management problem is already a problem with many different factors [4–6]. That is why research has shown increased interest in **non-centralised** approaches. In these approaches there are multiple decision systems that are each responsible for their own subsystem of the region, which can be a.o. a selection of trains, a route or a subregion.

One of the aspects of a non-centralised railway traffic management system is the **decomposition**: how the railway system, consisting of the infrastructure and the traffic travelling on it, is divided into subsystems. These subsystems can for example be a single train, a time interval or a subregion. Within the field of non-centralised approaches only a few publications have focused on the decomposition approaches, e.g. Lamorgese and Mannino (2015) [7] and Kersbergen et al. (2016) [8]. The railway system consists of two main components, the infrastructure and the train traffic, and within the few papers on decomposition approaches these two are not equally taken into account. As such, only a part of the railway system is taken into account in the decomposition. Moreover, no research could be found that studies what the individual impact of a decomposition approach can be on the efficiency and effectiveness of the conflict resolution. For example, Luan et al. (2018) [15] studies the presented decomposition approaches in combination with solution approaches, but does present the impact of solely the decomposition approaches.

Non-centralised approaches have not only been regarded in literature, but also in practice. In the current state of practice, the network is divided into regions to form three levels. The national control center views the entire Dutch network, 12 traffic control centers are responsible for their own region and those regions are subdivided into 4 to 12 subregions, or dispatcher posts. The state of practice of how a region is subdivided into subregions is essentially also a decomposition. This state of practice decomposition has come about historically and organically and has not yet been scientifically founded.

1.1 Research contribution

All of the above leads to the contribution of this thesis. The objective of this research is to develop a decomposition approach in non-centralised railway traffic management that incorporates both aspects of the railway system: the infrastructure and the train traffic. Additionally this research aims to provide insight into what the impact of a decomposition approach can be on the effectiveness and efficiency of the non-centralised conflict resolution.

This leads to the research question of this thesis:

What is the impact of an hourly traffic density based decomposition approach on non-centralised railway traffic management?

The main research question is answered by addressing the subquestions, which in turn are answered by following their respective action points.

1. *What is the current state of the art of centralised and non-centralised conflict resolution?*
   (a) Identify the current state of practice in railway traffic management
   (b) Research the centralised railway traffic management approaches in literature
   (c) Perform a literature review on non-centralised railway traffic management approaches with a focus on the decomposition approaches
   (d) Research the Key Performance Indicators used to assess non-centralised railway traffic management approaches
2. *How can a region of railway infrastructure be decomposed into subregions based on hourly traffic density?*
   (a) Identify the aspects and characteristics of decomposition models in literature
   (b) Formulate the criteria of an hourly traffic density based decomposition
   (c) Formulate a model for an hourly traffic density based decomposition
   (d) Verify the proposed decomposition approach
3. *How can a non-centralised conflict resolution model be designed?*
   (a) Formulate an approach for conflict resolution in non-centralised railway traffic management
   (b) Formulate a coordination scheme for non-centralised railway traffic management
   (c) Verify both the conflict resolution approach and the coordination scheme
4. *How can the impact of a decomposition approach on non-centralised railway traffic management be evaluated?*
   (a) Find a suitable case study to demonstrate the decomposition and railway traffic management
   (b) Apply the decomposition approach to the case study
   (c) Identify the baseline decomposition for the case study
   (d) Apply the non-centralised railway traffic management approach to both decompositions
   (e) Compare results of the conflict resolution using both decomposition for the chosen KPIs

1.2 Structure

This thesis answers the above research questions in the remaining five chapters. First the rest of this chapter discusses some background to non-centralised railway traffic management. Chapter 2 evaluates the state of the art on centralised and non-centralised conflict resolution approaches. Also this chapter will asses the current state of practice and the Key Performances Indicators used in literature. The decomposition model is discussed in Chapter 3. The other two elements of the non-centralised approach, the coordination and conflict resolution, are discussed in Chapter 4. The models established in the previous two chapters are implemented for a case study and experiments are done to asses the performance. This is presented in Chapter 5. Lastly in Chapter 6 a conclusion is drawn and recommendations for future research are given.

1.3 Background

This section discusses elements from railway traffic management, and specifically non-centralised approaches, that build the foundation for understanding the theory in the rest of the thesis.

1.3.1 Railway traffic management

In the Netherlands, rolling stock and train crew are managed largely by the Nederlandse Spoorwegen (NS), but also by smaller traffic operating companies such as Arriva and freight companies. ProRail is, amongst many other responsibilities, responsible for the managing of the railway traffic in the form of day-to-day operation and ensuring safety on the railways.
To ensure safety, the railway network is equipped with a safety system with three main functions [16]. The most relevant one for this research is the function which safeguards that no two trains can be in the same section at the same time. The current safety system ensures this via a lineside signalling system. The track is divided into sections called **block sections**. These block sections are used to maintain safe distances between trains to avoid any case of collision or other safety hazards. In practice, lineside signals that resemble traffic lights are used to ensure this safety. As illustrated in Figure 1.4, lineside signals can show one of three lights:

- **Green** shows that the section up ahead is clear and that the train can pass safely.
- **Red** means that the current train can not enter it. One the reasons for a red sign can be that there is already a train in the section ahead.
- **Yellow** is shown in the section before a red light. This means the train driver should slow down in order to be able to stop for that next red sign.

This form of signalling, where a safe distance is maintained using fixed section of the tracks, is called **fixed block signalling**.

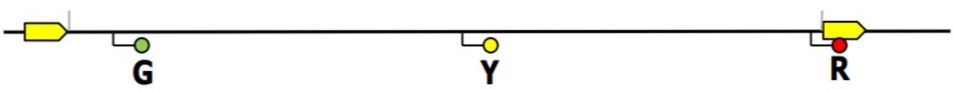


Figure 1.4: Lineside signalling [17]

In the case of a **conflict**, a train may already be occupying sections ahead, where another train is scheduled to be at that point. As a result, the first train cannot continue its schedule while maintaining a safe headway with the second train and thus will face a red sign. There are many situations that can cause these conflicts. The process of identifying and predicting them is called **conflict detection**. Within this thesis this process is left out of scope, or in other words, the conflicts are viewed as already detected. The next step is resolving the conflict to make the train traffic run as smooth as possible. This conflict resolution can take the form of a series of actions:

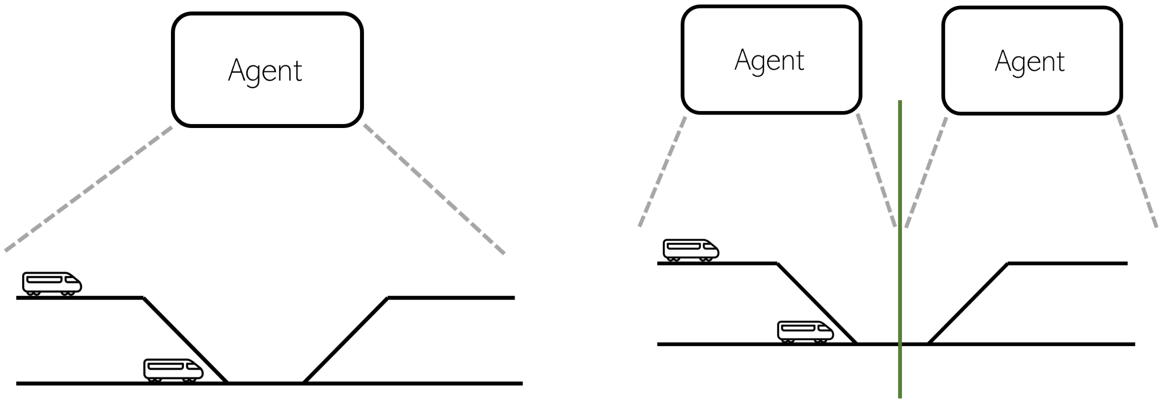
- **Reordering**: The order of the trains is changed. For instance, where a certain train A was scheduled to precede a certain train B, it may now be chosen that train B will precede train A.
- **Retiming**: The departure time, arrival time or passage time of a train is changed. This can be done by letting a train wait at a station or at a signal or by communicating the dynamic timetable changes to the train driver so they can react accordingly.
- **Rerouting**: The route of a train is changed. A train may be rescheduled to travel over a different track or to arrive at a different platform at the station.
- **Short turning**: The train will turn around sooner than it was initially scheduled, or in other words, a part of the route is cancelled. For example, if a train was initially scheduled to travel from Eindhoven to Rotterdam and back but a conflict has formed between Breda and Rotterdam, it may be chosen to let the train travel only until Breda and from there return its way back to Eindhoven.
- **Cancelling**: In this case the entire service of the train is cancelled and the train will not travel its route.

The decision of which action to take to which extent, whether made by an agent or the train dispatcher, is known as **railway traffic management (RTM)**. Many TMS are intended for **real-time RTM**, meaning the railway traffic is managed in real time, reacting and resolving situations as soon as they occur in the real-time process. Other terms to describe this decision making are **conflict resolution** and **rescheduling**. Railway traffic management can refer to the more general task of managing the railway traffic, whereas conflict resolution and rescheduling specifically refer to taking (a selection of) the above mentioned actions.

1.3.2 Agents in Traffic Management Systems

A TMS can consist of one or multiple **agents**: a (sub)system that can do computations and send and receive information to and from other agents [18]. In the case that the TMS consists of multiple agents, the **system architecture** describes how these agents communicate and relate to each other. For example, a system architecture can describe a hierarchy and/or an exchange of information.

In this research, two main system architecture types are distinguished: centralised and non-centralised, see Figure 1.5. In a **centralised** architecture, one central agent is responsible for the whole network [18], Figure 1.5a. A **non-centralised** architecture uses multiple agents [18], Figure 1.5b. The network is divided into subsystems and each agent is responsible for such a subsystem. The agents can then optionally have a form of communication with each other. The communication between the agents defines the type of non-centralised architecture. These types will be elaborated at a later point in this thesis.



(a) A centralised architecture

(b) A non-centralised architecture

Figure 1.5: Two types of system architectures

A TMS can be described as a system in a **cyber layer** interacting with a system in a **physical layer** [18]. The cyber layer contains the computation, such as a TMS, and the physical system contains the railway system with the infrastructure and the train traffic. Although the physical system can be the actual real-time train traffic, it is often a simulated physical system. In Figure 1.5, the agents are in the cyber layer and the trains and infrastructure make up the physical layer. Often these layers communicate with each other: the physical layer provides measurements to the cyber layer, such as delays of trains, and the cyber layer communicates actions to the physical layer, such as a reordering action.

A distinction can be made between railway traffic management and **railway traffic control**. Railway traffic management is mainly used as an advisory system. This means that the TMS will compute a solution to the conflict using its own simulated physical system, but the train dispatcher will still be in charge of taking the appropriate action(s) in the real-life physical system. In railway traffic control the TMS is designed to interact directly with the real-life physical system, without interference of the train dispatcher [18]. This research focuses on advisory railway traffic management TMS.

1.3.3 A non-centralised approach: (dis)advantages and components

Research has initially focused on the development of centralised TMS. This is an approach that can be fit for calculating smaller problems, for example small networks. However for larger and more complex situations the problem can become cumbersome, such that it becomes interesting to look into non-centralised approaches. Possible benefits of a non-centralised approach are:

+ **Decisional complexity** The problem viewed by a single agents is considerably smaller than in a centralised system. Because there are simply less factors to be taken into account, the complexity of the problem also decreases. This gives the agent more flexibility in finding solutions to conflicts.

+ **Computation time** Because in a non-centralised approach a single agent manages a smaller, and thus less complex, part of the network, the computation time will likely decrease with respect to a centralised approach. A shorter computation time means a faster response time, which could potentially make a non-centralised approach fit for real-time implementation.

+ **Flexibility** In a non-centralised architecture an agent encompasses a relatively small and local part of the network. This has a positive impact on the flexibility of the optimisation. If (small) changes are made to the infrastructure or timetable, in routine updates or maybe even real-time, this needs to be updated in the model of the respective agent. In a centralised approach a likely time-consuming global update would be required to update the entire model.

+ **Scalability** A non-centralised approach is more scalable, in the sense that the viewed problem can easily be expanded. This can mean expansion of the global problem, such that the region can easily be expanded by adding an extra agent to the system. Or, because agents in a non-centralised approach consider smaller and more local problems, this can mean the expansion of the problem that is viewed by a single agent. For example, local factors can be taken into account such as common train connections or intermodal connections, e.g. busses and trams.

However a non-centralised approach can also have several challenges:

- **Coordination** Every agents solves its individual local problem and all these local solutions of the agents must be combined to form a global solution to the whole problem. In order to do that it should be coordinated what happens at the boundaries of the local problems. Depending on the communication architecture this can range from a shared constraint to continuous updates to and from agents.

- **Consistency** Parts of the network are now managed by individual agents, which means that it can be difficult to maintain a consistent strategy throughout the entire network. For example, one agent may choose to prioritise one train, where the next agent may decide to deprioritise it as soon as it enters its subsystem.

- **Global feasibility/optimality** As a combination of the two above points, global feasibility and global optimality form a challenge within a non-centralised architecture. Each subsystem will manage up to its own optimality and inherently its own feasibility. This does not mean that this particular solution is globally feasible for the whole network, let alone globally optimal. This is a challenge that needs to be taken into account in the design of a non-centralised conflict resolution model.

A non-centralised approach to conflict resolution is in this thesis identified to consist of three main components:

- Recalling the explanations in subsection 1.3.2, the **decomposition** is how the physical system is subdivided into subsystems. Each of these subsystems will then be optimised by one agent. A subsystem can be any part of the railway system that has been decided upon as the end result of the decomposition, e.g. individual trains, a time interval or a subregion. The reason why one part of the system will belong to one subsystem and another part of the system will belong to another subsystem, is called the **criterion**. In other words, the criterion is the reasoning behind the decomposition.

- When the physical system is decomposed into agents, in the cyber system it can be evaluated how the agents communicate with each other. This is the **coordination**. The coordination establishes if, how and what the different agents will communicate with each other in order to work toward consistency, feasibility and optimality.

- It should also be determined what each agent will do. In this case each agent will be responsible for the conflict resolution in its designated part of the physical system. There are different types of conflict resolution approaches. These will be elaborated later on.

2. The State of the Art in Centralised and Non-Centralised Railway Traffic Management

The field of railway traffic management has been a lively field for several decades now. With new advancements in technology following each other rapidly, the possibilities in the field are constantly expanded. From this, new publications follow every year.
This chapter analyses the development in the field of railway traffic management, to support the rest of this study. First it starts with an analysis of the current state of practice of ProRail in real-time railway traffic management, section 2.1. It discusses the starting point of the railway traffic management literature, the centralised railway traffic management approaches, section 2.2 . Then, the non-centralised approaches are discussed, section 2.3. Concludingly, it is reviewed which Key Performance Indicators have been used in literature to evaluate non-centralised railway traffic management approaches, section 2.4.

In the literature analysis several scoping choices are made to ensure that the publications that are discussed are limited to those relevant to the research:

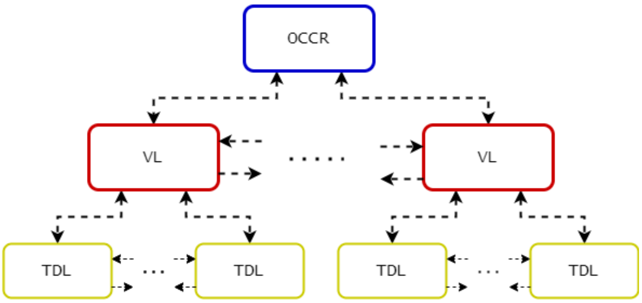
- The choice is made to limit the search to papers that have railway traffic management as their main focus or one of their main focuses. In other words, their focus should in line with the focus of this research. That means, for example, that papers focused on energy management within railway traffic are left out of scope, such as Khayyam et al.(2016) [19], Gu et al. (2013) [20] and Jiang et al. (2021) [21].
- This research investigates conflict resolution and thus papers discussing adjacent forms of railway traffic management are left out of scope for the literature review. The topics that have been excluded are, for example, conflict detection, scheduling and railway traffic control. These forms are steps that precede the conflict resolution and/or do not fit the real-time advisory TMS that is studied.
- This research focuses on fixed block signalling systems, which is the signalling system currently used. In the future a new safety system is scheduled to be introduced (ERTMS), which will use another form of signalling called moving block signalling. As a result, a research field has emerged studying how moving block signalling can be captured in TMS. Because this is still a growing field with a large focus only on the signalling system and centralised approaches, the publications in this field do not have a relevant contribution to this literature review focused on non-centralised approaches.
- Research in railway traffic management goes all the way back to 1994, as seen in Fang et al. (2015) [22]. However with the many developments in hardware and software since the 1990s, the older methods can not trustfully be compared to the newer methods [23]. That is why the literature review by Cacchiani et al. in [23] chooses to address only the publications from 2006 or later. This choice can be supported by the fact that the approaches before 2006 have seldom been cited in the recent literature. This shows that these older publications have become outdated and have little relevance in recent approaches. For these reasons, this review has confined the search to papers from 2006 and after.

2.1 State of practice of railway traffic management at ProRail

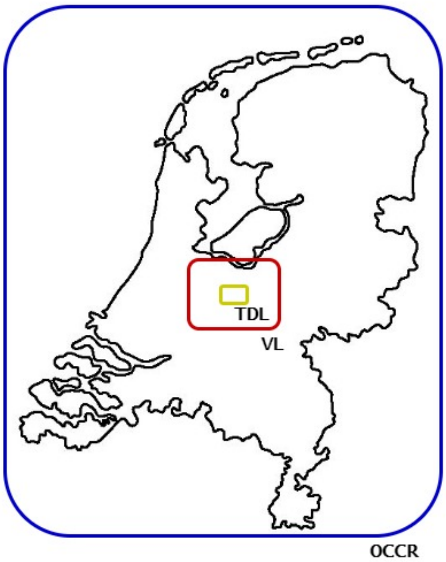
To be able to understand the theoretical side of railway traffic management, it is important to also be aware of the current state of practice in the field. This chapter discusses how the Dutch railway traffic is currently being managed by ProRail.

The responsibilities of the national railway traffic management are divided between three different entities. All three differ on the level where they operate, both from a geographical and operational standpoint as depicted in Figure 2.1. The highest level entity is the 'Operationeel Controle Centrum Rail' (Operational Control Centre Rail), or **OCCR** for short. The OCCR is responsible for coordination of the management of disruptions on a national level. They view the entire Dutch railway network and look at the traffic from a tactical level. Their job is solving initial nation-wide disruptions and keeping an eye on potential nation-wide consequences of local disruptions [24]. An example of when the OCCR came into play was when large portions of the train personnel were going on strike, which meant train traffic was not possible in large parts of the country [25]. At this time the OCCR was responsible for keeping a nation-wide view on the consequences of the strike and how regular traffic could be resumed as much as possible.

The railway traffic is managed at an operational level by the **'Verkeersleider' (VL)** and **'Treindienstleider' (TDL)** , in this report also referred to as the **traffic controller** and **train dispatcher** respectively. The traffic controllers are responsible for the railway traffic management on a regional level. Their scope is limited to their respective region and within this region they make the high level traffic management decisions. That means that they can take short turning and cancelling actions, as described in subsection 1.3.1. Also they can take global rerouting decisions, such as letting the high speed train use the regular infrastructure instead of using the separate high speed infrastructure. Corresponding to the level of the actions the traffic controllers use aggregated, general information, a.o. the position of a train between two stations and the arrival and departure times of a train. The train dispatchers are concerned with the more detailed operations. Their scope is limited to a subregion within the traffic controller region. During scheduled operations, they are responsible for route setting and enabling and ensuring safe train movements. In the case of conflicts, they are responsible for taking actions such as retiming, reordering and local route changes. Consequently, the information they manage is very detailed. They view all separate tracks, sections and switches of the infrastructure and the infrastructure occupation of each individual train that travels it. [11, 12]
This organisational structure ensures that the railway traffic can be kept up and running and safe 24/7.



(a) The hierarchy and communication of the entities



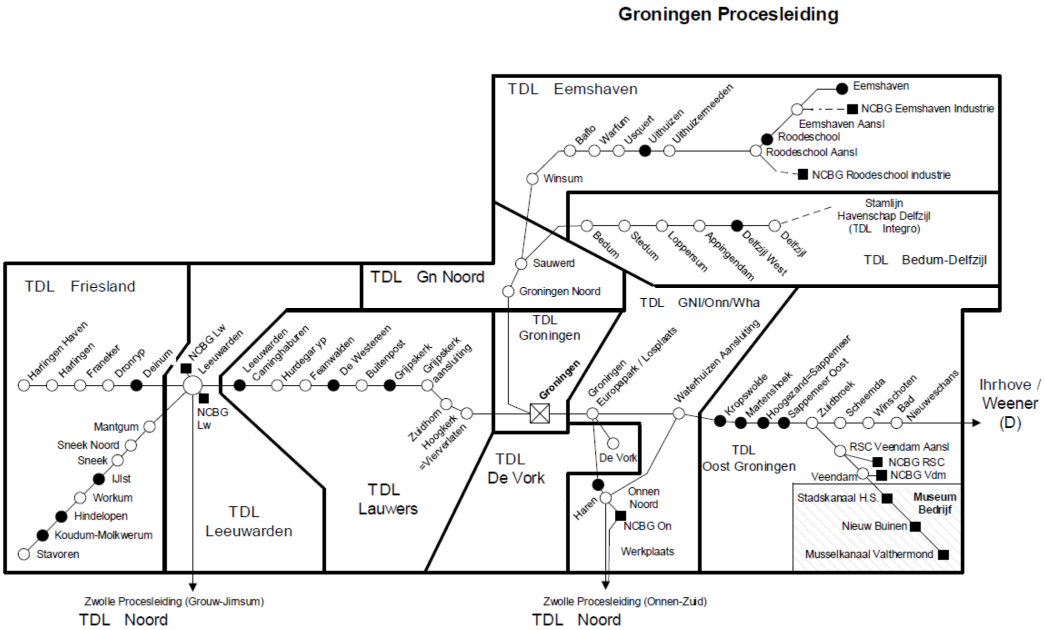
(b) A representation of the management regions of the entities

Figure 2.1: The entities involved in railway traffic management at ProRail

As shown in Figure 2.1b, the traffic controllers and train dispatchers are responsible for certain regions and subregions. The Dutch railway network is divided into 12 traffic controller regions, see Figure 2.2a, each built around a key point in the network [13]. This is often a station that forms an important part of the network. The 12 regions each contain 3 to 12 subregions [26]. Figure 2.2b shows how the Groningen traffic controller post is divided into 10 subregions. From another perspective this means that the Groningen traffic controller post has 10 train dispatchers present at all times, each to manage the traffic in one of those subregions. Essentially, this structure means that the current state of practice uses a non-centralised architecture.

This means the state of practice also has a decomposition, which is inherent to a non-centralised architecture. This determination of what infrastructure lies within those regions and subregions has come about historically and organically through the years. It ranges back to when different parts of the infrastructure used to have different types of safety systems and the (sub)regions from this division have been split and aggregated multiple times throughout the past to come to their current form [27]. In other words, the division has not yet been studied or scientifically founded as it is used in current practice. However, in the future a project has been scheduled at ProRail to start studying if the subregions can possibly be distributed more efficiently with an eye on improving the workload of the train dispatchers [27].

(a) The Dutch traffic controller posts [13]



(b) The Groningen post divided into train dispatcher regions [26]

Figure 2.2: Decomposition of Dutch railway traffic management

As can be seen in Figure 2.2b, infrastructure and thus trains cross the subregion borders. As a result conflicts, disruptions or measures taken will likely have consequences in neighbouring subregions. Or on the level of the traffic controllers, the cancelling or short-turning of a train will possibly have implications that need to be known by neighbouring regions. So it is inevitable that lines of communication are established between the different traffic controllers and train dispatchers. In both cases this communication is very much still informal and people oriented. With the help of an internal phonelist, the relevant traffic controller or train dispatcher can be contacted by telephone to update them on the current situation and possible future situations [26]. Contact between train dispatchers can take on a more informal form, because all train dispatchers of one train dispatcher region are stationed in a single room. The people managing neighbouring train dispatcher subregions will likely also literally have neighbouring desks. The communication between the subregions can thus be a simple nudge and conversation with a nearby co-worker [12]. Communication between either traffic controllers or train dispatchers is initiated because a situation or a measure in one (sub)region will have consequences for another (sub)region. In more technical terms, this communication can be described as **bilateral** and **event-driven**.

Besides communication between entities at the same level, there is also communication between the entities at different levels, i.e. between the OCCR and the traffic controllers, and the traffic controller and the train dispatchers. This can go in both directions: the lower level entities can update the higher level entities on the details taking place at the current moment or the higher level entities can inform the lower level entities on the high level decisions made. Again, communication will be initiated by an event. For example, during a large scale disruption the OCCR may communicate a mitigation strategy to the traffic controllers to be implemented on a regional level [28] or a train dispatcher may inform the traffic controller of a train that has departed considerably later than scheduled which might have large consequences [12]. Communication between the OCCR and the traffic controllers will take place mostly via phone conversations or internal systems [28], but the communication between the traffic controllers and train dispatchers is often more informal. The traffic controller is stationed in the same room as the train dispatchers of their region. This means that information can be exchanged simply by shouting something across the room. Especially in high pressure situations, this will be the main form of communication [12]. Summarising the communication between different levels, it can be interpreted again as **bilateral** and **event-driven** communication.
The communication is also reflected in Figure 2.1a, where the arrows indicate the direction of communication and the dotted lines indicate that the communication is event-driven.

As mentioned before, the train dispatchers take retiming, reordering and local rerouting actions. The decision on which action to take can be a difficult one to oversee. In order to aid the train dispatchers in making the decision that mitigates delay and other inconveniences as much as possible, several protocols are in place. In a document called 'Treindienst Afhandelings Document' or TAD, several delay scenarios are described including the desirable actions to take [29]. These scenarios encompass specific delays for specific, individual trains. Actions may be, for example, an order change or that another train will have to wait for connecting passengers. However the document is not able to capture all possible delay scenarios as there are many different possible scenarios that often involve multiple trains. That is why the decision making also for a large part relies on the knowledge and judgement of the train dispatcher themselves, following from their training and their experience [12]. Although these procedures will often lead to suitable and practical solutions, the complexity of the railway traffic management problem has as a result that it can not always be taken into account which actions will result in the most efficient or effective solution to the conflict.

Taking all aspects into account, the current state of practice is currently able to keep railway traffic running safely and smoothly. However with the expected increasing complexity, as mentioned in Chapter 1, this may become increasingly difficult, because the conflict resolution relies largely on human judgement. Additionally, the current division of regions and subregions could benefit from research if this is indeed an optimal division, something ProRail has also recognised.

2.2 State of the art of centralised railway traffic management

This section discusses the most influential papers on the subject of centralised railway traffic management. As mentioned before, in centralised approaches the physical system is solved as a whole by one single agent. The papers discussed in this section are the ones most cited and most found in the literature reviews in papers. The goal of this section is to give an overview of the field and the general trend it follows.

Overall, the centralised railway traffic management forms a strong basis with multiple approaches to rescheduling railway traffic in real-time. Within this basis, individual papers have brought their own focuses and improvements. For instance, Kecman et al. [30] and Törnquist and Persson [31] aim to look at conflict from a higher level, by establishing models that consider less detail and can compute larger networks. Corman et al. [32] later elaborate on this by developing a model and optimisation that can use characteristics of both a low level and a high level approach. Another collection of papers has explored different types of model formulations of the railway traffic management problem: D'Ariano et al. (2007) [10], Rodriguez (2007) [33], D'Ariano et al. (2008) [34] and Luan et al. (2018a) [6]. A focus of other authors has been the optimisation method with which the railway traffic management problem is solved. Lusby et al. [5] and Caimi et al. [35] apply an existing optimisation algorithm to the RTM problem for the first time. Other papers, those by Corman et al. [36] and Pellegrini et al. [4], develop a novelty optimisation algorithm. Meng and Zhou combine two focuses in [37], where they use both an exceptional model formulation and optimisation.

The above mentioned papers differ on a variety of factors in their overall approaches. The main distinguishing factors are discussed in the remainder of this section, elaborating on the specifics of the publications. These attributes will also be used in the concluding table of this chapter (Table 2.1) to categorise the literature.

2.2.1 Classification of conflict scenarios

A distinction can be made between two different scenarios that can lead to conflict: a disruption and a disturbance. Their distinctive factor is the magnitude of the delay in that scenario or in other words how much the operation of the train deviates from the original timetable. The term **disruption** is used to define large deviations of roughly 15 minutes or more. Disruptions have a considerably large effect on the train operation and will likely result in the cancellation or short turning of trains [11, 38]. In the case of small perturbations in the timetable, one defines it as a **disturbance**. These perturbations can be solved with changes in arrival- and departure times or in local routes [23].

Although disruptions are more exceptional, both disruptions and disturbances occur in day to day railway traffic management. However a disruption calls for more rigorous actions than a disturbance and thus forms a more complex problem. Essentially, during disruptions one needs to re-asses the timetable as a whole, whereas with disturbances small alterations suffice [37]. This has as a result that disruption management largely coincides with the papers on scheduling, discussing topics such as timetable stability and robustness [5]. Only a minority of the papers in this review is concerned with disruptions [5, 37].

This research will focus on the railway traffic management under disturbances. These are the most recurrent in day to day operation. Additionally because a new approach will be developed, it seems wise to use the less complex problem of disturbances for the proof of concept.

2.2.2 Dispatching problems in centralised railway traffic management

Publications can view different types of problems. The scope of the literature research is adjusted such that all presented papers view the conflict resolution problem, which encases one or multiple actions as presented in subsection 1.3.1: retiming, reordering, rerouting, short turning and cancelling. Most publications view retiming and reordering and the occasional paper will also encase rerouting. The last two actions are mostly only viewed in publications observing the larger scale (macroscopic) problem in the case of disruptions. The relation between different types of rescheduling as defined in Fang et al. (2015) [22] can be seen in Figure 2.3. The graph describes the problem type according to the time frame that is taken into account. The presented problem type will attempt to find a feasible schedule for the time frame $[t + a, t + b]$. A general conclusion can be drawn that the conflict resolution problem looks at the rescheduling within a small horizon.
Although all problems view the rescheduling problem, a few papers may view a more specific or additional problem type. For example, both Rodrigeuz (2007) [33] and Lusby et al. (2013) [5] observe the rescheduling problem specifically at a junction. Luan et al. do not only view the rescheduling problem in [6], but also include potential strategies for railway traffic control.

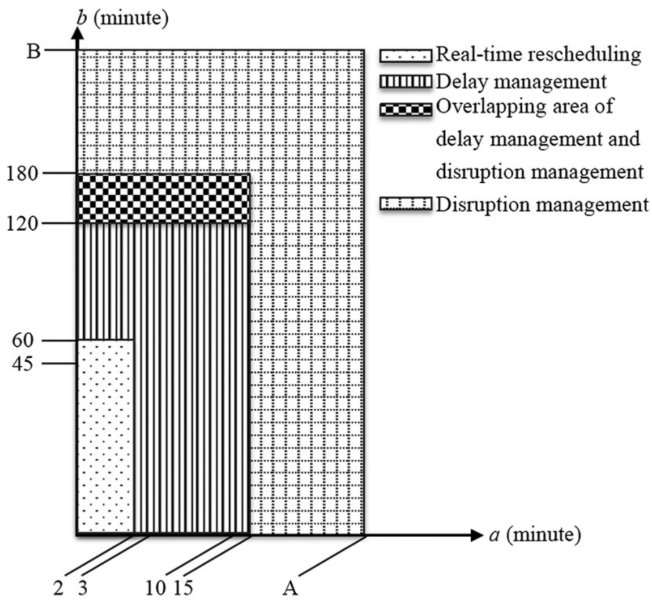


Figure 2.3: A visual description of different problem types [22]

2.2.3 Infrastructure detail level of rescheduling models

A distinction between different conflict resolution methods can also be made based on the scale of the infrastructure representation. The scale can be divided into three levels of detail: microscopic, mesoscopic and macroscopic. The difference is illustrated in Figure 2.4. A **microscopic** infrastructure considers all individual tracks, switches and lineside signals. A **macroscopic** infrastructure looks at the network from a higher level and considers the stations as nodes and the aggregated tracks between them as connecting arcs between those nodes. A **mesoscopic** infrastructure falls in between: The number of tracks per section is represented and the nodes now represent junctions.

Both the macro- and microscopic representation have been used in the found literature on railway traffic management. Microscopic approaches are used most often, because this is prevalent in current railway systems [23]. Macroscopic can also be convenient. For example, Kecman et al. (2013) [30] studies the entire Dutch railway network and uses a macroscopic infrastructure representation to reduce the amount of information within the computation. For similar reasons, Törnquist and Persson (2007) [31] also uses a macroscopic approach. Whereas Meng and Zhou (2011) [37] uses a macroscopic representation because it studies a large scale disruption that asks for the design of a new emergency timetable, for which macroscopic information suffices. However in most railway traffic management approaches the infrastructure representation is microscopic.

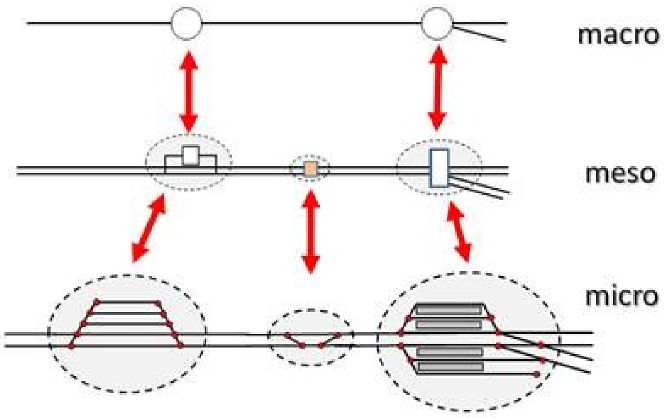


Figure 2.4: Micro-, meso- and macroscopic railway infrastructure representations [39]

A microscopic approach is also what is used in this research. A macroscopic approach will result in a conflict resolution solution that is feasible on the macroscopic level, but could just as well prove infeasible on the microscopic level. Moreover, for complete and thorough conflict resolution, detailed information is needed, such as arrival times of specific trains at specific parts of the infrastructure, such as block sections.

2.2.4 Conflict resolution method formulation

The railway traffic management problem can be modelled in different ways. The two prominent are the Alternative Graph (AG) formulation and Mixed Integer Linear Programming (MILP) (or a variation of it).

Alternative Graph model

This formulation was first presented by Mascis and Pacciarelli in 2002 [40]. It was later applied in D'Ariano et al. (2007) [10] in the final form it is now commonly used.

The basic principle of the model relies on the job-shop scheduling problem. Scheduling can be defined as "the allocation of shared resources over time to competing activities" [41]. In a job-shop scheduling problem jobs J are assigned to machines M at a certain point in time, where a machine can process only one job at a time. An example of a solution is shown in Figure 2.5, where jobs J_1, J_2 and J_3 are assigned to machines M_1, M_2 and M_3.This principle can be applied to railway by letting the trains resemble jobs and letting the machines resemble the block sections. In other words the block sections will be scheduled to process the "jobs" of a train running through the block section. There are some gaps between the job-shop scheduling problem and the railway problem that need to be bridged by additional constraints [10]. For example, a job has an infinite buffer of storing time when moving from one machine to the next, but for a train it is not possible to have an infinite buffer between one block section and the next. The implementation of a blocking constraint prevents this situation in the railway application.

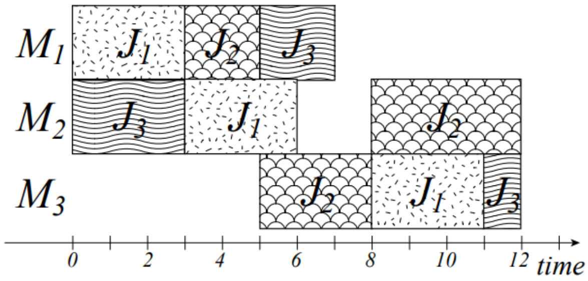
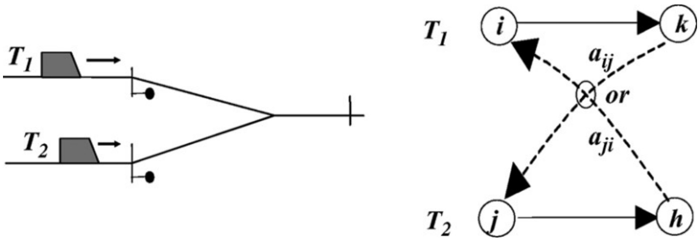
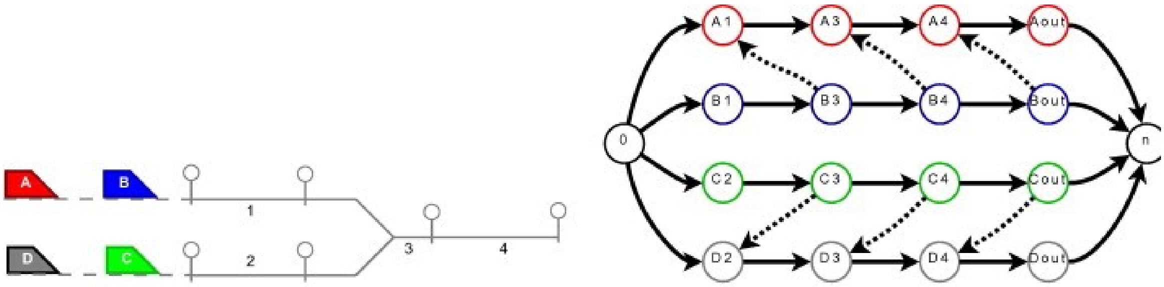


Figure 2.5: An example of an outcome to a job-shop scheduling problem [41]

The model can be translated into a graph representation. Here, a **node** represents an event, e.g. a train being at a certain place at a certain time. All arcs in the graph are **directed arcs**, meaning that besides showing a connection between two nodes they also show a succession between two nodes. For example, an arc from node A to node B, shows that event A will occur first and after that event B. For the interaction between two or multiple trains another element is added: **alternative arcs**. An alternative arc is added at a point of potential conflict, such as in Figure 2.6a where two tracks converge to one. At such a point an **alternative arc pair** is modelled: one arc resembling the precedence of one train, the other arc the precedence of the other. In the solution of the model, only one of these arcs can be chosen. This way the model resolves a conflict by choosing which train will precede the other. A solution to a more elaborate example can be seen in Figure 2.6b.



(a) A converging junction with two trains, as in [10]



(b) A converging junction with four trains [36]

Figure 2.6: Examples of Alternative Graph models

Mixed Integer Linear Programming

This formulation and adjacent formulations form the other prevalent model in centralised literature. At the foundation of the variants lies Linear Programming. The problem is formulated in three elements:

- **Objective function** This defines the objective of the model, often formulated as the minimisation of certain terms.
- **Decision variables** The value of these variables can be chosen by the model. They can be, amongst others, binary, integer or continuous variables. A **continuous variable** can take on any real number; an **integer variable** can be any real whole number and **binary variables** can have value 0 or 1.
- **Constraints** These equations define the boundaries of the problem. In the form of (in)equalities they set boundaries to which values the decision variables can take.

The types of decision variables dictate the name of the model. For the majority of the models multiple types of variables are used, which are called Mixed Integer Linear Programming (MILP) models. Because a variations in variables can be used, the model formulation is made easier. However, Caimi et al. (2012) [35] uses only binary variables and thus the model becomes a Binary Linear Programming (BLP) model. As the "L" in MILP suggests, all constraints and the objective function must be linear equations. If one or multiple equations is non-linear, it becomes a Non-Linear Programming (NLP) model, such as the MINLP in Luan et al. (2018a) [6]. In this paper the non-linear model is part of a study into the performance of three different models. The MINLP is not concluded to be a convenient formulation, because it tended to only solve to local optima.

There are two publications that use the Linear Programming foundation, but use a deviating form.
Meng and Zhou use stochastic programming, which is explained to be a MILP formulation that is given stochastic inputs, such as the time it takes a train to traverse a block section [37]. They use stochastic inputs as a means to determine the robustness of their model, i.e. the consistency of the model in the face of varying inputs. Rodriguez uses a formulation called Constraint Programming [33]. This is largely similar to a Linear Programming model, but this model does not contain an objective function. This way it focuses on feasibility and not on optimality. For this reason it can not find a solution to the problem, however it will decrease the solution space. Starting from this solution space, Rodriguez uses a simulation to come to a solution to the rescheduling problem.

On the topic of model formulations, one thing should be noted: MILP and AG are very closely related. This is because the Alternative Graph model leads to a mathematical model that is in practice most often implemented as a MILP model. So, in essence, Alternative Graph models are also MILP models. However, the Alternative Graph models all share the same mathematical foundation, whereas the MILP models have differing foundations. That is where the distinction is made.

2.2.5 Optimisation algorithms for centralised conflict resolution

As mentioned before, the conflict resolution problem is an optimisation problem. Especially for larger instances, the solving of the problem can get complex and time consuming. That is why the literature reflects a constant search for improving the solution method.

Many times researchers opt for a custom made solution method that is to speed up the solution process. For a part of those approaches the foundation is to split the solution process into two or more stages. These stages can, for example, be defined as a high level solver and a low level solver. The first solver will provide an initial general solution that decreases the solution space, after which the low level solver solves the problem considering more details. This process can be iterated to come to an optimal solution that is both feasible on the higher and lower level. Both Luan et al. [6] and Meng and Zhou [37] have developed such an approach. Luan et al. explore two different separations: high/low level and a presolving stage. Meng and Zhou choose to use three different stages: the first two steps eliminate suboptimal solutions and the third step comes to the final solution. Iterating over two solution levels, is the main focus of the approach presented by D'Ariano et al. [10] and Kecman et al. [30]. These two papers solve the overall conflict, plan the trajectory of individual trains and then reiterate to the overall conflict resolution. This aims to make sure the solution is feasible on both levels, while limiting the computation time. Separating the solution process into multiple stages can also be approached mathematically. Both Lusby et al. (2013) [5] and Rodriguez (2007) [33], apply a principle where the problem is solved by gradually adding more constraints to the problem instead of imposing all constraints initially.
Both Corman et al. (2017) [32] and Pellegrini et al. (2015) [4] focus on developing new optimisation approaches in their papers and therefore are particularly interesting. Corman et al. [32] study four different heuristics, all built from the foundation of fixing one variable within the problem. Pellegrini et al. [4] develop a new well-performing approach that solves the rescheduling problem with an algorithm based on the MILP formulation. This particular optimisation approached is well researched: it was first introduced in Pellegrini et al. (2014) [42], expanded in Pellegrini et al. (2015) [4] and later reformulated in Pellegrini et al. (2019) [43].
Some research has also used existing methods for solving the rescheduling problem. For example, Törnquist and Persson (2007) [31], D'Ariano et al. (2008) [34] and Corman et al. (2011) [36] use the Branch and Bound approach, where different solution to the problem are explored and "branches" leading to suboptimal solutions are cut off. Corman et al. have a slight addition to the formulation in D'Ariano et al.(2008) [34] as they add the constraint that the conflicts are to be solved whilst considering several train priorities. Caimi et al. use the rolling horizon approach, as used in Model Predictive Control, to solve the rescheduling problem in Caimi et al. (2012) [35].

2.2.6 Conclusion on centralised railway traffic management

The publications mentioned in this section are summarised in Table 2.1. Looking at this table, one interesting remark is that there have not been publications in the past few years. This observation confirms a trend that can also be observed in the individual subsections: the basis of centralised conflict resolution has been developed. The general approach uses a microscopic scale, studies disturbances in a rescheduling problem and formulates that using either a MILP or Alternative Graph formulation. The presented papers have been cited many times in more recent approaches that use this basis and build onto it with different specialisations. These specialisations include the studying of a new moving block signalling system, such as Wang et al. (2014) [44], and extensively incorporating passenger connections into conflict resolution, such as Corman (2020) [45].

One concern that surfaces, amongst others mentioned in Pellegrini et al (2015) [4], Lusby et al. (2013) [5] and Luan et al. (2018a) [6], is that the computation time of a centralised conflict resolution can become very large, especially for complex situations. This is undesirable if the application is intended for using in real-time which is the case for the majority of the approaches. To improve this a selection of the papers seems to have tried to improve the solution methods. However in the current literature a suitable approach has not yet been adopted. To overcome this problem, a large part of the researchers have started studying non-centralised approaches. These approaches are discussed in the next section.

Author	Scale	Scenario	Model formulation	Solution method
D'Ariano et al. (2007) [10]	Micro	DB	AG	Iterative rescheduling algorithm
Rodriguez (2007) [33]	Micro	DB	Constraint programming & Simulation	Three mechanism approach
Törnquist & Persson (2007) [31]	Macro	DB	MILP	B&B
D'Ariano et al. (2008) [34]	Micro	DB	AG	B&B
Corman et al. (2011) [36]	Micro	DB	AG	Priority based B&B
Meng & Zhou (2011) [37]	Macro	DP	Stochastic programming	Custom B&B
Caimi et al. (2012) [35]	Micro	DB	Binary Linear Programming (BLP)	MPC framework
Lusby et al. (2013) [5]	Micro	DP	MILP	Custom Branch & Price
Kecman et al. (2013) [30]	Macro	DB	AG	Iterative rescheduling algorithm
Pellegrini et al. (2015) [4]	Micro	DB	MILP	RECIFE-MILP
Luan et al. (2018a) [6]	Micro	DB	MILP & MINLP	Two step approach

Micro: microscopic; Macro: macroscopic; DB: Disturbance; DP: Disruption; AG: Alternative Graph; MI(N)LP: Mixed Integer (Non-)Linear Programming; B&B: Branch & Bound; MPC: Model Predictive Control

Table 2.1: Centralised railway traffic management approaches

2.3 State of the art of non-centralised railway traffic management

Non-centralised railway traffic management approaches have become of interest in literature. In a non-centralised approach, the physical system is divided into subsystems that are each solved by an individual agent. This section reviews the literature on these non-centralised approaches.

Reflecting the conclusions in centralised literature, the main motivation for looking into non-centralised approaches is mentioned to be shortening computation time [7, 46–48]. The shortening of the computation time is often also achieved [7, 9, 38, 46, 49–51], but not in all cases. In Sinha et al. (2016) [52], the centralised approach has a faster computation. Often the non-centralised also performed better in terms of delay reduction [38, 46, 49, 53, 54]. However, Kersbergen et al. (2014, 2016) [8, 55] find that non-centralised and centralised approaches show comparable results and Keita et al. (2020) [50] finds that a non-centralised approach only performs better in some cases.
The non-centralised approach consists of several components, as also mentioned in subsection 1.3.3. Similar to centralised railway traffic management, papers generally focus on one or two of these components. For example, some focus on developing new decomposition approaches: how to divide the system into subsystems [7–9, 50, 51, 56]. Another used focus is on the introduction of coordination approaches: how the subsystems communicate with each other [15, 46, 52, 53, 55, 57, 58]. Some papers mainly present a new approach to conflict resolution [38, 44, 47, 59]. Other publications have a more specialised focus such as the optimisation problem [49, 60] and the system architecture [48, 61].

As can maybe already be deduced, the field of non-centralised literature has received considerable attention, and especially in recent years. This literature is discussed in the remainder of this section on the basis of some of the most distinguishing factors between the papers. The discussion is concluded by summarising the publications in a table according to these factors (Table 2.3).

2.3.1 Dispatching problems in non-centralised railway traffic management

Identical to the review for centralised railway traffic management, the scope of the review of non-centralised railway traffic management approaches is adjusted such that it encases only papers with a focus on the rescheduling problem. Again some publications view other additional problem types. For instance, Wang et al. choose to also encompass the railway traffic control problem next to the delay management problem in [54]. Zhang et al. include the additional problem of emergency track maintenance in [51]. Lastly, in [59] Chen et al. view the rescheduling problem specifically in the face of emergencies.

2.3.2 Non-centralised system architectures in railway traffic management

The system architecture refers to the conceptual model of the behaviour and structure of a system [62]. Within the non-centralised system architecture, more specific architectures can be defined that are encountered in literature: decentralised, distributed, self organising and Multi Agent systems. The definition of these terms does not have a clear consensus in the field and publications contradict each other in the definitions. What one calls decentralised another calls self organising. And what one calls distributed another calls decentralised. For clarity, this research defines the system architectures as follows:

- **Decentralised**: The agents do not communicate with each other.
- **Distributed**: The agents do communicate with each other.
- **Multi Agent System (MAS)**: The system consist of multiple agents with unspecified communication.
- **Self organising**: The agents operate directly in the physical layer of the system and do communicate.

The literature is categorised according to this classification, even though the publication itself might mention something else. The last category, self organising, is left out of scope for this review. This category sees parts of the physical systems as individual agents, in this field that oftentimes means that trains are the agents which coordinate with each other to organise the physical system, such as in Su et al. (2015) [63] and Yong et al. (2017) [64]. This is a futuristic and conceptual system architecture of which research is in its infancy [65]. Because the publications are generally on a very conceptual and explorational level without the mentioning of concrete models, they are left out of the review. For more information on self organising literature, consult the review by Marcelli and Pellegrini [65].

The Multi Agent System architecture is mostly used in the case where agents compute on different operational levels and time periods. For example, in Lamorgese and Mannino (2015) [7] the system is divided into agents on two levels: on the level of individual stations and on the level of train lines. The train line agent computes arrival and departure times for trains at the stations and the station agents will then find a route within the station that will fit with these times. Letia et al. also use agents on similar levels in [66]. A train agent has to find a path that corresponds to the arrival and departure times issued by the scheduling agent and along that path block section agents will clear the infrastructure for trains to pass over. Narayanaswami and Rangaraj present an elaborate MAS in [47]. The systems consists of four different agents: a train agent for every individual train; a station agent for every individual station; an auctioneer clearing the routes for the trains and a central supervisor agent overseeing the system. Multi Agent Systems are also seen in the literature on railway traffic management control, where the running details of trains, such as speed profiles, receive more attention. Examples of such systems can be found in Hassanbadi et al. (2015) [61] and Guo et al. (2021a, 2021b) [48, 60].

The decentralised and distributed architecture are used in interchangeable instances, meaning it can not be said that one is used in a certain instance more often than the other. However the decentralised is most commonly used. An interesting comparison between the two architectures is done in Kersbergen et al. (2014) [55]. The authors find that the distributed approach performs better in terms of delay reduction, but the decentralised approach computes faster. A difference that can be found between the two approaches is that decentralised approaches are mostly used in combination with a higher level coordinator in order to be able to reach a global feasible or even optimal solution.

For instance, Kersbergen et al. use a coordinator supervising central cost function in [8, 55] and Corman et al. use a coordinator graph that dictates the relations between different regions in [46, 49, 53]. (The coordination strategies themselves will be elaborated on in subsection 2.3.3.) The coordinator is not responsible for optimising the conflict resolution, but solely for initiating and organising coordination between the agents. In that way it still is a decentralised system and not a centralised one. The suspected reason for why a decentralised architecture is seen predominantly in non-centralised railway traffic management literature is that it is the easier architecture to implement. Because much research is still being done on different aspects of the non-centralised approaches, the choice for a decentralised approach may be made if that is not where the main focus of the publication lies.

2.3.3 Coordination within non-centralised railway traffic management systems

The coordination within a non-centralised railway traffic management system is essential in working towards global feasible and optimal solutions, one of the challenges of a non-centralised approach. This challenge has been tackled in different ways by different authors. This subsection mainly discusses the papers that have focused on the coordination problem.

Corman et al. introduce the coordinator graph in [53] and also use it in their later papers Corman et al.(2012, 2014) [46, 49]. An example is shown in Figure 2.7. The coordinator graph is formulated similarly to the Alternative Graph used in conflict resolution, such that the nodes represent the entering of a block section and the arcs represent the time between two nodes. The border is defined by a block section so the border graph contains two nodes for every train: the train entering the border section and the train entering the next section. The local solutions are incorporated into the border graph by using the arrival time of a train at the border section from the local solution. By combining the local solutions, the coordinator graph can check the global feasibility. Kersbergen et al. introduce a central cost function to approach feasibility and optimality in [55] and use it later in Kersbergen et al. (2016) [8]. The central cost function describes the cost function of the overall system. The agent will then find a solution within its own subsystem that also minimises the central cost function, while taking into account the effects of its decision on the agents of neighbouring regions. This will lead to a globally feasible solution that will also approach optimality.

Another used coordination approach is boundary constraints, where coupling constraints that couple the regions are introduced. The content of the constraint partially depends on the type of decomposition used. For a geographically decomposed system, the main coupling constraint is the time transition constraint as used in Luan et al. (2018b, 2020) [9, 15] and Toletti et al. (2020) [57]. This constraint formulates that the time at which a train leaves its current region is also the time at which the train will enter the next region. In addition to the time transition constraint Luan et al. (2018b) [15] also uses a speed transition constraint, which works similarly to the time transition constraint to ensure that the speed at which a train leaves a region will also be the speed at which it will enter the next one. Toletti et al. present an elaborate coordination scheme that reiterates over the constraint in [57], see Figure 2.8.

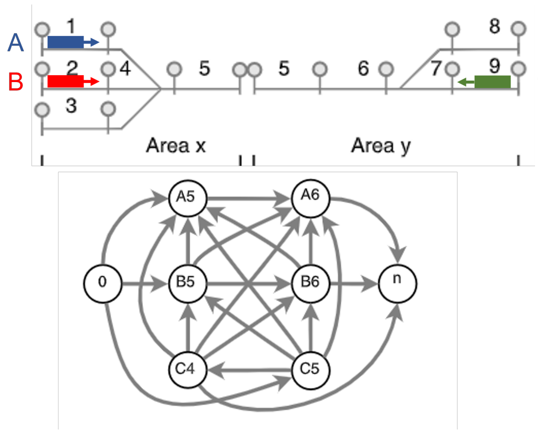


Figure 2.7: A coordinator graph in [53]

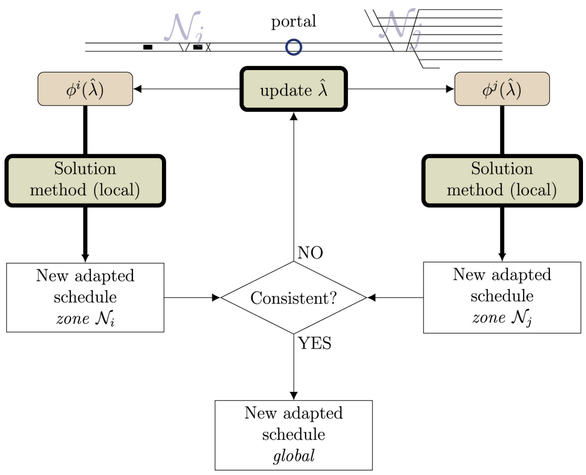


Figure 2.8: The iterative coordination strategy presented in [57]

Besides the time transition and speed transition constraint Toletti et al. also include a constraint that if the whole route of a train is cancelled in one region it will also be cancelled in the next. All these mentioned constraints ensure continuity: they ensure that a train can not disappear at a region border for a certain amount of time; that a train can not change its speed at an infinitesimal time instance between two regions and that a train can not suddenly reappear in a region after its route was cancelled in a previous one. Luan et al. (2020) [9] uses different boundary constraint for the two other decompositions it discusses: a train based decomposition and a time interval decomposition. The coupling constraint in the train based decomposition is a capacity constraint, which establishes that a block section can only be used by a single train. In the time interval decomposition, both the time transition and capacity constraint are used.

Within the usage of boundary constraints, the coordination approach is refined by how these boundary constraints are solved. Luan et al. (2018b) [15] presents both the Alternating Direction Method of Multipliers algorithm and Priority Rule based algorithm. The Priority Rule based algorithm solves each region sequentially in order of priority, in which the region with the largest delay in an unrescheduled situation has the highest priority. The outputs of one region are then imposed on the next region and so forth. If this way a feasible solution is not found, the process is reiterated with an alternate order. The Alternating Direction Method of Multipliers algorithm (ADMM) solves the regions iteratively and by doing so iterates upon the shared variables at the boundaries. In addition to these algorithms, Luan et al. (2020) [9] applies the Cooperative Distributed Robust Safe But Knowledgeable algorithm (CDRSBK). This algorithm solves a region and its neighbouring region in one iteration and then iterates over the different regions. Toletti et al. (2020) [57] implements an approach that does a consistency check at the boundary. In other words, the regions are solved individually taking the coupling constraints into account. It is then checked if the values at the boundary satisfy these constraints and if the local solutions can be adapted into the global solution.

2.3.4 Decomposition of the railway system in non-centralised railway traffic management

The decomposition formulates how the system is divided into subsystems. Within literature several different types of decomposition are found.

- **Geographical (GEO)**: The infrastructure region is divided into subregions
- **Mathematical (MAT)**: The mathematical conflict resolution problem is divided into mathematical subsystems. For example, Keita et al. (2020) [50] defines the rerouting problem separate from the retiming problem and Kersbergen et al. (2016) [8] separates all the constraints to be solved individually.
- **Train based (TRA)**: Each train forms its own subsystem.
- **Time interval (TIN)**: The problem is divided into certain time instances.
- **Macro-Micro (MM)**: The macro problem is separated from the micro problem, meaning the problem is first solved on the high level macroscopic scale and then solved on the detailed microscopic level.
- **Agent based (AB)**: The system is divided into agents, that are each responsible for their own task within the system. For example, Narayanaswami and Rangaraj (2015) [47] defines train agents, station agents, an auctioneer agent and a coordinating agent.

Although every publication on non-centralised railway traffic management uses a decomposition, as it is an inherent part of the approach, not many focus on the decomposition itself. In the majority of publications the decomposition is taken as a prerequisite and/or is not founded scientifically. Often the decomposition is used as seen in the state of practice, e.g. in Corman et al. (2012) [46], Toletti et al. (2020) [57] an Cavone et al. (2020) [38]. For the state of practice see section 2.1.

The publications that do focus on decomposition, are elaborated in Table 2.2. The table discusses the type, criteria and result of the decomposition. The criteria of the decomposition dictate why an element will be assigned to one subsystem or another and the result of the decomposition is what a subsystem contains. For example, Lamorgese and Mannino (2015) [7] applies a Macro-Micro decomposition to a train network. As a criterion the problem scale is used, meaning that if the problem is microscopic or macroscopic dictates to which subproblem it is assigned. The result is an agent for every station and every line between stations, in which the line agents solve the macroscopic problem while the station agents solve the microscopic problem at the stations. The approach is inspired by the Benders cut algorithm which divides the problem into solving binary variables and continuous variables. Another paper that was inspired by Benders cut is Keita et al. (2020) [50]. Instead of the two step approach they add a third step. First, the routing problem is solved. Secondly, the optimal reordering and retiming actions are found. Lastly, the delays that correspond to the decided actions are computed.

Kersbergen et al. [8] mathematically decompose the system by making subsystems consisting of constraints. The criteria that they use are: the maximum difference in number of contraints between subproblems and the amount of constraints not in the current subproblem connected to a constraint in the current subproblem, or in other words problem size and connectivity. Connectivity is used in another sense in the geographical decomposition presented in the research by Chen et al. in [59]. They define nodes within the railway network at every lineside signal. These nodes are then given a degree of connectivity and centrality in the form of two factors: how many trains travel over that node and how many block sections are connected to that node. The combination of these two factors signifies the node's importance and a node is labeled as important if its importance is above a certain threshold. In the decomposition it is then considered that a region contains only two important nodes, resulting in nodes being assigned to a certain subarea. The paper in this selection that pays the most attention to the decomposition is Luan et al. (2020) [9]. In this paper three types of decompositions are presented and compared. Firstly they present a geographical decomposition that divides the railway network into subregions considering two criteria: the deviation of the size of a region, in terms of amount of block sections in the region, compared to the mean region size and the amount of train services crossing a region border. Secondly they present a train based decomposition where every train is viewed as a subsystem. Thirdly, they show a time interval based decomposition, where the rescheduling problem is solved for certain time intervals. Whether an occurrence belongs to a time interval is determined by the estimated occurrence time. In the conclusion of the paper it is mentioned that the geographical decomposition is offline, whereas the train based and time interval based decompositions are online, meaning they need to be updated as the schedule changes. This means that these two decompositions require computational power besides the computational power needed for the conflict resolution problem. When trying to reduce computational power, using an offline decomposition thus forms a more logical choice.

Authors	**Type**	**Criteria**	**Result**
Lamorgese & Mannino (2015) [7]	MM	Problem scale	Station & Line agents
Kersbergen et al. (2016) [8]	MAT	Problem size, Connectivity	Constraints
Chen et al. (2020) [59]	GEO	Connectivity, Centrality	Subregions
Luan et al. (2020) [9]	GEO	Region size, Connectivity	Subregions
	TRA	Singular train	Trains
	TIN	Occurence estimate	Time intervals
Keita et al. (2020) [50]	MM	Problem type	Subproblems

Table 2.2: Decomposition in non-centralised railway traffic management approaches

2.3.5 Optimisation algorithms in non-centralised conflict resolution

This subsection discusses the optimisation methods used to solve the conflict resolution problem in non-centralised approaches. A selection of the optimisation methods have also been used in the centralised and have thus already been discussed in subsection 2.2.5. These methods include: Branch and Bound (B&B), the MPC framework, the bi-level heuristic and RECIFE-MILP.

Some different optimisation approaches are also used. For example, Letia et al. make use of a simple algorithm for the conflict resolution in [66], called the Earliest Deadline First or EDF algorithm. In the case of a conflict, the algorithm will let the train with the earliest deadline precede. Luan et al. (2018b, 2020) [9, 15] and Zhang et al. (2021) [51], use a CPLEX algorithm. This method uses the simplex algorithm which searches the boundaries formed by the model constraints to find the optimal solution. Narayanaswami and Rangaraj [47] take another approach and use a bidding heuristic, where train agents have to bid on preceding in the case of conflict. The auctioneer agent in their modelling will then make a decision on which train precedes.

Two papers focus on implementing an optimisation approach. Meng and Zhou [37] use a Lagrangian relaxation. In this approach the problem is first solved with relaxed constraints, meaning a violation of a constraint leads to a penalty instead of model infeasibility. This initial solution is then used to solved the fully constrained problem. Corman et al. introduce the Job Greedy Heuristic in [49]. This paper also uses the Alternative Graph formulation and it has been discussed earlier in this thesis that this formulation is based on the job shop scheduling problem. This is what the term 'Job Greedy Heuristic' refers to. One by one, a job is added to the problem and solved. Or in other words, for each solution iteration a train is added to the problem.

2.3.6 Conclusion on non-centralised railway traffic management

Looking at the presented non-centralised approaches, several conclusions can be drawn. Firstly, the field on non-centralised approaches to the railway traffic management problem contains a fair amount of research and has also received many new publications in recent years. It can be seen that the methods used in non-centralised approaches do not differ considerably from the methods used in centralised approaches for some aspects, for example for the dispatching problem, conflict resolution methods and optimisation. Naturally differences can be seen in the aspects inherent to non-centralised approaches. Several different methods have been introduced to coordinate different subsystems, such as a coordination graph, a central cost function and boundary constraints. Also, most papers opt for introducing a decentralised approach, instead of a distributed or MAS approach. One key aspect of a non-centralised approach that has however received limited attention is the decomposition. Only a few papers have elaborated on the decomposition approach and have used mathematical or scientific methods to support the decomposition. A different trend can be observed in the field of road traffic management. Publications such as Anwar et al. (2017) [67] and Lin and Xu (2020) [68] have focused on decomposition and more specifically how road networks can be divided into subregions based on traffic patterns. This means that focusing on decomposition can be interesting for railway networks too.

Author	**Architecture**	**Decomposition**	**Coordination**	**CR Model Formulation**	**CR Optimisation**
Letia et al. (2008) [66]	MAS	AB	Communication	Constraint Logic Programming	EDF algorithm
Corman et al. (2010) [53]	De	GEO	Constraining arcs in coordination graph	AG	B & B
Corman et al. (2012) [46]	De	GEO	Constraining arcs in coordination graph	AG	B & B
Meng & Zhou (2014) [69]	De	TRA	Priority rules	Rule Based Programming	Lagrangian Relaxation
Corman et al. (2014) [49]	De	GEO	Constraining arcs in coordination graph	AG	Job Greedy Heuristic
Kersbergen et al. (2014) [55]	Di & De	MAT	Central cost function and/or constraints	Discrete Event System	MPC framework
Lamorgese & Mannino (2015) [7]	MAS	MM	Feasibility cuts	AG	B & B
Narayanaswami & Rangaraj (2015) [47]	MAS	AB	Communication	Agent Based Simulation	Bidding Heuristic
Kersbergen et al. (2016) [8]	Di & De	MAT	Central cost function and/or constraints	Discrete Event System	MPC framework
Luan et al. (2018b) [6]	De	GEO	Boundary constraints	MILP	CPLEX
Luan et al. (2020) [9]	De	GEO, TRA & TIN	Boundary constraints	MILP	CPLEX
Chen et al. (2020) [59]	De	GEO	Boundary constraints	MILP	Commercial solver
Toletti et al. (2020) [57]	Di	GEO	Consistency check at boundary	Resource Graph	Commercial solver
Cavone et al. (2020) [38]	De	GEO, MAT	Not specified	MILP	MPC framework
Keita et al. (2020) [50]	Mat	MM	Feasibility cuts	MILP	Bi-level heuristic & RECIFE MILP
Zhang et al. (2021) [51]	Mat	MAT, TIN	Shared constraints	MILP	CPLEX
Wang et al. (2022) [54]	Mat	MAT	Recommended & Realised solutions	MILP	MPC framework

MAS: Multi-Agent System; De: Decentralised; Di:Distributed; Mat: Mathematical; AB: Agent-Based; GEO: Geographical; TRA: Train Based; MAT: Mathematical; MM: Macro-Micro; TIN: Time Interval based; AG: Alterntive Graph; MILP: Mixed Integer Linear Programming; EDF: Earlier Deadline First; B&B: Branch & Bound; MPC: Model Predictive Control

Table 2.3: Non-centralised railway traffic management approaches

2.4 Key Performance Indicators used in non-centralised railway traffic management

To quantify the performance of an approach, research generally uses one or multiple **Key Performance Indicators (KPIs)**. More specifically, a KPI is a measurable value that demonstrates the effectiveness of an approach. This section studies the KPIs used in the non-centralised railway traffic management literature discussed in section 2.3. This will provide insight into the KPIs that might be relevant to the current research. Figure 2.9 summarises the KPI usage by showing how many papers measure their performance according to certain KPI categories. The KPIs are categorised to provide a compact and effective overview. These categories are elaborated below.

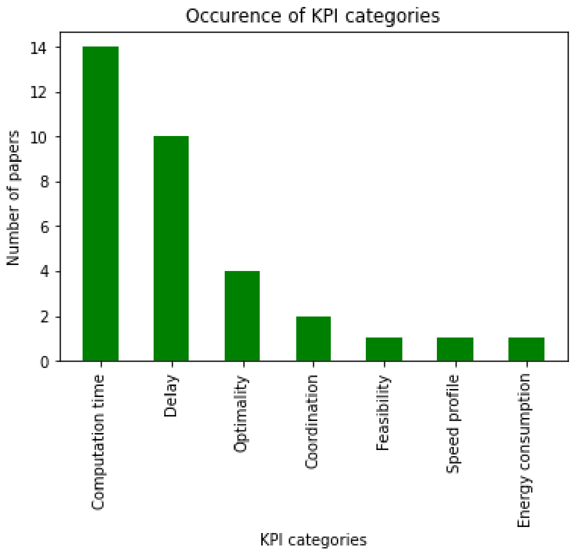


Figure 2.9: KPI categories used in the non-centralised literature in Table 2.3

First of all, it is apparent that computation time and delay are both used frequently. This is in line with the main motive of investigating non-centralised railway traffic management being a reduction of computation time. Comparably, using delay KPIs is a logical choice, because it is one of the main ways to measure the effectiveness of railway traffic management: Good management will keep delays to a minimum. The delay is often defined as the total delay over the studied network. For both computation time and delay, similar variations of the KPIs are used. Both see the usage of a maximum and average value, to characterise the magnitude as well as the distribution [38, 46, 49, 55]. To further characterise the distribution of the computation time and delay, Cavone et al. (2020) [38] also uses the standard deviation of both parameters as KPIs.
Two other interesting KPI categories are the ones considering optimality and feasibility. These are used to study the quality of the solution finding algorithm and thus are often used in the papers focused on optimisation or coordination, like Corman et al. (2012, 2014) [46, 49] and Meng and Zhou (2014) [69].
The remaining KPI categories are often chosen considering the specific goal of the research. For example, Wang et al. (2022) [54] focuses partly on railway traffic control, which besides railway traffic management also studies the enforcement of the actions resulting from the RTM approach. A reflection of the behaviour of a train can be found in its speed profile: a graph that plots the train velocity against the travel distance. The adherence of a train to its scheduled speed profile, can therefore be an important KPI when studying railway traffic control, reflected in the Speed Profile KPI Category. Within railway traffic control the publication also studies energy consumption and thus also uses that as a KPI. Two papers focused on coordination show other specific KPIs. In Luan et al. (2018b) [15] a priority based coordination algorithm is tested and the adherence to that priority is used as a KPI. Similarly, the performance of the iterative coordination approach presented in Toletti et al. (2020) [57] is measured using the KPI of how many iterations are needed to find an optimal solution.

Concludingly it can be said that computation time and delay related KPIs are used most often and would probably also be interesting to use in the current research. Other KPIs are used out of interest for specific behaviours. Papers focused on decomposition generally have not been found to share a specific KPI.

2.5 Conclusion on the state of practice and state of the art

The goal of this chapter was to answer the first research question: *What is the state of the art in centralised and non-centralised conflict resolution?*. To answer this question, research was done into the current state of practice, the state of the art in centralised approaches, the state of the art in non-centralised approaches, the usage of decomposition in non-centralised approaches and the KPIs used in non-centralised approaches.

This leads to the following conclusion. In the state of practice conflict resolution decisions are largely based on human judgement, which does not necessarily lead to the most effective solution. Research into centralised Traffic Management System approaches has provided insight into how the decision process can be aided digitally. However in centralised approaches the computation time tends to increase with larger problems such that it is no longer a sufficient solution for real-time implementation. That is why researchers have looked into non-centralised approaches and have found that these often perform faster and better. But within these approaches not much attention has been paid to the decomposition approach. Implementing a scientifically founded decomposition into a non-centralised approach could lead to better performance. Something which is supported by the interest in the field of road traffic management.

On top of that, the organisational structure of the current state of practice already represents a non-centralised system. Similarly to the state of the art, the decomposition has not been founded scientifically. Doing research into a potentially more efficient decomposition, could also give valuable insight to the current division of subregions.

For those two reasons this thesis will research a new decomposition approach in non-centralised railway traffic management and how this will impact the performance of the conflict resolution.

3. An Hourly Traffic Density Based Decomposition approach for Non-Centralised Railway Traffic Management

This chapter discusses the development of a decomposition approach for non-centralised railway traffic management. The concepts of decomposition are introduced, together with the design choices of the model, section 3.1. The mathematical formulation in literature is elaborated, section 3.2. After that the proposed model is discussed, section 3.3. The chapter is concluded with the model verification and applying the model to an example, section 3.4.

3.1 Introduction to decomposition approaches in railway traffic management

The decomposition aspect of a non-centralised railway traffic management problem concerns how the physical systems will be divided among the agents in the cyber layer. This decomposition is done on the basis of certain criteria signifying the characteristics considered in the decomposition process. The previous chapter has described the current state of the art on the subject and this research aims to expand and improve that current knowledge.

The goal of developing a decomposition approach is to gain insight on the impact a new decomposition would make on a non-centralised railway traffic management system, potentially it could improve computation time and performance. Also, a theoretically based decomposition approach can give valuable insights in the current state of practice. The idea is to create a decomposition that will *equally divide decisional complexity among the agents*, to improve computation, performance and flexibility.
As explained in subsection 2.3.4, different decompositions aim for different end results: subproblems, time intervals, station and line agents etc. This research develops a decomposition that results in subregions. One main reason for this choice is that this is also how the railway traffic management problem is decomposed in the current state of practice. Having the same end result makes the research comparable to the state of practice and gained insights can more quickly and easily be applied to the current practice. Another main reason is that this leads to a static decomposition, in the sense that once the problem has been decomposed the decomposition will not change [9]. For example, a train based or time interval based decomposition will need to be updated frequently to stay relevant. A train based decomposition is different for every timetable. A time interval based decomposition is even different for every scenario, which means it will have to be updated for every delay scenario. Such an approach may be interesting when, for example, the traffic within the network varies strongly throughout the day. However, this updating will require computational power and thus be at the expense of the computation time. Because this research looks at the real-time implementation of a traffic management system, it would not be beneficial to sacrifice computation time for a dynamic decomposition approach. Furthermore, the Dutch timetable uses hourly cycles that are repeated throughout the day and as such the schedule does not contain large irregularities.

As stated, the aim of developing a decomposition approach is to equally divide decisional complexity among the agents, to improve computation, performance and flexibility. It is on the basis of this aim that the criteria for the decomposition are defined. Previous research on decomposition approaches resulting in subregions is done in Chen et al. (2020) [59] and Luan et al. (2020) [9], which have both been discussed in the previous chapter. Both approaches have mainly focused on the infrastructure and have not equally considered train traffic in their decomposition approaches.

The model in both approaches is formulated as an optimisation problem. The decomposition approach in the current research also takes the form of an optimisation model that assigns parts of the infrastructure to certain subregions, drawing inspiration from the geographical decomposition presented in Luan et al. (2020) [9]. However it will include the railway traffic system as a whole and not only the infrastructure, by aiming to divide the decisional complexity among agents.

3.2 Mathematical decomposition modelling for railway traffic management

A decomposition model has the purpose of dividing the physical system between the agents. Because in this thesis railway systems are being viewed, the physical system consists of the infrastructure and the trains travelling through it. Also, it has been determined that the end result of the decomposition are subregions. This means that the regarded region of infrastructure will be divided among the agents in the form of subregions. To do this the decomposition model is formulated as an optimisation model, where the optimal division is found according to the determined criteria. An optimisation model consists of three main elements: the objective function, the constraints and the decision variables. The objective function describes the terms to be minimised or maximised; the constraints form the conditions which must be fulfilled during that optimisation and the decision variables are the variables that can be varied by the model to comply with the constraints and optimise the objective function.

The decomposition model presented in this thesis is inspired by the geographical decomposition model presented in Luan et al. (2020) [9]. This is because the paper is one of the few that have focused on the decomposition aspect of non-centralised railway traffic management. Within those publications, it exceeds in the fact that the decomposition model is discussed in an elaborate and detailed manner. Also, being published in 2020, it is part of the most recent developments and since then has received a considerable amount of citations. This makes it a good candidate to form the basis of the current decomposition model. The general idea of the geographical decomposition model presented in Luan et al. (2020) [9] is described below, to provide context to the proposed decomposition. The aim of their presented decomposition is to divide a region into subregions, by assigning block sections within the region to a subregion. A block section is section of tracks in the infrastructure, that are used by trains to maintain safe distances, as discussed in subsection 1.3.1. A **region** refers to the whole network that is viewed within the decomposition and a **subregion** is a region within that region which is constructed during the decomposition.

As an input the model uses: the block sections within the region; the route of a train service formed by the consecutive block sections and the desired amount of subregions. This results in an output of block sections being assigned to a certain subregion. The main decision variable in the model is the assignment of a block section to a subregion. The assignment is based on two criteria: (1) the amount of subregions a train service passes through in its route and (2) the balancing of the subregion sizes. These criteria are incorporated as two individual terms in the objective function, as follows:

(1) how many times train services cross a border between two subregions within their routes

(2) the difference between the amount of block sections assigned to subregions and the average amount of block sections per subregion

The constraints in the model are used to imply relations between the main decision variable and auxiliary decision variables. Also, there is a constraint that dictates that every subregion should contain at least one block section, making sure that the region is indeed divided into the desired amount of subregions.

The model by Luan et al. [9] shows how a decomposition model can be set up, but also has limitations. It considers trains services and not individual trains, but different train services can travel at different time intervals ranging from a few times a day to once every ten minutes. Both frequencies are regarded the same in the model, whereas the latter will have substantially more effect on the network. Therefore a distinction must be made between train services and individual trains. Also, the second criterion only considers region size in terms of block sections. Considering only the block sections means only the infrastructure is regarded and not the other part of the physical railway system: the train traffic. Because the decomposition looks at block sections it can be said it operates on a microscopic scale. Implementing a non-centralised railway traffic management system generally becomes of interest when regarding larger networks, meaning that conducting a decomposition on a microscopic scale can result in large amounts of data. A decomposition on a meso- or macroscopic scale could significantly decrease that. In general, the research presented by Luan et al. [9] looks at the performance of the different decompositions they regard, among which the discussed geographical decomposition, only in combination with coordination optimisation algorithms and measure the performance only according to computational KPIs such as the optimality gap. Thus the research presents little knowledge on the impact of the different decompositions on the conflict resolution problem.

3.3 The proposed decomposition model based on hourly traffic density

The decomposition proposed in this thesis has as a goal to incorporate both the infrastructure and the railway traffic. By doing so it aims to divide the decisional complexity evenly among the subregions, something that has not yet been extensively studied in literature or practice. This is done by including the hourly traffic density in the decomposition as a criterion, which aims to effectively capture the complexity of the railway traffic management problem at that part of the infrastructure.

3.3.1 The criteria of a decomposition approach distributing decisional complexity

The choice for hourly traffic density as a criterion is a result of a series of experiments with different criteria that are explained in this subsection. The Delft train station is used as an example throughout the elaboration of the experiments. This station forms a part of the Rotterdam Centraal - Den Haag Centraal corridor, see Figure 3.1, which for the purpose of these experiments is decomposed. In this decomposition the stations in the corridor are assigned to 4 subregions, the same amount as the train dispatcher subregions it encompasses.

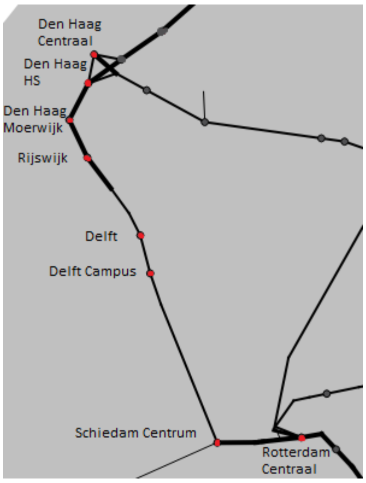


Figure 3.1: The Rotterdam Centraal - Den Haag Centraal corridor, edited from [70]

The first experiments are done with the amount of parallel tracks of a station as a criterion, which mainly reflect the infrastructure. This means that the decomposition is instructed to divide the stations in the corridor among the subregions while minimising the variance in size of the subregions, so that the subregions are around the same size. In the amount of parallel tracks criterion, large stations with a large number of tracks are weighed as very complex in the subdivision. However a large number of tracks does not say everything on the complexity of the railway traffic management problem. The Delft train station currently has only two parallel tracks, meaning that the criterion of amount of tracks will consider it much smaller than, for example, Rotterdam Centraal which has ten parallel tracks. However the Delft train station forms a bottle neck within the Rotterdam Centraal - Den Haag Centraal corridor, so the 'amount of tracks' criterion would not do the complexity of the RTM problem justice. Leading from this conclusion, experiments are done based on train traffic using hourly traffic as a criterion, i.e. trains per hour. This means that for every station it is considered how many trains cross it every hour. For the Delft station this means 10 trains per hour. Although this number does give insight into how busy a station is, it can still highly vary how those 10 trains interact within the infrastructure. To take it into extremes: If those 10 trains each have a separate track, so 10 parallel tracks, it would need little to no rescheduling actions in the case of delay, but if those 10 trains all have to share one track a small delay can already have large implications. That is why the experiments lead to the final criterion of **hourly traffic density** in terms of trains per track per hour. This combines both the infrastructure and the train traffic travelling over it. This describes the RTM problem complexity in the sense that it shows how likely conflicts are to occur. A high ratio of hourly traffic density means that these trains will not have large headways between them and that a conflict is likely. Another way of looking at it, is by describing it as the ratio between supply and demand. It describes to what extent the supply of infrastructure can satisfy the demand of the traffic operating companies. From this view, a high hourly traffic density means that the infrastructure is not fully able to satisfy the demand of the traffic flow. Looking at the example, Delft has a hourly traffic density of 5 trains per track per hour, the highest of the Rotterdam Centraal - Den Haag Centraal corridor. Thus reflecting that it forms a bottle neck. As a result it is weighed as complex in a decomposition.

It is interesting to remark that this observation is related to current developments where Delft station is being renovated to include an extra set of tracks to come to a total of four parallel tracks.
In subsection 2.3.6 it was mentioned that the road traffic domain has already considered decomposition based on traffic patterns. Similarly, the field has publications already mentioning the term hourly traffic density. Here it is defined as the number of vehicles per kilometer averaged over an hour. For example Williams et al. (2016) [71], Sayegh et al. (2016) [72] and Son et al. (2018) [73] use the term to research the influence of traffic density on air pollution. Sørensen et al. (2021) [74] use the hourly traffic density to study characteristic of nearby charging facilities for electric vehicles. Although these applications do not necessarily align with the railway system decomposition of the current research, this does show that hourly traffic density is a common term that can also be used in this thesis.
The hourly traffic density is seen as a way to capture the complexity of the RTM problem. In essence reflecting how busy that piece of infrastructure is, the criterion can give the first insights into the potential of decisional complexity based decomposition and the impact that such a decomposition has on the conflict resolution problem. Working from these insights, it could be considered in later studies to research other characteristics that reflect the decisional complexity.

Aside from hourly traffic density, the decomposition uses a second criterion: the **subregion crossings**. The subregion crossings refers to how many individual trains cross from one subregion to another. For example, if a train service passes through three different subregions and has a frequency of 4 trains per hour, 4 trains are passing two subregion borders every hour. This contributes to 8 subregion crossings. This is one of the criteria also used in Luan et al. (2020) [9] and it serves two main functions. The first is that it minimises the amount of interactions between agents. Each agent is responsible for a subregion and if one train crosses from one subregion to another, that may be cause for an interaction between agents. An interaction will cost computation time and efficiency, thus they are aimed to be kept at a functional minimum. Secondly, this criterion serves the purpose of keeping the subregions together and avoiding that different parts of a subregion are scattered around the network. This is because in order to minimise the number of subregion crossings, the model will assign neighbouring stations to the same subregion.

3.3.2 Inputs and outputs to the proposed decomposition model

Before diving into the specific formulation of the decomposition model, its inputs and outputs are explained to give a general impression. This visualised in Figure 3.2.

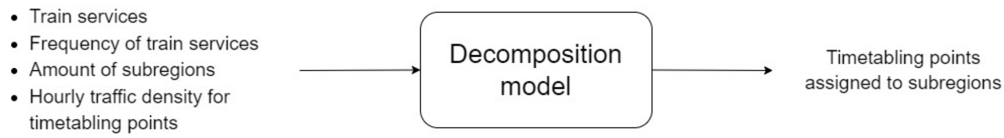


Figure 3.2: The inputs and outputs to the proposed decomposition model

The output of the model is that a region is divided into subregions, or in other words, that elements of the infrastructure are assigned to subregions. The question of what these infrastructural elements entail, directly leads to what the inputs are to the model. Contrary to previously presented microscopic decomposition models, this approach will take a macroscopic approach. Theoretically speaking the subregions could be composed of elements any size: block sections, junctions, stations, the links between stations or even stations with a corresponding link. The infrastructural elements in this model are based on 'dienstregelpunten' (**timetabling points**). These are points in the infrastructure that are recognised to need scheduling actions by network administrator ProRail. These actions can be arrivals, departures, stops or pass throughs, so generally the timetabling points are the extraordinary parts of the infrastructure such as stations and junctions. Often, two timetabling points are connected by a link called 'vrije baan', where the train dispatchers do not have jurisdiction to manage the traffic. On these links the traffic does not need to be managed, because they often consist of parallel tracks with no option to switch between them and the lineside signals work automatically. To form the infrastructural elements used in the proposed decomposition, the links are aggregated with the timetabling points. Because by definition there is no railway traffic management in these links, which timetabling point they are attributed will likely be of no large consequence to the effect of the decomposition on the conflict resolution. In the remainder of this thesis the term timetabling point will refer to the aggregated infrastructural element of a timetabling point and its respective link.

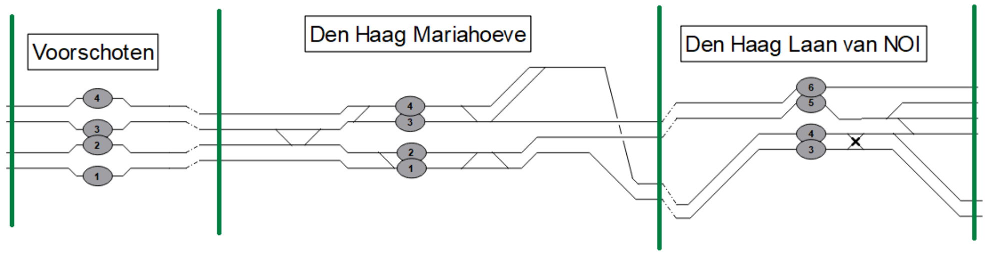


Figure 3.3: The timetabling points Voorschoten, Den Haag Mariahoeve and Den Haag Laan van NOI and their borders, edited from [70]

As the timetabling points are also the points of interest for railway traffic management, they form a suitable definition of the infrastructural elements in the decomposition. Additionally, this definition will lead to subregions of similar size as the TDL regions discussed in section 2.1. Making a comparable end result to the state of practice, will allow gained insights to be quickly and easily applied to the current practice and other parts of the network. An example of these timetabling points is shown in Figure 3.3.

An important input is the hourly traffic density. The hourly traffic density is a characteristic given to each timetabling point, thus representing the complexity of the railway traffic management problem in that area. It is calculated by dividing the hourly traffic flow in the area by the amount of tracks the area has. In other words it is the trains travelling through the area every hour divided by the number of tracks. If trains stop at a timetabling point, because it is for example a station, these stopping trains are taken into account with a factor 2. The distinction is made in the first place, because not all trains stop at all stations. Intercity trains will only pass through the smaller sprinter stations and freight trains only occasionally stop at a station. A stop means that a train will require an additional scheduling action and therefore forms a complexer part of the railway traffic management problem. This additional scheduling action is why a stopping train is included in the hourly traffic with a factor of 2. The amount of tracks of the area that is taken into account for the hourly traffic density is the constraining number of tracks. Using the constraining amount of train tracks will lead to the maximum value of hourly traffic density in that area and will thus reflect the most complex situation in that area.

Also, the routes of the train services within the region are introduced into the model. This is done by stating whether a train service crosses an individual timetabling point. Because the timetabling points are ordered in the input, the aggregation of this information leads to the route of the train service. In this parameter, no distinction is made in the direction of the train, because a distinction in heading is not made in the model. Moreover, a distinction does not need to be made as long as the same scheduling points are crossed in both directions.
Another input, is the frequency of the aforementioned train services. This gives the model insight into which train services might have a higher importance in the decomposition. This frequency is defined as number of trains per hour, considering the worst case number of trains. This means that if a train travels once an hour during rush hour, it is taken into account as one train per hour in the frequency. This is also where the directions of trains come into play. If a train service travels twice per hour one way and twice per hour the other way, it is noted as four trains per hour travelling for that train service.
The last input into the decomposition model is the amount of subregions the region is desired to be divided into. This is a prerequisite to the model and thus needs to be determined beforehand. The chosen number of subregions can, for example, be based on the amount of train dispatcher subregions used in the current state of practice for the respective region.

3.3.3 Model formulation of a hourly traffic density based decomposition approach

The decomposition model is based on the decomposition model presented in Luan et al. (2020) [9]. Compared to their model, the proposed models has as an altered objective function. Also, including an extra constraint (Equation 3.8) was found necessary for desired functioning of the model.

The aim of the decomposition model is to assign timetabling points to subregions. This assignment is captured in decision variable μ_i, which has as a value the number of the subregion that timetabling point i is assigned to. This assignment is based on the the minimisation of the objective function (Equation 3.4), which consists of the weighted sum of two criteria: the size variance between the subregions in terms of the hourly traffic density and the number of trains crossing subregion borders per hour.

This minimising is enabled by choosing the values of two auxiliary decision variables: γ_{fj} and α_{ri}. Where γ_{fj} represents whether train service f crosses a subregion border at point j and α_{ri} describes whether timetabling point i is assigned to subregion r. With these two decision variables the two criteria are formulated as follows. The number of trains crossing a subregion per hour is computed by multiplying γ_{fj}, whether train service f crosses a subregion border, with the frequency of train service f, ρ_f. The minimisation of this criterion, leads to as little trains as possible crossing subregion borders and subsequently that the model will choose neighbouring timetabling points to belong to the same subregion as much as possible. The size variance in terms of hourly traffic density is formulated by, for every subregion, calculating the difference between the size of subregion r and the average size of a subregion: the sum of the hourly traffic density t_i of all timetabling points i divided by the number of subregions. The sum of all these differences formulates the criterion. By minimising this size variance, the hourly traffic density will be divided as equally as possible over all the subregions.

The model is constrained by a series of constraints, which serve one of two functions: either defining a relation between the decision variables γ_{fj}, α_{ri} and μ_i, or constraining the assignment problem. The constraints in Equation 3.5 and 3.6 belong to the first category. The decision variable γ_{fj} and α_{ri} are auxiliary decision variables that help to formulate the objective function, but essentially both decision variables directly relate to the assignment of timetabling points to the subregions captured in μ_i.
Equation 3.5 relates the assignment of the timetabling point i in μ_i to if train service f crosses a subregion at point j. This is done by introducing the parameter B_{fji}, which describes elements of the route matrix B_f of every train service f. In this matrix every row represent a trip from one timetabling point to another and every column represents a timetabling point, such that the size of route matrix B_f is $(||E_f|| - 1\,x\,E)$, where E describes the set of all timetabling points i and E_f is the set of all timetabling points included in the route of train service f. The functioning of these parameters is illustrated by an example. Considering a region containing timetabling points A through E and a train service 1 that travels over all timetabling points except C, as in Figure 3.4. The set E and parameters E_f and B_f can be defined as in Equation 3.1, 3.2 and Equation 3.3 respectively. In Equation 3.3, the elements outside the matrix show what timetabling points the columns refer to and which trips the rows refer to.

$$E = \{A, B, C, D, E\} \tag{3.1}$$

$$E_1 = \{A, B, D, E\} \tag{3.2}$$

$$B_1 = \begin{array}{c} \\ \text{A} \to \text{B} \\ \text{B} \to \text{D} \\ \text{D} \to \text{E} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{A} & \text{B} & \text{C} & \text{D} & \text{E} \end{array} \\ \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \end{array} \tag{3.3}$$

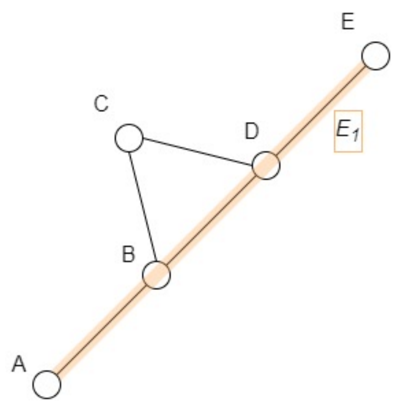


Figure 3.4: Example region A-E and train service 1

Equation 3.6 relates the assignment vector μ_i to decision variable α_{ri}, which shows if timetabling point i is assigned to subregion r. The large difference between the two decision variables is that μ_i shows the subregion number every timetabling point i is assigned to and α_{ri} shows for every subregion r and all timetabling points i if the timetabling point is assigned to the subregion or not.
The next two constraints have as a main function to constrain the assignment problem. Firstly, Equation 3.7 ensures that every subregion r contains at least one timetabling point i. This is to make sure that there will be no empty subregions and that the model uses the amount of subregions it is given at the initialisation. The second constraint in Equation 3.8 ensures that all timetabling points i are assigned to a subregion r. Omission of this constraint can leave one or multiple timetabling points that are not assigned to a subregion.

The mathematical formulation can be summarised as below. In this formulation the notations $|x|$ and $||X||$ are used. With $|x|$ it is indicated that the absolute value of x is taken. The notation $||X||$ indicates the cardinality of set X, or in other words the size of set X. Also in this formulation a distinction is made between functional constraints and additional constraints. The **functional constraints** serve to constrain the solving space of the optimisation problem. The **additional constraints** serve to define the values that the decision variables can take.

Indices

- f for train service f in train service set F
- i for timetabling point i in infrastructure set E
- j for timetabling point j in train service route E_f of train f
- r for subregion r in subregion set R

Sets

- F set of train services $\{trainservice_1, ...\}$
- E set of timetabling points $\{e_1, ..., i\}$
- R set of subregions $\{1, ..., r\}$

Parameters

- B_{fji} the route matrix for every train service f indicating which timetabling points the train service crosses in its route, with $B_{fji} = 0$ if the timetabling point i is not used in the j^{th} trip in the route of train service f; $B_{fji} = 1$ if train service f leaves timetabling point i in the j^{th} trip in the route and $B_{fji} = -1$ if train service f enters timetabling point i in the $(j+1)^{th}$ trip in the route $B_{fji} \in \{-1, 0, 1\}$, [-]
- E_{fj} the timetabling point i used as the j^{th} timetabling point in the route of train service f $E_{fj} \in E$, [-]
- t_i the hourly traffic density of timetabling point i $t_i \in \mathbb{R}^+$, [trains/track/hour]
- ζ the weight factor balancing the two terms in the objective function $\zeta \in (0, 1) \subset \mathbb{R}$, [-]
- ρ_f the frequency of train service f $\rho_f \in \mathbb{R}^+$, [trains/hour]

Decision variables

- α_{ri} the assignment variable of timetabling point i to subregion r: $\alpha_{ri} = 1$ if timetabling point i belongs to subregion r, else $\alpha_{ri} = 0$ [-]
- γ_{fj} if two consecutive timetabling points j and $j+1$ in the route of train service f belong to different subregions: $\gamma_{fj} = 1$ if E_{fj} and $E_{f(j+1)}$ belong to a different subregion, else $\gamma_{fj} = 0$ [-]
- μ_i the region from set R that timetabling point i belongs to [-]

Objective Function

The objective function consists of two elements. The first element is to minimise the amount of subregions train service f crosses when travelling over all timetabling points j in its route. The second element is to balance the hourly traffic density among all subregions r, more specifically to minimise the difference between the hourly traffic density over all timetabling points i in subregion r and the average hourly traffic density per subregion. The objective is to minimise the weighted sum of both elements.

In the first term, the sum is taken of binary decision variable γ_{fj} over all trips from timetabling point j to timetabling point $j+1$ in the route of train service f to give the amount of subregion borders crossed by that train service. Multiplying that by frequency ρ_f, gives the amount of individual trains crossing subregion borders per hour. This is then summed over all train services. In the second term, the size of subregion r in terms of hourly traffic density is calculated by multiplying the decision variable α_{ri} with the hourly traffic density t_i of timetabling point i and taking the sum over all timetabling points i. This is subtracted by the average subregion size: the sum of the hourly traffic density t_i over all timetabling points i divided by the number of subregions, $||R||$. The absolute value is taken of this difference between subregion size and the average subregion size and summed over all subregions r.

$$\min \left[\zeta \cdot \sum_{f \in F} \left(\sum_{j \in [1, ||E_f||-1]} (\gamma_{fj}) \cdot \rho_f \right) + (1-\zeta) \cdot \sum_{r \in R} \left| \sum_{i \in E} (\alpha_{ri} \cdot t_i) - \frac{\sum_{i \in E} t_i}{||R||} \right| \right] \tag{3.4}$$

Functional Constraints

1. This constraint defines the value of binary variable γ_{fj} for all train services f and all timetabling points j in the route of train service f. If the timetabling point j and $j+1$ in the route of train service f are in the same subregion the term $\left|\sum_{i\in E}(B_{fji}\cdot\mu_i)\right|$ will be zero, because the positive and negative values of the multiplication cancel each other out in the summation. Therefore the entire fraction will be zero and thus so will γ_{fj}. In contrast when timetabling point j and $j+1$ belong to a different subregion, the term will become a positive, nonzero value which will give the fraction a value higher than 0 but less than or equal to 1. Therefore γ_{fj} will be 1. Summarising, if timetabling point j and $j+1$ in the route of train service f are in different subregions $\gamma_{fj}=1$, else $\gamma_{fj}=0$

$$\frac{\left|\sum_{i\in E}(B_{fji}\cdot\mu_i)\right|}{||R||-1}\leq\gamma_{fj}\qquad\forall j\in[1,||E_f||-1]\;\forall f\in F \tag{3.5}$$

2. This constraint defines the value of binary variable α_{ri} for all subregions r and all timetabling points i. If a specific timetabling point i belongs to a specific subregion r then $\mu_i=r$, as a consequence $|\mu_i-r|=0$. Therefore the fraction will also be equal to zero and thus α_{ri} will be equal to 1. Contrary, if a specific timetabling point i does not belong to a specific subregion r, the fraction will have a positive, nonzero value not larger than 1. This will ensure α_{ri} equals zero. Summarising, if timetabling point i belongs to subregion r $\alpha_{ri}=1$, else $\alpha_{ri}=0$.

$$\alpha_{ri}\leq 1-\frac{|\mu_i-r|}{||R||}\qquad\forall r\in R, i\in E \tag{3.6}$$

3. This constraint ensures that every region r should contain at least one timetabling point, by stating that the sum of α_{ri} over all timetabling points i should be at least one.

$$\sum_{i\in E}\alpha_{ri}\geq 1\qquad\forall r\in R \tag{3.7}$$

4. This last constraint dictates that the sum of α_{ri} over all subregions r should be exactly one. This entails that all timetabling points are assigned to exactly one subregion, so that no timetabling point is left unassigned.

$$\sum_{r\in R}\alpha_{ri}=1\qquad\forall i\in E \tag{3.8}$$

Additional Constraints

The binary variables can have a value of either 0 or 1:

$$\alpha_{ri}\in\{0,1\}\quad\forall r\in R, i\in E \tag{3.9}$$

$$\gamma_{fj}\in\{0,1\}\quad\forall j\in[1,||E_f||-1]\forall f\in F \tag{3.10}$$

The variable μ_i can have the value of the subregions in set R:

$$\mu_i\in R\quad\forall i\in E \tag{3.11}$$

3.3.4 Assumptions of the proposed decomposition approach

The mathematical model, as is, makes a series of assumptions. Each of which is discussed below:

- *This model assumes a fixed number of subregions.* This means that the number of subregions has to be predetermined in set R. The current formulation does not include finding which amount of subregions would be most effective.
- The infrastructure is considered at macro scale. This means that *microscopic details are not included in the decomposition.* Defining the timetabling points by means of timetabling points means that the decomposition can only assign those timetabling points to a subregion. It is for example not considered if half a station could belong the another subregions than the other half, or when considering parallel tracks, that one track could belong to one subregion and the other track to another.

- The current formulation of the parameters *relies on a (relatively) regular hourly schedule*, more specifically the hourly traffic density t_i and the train service frequency ρ_f. Considerable assumptions would need to be made for those parameters if the model were to be applied to a highly irregular schedule. The Dutch schedule, the scope of this research, does consist of an hourly schedule, meaning that roughly the same schedule is repeated every hour. Some discrepancies may occur during rush hour or night time, when the schedule is adjusted. In this research, for both the hourly traffic density as the train service frequency, the worst case scenario is considered, meaning the highest possible amount of trains in the day, which is often during rush hour.

3.3.5 Implementation of the mathematical decomposition formulation

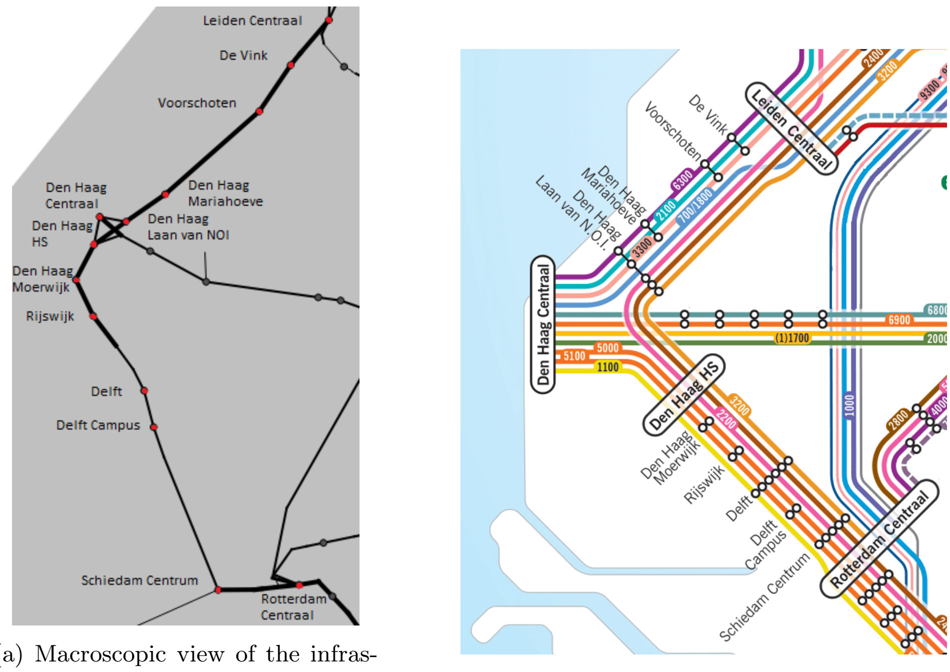
The above mathematical model has been implemented in Python using the Gurobi commercial solver, which uses the simplex optimisation algorithm. Because the Gurobi solver can only accept equations in certain format, the implemented model contains some auxiliary variables and constraints. The specifics are as follows:

Constraints Equation 3.6 and Equation 3.5 and objective function Equation 3.4, contain absolute values of terms with a variable. This can not be implemented as is into Python and Gurobi. For each term two additional variables and constraints have to be made. One variable to describe the respective term and a constraint to couple the variable to the term. An additional variable is needed to then describe the absolute value of the previous variable and another constraint will perform the operation. This last variable represents the absolute value of the term and can be used in the equation. Details on the implementation of the decomposition model are provided in Appendix C.

3.4 Verification of the hourly traffic density based decomposition approach

To check if the mathematical model has been implemented correctly, the implementation is verified. Also the model is illustrated by applying it to an example regions.

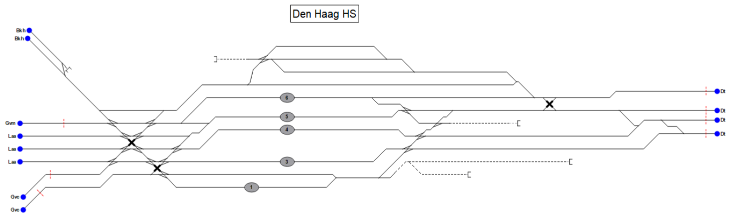
The example region that is discussed is the region from Rotterdam Centraal to Leiden Centraal, see Figure 3.5a. This region forms a part of the Oude Lijn corridor, which reaches from Amsterdam Centraal to Rotterdam Centraal via Haarlem. The train services are visualised in Figure 3.5b. In the figure, a line represent a train service travelling twice per hour. It should be noted that freight trains are not included in this visualisation.



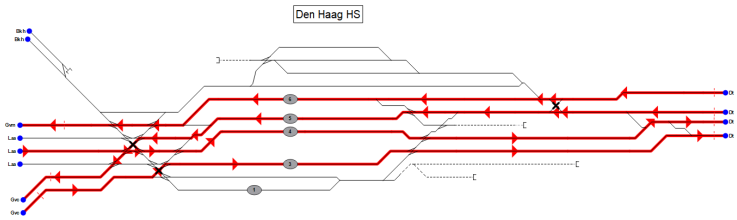
(a) Macroscopic view of the infrastructure, edited from [70]

(b) Visualisation of the train services [14]

Figure 3.5: The Rotterdam Centraal-Leiden Centraal corridor



(a) The infrastructure at Den Haag Holland Spoor station



(b) The traffic flow through Den Haag Holland Spoor station

Figure 3.6: The situation at Den Haag Holland Spoor station [70]

How the hourly traffic density is determined is illustrated by means of an example of Den Haag Holland Spoor station, which is seen as a timetabling point in the decomposition. Figure 3.6a demonstrates the infrastructure at the railway station, where the grey circles show platforms. Here it can already be seen that the amount of parallel tracks varies throughout the timetabling point: The most left counts 7 tracks, the most right 4 and the middle 10. Understandably, the hourly traffic density will vary accordingly, with the choice for either of these numbers. The choice is made to use the constraining number of tracks, because this is the bottleneck of the timetabling point. In the case of the example this means the 4 tracks on the most right side are used.
Looking at the traffic flow data, this station, similar to the entire Dutch network, has an hourly cyclic schedule. In this part of the network the train traffic per hour is the same every hour throughout the day. There are no additional passenger trains during rush hour and no additional freight trains outside of rush hours. Therefore a regular hour can be used to determine the hourly traffic flow value to be used in the hourly traffic density. A total of 28 trains travel through the timetabling point every hour, see Figure 3.6b. Of these trains all intercity's and sprinters, 26, make a stop at the station and the 2 freight trains pass through without stopping. The factor 2 for stopping trains is taken into account by multiplying the trains that only pass through with 0.5 . This means the hourly traffic, or trains per hour, comes to 27, see Equation 3.12. This gives a final hourly traffic density of 6.75 trains per track per hour as calculated in Equation 3.13.

$$hourly\,traffic = 26\,stopping\,trains + 0.5 \cdot 2\,passing\,trains = 27\,trains/hour \tag{3.12}$$

$$hourly\,traffic\,density = 27\,trains/hour\;/4\,tracks = 6.75\,trains/track/hour \tag{3.13}$$

Doing this for all timetabling points in the region, gives the hourly traffic density values as an input to the model as in Table 3.1. Using these inputs the region Rotterdam Centraal - Leiden Centraal is decomposed with the proposed model. This leads to the decomposition as shown in Figure 3.7. In this decomposition the criteria have the following values: the sum of the difference between the hourly traffic density in the subregion and the average hourly traffic density per subregion is 14.9 and the number of trains every hour crossing a subregion in their route is 92. The weight factor is chosen such that the overal solution has a low objective value. The solution for which the total objective value is lowest, is from $\zeta = 0.3$ up till $\zeta = 0.7$, which all give the same solution. Following this observation, $\zeta = 0.5$ is used as the weight factor for this example and verification.

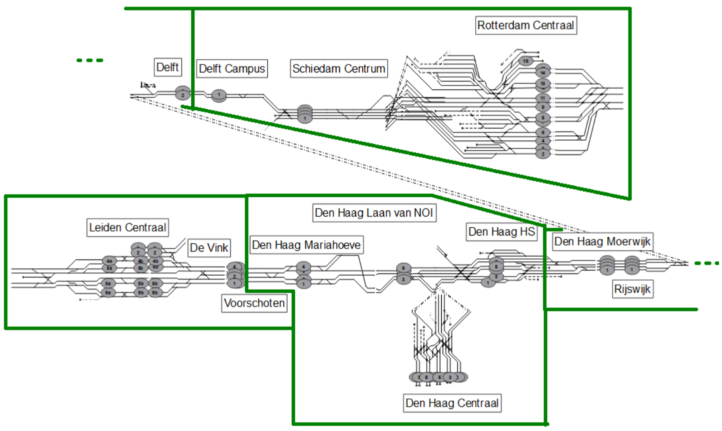


Figure 3.7: Decomposition of the Rotterdam Centraal - Leiden Centraal corridor, edited from [70]

Timetabling point	**Hourly traffic density [trains/track/hour]**
Rotterdam Centraal	6.75
Schiedam Centrum	8.33
Delft Campus	9.00
Delft	13.50
Rijswijk	4.50
Den Haag Moerwijk	4.50
Den Haag HS	6.75
Den Haag Centraal	6.00
Den Haag Laan van NOI	7.75
Den Haag Mariahoeve	6.00
Voorschoten	6.00
De Vink	6.00
Leiden	8.67

Table 3.1: Hourly traffic density values for the timetabling points in the Rotterdam Centraal - Leiden Centraal corridor

Using the example of the Rotterdam Centraal - Leiden Centraal region the model is verified. This is done by changing parameters to extremes and comparing the output to the baseline, as seen in Table 3.2. The decomposition described above is used as the baseline. It is hypothesised how the actions will impact the outcome of the model on the basis of the mathematical model. The outcome is signified by the two terms of the objective function: amount of total subregion crossings and the deviation in size between the subregions. Then it is determined if the implementation passes or fails the verification step, according to if it corresponds to the hypothesis.

Verification step	**Hypothesis**	**Outcome**	**Pass or Fail**
ζ=0	SC > 92; SS < 14.9	SC = 360; SS = 2.2	Pass
ζ=1	SC < 92; SS > 14.9	SC = 76; SS = 69.8	Pass
$\|\|R\|\| = \|\|E\|\|$	SC >92; SS > 14.9	SC = 360; SS = 44.8	Pass
$t_{Voorschoten} = 40$	SC $\geq$ 92; SS > 14.9	SC = 92; SS = 30.9	Pass
$\rho_{Rotterdam-DenHaag} = 60$	SC > 92; SS $\geq$ 14.9	SC = 148; SS = 49.9	Pass

Table 3.2: The verification steps, hypotheses and outcomes for the decomposition model
SC: Subregion crossings; SS: Subregion size distribution

In the first two verification steps the weight factor ζ is set to two extremes, 0 and 1. For these cases the two objectives are alternatingly the sole objective of the optimisation. Therefore it would be expected that the objective that is included will be minimised, while the minimisation of the other objective is disregarded. This would mean that for $\zeta = 0$ only the balancing of the subregion size is minimised and that for $\zeta = 1$ only the subregion crossings is minimised. In the third verification step, the amount of subregions is set equal to the number of timetabling points in the region, or mathematically $||R|| = ||E||$. Because the model contains a constraint that no subregion should be left empty (Equation 3.7), the expectation is that each subregion will only contain one timetabling point. The next step tests the case where the hourly traffic density of one individual timetabling point t_i is set to be much larger than the coefficient of all the other timetabling points. This is expected to primarily have an effect on the balancing of the subregion size, because the subregion that this timetabling point is assigned to will likely deviate much from the average subregion size. As a secondary effect, the number of subregion crossings may increase. The timetabling point with a high hourly traffic density is expected to lead to a decomposition where it is included in a different subregion than in the baseline. Most likely the subregion it belongs to will now contain less timetabling points to maintain a balancing of the subregion sizes. Less timetabling points in a subregion may lead to more subregion crossings, but it may also have no effect at all. Following the same line of thought it is tested what happens if the frequency of one train service ρ_f is set much larger than the frequency of other train services. According to expectation, the amount of trains crossing subregions will increase because there are now many more trains travelling within the region. Also it is likely that the decomposition will change as a result of this high frequency, such that the amount of subregion crossings in this high frequency train service is kept to a minimum.

Table 3.2 shows that the decomposition model has passed all the verification tests. During the verification some interesting observations were made that should be taken into account when using the proposed decomposition model in addition to the assumptions mentioned in subsection 3.3.4. The considerations are:

- With a low value of ζ, the amount of subregion crossings becomes less important in the optimisation. This means that the optimisation will focus on balancing the subregion sizes, besides that another effect is also observed. This first term in the objective is not only a criterion to minimise the amount trains crossing subregion, but it also serves a purely functional purposed. By minimising the amount of subregion crossings this term ensures that the subregions can only consist of neighbouring timetabling points. In other words it prevents that a subregion consists of timetabling points scattered round the network. In that case the model would mathematically be feasible, but it would be nonsensical in practice. So it should be considered when implementing the decomposition model that the subregion crossing is minimised a significant enough amount such that it can ensure that the subregions are not scattered around the network. This way the decomposition output can be useful in practice.
- The decomposition model does not act in the case of an infeasible route of a train service, because it is not linked to the connectivity of the infrastructure. For example, it could be inputted that a train would travel from Rotterdam Centraal to Delft without crossing Schiedam Centrum. With the current infrastructure this would be physically impossible to occur in practice. However the knowledge of the decomposition model on the routes of the trains is limited to the routes it is inputted. So if a physically infeasible route is given as an input, the model will not be able to distinguish this physical infeasibility in any way and simply consider it possible. Therefore it is important to see to it that all routes are physically feasible before providing them to the decomposition model.

3.5 Conclusion on an hourly traffic density based decomposition approach for non-centralised railway traffic management

This chapter has answered the research question: *How can a region of railway infrastructure be decomposed into subregions based on hourly traffic density?*. The concepts of decomposition in non-centralised railway traffic management were introduced and the design choices of the proposed model were discussed. The general mathematical modelling of a decomposition model was elaborated, supported by the model formulation in Luan et al. (2020) [9]. A new decomposition approach using the hourly traffic density as an important criterion was proposed. The model was verified and demonstrated on an example.

By using the hourly traffic density, the infrastructure, train services and train service frequency, the decomposition model provides a division of a region into subregions, according to the criterion of hourly traffic density. The model assumes a fixed number of subregions, infrastructure on a macroscopic scale, a (relatively) regular hourly schedule and known infrastructure and train schedules. After implementation in Python using the Gurobi commercial solver, the decomposition proves to be able to efficiently decompose a part of the Oude Lijn corridor. During verification the decomposition behaves as expected. It surfaces that it is important to supervise the practical feasibility of the inputted routes and the outputted decomposition.

The proposed decomposition model contributes to research and practice by presenting a decomposition based on the decisional complexity in terms of hourly traffic density. It shows to have results compliant with observations in practice, such as Delft station forming the bottleneck of the Rotterdam Centraal - Den Haag Centraal corridor. These types of observations seem to not yet been included in current research on decomposition in railway traffic management or in the decomposition of the train dispatcher subregions used in practice. Therefore the proposed decomposition model can give interesting insights.

It is expected that the proposed decomposition model will provide the most insight in difficult to oversee decomposition situations. These situations can for example be when many different routes are to be taken into account, or where traffic flow has been increasing without changes in infrastructure. These cases are most likely to be inefficiently decomposed in practice, because of the human eye can not easily take into account all the different factors.

Overall the proposed decomposition model shows to be promising, both for research and practice. The next chapters of this thesis regard the other aspects of the non-centralised railway traffic management problem and the performance assessment of the decomposition compared to the decomposition currently used in practice.

4. A Non-Centralised Railway Traffic Management Model

The non-centralised conflict resolution model consists of several aspects: the decomposition, the coordination and the conflict resolution. All aspects refer to the agents of which the non-centralised model is built up. The decomposition determines which parts of the network should be overseen by an agent. The coordination will ensure communication between these agents and the conflict resolution is what is happening inside the agent. The first aspect, the decomposition, was discussed in the previous chapter and the remaining two, the conflict resolution and coordination, are elaborated in this chapter.
First the conflict resolution model is introduced, section 4.1. Then the coordination approach is discussed, section 4.2. Lastly, both approaches are illustrated using an example, section 4.3. This example also serves as a frame work for the verification of both approaches.

The focus of this thesis is on the proposed decomposition approach, meaning that the majority of the research was spent on developing this approach. That is why for the coordination and conflict resolution relatively uncomplex approaches are used. These approaches do not need to be the latest developments, but need to suffice to test the decomposition in a non-centralised conflict resolution model.

4.1 A railway traffic management approach for non-centralised conflict resolution

Within conflict resolution modelling, the two prominent approaches are the MILP and Alternative Graph formulation, as discussed in subsection 2.2.4. This research chooses to use the Alternative Graph formulation as used in, amongst others, D'Ariano et al. (2007) [10], Corman et al. (2010) [53] and Lamorgese and Mannino (2015) [7]. There are two main reasons for the choice of using the Alternative Graph model. Firstly, it is a fast and efficient model, which are characteristics that are beneficial for the real-time implementation envisioned for Traffic Management Systems. Additionally, the model has a compact formulation. Generally, non-centralised railway traffic management approaches view relatively large networks, so models with many aspects can become unmanageable. A compact formulation is therefore desirable.

4.1.1 Alternative Graph modelling details

The foundations of the Alternative Graph formulation were already established in subsection 2.2.4. These foundations will be re-established and elaborated. All the principles that are introduced, are illustrated in the example in Figure 4.1.
The Alternative Graph model formulates the conflict resolution model as a graph. Within this graph the nodes represent events of a train doing a certain action at a certain time, e.g. entering a block section or departing from a station. The arcs connecting these nodes, or events, show two things. First, they indicate the succession between two events. Secondly, the weight assigned to the arc specifies how much time must be between those two events at the least. For instance, this weight could entail the running time it takes a train to reach the beginning of the next block section, the dwell time a train has to wait at a station or the time headway that should be between two trains entering the same block section. The principle of the Alternative Graph model relies on the introduction of alternative arcs. The alternative arcs show the succession of events of different trains, i.e. the precedence between two trains. These arcs always come in pairs called alternative arc pairs, where one arc ensures the precedence of one train and the other arc ensures the precedence of the other train. The model states that in the final conflict resolution solution only one arc of this pair can be chosen, ensuring the precedence of only one of the trains.
The previous paragraph forms the foundation of the formulation, but it also contains several other aspects.
The first being that the graph contains two special nodes, which have been added for modelling purposes: the source Node 0 and the sink Node n. **Source Node** 0 represents the starting time stamp, often $t = 0$. From that node an arc goes to the first node of every train. The weight of this arc shows how long it takes from $t = 0$ for the event to occur, so more concretely this weight is the time at which the train is planned to enter the network including its entry delay. **Sink Node** n is used to model the maximum **secondary delay** within the network, which is the maximum delay over all the trains at the scheduled arrival points that is removable with rescheduling also called the **consecutive delay**.

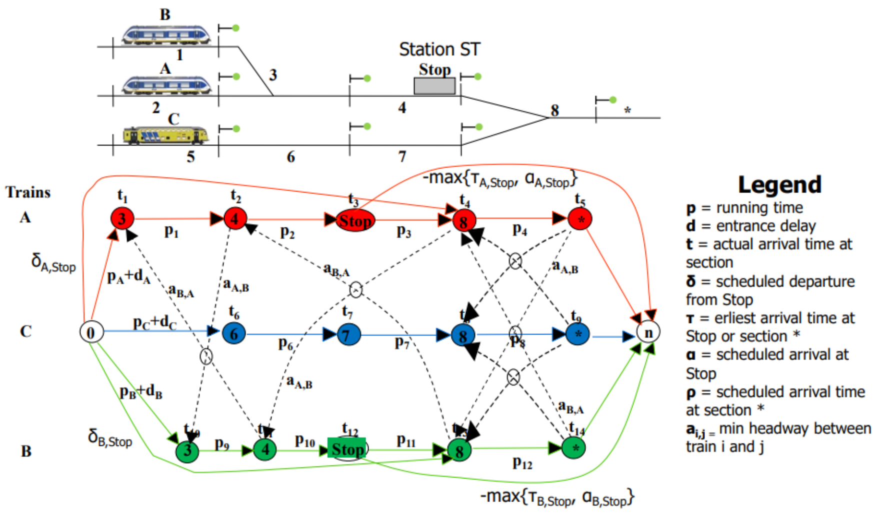


Figure 4.1: An example of an Alternative Graph model [75]

Next the concept of arrival and departure arcs needs to be introduced. These arcs are also known as **timetabling arcs**, because the arrivals and departures are the elements fixed in the timetable. For every departure, an arc goes from Node 0 to the departure node, whose weight is the scheduled departure time. For every arrival, an arc goes from the arrival node to Node n, whose weight of the arc is the negative value of the scheduled arrival time. If initial delays are applied at these arrival points, the weight of the arc is the negative value of the maximum of the scheduled arrival time and the earliest possible arrival time taking into account the *initial* delay.

4.1.2 Conflict resolution with the Alternative Graph model

The conflict resolution has the form of an optimisation problem, meaning it chooses decision variables as to optimise a given objective function. For the Alternative Graph model the objective is to:

choose a selection S of alternative arcs that has no positive length cycle and minimises the length of the longest path from 0 to n, $l^s(0, n)$ [11]

This statement calls for elaboration. The main decision mechanism of the model is making a selection of alternative arcs. As previously mentioned, the alternative arcs come in pairs of which only one arc can be used. The selection of alternative arcs refers to the selection of one of those arcs in the pair. This selection is made to resolve the conflicts between trains with as little delay as possible. To that effect, the model searches for the longest path using all the given nodes and all arcs: fixed arcs, alternative arcs and timetabling arcs. Resulting from the definition of these arcs and nodes, this longest path represents the maximum secondary delay and the sink node Node n will take the value of this maximum secondary delay. Inherent to searches for the longest path, is that the selection of alternative arcs does not contain any positive cycles. An example of a positive cycle would be if the model would at first choose an alternative arc such that train A precedes B, but for the next alternative arc pair chooses such that train B precedes train A, when there is no opportunity in the infrastructure to reorder. Besides that this is physically not possible, the model can then also not identify a longest path. Therefore no positive cycles are included in the selection of alternative arcs S.

In this research the Alternative Graph model is used as presented in D'Ariano et al. (2007) [10], with some notation changes. These changes are made because the conflict resolution is implemented as a MILP problem. The original formulation in [10] uses a constraint notation using "if.. then..." statements and a disjunctive ($\vee$). This is not in line with the Linear aspect of a MILP formulation, so for it to be implemented it would have to be reformulated. As a consequence, it is chosen to present the conflict resolution in this reformulation to have little discrepancies between the mathematical and implemented model. Additionally the reformulated version was deemed as a clearer representation of the workings of the model. Specifically the reformulation applies to: the introduction of decision variable z_{ij}, the reformulation of the constraint in Equation 4.3 and the addition of the constraint in Equation 4.4.

The main objective is to minimise the maximum consecutive delay, which is represented in the equation $t_n - t_0$, meaning the time stamp of sink node n minus the time stamp of source node 0, see Equation 4.1. This objective can be reached by influencing the two decision variables t_i and z_{ij}. The decision variable t_i signifies the starting time of the operation at node i, e.g. entering a block section, a departure or an arrival. If a train is delayed, this is reflected by the value of t_i for the respective nodes i of that train increasing with that delay. The decision variable z_{ij} is used to show which arc out of the alternative arc pair is selected. The indices i and j indicate that the variable refers to alternative arc (i, j) from node i to node j , which is part of alternative arc pair $((i, j), (h, k))$ in the set of alternative arc pairs D. The decision variable z_{ij} influences the precedence of trains and by choosing that precedence has an influence on the variable t_i, such that this can then influence the value of the sink node n.

The problem is constrained by three constraints. The first constraint in Equation 4.2 constrains the consecutive actions of a train by setting a minimal time that should be between two operations. This time is the minimal technical time, for example, it can represent the technical running time: how fast a train can technically get from one point to another without taking into account buffer times. In terms of graph theory, the equation constrains the arc length between two nodes i and j by setting a minimal arc length f_{ij}. The constraint only applies to the fixed arcs in set F. Contrary to the alternative arcs, all the arcs in set F are used.
A similar constraint is formulated for the alternative arcs in Equation 4.3, but because not all the alternative arcs are used, it is defined slightly different. This equation constrains two consecutive actions of two different trains by again setting a minimal technical time that should be between these actions. This is the minimal headway time that should be maintained between two trains, which in terms of graph theory is a minimal arc length f_{ij} between two nodes i and j. The constraint should only be enforced for the alternative arc of that pair that is in fact used, so only one of the arcs of the alternative arc pair $((i, j), (h, k))$ in set D. This is done by including both alternative arc (i, j) and (h, k) and then the variable z_{ij} ensures the constraint is only enforced for the selected arc.
Lastly, the constraint in Equation 4.4 dictates that only one arc out of the alternative arc pair $((i, j), (h, k))$ is used. With the choice of one of those arcs, the model chooses the precedence between two trains.

The mathematical formulation of the Alternative Graph model in this research is notated as follows. The same distinction between functional and additional constraints is used as in the decomposition model in subsection 3.3.3.

Index

i,...	for node i in node set N

Sets

N	set of nodes	$\{0,i,...,n\}$
A	set of all arcs	$\{(i, j),...\}$
F	set of fixed arcs	$\{(i, j),...\}$
D	set of alternative arc pairs	$\{((i, j), (h, k)),...\}$

Parameters

f_{ij}	the arc weight of arc (i, j) from node i to node j	$f_{ij} \in \mathbb{R}^+ \; \forall (i, j) \in A$, [s]

Decision variables

t_i	the starting time of the operation at node i	[s]
z_{ij}	if alternative arc (i, j) is selected: $z_{ij} = 1$ if alternative arc is used, if not $z_{ij} = 0$	[-]

Objective Function

The objective of the model is to minimise the maximum secondary delay. This value is represented in $t_n - t_0$, where t_0 is the overall starting time which is often equal to zero. Therefore the objective function can often be reduced to $min\, t_n$.

$$min[t_n - t_0] \tag{4.1}$$

Functional Constraints

1. This constraint determines how much time should be between two consecutive actions connected with a fixed arc. The greater than or equal sign ensures that the time between two events must be at least this technical time, however this time can also become larger due to delays.

$$t_i - t_j \geq f_{ij} \quad \forall (i,j) \in F \tag{4.2}$$

2. This constraint ensures the length of the alternative arcs represents the minimal headway between two trains. An extra consideration for the alternative arcs is that they come in pairs and that this constraint should only be enforced for the alternative arc of that pair that is in fact used. This task is fulfilled by the variable z_{ij} which is equal to one only is the alternative arc (i,j) is used, else it is zero. Multiplying with the respective variable z_{ij} or z_{hk}, ensures that the only non-zero terms in the equation enforce the minimal technical running time only for the used alternative arc.

$$z_{ij} \cdot (t_j - t_i) + z_{hk} \cdot (t_k - t_h) \leq z_{ij} \cdot f_{ij} + z_{hk} \cdot f_{hk} \quad \forall ((i,j),(h,k)) \in D \tag{4.3}$$

3. This constraint dictates that exactly one of the alternative arcs in an alternative arc pair is used. This is done by saying that the sum over the variable z_{lm} for both the alternative arcs in the pair should be equal to one, such that exactly one value of z_{lm} can be equal to one.

$$\sum_{(l,m) \in \{(i,j),(h,k)\}} z_{lm} = 1 \quad \forall ((i,j),(h,k)) \in D \tag{4.4}$$

Additional Constraints

The time decision variable t_i can take on all real positive values:

$$t_i \in \mathbb{R}^+ \quad \forall i \in N \tag{4.5}$$

The binary decision variable for the alternative arc pairs z_{lm} can have a value either 0 or 1:

$$z_{lm} \in \{0,1\} \quad \forall (l,m) \in \{(i,j),(h,k)\}, \forall ((i,j),(h,k)) \in D \tag{4.6}$$

4.1.3 Alternative Graph conflict resolution in a non-centralised approach

It should also be mentioned how Alternative Graph modelling is influenced by being used in a non-centralised approach. In a centralised approach one Alternative Graph would be used to describe the whole region. In a non-centralised approach that graph is split up into separate graphs for the respective subregions. This principle is illustrated in Figure 4.2. The Alternative Graph is split at the subregion border and a new sink node and source node are inserted to form two self-contained graphs. Also nodes can be added to signify the event of a train leaving or entering a subregion, if the subregion border does not coincide with the start of a block section.

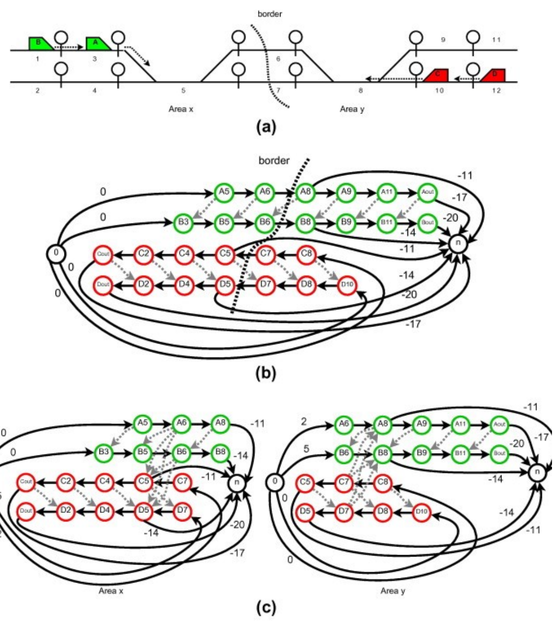


Figure 4.2: Alternative Graphs in a non-centralised approach [46]

In the non-centralised railway traffic management architecture, the conflict resolution is done per subregion by its respective agent. This can be subdivided into two alternating processes. One process chooses an alternative arc selection within the Alternative Graph. The other process monitors the delay trains in the subregion of the agent. This is visualised in Figure 4.3. A set of alternative arcs S_r is chosen for subregion r. Subsequently, it is monitored how this selection affects all trains t within the subregion and how much delay it will lead to. The delay d_{tr} of all trains t in subregion r is then used to review the selection of alternative arcs. This process is repeated until a selection of alternative arcs is found that minimises the maximum secondary delay.

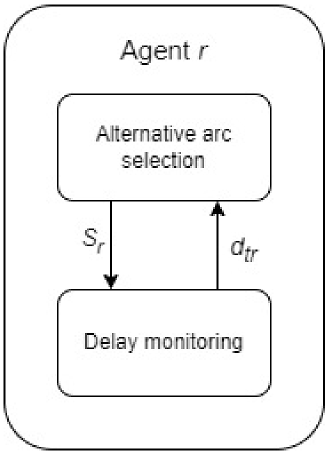


Figure 4.3: The two processes within an agent for subregion r

4.1.4 Limitations of the conflict resolution approach

The Alternative Graph model has proven to be a fast and efficient approach to conflict resolution, but it also has a limitation. This model formulation is only equipped to do retiming and reordering as rescheduling actions. In other words it is only able to change the times and orders of trains, but it is not able to reroute, short turn or cancel trains. This should be considered when applying the model, in the sense that it is most applicable to resolving disturbances. To solve disruptions a model adaptation, like D'Ariano et al. (2008) [34], or another model entirely can be chosen. In this research, the conflicts will be kept to disturbances.
In the process within the agent, the agent will only view what is happening in its own subregion at that time. It might be considered to include elements such as forecasting the impact of delays occurring in other subregions on the situation in the agent's own subregion. However keeping an eye on the main goal of the non-centralised railway traffic management approach, namely testing the impact of different decompositions, this is not considered in the current research. Coordination approaches along a similar line of thought are elaborated in subsection 4.2.3.

4.2 A coordination approach in non-centralised railway traffic management

The coordination model can take on different forms, which have been discussed in subsection 2.3.3. For example it can use a coordinator graph, shared parameters or boundary constraints. In this research it is chosen to use boundary constraints as a coordination method. This approach is used and mentioned in several publications [9, 15, 57]. On top of that its complexity is in line with the focus of this thesis.

4.2.1 Boundary constraints in coordination within non-centralised railway traffic management

A commonly used boundary constraint is the time transition constraint. This constraint dictates that when a train leaves one subregion at a certain time, it will enter the next at the same time. Essentially, the constraint ensures continuity between the subregions. The time constraint could be formulated mathematically as:

$$dp_{tr} = arr_{tq} \quad \forall t \in T_{rq}, r \in R, q \in Q_r \tag{4.7}$$

In this equation, R is the set of subregions R and Q_r is the set of neighbouring subregions q to subregion r. T_{rq} is the set of trains t that travels over the boundary between subregions r and q. The departure time of train t at the border of subregion r is depicted by dp_{tr} and arr_{tq} depicts the arrival time of the same train t at neighbouring subregion q.

This constraint fulfils a main concern in a non-centralised systems: that trains travel within the simulated system just as they would in the physical system. Without the constraint a train could seemingly disappear at a subregion border for a certain amount of time or enter a subregion earlier than that it departs from the previous one. Time continuity is one form, but continuity has also been taken into account in other forms in previous approaches, such as speed continuity and continuity in the cancelling of trains [15, 57]. But because the conflict resolution model will neither take the train speed into account nor be able to cancel trains, this type of continuity is not relevant in this approach.

Implementing the time transition constraint as stated in Equation 4.7, means coordination will need to take place at every border, although it may not always be necessary. The time of crossing a subregion border will need to be communicated even if the train is running according to schedule and the agent is already aware of all scheduled times within its subregion. Also the time transition constraint means that every subregion will need to be solved in every scenario. However, this is not necessary if a subregion is not affected by the current delay scenario. Similar to the argument above, the agent knows the scheduled operations of the train. Asking it to solve the scheduled scenario would only take unnecessary computation time.
Therefore this research proposes an alternative to the time transition constraint: a delay transition constraint. This entails that instead of the arrival and departure times, the delay of a train is coordinated between agents. The constraint can be reformulated as:

$$d_{final,tr} = d_{entrance,tq} \quad \forall t \in T_{rq}, r \in R, q \in Q_r \tag{4.8}$$

This constraint dictates that the delay with which train t leaves subregion r, i.e. $d_{final,tr}$, is equal to the delay with which that same train t will enter the neighbouring subregion q, i.e. $d_{entrance,tq}$. Using a boundary constraint in this formulation, opens up the possibility of communicating only in the case of delay. How such a delay driven coordination is implemented is discussed in the next subsection.

4.2.2 A coordinator within a non-centralised coordination approach

Communicating only in the case of a delay, can be categorised as event-driven communication. Meaning the communication is driven by an event, in this case the event being a delay. To enforce it, it must be coordinated in some way within the cyber layer when agents are expected to solve their subregions and communicate with other agents. That is why a coordinator is added as an additional entity. The coordinator is placed in a higher layer than the conflict resolution agents, which allows it to oversee the global conflict resolution task. As a coordinator the entity has three main tasks:

1. **Communicating initial delay to the agents.** The initial delay is given as an input to the coordinator at the start of the railway traffic management problem. The existence of an initial delay and thus a deviation from the scheduled situation is the instigator for applying a railway traffic management approach. For instance, an initial delay can be that a train enters the region with a five minute delay. This kind of delay is different from the secondary delays, which are all the delays that are a consequence of that initial delay. Because the coordinator is on a higher level, it can see in which subregion(s) the initial delay occurs. Therefore part of its task is to communicate the initial delay to the relevant agent or, if the initial delay occurs in multiple places in the region, the relevant agents. The respective agent(s) can then use this initial delay to calculate the impact of that delay on their subregion.

2. **Assigning the conflict resolution task to the agents.** If trains enter a subregion with a delay, be it an initial or secondary delay, the situation within that subregion will differ from the original schedule. So in the case of delayed trains entering a subregion, the coordinator will alert the respective agent and tell it to solve its subregion. If there are no delayed trains entering that subregion then it will not need to be solved, because the situation is simply the same as the original schedule. This method differs from conflict detection. The coordinator does not know if the delayed trains will lead to a conflict in a certain subregion, all it knows is that there is a probability for a conflict and thus that the subregion should be calculated by its agent. The coordinator does not only determine if a subregion should be solved, but also in what order the subregions should be solved. This is of importance, because the solution of one agent could have consequences for other agents.

3. **Prompting communication between agents, or not.** The first part of this coordination task is that the coordinator determines if the trains leaving a subregion are delayed. If one or more of the trains leaving the subregion are delayed, this will have an impact on other subregions and thus this delay is to be communicated. The second part of this task is to determine to which subregions these trains will travel and thus to which agents this delay ought to be communicated. The coordinator will then instruct the agent to communicate the delays to the respective agent(s), or not.

As is reflected in these tasks, the coordinator has continuous communication with all agents. This communication is also bilateral, meaning that it goes both from the coordinator to the agents as the other way around: The coordinator gives the agent a solving or communication task and the agent returns the result of the conflict resolution of the subregion to the coordinator. Shown schematically, the system for four agents can be drawn as in Figure 4.4. Here $d_{in,r}$ signifies the initial delay for subregion r and d_{tr} stands for the delay of all trains t leaving subregion r.

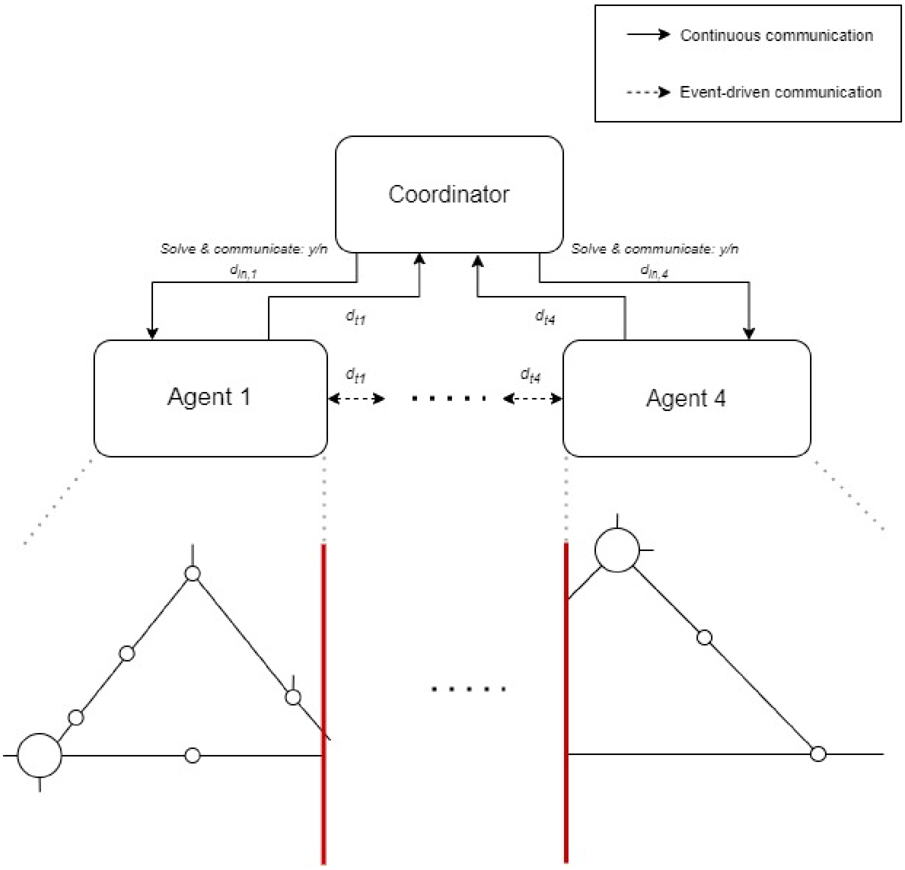


Figure 4.4: Schematic representation of the coordination scheme

It should be noted that there is a difference between the coordinator and an agent. The coordinator, contrary to an agent, is not tasked with solving or optimising a region. It is not computing a solution to the conflict resolution, but assigning the tasks of the computation to the agents. This difference is also discussed in Corman et al. (2014) [49]. Because the coordinator differs from an agent, the system architecture remains non-centralised. More specifically, looking at the coordination scheme the system has a distributed system architecture. As defined in subsection 2.3.2, this means that within the non-centralised architecture the agents have communication with each other.

4.2.3 Limitations of an event-driven delay transition coordination scheme

As mentioned, the boundary constraint will ensure continuity between the subregions. This is deemed to suffice for the current study, where the main goal of the non-centralised railway traffic management approach is to support the validation of the decomposition. However it only tackles one of the challenges of non-centralised railway traffic management systems. Other challenges are not included in this model and could lead to complications for other implementations. For example, it is not taken into account if the optimal solution found by an agent is indeed also the optimal solution for the global problem, i.e. global optimality, or maintaining precedence choices made in previous regions, i.e. consistency. Also, in the field of coordination the possibility lies for implementing cooperation between agents to a common solution or prediction of the impact of delays in other subregions on an agent's own subregion.
The choice of including only a simple form of coordination could have several implications. It could for example be that the solution found is not the global optimal solution and that a solution is not found in the least amount of time possible. It could also occur that a choice made by one agent is reversed by the next agent. However, reminded of the fact that this non-centralised model is mainly a means to test the decomposition model, this approach is deemed sufficient.

4.3 Example and model verification of a non-centralised railway traffic management approach

Combining the conflict resolution model and the coordination scheme, a full non-centralised railway traffic management approach can be simulated. Both are implemented in Python and to solve the conflict resolution the commercial solver Gurobi is used. Details on this implementation can be found in Appendix C.
This section will introduce an example to demonstrate the functioning of the established non-centralised approach. A fictitious example is chosen that emulates elements the implemented non-centralised approach may encounter. The focus of this section are the conflict resolution and the coordination, so the decomposition is taken as a prerequisite. The example is shown in Figure 4.5.

In the figure the agents are shown and the railway system that is being managed. The fictious railway system consists of infrastructure, trains and a schedule. This has been decomposed into three subregions. The decomposition has not been done according to the proposed decomposition model, but was determined to best serve the illustrative purposes. The coloured arrows indicate the routes of the respectively coloured trains, such that train A and B travel from left to right and train C from right to left. The houses on the infrastructure indicate stations. The most left station is smaller, because it represents a smaller station where only an arrival time is scheduled. The larger station has both a scheduled arrival and departure time for all trains. There is one agent for every subregion, that alternates the processes of finding a selection of alternative arcs and monitoring the subregion when assigned the conflict resolution task by the coordinator. Each agent communicates with the relevant other agents and the coordinator.

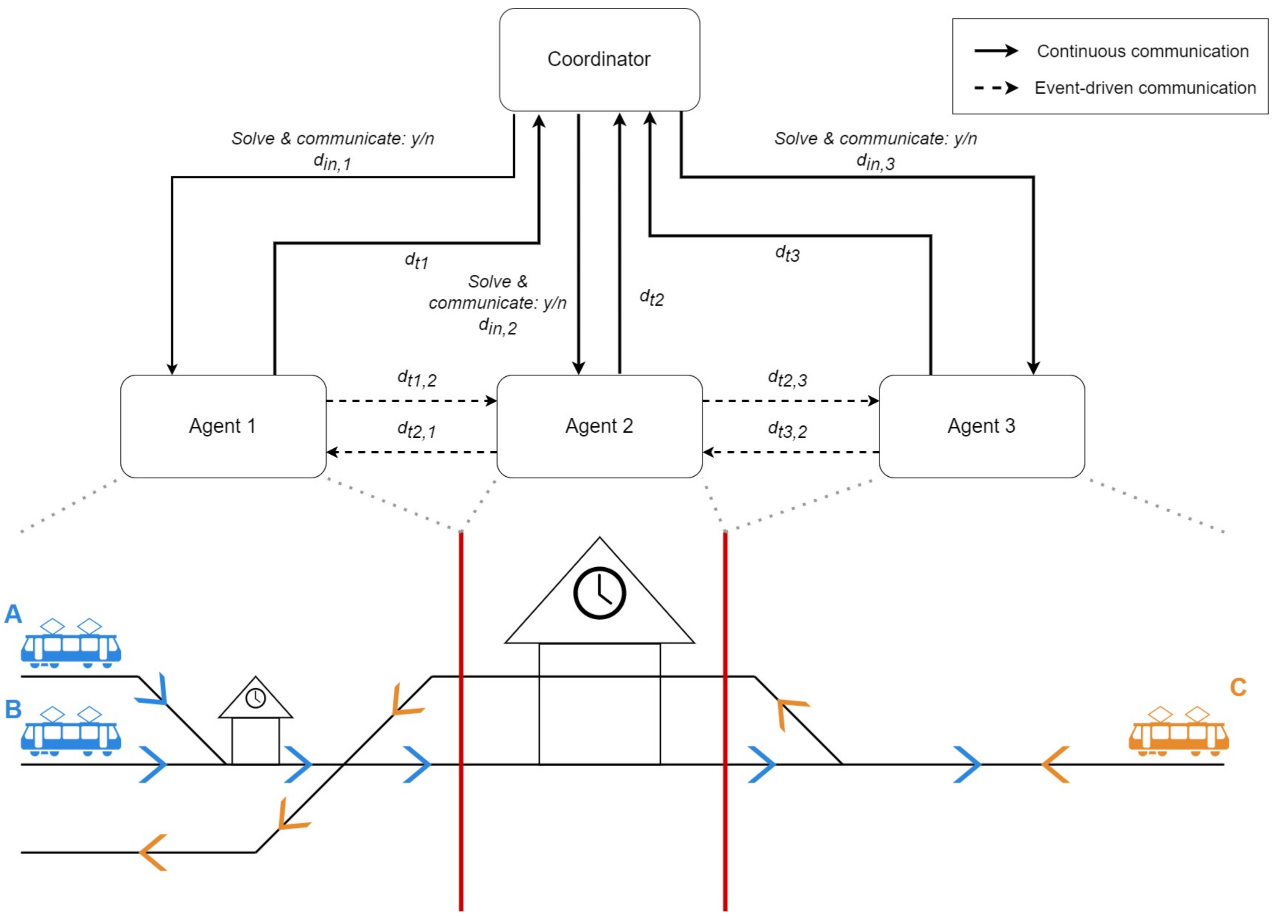


Figure 4.5: Demonstrative example of a non-centralised approach

The scheduling times of this example are shown in Table 4.1. These times are simply for illustrative purpose and for the sake of the example they are assumed to be in minutes. Also, note that train C travels in the opposite direction, such that its route differs from train A and B.

Subregion	**Node**	$t_{scheduled}$ [min] **Train A**	**Train B**	**Operation**
1	A	0	7	Entry
	C	2	9	Arrival
	Z	13.5	20.5	Exit
2	A	13.5	20.5	Entry
	C	15.5	22.5	Arrival
	D	20.5	27.5	Departure
	Z	24	31	Exit
3	A	24	31	Entry
	Z	32.5	39.5	Exit

(a) Trains A and B

Subregion	**Node**	$t_{scheduled}$ [min] **Train C**	**Operation**
3	H	0	Entry
	Y	8.5	Exit
2	L	8.5	Entry
	J	10.5	Arrival
	I	13.5	Departure
	Y	17	Exit
1	T	17	Entry
	Y	28.5	Exit

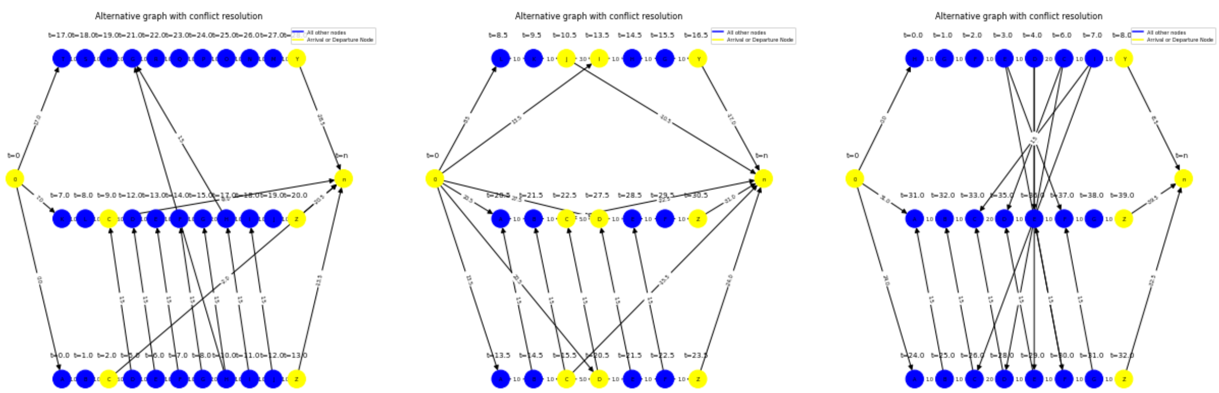
(b) Train C

Table 4.1: Scheduled times in the example

The Z and Y nodes indicate the reaching of the subregion border and the arrival and departure times are at the stations indicated in the infrastructure in Figure 4.5. In the schedule a buffer of 5 % is implemented on top of the technical running times. The minimal headway is set to be 1.5 minutes .

4.3.1 Verification of the Alternative Graph conflict resolution

The conflict resolution model is demonstrated and verified on the introduced example. The railway management problem is formulated as an Alternative Graph, as elaborated in section 4.1, for each subregion. Without introduced delays, and consequently no rescheduling, the Alternative Graphs of the subregions are visualised as in Figure 4.6.



(a) Subregion 1 (b) Subregion 2 (c) Subregion 3

Figure 4.6: The unrescheduled Alternative Graphs for the introduced example

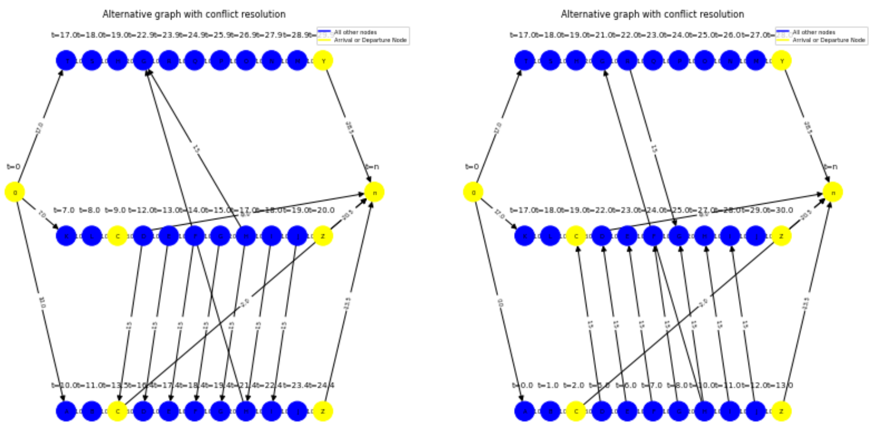
In these graphs the upper row of nodes are the events of train C, the row below that the events of train B and the lower row those of train A. The yellow nodes are special nodes, i.e. the sink and source node, arrival nodes and departure nodes. The blue nodes represent all other operations.

For the verification of the implementation, a series of verification steps are applied as in Table 4.2.

Verification step	**Hypothesis**	**Outcome**	**Pass or Fail**
$t_{C1_A} >> t_{C1_B}$	Train B precedes A	Train B precedes A	Pass
$t_{G1_B} >> t_{G1_C}$	Train C precedes B	Train C precedes B	Pass
$d_B = 1$	$d_{B1} < d_B$	$d_{B1} = 0.5, d_B = 1$	Pass
$Headway = 5$	$d_{final_A}, d_{final_B}, d_{final_C} \geq 0$	$d_{final_A} = 0, d_{final_B} = 0.5, d_{final_C} = 1.5$	Pass

Table 4.2: The verification of the conflict resolution model

The first two steps are to determine if the reordering of the model is functioning. In the unrescheduled situation in subregion 1 train A precedes train B, which in turn precedes train C. In the verification steps the trains are delayed such that the time at which they enter a shared block section is increased. When train A arrives at the convergence (in front of the small station in Figure 4.5) significantly later than train B, it is expected that the order between the trains is changed such that train B will now precede train A instead of vice versa as in the scheduled situation. Similarly, when train B is delayed such that it arrives at the crossover (just after the small station in Figure 4.5) later than train C it is expected that the order is changed accordingly. The rescheduled situations can be seen in Figure 4.7a and Figure 4.7b respectively. Comparing these graphs to the scheduled situation in Figure 4.6a the changes in precedence are clearly reflected in a different selection of alternative arcs.



(a) Train B precedes train A　　(b) Train C precedes train B

Figure 4.7: Precedence changes in subregion 1

In the third step, it is tested if the delay is partially absorbed by the implemented buffer times. When delaying train B slightly it is expected that a part of this delay is absorbed by the buffers, such that the delay with which train B leaves subregion 1 , d_{B1}, is smaller than the imposed initial delay initial delay d_B. In the outcome it can be seen that this is indeed the case.
Lastly, the functioning of the headway is inquired. Because the headway is increased from 1.5 to 5 minutes, the trains are forced to increase the distance between them. Because this increased time is not accounted for in the schedule, the scheduled situation will likely be endangered such that the trains are not able to drive according to the original schedule. This can lead to one or multiple trains experiencing a delay because they are forced to maintain more distance to the preceding train. This is also what is observed in the outcome. Both train B and C experience a final delay, the delay the trains have when they leave the network. Train A does not as it is the first train in this situation and does not have to account for preceding trains.

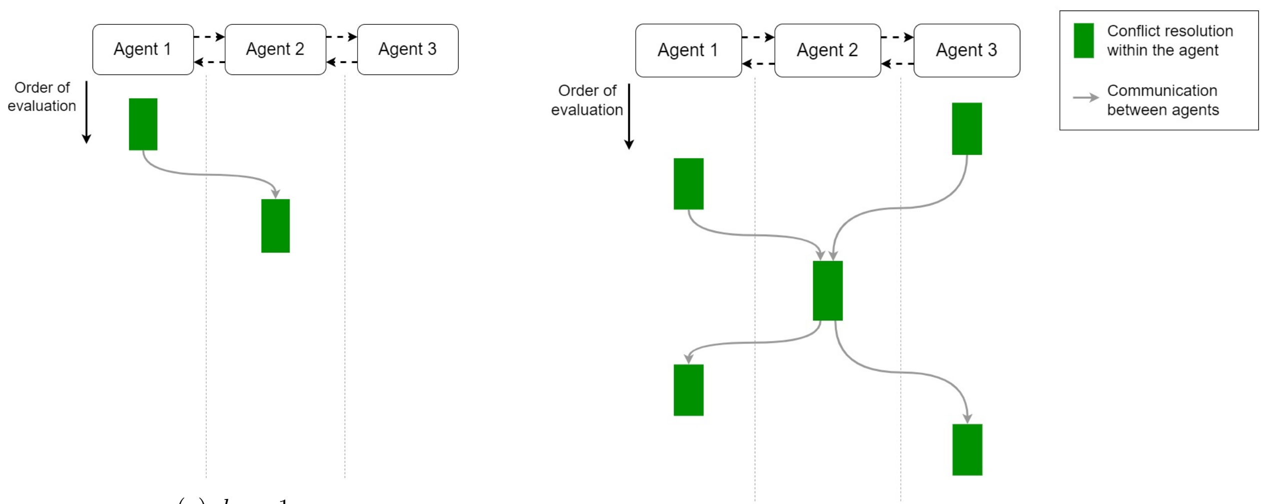
4.3.2 Verification of delay-driven communication with a coordinator

The next step is to verify the coordination scheme. It has two main aspects that are of interest for verification. The first is the boundary constraint and the second is the event-driven communication. These aspects are tested in Table 4.3.

Verification step	Hypothesis	Outcome	Pass or Fail
$d_A = 9$	$d_{A1} = d_{entry_{A2}}$	$d_{A1} = d_{entry_{A2}} = 11$	Pass
$d_C = 0.5$	Only Agent 3 runs, no interactions	Interactions = 0, only Agent 3 runs	Pass
$d_B = 1$	Only Agent 1 & 2 run and coordinate	Interactions = 1, Agent 1&2 run	Pass
$d_B, d_C > 0$	Reiterate running Agent 1 & 3	Runs Agent 1, 3 & 2 and reiterates Agents 1 & 3	Pass

Table 4.3: The verification of the coordination scheme

In the first step it is tested if the boundary constraint is functioning, by testing what happens if train A is delayed. In this case it is desirable that the delay with which train A leaves subregion 1, then becomes the delay with which train A enters subregion 2. The coordination behaves as expected. The next aspect, tested in the following steps, is the event-driven communication. The coordination should only allow communication between agents if a delayed train crosses their subregion border. When train C is given an initial delay of only 0.5 minutes, this is fully absorbed by the buffers such that at the border of subregion 3 train C is no longer delayed. In this case, subregion 3 is the only subregion where the situation differs from the scheduled situation thus only Agent 3 should be run. Similar behaviour is expected in the case where train B is given a delay of 1 minute. The reaction of the coordination scheme to this situation is visualised in Figure 4.8a. The figure, going downward, presents the order in which (and if) the agents evaluate their respective subregion. In this specific case of delaying train B, it can been seen that first Agent 1 is evaluated and that it then communicates with Agent 2 before that is evaluated. After that Agent 3 is not evaluated because all delay of train B was absorbed. In the last case two trains are issued an initial delay, namely train B and train C. This means that two subregions will have an initial delay. Also because the trains are travelling in opposite directions, an iteration is necessary. This is reflected in Figure 4.8b. First Agent 1 and Agent 3 are evaluated sequentially, providing the information for Agent 2 to compute its subregion. From this point a delayed train C will travel from subregion 2 to subregion 1 and a delayed train B will travel from subregion 2 to 3. These trains with secondary delay will have an impact on the respective subregions, that was not taken into account the first time they were solved. This means that Agent 1 and 3 need to be reiterated with the information of both the initially delayed train and a secondary delayed train entering the subregion.



(a) $d_B = 1$

(b) $d_B, d_C > 0$

Figure 4.8: Order of evaluation for two verification steps

4.4 Conclusion on the non-centralised railway traffic management approach

This chapter has answered the research question: *How can a non-centralised conflict resolution model be designed?*. The conflict resolution and coordination approaches were introduced and verified.

For the conflict resolution, the Alternative Graph model was discussed as presented in D'Ariano et al. (2007) [10]. The model was slightly reformulated compared to that presentation as to better represent the implementation. The coordination approach uses a delay transition constraint at subregion boundaries, event-driven communication and a coordinator. Boundary constraints have been used in several publications. However the concept of using a delay transition constraint to enable event-driven communication between agents has not been previously encountered in literature. The verification shows that the presented approaches behave as expected. Also no major limitations were encountered in the verification process, which means the approaches are suitable for using in a test case.

In both approaches it was considered that they fit the main purpose of this research: to test and support the proposed decomposition approach. Both are efficient and implementable. Although the Alternative Graph model has been established in literature, the coordination approach could potentially be useful in future research.

Combining the conflict resolution with the coordination scheme gives a non-centralised railway traffic management approach. This combination provides the tools that allow the testing of the decomposition approach in simulated delay scenarios. This is what is discussed in the following chapter.

5. Case Study Analysis

This chapter discusses the evaluation of the impact of the proposed decomposition model on non-centralised conflict resolution. In other words, the chapter concerns the validation of the proposed decomposition model. For this purpose a test case is chosen from the existing Dutch railway network, section 5.1. For this test case the inputs are determined both to the hourly traffic density based decomposition approach and the non-centralised railway traffic management approach, section 5.2. The proposed decomposition is compared to the baseline decomposition, the train dispatcher subregions currently used at ProRail. This is first done by comparing them purely on the basis of the criteria in the decomposition. After that the non-centralised railway traffic management approach is applied to both decompositions. This allows the decompositions to be compared on how they impact the conflict resolution in terms of delay and decisional complexity. The results of these comparisons are shown in a set of graphs, section 5.3.

The source of comparison is the different decompositions. Both decompositions will be subjected to identical coordination and conflict resolution schemes as shown in Chapter 4. The subregions of the respective decomposition will all be managed by an agent. For the baseline decomposition, this means that every train dispatcher subregion is managed by an agent. It could also be seen as that the human train dispatcher is, for the sake of comparison, substituted by the Traffic Management System.

5.1 Roosendaal-Liempde region test case

To validate the decomposition approach a test case is chosen that can illustrate its impact. Several factors are taken into account during this choice:

- **Part of a ProRail corridor** As shown in section 2.1, in the current state of practice geography has been the main driver in decompositions. However experts at ProRail have expressed that looking at the network from a traffic flow perspective may yield more insightful results [27]. More specifically, they advised to decompose (part of) a corridor instead of, for example, a 'Verkeersleider'-region. Such a corridor is a part of the network that contains many trains travelling roughly the same route along this path. Examples are the High Speed corridor consisting of the high speed train route from Breda to Amsterdam Centraal and the A2 corridor all the way from Maastricht and Heerlen to Den Helder and Enkhuizen via Eindhoven, Utrecht and Amsterdam. Their experience has shown that managing railway traffic based on these corridors can be more effective than based on the regions. As a result, they are considering to include the corridors more significantly in future decompositions. This research will do so as well, to provide insight and relevance to the state of practice.
- **Complex decomposition situation** In the conclusion of Chapter 3 it is mentioned that the decomposition approach will likely provide the most insight in difficult to oversee decomposition situations. Therefore such a situation is preferably used to demonstrate the decomposition approach. This could, for example, be a region where many different routes are involved.
- **Significant size** The chosen test case should be large enough to be able to present the decomposition approach and the impact of the decomposition on the conflict resolution. This size is estimated to be around the size of smaller traffic controller region, i.e. around four to five train dispatcher subregions.
- **Traffic heterogeneity** The second aspect of the test case is that it is preferred to provide interesting conflict situations that can support the testing of the decomposition approach. Therefore the test case should preferably contain traffic heterogeneity, meaning that different types of trains travel the infrastructure, such as sprinters, intercity's and freight trains. These train types differ on several characteristics, such as length, speed and the stops they make. Combining different types leads to dynamic situations that are interesting cases from simulating train conflicts. From another perspective, freight traffic often has routes that differ from those of passenger traffic which then in turn becomes an interesting factor in the decomposition. So this factor has an influence on two aspects: firstly the differing routes of the freight traffic are taken into account in the decomposition, secondly the traffic heterogeneity will provide dynamics that enable the application of a conflict resolution approach to interesting conflicts.

- **Point of Interest for conflicts** Continuing the train of thought of the previous point, the test case should contain Points of Interest for conflict to provide interesting conflict situations to test the decomposition. These Point of Interest can be different kinds of infrastructural phenomena, for example, multiple tracks converging to one, one track diverging to multiple tracks or tracks crossing over other tracks.

Taking all the factors into account, the test case is chosen as the network between Roosendaal and Lage Zwaluwe, and Den Bosch and Liempde, see Figure 5.1. This is part of the IJssellijn corridor and the freight route 'de Brabant route', shown in Figure 5.2.

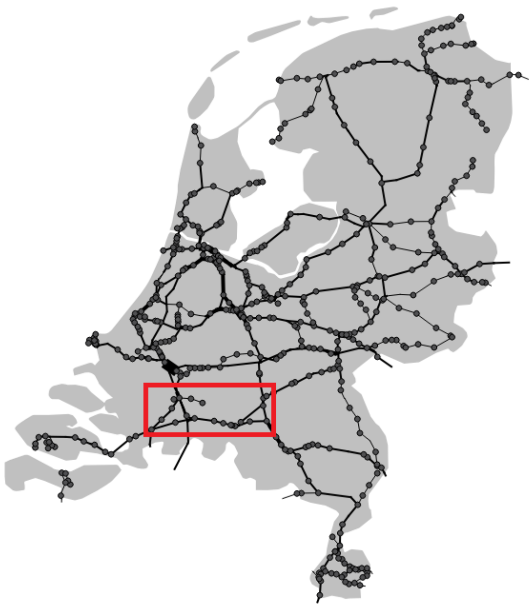


Figure 5.1: The selected test case, edited from [70]

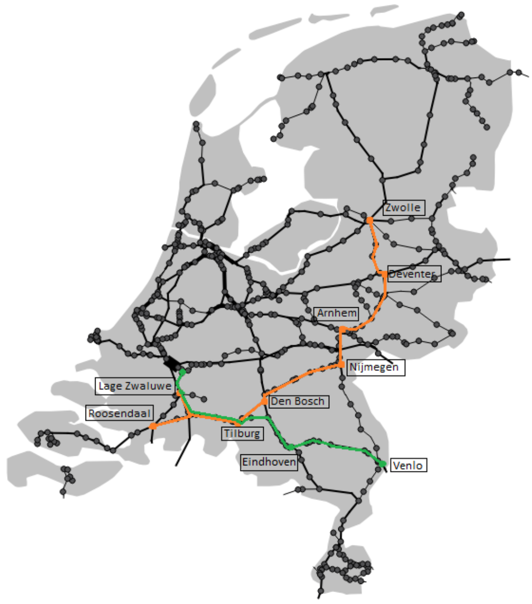


Figure 5.2: The IJssellijn corridor (orange) and Brabant route (green), edited from [70]

When viewing only the corridors, the test case would only contain the routes in Figure 5.2. However one key aspect of the test case is that it will allow the comparison between the baseline decomposition currently used at ProRail to the decompositon following from the proposed decomposition approach. To make a proper comparison the current arrangement of the subregions in these parts of the corridors must be taken into account. The corridors pass through the Roosendaal Verkeersleider region and the Eindhoven Verkeersleider region. It is chosen to let the outline of the test case be defined by the current train dispatcher subregions. This means the test case will encompass the train dispatcher subregions of Roosendaal, Breda, Den Bosch and Tilburg. The individual infrastructure for the high speed trains is excluded from the test case, because in the state of practice the high speed infrastructure forms a separate subregion from Breda to Rotterdam Centraal. At Zevenbergsehoek aansluiting the high speed trains converge from the high speed infrastructure onto the regular infrastructure.

This test case contains some interesting aspects. First of all, at this point in the Dutch network many different routes come together, which makes it interesting for the decomposition approach. Additionally, the region has heterogeneous traffic in the sense that besides the intercity, sprinter and high speed passenger trains also freight trains travel through the region. The varying characteristics of the traffic result in interesting factors in the conflict resolution and the decomposition. Morevover the triangular formation of the infrastructure means that it contains many different Points of Interest for conflict.

5.2 Inputs to the Roosendaal-Liempde region case study

Both the decomposition approach and the non-centralised railway traffic management approach are applied to the Roosendaal-Liempde region. This requires the determination of the inputs to both approaches, which is done in this section.

5.2.1 Inputs to the hourly traffic density based decomposition

To come to the decomposition of this test case, the inputs to the decomposition model need to be determined. First this means viewing the infrastructure. The timetabling points are the elements that can be assigned to subregions. These timetabling points can be seen in Figure 5.3. Most timetabling points are stations, e.g. Roosendaal, Oudenbosch and Zevenbergen, while some form important junctions, such as Zevenbergschenhoek aansluiting, Boxtel and Liempde. In total there are 19 timetabling points.

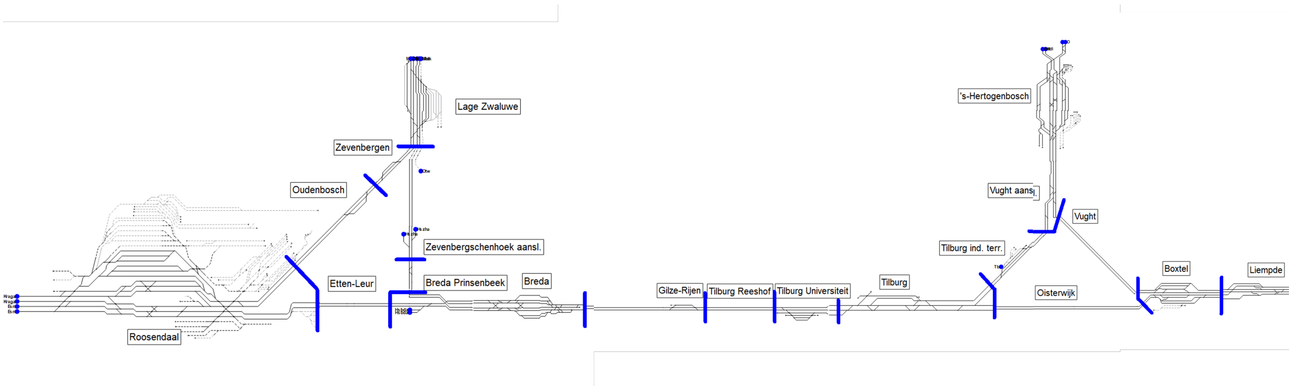


Figure 5.3: The timetabling points within the test case, edited from [70]

For determining the hourly traffic density for every timetabling point, their respective constraining number of tracks is established. This is then combined with the number of trains travelling through the timetabling point every hour, which is determined based on all potential traffic in the basic hourly pattern of the 2022 timetable. To understand what this means, it first needs to be established that the Dutch timetable is cyclical, meaning that an hourly pattern is constructed which is then repeated every hour to build a 24 hour timetable. Deviations from this pattern can happen, for example, during rush hour, night time or official holidays by adding or removing trains. Secondly, it needs to be established how freight trains are scheduled into the timetable. Freight companies reserve time slots within the passenger train timetable, which can then be used when needed. In practice only a fraction of these reservations are used, but in theory every single one can be used. So when stating that all potential traffic is considered, this means that the rush hour pattern is regarded including all freight train reservations. This choice is made to simulate the worst case of hourly traffic density that can occur, which will give the best insight into the supply/demand ratio and decisional complexity of a timetabling point.

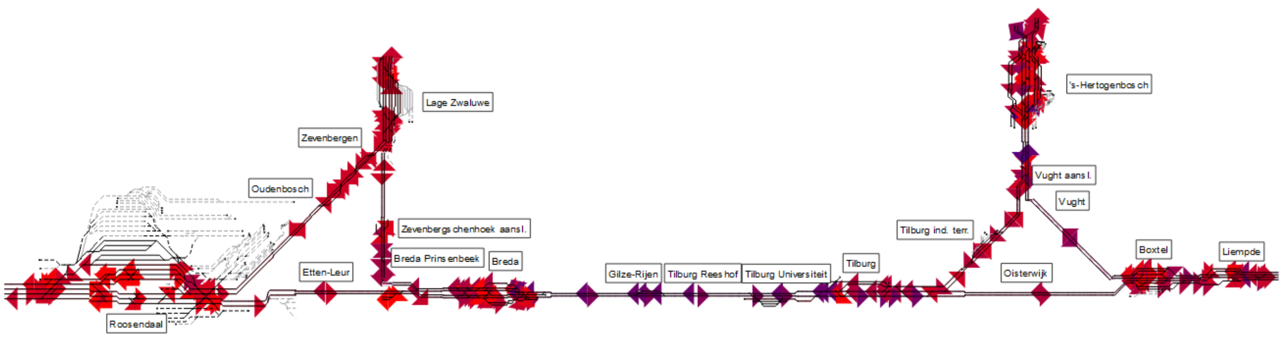


Figure 5.4: The train services within the Roosendaal-Liempde region

Other inputs required for the decomposition model are the train services and their frequency. The train services within the region are visualised in Figure 5.4, where the arrows indicate train services travel in that direction and the gradient from red to purple indicates a gradient of amount of trains travelling that route. It can be seen that the busiest part of the region is from Zevenbergschenhoek aansluiting to Den Bosch and especially between Breda and Tilburg. The frequency of a train service is again determined based on all potential traffic.

Lastly, the amount of subregions the region is decomposed into needs to be determined. It is chosen to use the same amount of subregions as are currently used in the State of Practice. In other words, the chosen region consists of 4 train dispatcher subregions, so the proposed decomposition approach is tasked with dividing the region into 4 subregions based on hourly traffic density.

Details on the timetabling points, trains services and their frequencies can be found in section A.2.

5.2.2 Inputs to the railway traffic management approach

To determine how the decomposition affects the conflict resolution, the test case must be expanded to include potential conflicts. In other words, train traffic with conflicts can be simulated on a microscopic level within the region to assess the conflict resolution under the influence of the decomposition.

Before conflicts can be simulated, an initial timetable needs to be chosen. This timetable is a 15 minute interval of a basic timetable for Tuesday in 2022 by ProRail [70]. As explained previously the Dutch timetable consists of an hourly repeating pattern. So a researched choice for a quarter of this hour can give insights valuable for the whole day. The selection for the schedule was based on the fact that it would as susceptible to conflict as possible, or in other words where the traffic flow is as busy and complex as possible. This means that there are many options for potential conflicts. As a result, the conflict resolution problem will be the most complex at this point in the timetable. So this is where the largest difference is expected to be observed between the effects of both decompositions. In essence this means that looking at the available timetable the following factors must be taken into account:

- **Number of trains** The goal is to find a place in the timetable which forms a complex conflict resolution problem and one large factor in that is the number of trains that are involved. In practice this means that the timetable from 23:00 to 06:00 is not of interest for this application, because only a small amount of trains are present within the region during these hours.
- **Traffic heterogeneity** The presence of traffic heterogeneity, i.e. different train types, was also a factor in choosing the Roosendaal - Liempde region. Within the conflict resolution, the traffic heterogeneity is interesting on a microscopic scale. The differing characteristics between train types, such as speed, acceleration and deceleration, bring complexity to the conflict resolution problem. Therefore it is desired to pick a part of the timetable when differing train types are present in the region.
- **Trains at Points of Interest for conflict** The region was chosen as to contain points where interesting conflicts can occur and it is interesting to include these potential conflicts within the selected 15 minutes of timetable. For that reason it is analysed when trains are present at the identified Point of Interest for conflict and more importantly that the present trains have shared infrastructure. This gives a basis of potential conflicts, that can be used to test the decompositions.

Taking all these considerations into account it is chosen to use the selection between 20:25 and 20:40 of the basic Tuesday timetable. In Appendix B some data analysis is shown to support this selection.

Within the scheduled timetable, there are still no conflicts to be resolved. To that end, delay scenarios are created. In these delay scenarios the aim is that conflict is created which ideally has an effect in multiple subregions. Not only should the delay scenario give the consequences of the conflict time and space to propagate, the conflict resolution algorithm should also have the decision space to have the opportunity to relieve the consequences of the conflict.

Five different initial delay types are chosen for this purpose. Each delay type is the delaying of one train upon entering the region at one of the five entry points, that is also early on in the 15 minute time frame. This way the delay has time to propagate but the algorithm also has time to resolve it. The five entrypoints of the region are: Roosendaal, Lage Zwaluwe, Zevenbergschenhoek aansluiting, Den Bosch and Liempde, see Figure 5.5. The size of the delay is set to the planned headway the train has to the train in front of it. This way two trains are planning to occupy the same block section(s) at the same time, unavoidably leading to a conflict.

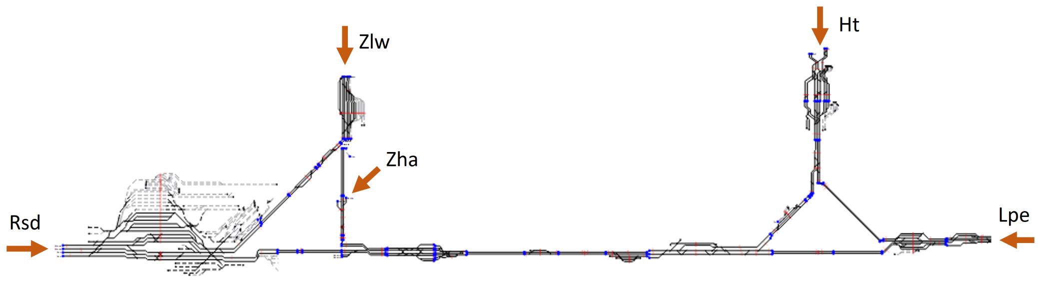


Figure 5.5: Entry points of the Roosendaal-Liempde region, edited from [70]

In total, 17 delay scenarios are used, where each delay scenario consists of one or multiple delay types. The delay scenarios are summarised in Table 5.1. An abbreviation refers to the entry point where a train is given an initial delay. It may be observed in the table that for the delay scenarios that multiple delay types, not all possible combinations of entry points are studied. The choice was made to only combine delay types that would influence a common part of the network at some point. For example, the delayed train at Roosendaal and Lage Zwaluwe will meet at Breda and the delayed train at Den Bosch and Zevenbergschenhoek aansluiting will both have an effect on the route between Breda and Tilburg. When, for instance, delaying a train at Roosendaal and Liempde, these two trains will not meet in a common part of the network within the 15 minute time frame and thus this delay scenario would simply be the cumulative of the individual delay scenarios of a delayed train at Roosendaal and a delayed train at Liempde. So to keep the results relevant and interesting, only these 17 delay scenarios are studied. Besides these delay scenarios, the scenario where there is no delay is also studied. This is when all trains travel according to schedule.

One delayed train		**Two delayed trains**		**Three delayed trains**	
Entrypoint	Delay [s]	Entrypoints	Delay [s]	Entrypoints	Delay [s]
Rsd	203	Rsd, Zlw	733	Rsd, Zha, Ht	543
Zlw	530	Zlw, Zha	705	Rsd, Zha, Lpe	681
Zha	175	Rsd, Zha	378	Zlw, Zha, Ht	870
Ht	165	Zha, Ht	340	Zlw, Zha, Lpe	1008
Lpe	303	Zha, Lpe	478	Zha, Ht, Lpe	643
		Ht, Lpe	468		

Table 5.1: The delay scenarios categorised by the amount intially delayed trains with the entry point(s) of the delayed train(s) and the total intial delay

Rsd: Roosendaal; Zlw: Lage Zwaluwe; Zha: Zevenbergschehoek aansluiting; Ht: Den Bosch; Lpe: Liempde

In the implementation of the test case a series of assumptions are made. Firstly, schedule and route information is not available for all train types. As a result a few train types are not considered: international trains, video inspection trains ('video schouw treinen') and empty trains ('leeg materieel'). These train types are responsible for a handful of (generally small) movements within the viewed time frame and region. All major train types are considered: intercity trains, sprinter trains, high speed trains and freight trains. Secondly, in the Alternative Graph formulation every alternative arc will strictly speaking have its own individually calculated weight. In other words, the minimal headway is different for all combinations of individual trains and is also dependent on which train precedes the other. These values can be calculated for example by the compression method where the timetable is squeezed together to find the critical headway on the shared infrastructure of two trains, see Pachl (2014) [76] for more information. However the scope of this research combined with the size of the test case, does not allow such extensive computations. Therefore it is chosen to assume the same minimal headway for all train combinations and orders. A minimal headway of 150 seconds was chosen based on calculating the critical headway for a small sample of train combinations, experimenting with the feasibility of the original schedule with different headways and the research in Koning (2002) [77].
The impact of these assumptions on the relevance of the results is expected to be minimal. This is because the results consist of a comparison between the proposed decomposition and baseline decomposition both with the same non-centralised railway traffic management approach. So the same assumptions apply to both cases. As long as the results maintain the relative comparison between the decompositions, these assumptions will likely have a minimal effect.

5.3 Results of applying the proposed decomposition and the non-centralised railway traffic management approach to the case study

With all the inputs established, the results can be generated. First the region is decomposed using the hourly traffic density based decomposition approach. After which the proposed decomposition and the baseline decomposition are compared to each other in themselves. Then the railway traffic management approach is applied to the decomposition and the results using both the decompositions are compared. First this is done based on the decisional complexity and interactions and then based on the delay performance. In Appendix D the detailed results can be found of the conflict resolution experiments.

5.3.1 Outcome of an hourly traffic density based decomposition on the Roosendaal-Liempde region

Using the inputs, a decomposition was found by the decomposition approach presented in Chapter 3. For the weight factor, $\zeta = 0.6$ is used. How this value was chosen is elaborated in Appendix A. In Figure 5.6, the orange line shows the border of the subregions in the proposed decomposition and the blue lines show the border of the baseline decomposition. The baseline decomposition refers to how the train dispatcher subregions are currently defined. The subregions are referred to according to the intercity station it contains or the abbreviation of that station. This means the four subregions are: Roosendaal or Rsd; Breda or Bd; Tilburg or Tb and Den Bosch or Ht. This name refers to the subregion in the respective decomposition.

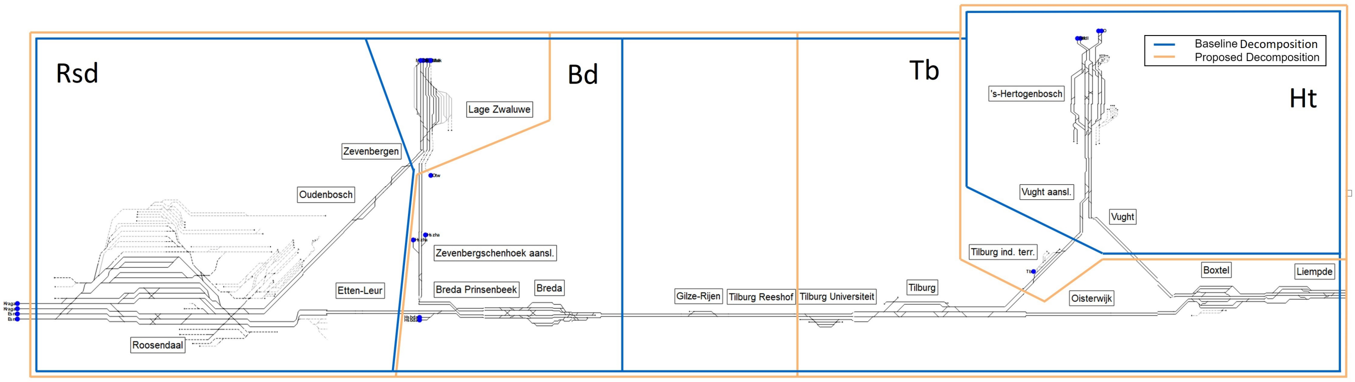


Figure 5.6: The proposed decomposition and baseline decomposition, edited from [70]
Rsd: Roosendaal; Bd: Breda; Tb: Tilburg; Ht: Den Bosch

First of all, looking only at the proposed decomposition a few observations can be made. All subregions contain 4 to 5 timetabling points. Also, all subregions contain exactly one intercity station.
Secondly, the proposed decomposition can be compared to the baseline decomposition. One large difference can be seen in the size of the Tilburg subregion. In the baseline this subregion is quite large, reaching from Gilze Rijen to Tilburg Industrie and Liempde and containing 8 timetabling points. Using the hourly traffic density, Tilburg is identified as the timetabling point with the highest value of the criterion, being two times larger than the average value. As a result the decomposition model chooses to include less timetabling points in the subregion containing Tilburg, to come to a more even distribution of hourly traffic density over the subregions. Another difference in subregion size can be seen in the Roosendaal subregion, where the proposed decomposition includes Lage Zwaluwe contrary to the baseline decomposition. This is again a result of the hourly traffic density criterion. Roosendaal and the surrounding timetabling points, Etten-Leur, Oudenbosch and Zevenbergen, have a relatively small hourly traffic density. So to evenly distribute the hourly traffic density over the subregions, the proposed decomposition approach chooses to also include Lage Zwaluwe in that subregion. Overall, at first glance, the subregions seem to be more evenly distributed in terms of hourly traffic density. To further study this, the subregions are compared quantitatively.

To be able to compare the two decompositions quantitatively, the baseline decomposition needs to be quantified. Because the baseline decomposition is acquired in a qualitative format this is done using post processing. The baseline decomposition can be expressed in timetabling points, in the same manner as the proposed decomposition. By identifying which timetabling points belong to a subregion subsequently the objective values can be determined for the baseline decomposition. The term objective values refers back to the objective function of the decomposition model, in which the weighted sum of two objectives is to be minimised. These two terms represent the two criteria of the decomposition model: the amount of trains crossing subregion borders and the distribution of subregion size in terms of hourly traffic density. Mathematically, these objective values are represented as their respective term in the objective function (Equation 3.4), i.e.:

- **Subregion crossings:** The number of trains crossing a subregion border every hour, abbreviated as SC

$$SC = \sum_{f \in F} \left(\sum_{j \in \{1, ||E_f|| - 1\}} (\gamma_{fj}) \cdot \rho_f \right) \tag{5.1}$$

- **Subregion size variance:** The difference between the size of the respective subregion and the average subregion size, in terms of hourly traffic density, summed over all subregions, abbreviated as SS

$$SS = \sum_{r \in R} \left| \sum_{i \in E} (\alpha_{ri} \cdot t_i) - \frac{\sum_{i \in E} t_i}{||R||} \right| \tag{5.2}$$

In the decomposition approach these terms are to be minimised, so a lower value of these terms would mean that the decomposition is more desirable based on those criteria. In Figure 5.7 the objective values are compared for the two decompositions, including the total objective value which is the sum of Equation 5.1 and 5.2.

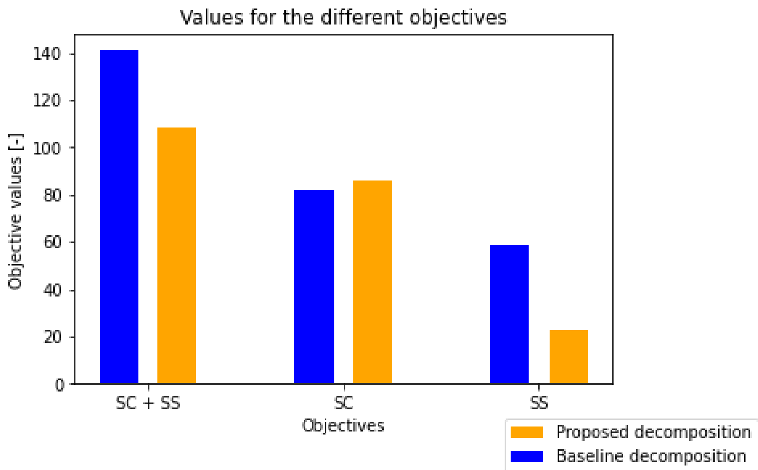
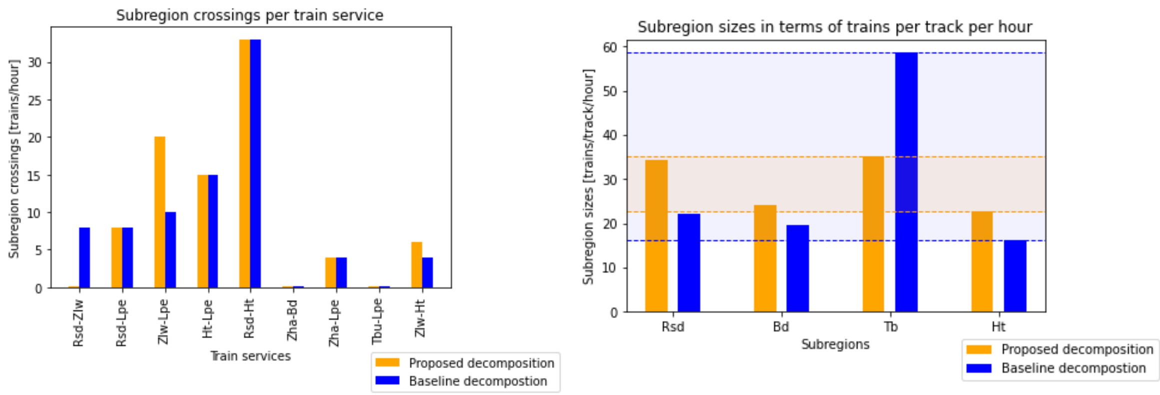


Figure 5.7: the objective values of the respective decompositions

The graph shows that according to the defined criteria, the proposed decomposition is more desirable. Looking at the subregion crossings, SC, the two decompositions have similar values, in fact the proposed decomposition has a slightly larger value for this criterion. This means the reduction of the total objective value is due to the fact that the value for subregion size variance is significantly lower. This supports the qualitative observation made previously. On top of that, this supports that using the proposed decomposition approach can considerably reduce the variance in subregion size. The composition of the individual objectives is presented in Figure 5.8.



(a) Number of subregion crossings per train service for the respective decompositions

(b) Subregion sizes of the respective decompositions in terms of hourly traffic density

Figure 5.8: Details on the objective values of the respective decompositions

In Figure 5.8a, it can be seen how many trains cross a subregion for every train service. Differences between the two decompositions can be seen in a few train services. In the train service from Lage Zwaluwe to Liempde (Zlw-Lpe) and Lage Zwaluwe to Den Bosch (Zlw-Ht) the proposed decomposition leads to more subregion crossings. On the other hand, in the train service from Roosendaal to Lage Zwaluwe (Rsd-Zlw) the proposed decomposition leads to no subregion crossings, while the baseline decomposition does. All three of these differences can be attributed to the fact that the proposed decomposition approach has chosen Lage Zwaluwe to be part of the Roosendaal subregion, instead of the Breda subregion as in the baseline decomposition. This consideration is made by the model to be able to reduce the subregion size variance. In the proposed decomposition the size variance in terms of hourly traffic density is considerably lower than in the baseline decomposition. The nature of this difference is reflected in Figure 5.8b, where the size of every subregion is shown in hourly traffic density and the variance is indicated by a band between the minimum and maximum value. Here it becomes visible that the subregions are indeed less varying in size, compared to the baseline decomposition. In the baseline decomposition especially the Tilburg subregion is significantly larger than the rest of the subregions. In the proposed decomposition the size of the Tilburg subregion is reduced by distributing the hourly traffic density more evenly over all the subregions.

Concludingly, the proposed decomposition appears more desirable in terms of the defined criteria. However, this does not yet give insight into what the impact of the proposed decomposition is on the conflict resolution, which is the question this thesis is aiming to answer. The next subsection investigates this question further.

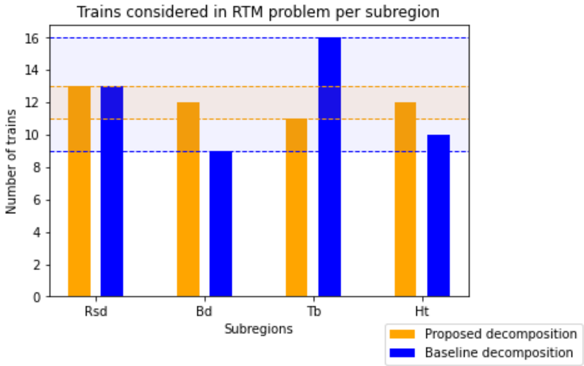
5.3.2 Decisional complexity and interactions in a non-centralised railway traffic management approach using two different decompositions

The objective of the decomposition is to minimise two main key elements: subregion size variance and subregion crossings. The overall objective of these two elements is to reduce decisional complexity and keep interactions between agents to a minimum. Combining the decompositions, baseline and proposed, and the simulated train traffic, it can be measured what the impact of these decomposition criteria is.

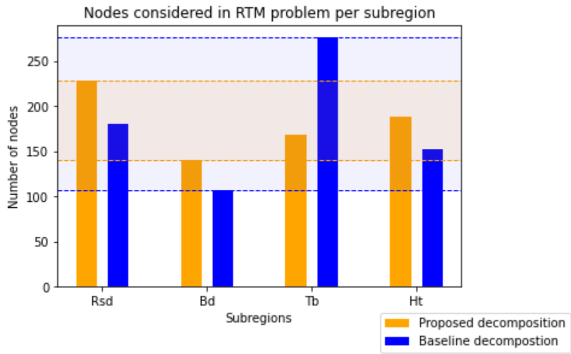
Decisional complexity

The criterion to be reviewed first is the subregion size variance. This criterion has as a main goal to reduce the decisional complexity of the railway traffic management problem in the subregions by distributing it. To measure the decisional complexity within the conflict resolution problem, two different things are measured: the size of the Alternative Graph problem and the computation time of the conflict resolution problem.

As stated in section 4.1 the conflict resolution problem is modelled in the form of an Alternative Graph, which consists of nodes representing a certain operation of a certain train. The size of the Alternative Graph in a subregion is representative for the size and complexity of the conflict resolution problem in that region. It is chosen to reflect the size of the Alternative Graph in two parameters: the number of trains considered in the Alternative Graph in a subregion and the number of nodes in the Alternative Graph, excluding the sink and source node. Because the proposed decomposition aims to distribute the size of the conflict resolution problem, it is expected that these parameters are more distributed among the subregions than when using the baseline decomposition.



(a) Number of trains considered in the conflict resolution problem of each subregion for the respective decompositions



(b) Number of nodes considered in the conflict resolution problem of each subregion for the respective decompositions

Figure 5.9: Size of the conflict resolution problem in the subregions for the respective decompositions

The size of the Alternative Graph problem formulation for both decompositions is shown in Figure 5.9, where the bands range from the minimum to the maximum value to visualise the variance. Subfigure 5.9a, shows the trains considered in the Alternative Graph in each subregion of the respective decompositions. Note that if a train travels from one subregion to another subregion, it is counted in both subregions as it will contribute to the problem size in both of those subregions. From this it becomes visible that, using the proposed decomposition, every subregion considers around 12 trains. Whereas when using the baseline decomposition the subregions consider from 9 to 16 trains. Computing the standard deviation of the number of trains considered in every subregions gives 0.71 when using the proposed decomposition and 2.7 using the baseline decomposition.
The number of nodes considered in the subregions in the respective decompositions are shown in subfigure 5.9b. Using the proposed decomposition the Alternative Graphs of the subregions contain 140 to 230 nodes, while using the baseline decomposition the subregions contain 100 to 270 nodes. Computing the standard deviation of the number of nodes considered in every subregions gives 32.0 when using the proposed decomposition and 61.8 using the baseline decomposition.
Combining the two graphs, it can be concluded that the size variance of the Alternative Graph problem between the subregions is smaller when using the proposed decomposition than when using the baseline decomposition. *The standard deviations of the problem size, in terms of trains and nodes considered in the conflict resolution is decreased drastically with a factor of almost 4 and 2 respectively.* In other words, the problem size of the conflict resolution is more evenly distributed when using the proposed decomposition.

The computation time of the conflict resolution problem is also used to measure the distribution of decisional complexity. How long the computer takes for calculation is related to how complex the problem is to resolve. The computation time of the conflict resolution is compared for the two decompositions in terms of how many trains are initially delayed. In the previous graphs, it could be seen that the size of the conflict resolution problem is more evenly distributed across the subregions when using the proposed decomposition. As a result it is expected that the decisional complexity is also more equally distributed among the subregions using the proposed decomposition, which would result in a lower computation time.

In Figure 5.10, the relative computation time is plotted against the number of initially delayed trains. The relative computation time entails the computation time of the conflict resolution using the proposed decomposition divided by the computation time of the conflict resolution using the baseline decomposition. A relative computation time lower than 100 % means that the computation time when using the proposed decomposition is lower than when using the baseline decomposition. Each of the dots can be translated back to one of the delay scenarios shown in Table 5.1. Between these dots a trendline is fitted, which can visualise the general trend in the data.
In the absolute data, it is observed that using the non-centralised railway traffic management approach leads to a computation time lower than 1 second using either decomposition. This can be considered fast computation and to be suitable for real-time implementation as a digital aid for train dispatchers.

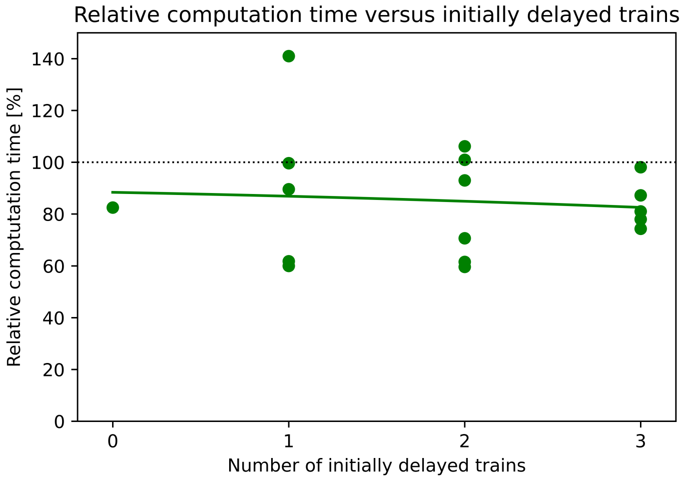


Figure 5.10: The relative computation time of the conflict resolution problem using the proposed decomposition compared to the baseline decomposition versus the number of initially delayed trains

Secondly, in the graph it can be observed that the computation time of the conflict resolution using the proposed decomposition is generally lower than when using the baseline decomposition. Moreover, when more trains are delayed, so the conflict resolution problem becomes more complex, all relative computation times are lower than 100 %. The trendline indicates that the relative computation time decreases to *an average computation time reduction of 20 %*. Meaning that the proposed decomposition allows significantly faster computation of the conflict resolution, especially for more complex formulations. This behaviour can be especially beneficial in complexer problem formulations, such as larger networks, more complex delay situations or additionally incorporated local constraints.

Overall, it can be concluded that the size and decisional workload of the conflict resolution in subregions can be improved by using the proposed decomposition.

Interactions

The second criterion used in the proposed decomposition approach is to minimise the amount of trains crossing subregion borders. It has as a main goal to reduce the amount of interactions between different agents. To study the effects of this criterion, the interaction between the agents is measured when solving the conflict resolution problem using the different decompositions.

Two agents interact when a delayed train travels from one subregion to another. So the trains that cross a subregion border as a consequence of the decomposition of the region, all form a potential interaction between agents. Because these interactions take up computational time and efficiency, they are desired to be minimal. Reflecting back on Figure 5.7, it is already known that there is only a slight difference between the amount of trains crossing a subregion border in the different decompositions. With this knowledge it is not expected that large distinctions are observed between the interactions of agents in the conflict resolution problem using the baseline decomposition and the proposed decomposition.

In Figure 5.11 the difference in the number of interactions can be seen during solving the conflict resolution problem using the proposed decomposition compared to the baseline. Subtracting the number of interactions in the conflict resolution using the proposed decomposition from those using the baseline decomposition, gives the difference in number of interactions. A positive value of this difference means that the proposed decomposition leads to more interactions between agents in the conflict resolution than the baseline decomposition. Each dot references the measurement of a specific delay scenario for a different number of trains in the initial delay scenario and the size of the dot signifies how many delay scenarios return that value.

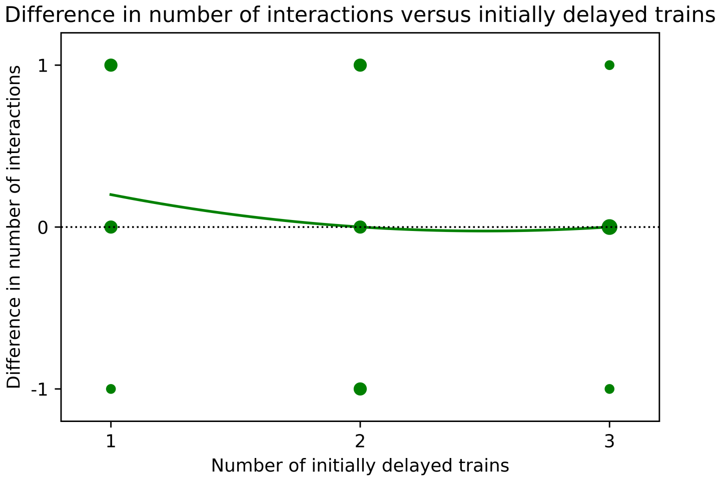


Figure 5.11: The difference in the number of interactions between agents using the proposed decomposition compared to the baseline decomposition versus the number of initially delayed trains

Observing the measurements and the trend line, no real distinctions can be seen between the two decompositions. In some cases the proposed decomposition leads to an additional interaction compared to the baseline decomposition in the same delay scenario, but in other cases the opposite is true. The trend line supports this by showing a trend around the value of zero difference in interactions. As a conclusion, the proposed decomposition *does not lead to an increase or decrease in interactions between agents* in the conflict resolution problem compared to the baseline decomposition. This result is consistent with the fact that both decompositions have around the same amount of subregion crossings.

5.3.3 Delay performance of a non-centralised railway traffic management approach using two different decompositions

In the literature review on the KPIs used to review non-centralised railway traffic management approaches (section 2.4), it is observed that every studied paper uses delay related KPIs to measure their contribution. Evidently, delay is a main performance indicator when studying non-centralised railway traffic management approaches. In this study the delay is also used as a performance indicator in several different forms, to observe what impact the decomposition has on the delay mitigation of the non-centralised conflict resolution.

Total output delay

One form of the delay that is studied, is the relative total output delay of the conflict resolution using the proposed decomposition compared to the baseline decomposition. Before elaborating on the delay reduction, it needs to be established what is understood by the term output delay. The **output delay** is the delay a train has when it leaves the simulated traffic, which can be for two reasons: the train leaves the region or the end of the simulated 15 minute time period is reached. The total output delay is the output delay summed over all trains.

The relative total output delay is defined as the total output delay of the conflict resolution using the proposed decomposition divided by the total output delay using the baseline decomposition. Making this comparison gives insight in the impact of the decomposition on the absorption of delay. The more delay that is absorbed, the less impact the initial delay, or in other words, the lower the total output delay, the better the delay performance. Resulting from the formulation of the relative output delay, a value lower than 100 % means that there is less output delay when using the proposed decomposition.

Figure 5.12 plots the relative total output delay against the number of initially delayed trains. The measurements of the delay scenarios are represented by the dots and the size of the dot signifies how many delay scenarios return that value. The line visually represents the trend in the data.

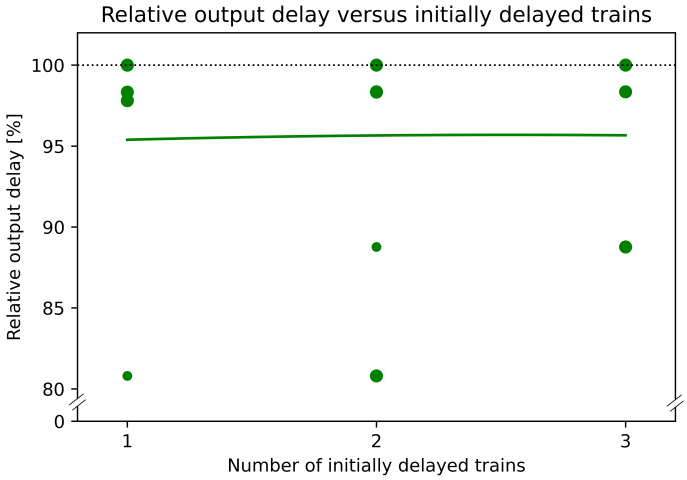


Figure 5.12: Relative total output delay of the conflict resolution using the proposed decomposition with respect to the conflict resolution using the baseline decomposition versus the number of initially delayed trains

From this plot two conclusions can be drawn. Firstly, all values of the delay reduction are equal to or smaller than 100 %, which means when using the proposed decomposition the conflict resolution solution resulted in the same if not less output delay than when using the baseline decomposition. This means the proposed decomposition allows the conflict resolution to generally absorb more delay. The proposed decomposition allows the conflict resolution to *reduce the output delay up to 19 %*, compared to when using the baseline decomposition. Secondly, the trendline shows a fairly consistent trend around the value of 95 % relative total output delay over the different values of initially delayed trains. Concludingly, *the proposed decomposition allows the conflict resolution to mitigate on average 5 % more delay* than the baseline decomposition. This means the proposed decomposition can be able to considerably reduce the total output delay of the conflict resolution.

Maximum and minimum delay

Another studied characteristic of the delay is minimum and maximum values of the output delay of the conflict resolution solution using the proposed decomposition compared to the baseline decomposition. That means the minimum is taken of the output delay of all trains that have an output delay. So the trains that have no output delay are not considered in the calculation of the minimum. Consequently when taking the maximum output delay, this is the maximum delay over all the delayed trains.

Having the minimum and maximum measurement of the output delay over all trains between the conflict resolution using the different decompositions, can show additional insight into how the decomposition affects the conflict resolution. More specifically it can give insight into how the total delay is distributed among the trains. It is desirable that the variance of output delay over the trains is low, because that means that there is not a large distribution between the minimum and maximum output delay.

Figure 5.13 shows the relative distribution of the output delay versus the number of initially delayed trains. The distribution of output delay is defined as the maximum value of output delay minus the minimum value. To form the relative distribution of output delay, the output delay distribution using the proposed decomposition is divided by the output delay distribution using the baseline decomposition. A value lower than 100 % thus means that the proposed decomposition allows a smaller distribution of output delay values. The dots in the graph show the measurements for the different delay scenarios and the size indicates for how many delay scenarios that value is measured.

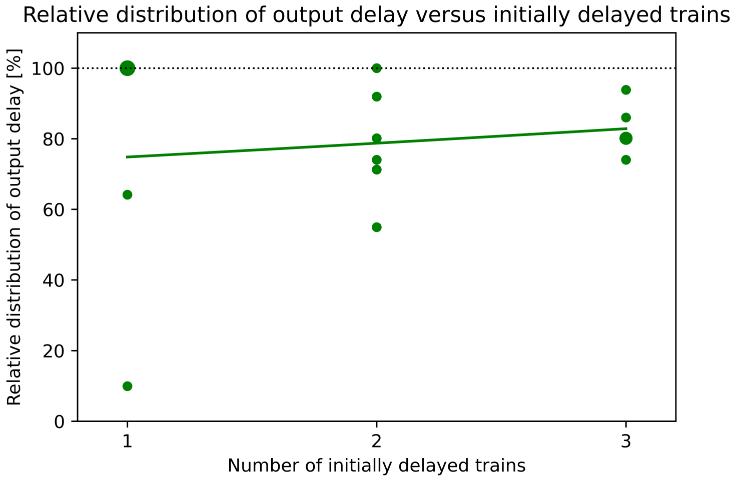


Figure 5.13: Relative distribution of the output delay of the conflict resolution using the proposed decomposition compared to the baseline decomposition

The graph shows that generally the values of the relative distribution of output delay are lower than 100 %. More specifically, *the difference between the maximum and minimum output delay is up to 90 % smaller and on average 21 % smaller.* This means that the proposed decomposition is able to significantly reduce the distribution of output delay. This is especially the case for an increasing number of initially delayed trains. A general observation outside of the graph is that the maximum output delay is identical when using the proposed decomposition and the baseline decomposition in all delay scenarios. This means that the proposed decomposition is not able to reduce the maximum output delay given by the conflict resolution solution compared to the baseline decomposition. However the minimum output delay does vary. From another perspective, the proposed decomposition gives the conflict resolution the opportunity to increase the minimum value of output delay and thus better resolve the smaller delays.

Number of delayed trains

The last measurement with respect to the output delay is the number of delayed trains, i.e. the number of trains that have a delay larger than zero when leaving the region or at the end of the 15 minute time frame. Subtracting the number of delayed trains in the conflict resolution using the proposed decomposition from those using the baseline decomposition, gives the difference in number of delayed trains. A negative value for difference in number of delayed trains means that the initial delay has had a lasting impact on less trains. Also this measurement is insightful to study over how many trains the total delay is distributed.

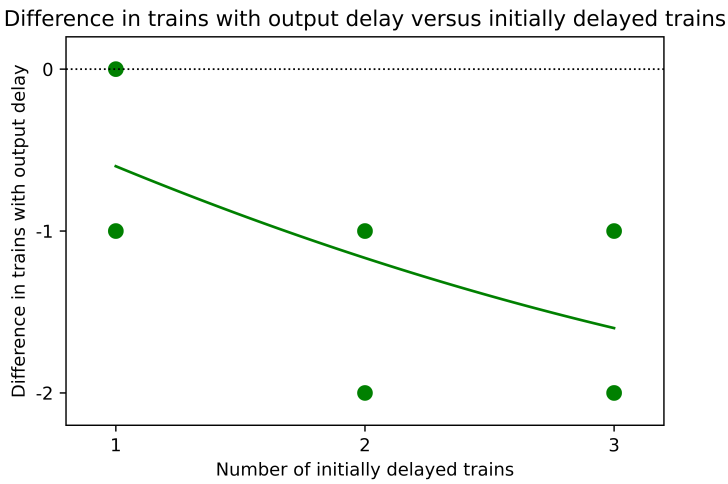


Figure 5.14: Difference in number of trains with an output delay versus the number of initially delayed trains for the conflict resolution using the proposed decomposition compared to the baseline decomposition

In Figure 5.14 the difference in number of delayed trains is plotted against how many trains are delayed in the initial delay scenario. Every dot signifies a measurement from a specific delay scenario and the size of the dot signifies for how many delay scenarios that value occurs.
In the plot it can be seen that generally the difference in delayed trains is a negative value. This means the proposed decomposition leads to the conflict resolution delaying consistently less trains than the baseline decomposition. More specifically, the proposed decomposition allows the conflict resolution to *decrease the number of delayed trains with on average 17 %*. Moreover this trend seems to strengthen as the complexity of the delay scenario increases. Meaning that for increasingly complex initial delay scenarios the proposed decomposition allows for a decreasing amount of delayed trains. This is especially observed going from the delay scenarios with one delayed train to the delayed scenarios with two delayed trains.

Combining all the three graphs on delay KPIs, a conclusion can be drawn on how the proposed decomposition affects the delay performance of the conflict resolution. Using the proposed decomposition the conflict resolution performs better in terms of delay, resulting in less output delay, a smaller distribution within the delay and less delayed trains. So in other words, the proposed decomposition allows the conflict resolution to absorb delays better, especially for the smaller delays.

5.4 Conclusion on the case study analysis of the Roosendaal-Liempde region

This chapter has answered the research question: *How can the impact of a decomposition approach on non-centralised railway traffic management be evaluated?*. It was chosen to test the decomposition on the region from Roosendaal to Liempde. Both decompositions were then subjected to the same non-centralised conflict resolution scheme and the same delay scenarios. It was chosen to simulate the 15 minute time window between 20:25 and 20:40 of a basic timetable for Tuesday in 2022. In the delay scenarios a train was delayed upon entering the region at one or multiple entry points. The impact of the proposed decomposition on the conflict resolution was quantitatively evaluated by measuring a collection of KPIs:

- the trains and nodes considered in the Alternative Graphs of the subregions
- the computation time of the conflict resolution problem
- the number of interactions between agents when solving the conflict resolution problem
- the total output delay of the conflict resolution problem
- the maximum and minimum values of the output delay taken from the trains that experience an output delay
- the number of delayed trains

By analysing these measurements in a series of graphs several conclusions could be drawn. Studying the objective values of both decompositions showed that the proposed decomposition had a more even size distribution over the subregions, in terms of hourly traffic density. This was supported by the fact that *the standard deviations of the amount of trains and nodes considered in the Alternative Graphs of subregions was a factor of almost 4 and 2 smaller, respectively,* in the conflict resolution problem using the proposed decomposition. The trend in the computation time showed that the proposed decomposition allows up to *20 % faster computation* of the conflict resolution problem. This suggests that the proposed decomposition is able to more equally distribute the decisional complexity of the conflict resolution problem over the subregions, than the baseline decomposition. Thus using the size variance in hourly traffic density as a criterion in decomposition forms a way to reduce overall decisional complexity of the conflict resolution problem. With respect to the other criterion in the decomposition, the subregion crossings measured by the *interactions between agents, no real distinction could be found* between using the proposed decomposition and the baseline decomposition.

Three different delay KPIs were studied. The relative output delay showed that the proposed decomposition allows the conflict resolution to mitigate more delay than when using the baseline decomposition, leading to *on average 5 % less output delay*. The maximum and minimum output delays showed that the proposed decomposition leads to less difference between the maximum and minimum output delay of the conflict resolution. *The difference between minimum and maximum is 21 % smaller on average and up to 90 % smaller*. In combination with the knowledge that the maximum output delay of the conflict resolution was the same using both decompositions, this shows the proposed decomposition allows the conflict resolution to mitigate the smaller delays. Additionally, the proposed decomposition is able to allow the conflict resolution to *reduce the number of delayed trains with 17 %*.

All together this shows that using the proposed decomposition, the conflict resolution has a better delay performance by seeming to be able to absorb the delay better especially for the trains with smaller delays.

An observation over all the graphs is that the measurements seem to converge to a certain value with an increasing amount of initially delayed trains, specifically the relative computation time, the interactions, the relative output delay, the relative distribution of output delay and the number of trains with output delay. This is likely because, as more trains are given an initial delay, the results become less incidental to a certain delay scenario. This seeming convergence gives the impression that the results generated in this case study will remain significant and consistent over different types of delay scenarios and when expanding the case study to a larger region or time frame. This is especially relevant to a non-centralised approach, as they are commonly used to tackle these larger and complexer problems. Based on the observations, with the expansion of the number of initially delayed trains the relative numbers will stay consistent leading to increasingly large absolute reduction values. For example, 20 % reduction of computation time can lead to 0.2 seconds less computation time on the basis of a baseline of 1 second of computation time, but for a baseline of 10 seconds this can lead to on average 2 seconds less computation time.

Overall, this evaluation has shown that the proposed decomposition is able to improve the non-centralised conflict resolution in terms of decisional complexity and delay performance. Thus the approach can give valuable insights both to literature and to practice.

6. Conclusions and recommendations

This thesis aims to answer the research question: *What is the impact of an hourly traffic density based decomposition approach in non-centralised railway traffic conflict resolution?*. A new decomposition approach was presented that is able to decompose a region into subregions based on hourly traffic density. A coordination scheme and conflict resolution method were presented to construct a non-centralised railway traffic management approach. To study the impact of the proposed decomposition on the conflict resolution compared to the baseline decomposition from ProRail, both decompositions were subjected to this non-centralised railway traffic management in the case study on the region Roosendaal-Liempde.
In this final chapter, conclusions are drawn on the basis of the sub research questions, section 6.1. Then a main conclusion is drawn based on the main research question, section 6.2. After that the recommendations for future research are discussed, section 6.3.

6.1 Sub conclusions

The sub conclusions in this section are drawn based on the sub research questions and their action points, as they are defined in section 1.1.

6.1.1 What is the current state of the art of centralised and non-centralised conflict resolution?

There are multiple approaches for centralised conflict resolution that have been published, but the specific research field has stopped rapidly developing. Most options within conflict resolution have been explored and the interest has shifted into exploring the adjacent research fields, such as passenger satisfaction, energy management and non-centralisation. Researching non-centralised railway traffic management is driven by an ask for problems with more constraints, more flexibility and less computation time.
The non-centralised approaches have received increasing attention in recent years. The different papers have focused on the different elements of the non-centralisation such as the coordination, the solving algorithm of the conflict resolution, the architecture and the decomposition. Although various decompositions have been researched, the topic has not yet seen much in-depth attention. There has been little research into how a railway system can be decomposed based on decisional complexity and what the impact of a decomposition is on the conflict resolution problem.
In the research on non-centralised approaches the most common KPIs concern delay performance and computation time.
The state of practice already uses a non-centralised system architecture and thus also uses a decomposition. This decomposition has come about organically and historically with changes being implemented over the years, but the division of train dispatcher subregion has not been researched scientifically.
Overall it seems valuable to research how a railway system can be decomposed based on decisional complexity and what the impact of such a decomposition is on the conflict resolution problem.

6.1.2 How can a region of railway infrastructure be decomposed into subregions based on hourly traffic density?

This research studies a decomposition of a region into subregions based on decisional complexity. This is done by decomposing based on hourly traffic density, using the hourly traffic density as a main criterion, more specifically minimising the size variance between subregions in terms of hourly traffic density. The decomposition model is an optimisation model that assigns timetabling points to subregions while minimising two criteria: the size variance and the amount of trains crossing subregions. The model considers the train services, the frequency of these train services and the timetabling points. It was shown that this decomposition model can decomposed the Rotterdam Centraal - Leiden Centraal corridor by taking both the size variance and the subregion crossings into account.

6.1.3 How can a non-centralised conflict resolution model be designed?

The non-centralised conflict resolution model consists of two elements: The conflict resolution method and the coordination scheme. The conflict resolution method used in this thesis is the Alternative Graph method, which formulates the conflict resolution problem using graph theory. Using that formulation the model can perform reordering and retiming measures in the case of conflict. The used coordination scheme is delay driven communication using a coordinator and a delay transition constraint. This non-centralised conflict resolution model has shown to be able to perform reordering and retiming measures and coordinate effectively between agents and in a delay driven manner.

6.1.4 How can the impact of a decomposition approach on non-centralised railway traffic management be evaluated?

The non-centralised conflict resolution was subjected to two different decompositions: the proposed hourly traffic density based decomposition and the baseline decomposition of the current train dispatcher subregions. The conflict resolution was imposed on the Roosendaal - Liempde region using a ProRail timetable and delay scenarios with delayed trains at the entry points. This case study showed that the decomposition is able to better distribute the decisional complexity of the conflict resolution problem among the subregions compared to the baseline decomposition, which is reflected in a lower standard deviation in size and up to 20 % reduction of computation time. Introducing the proposed decomposition did not have a discernible impact on the amount of interactions between agents. The delay performance of the conflict resolution is improved by using the proposed decomposition compared to the baseline decomposition: the difference between the maximum and minimum delay is on average 21 % smaller, the amount of delayed trains is reduced by 17 % on average and the total output delay is reduced by on average 5 %. Overall it was concluded that using the proposed decomposition for the non-centralised conflict resolution, reduces the overall decisional complexity and improves the mitigation of delay.

6.2 Main Conclusion

The main research question is: *What is the impact of an hourly traffic density based decomposition approach on non-centralised railway traffic management?*. To answer this question, firstly, a hourly traffic density based decomposition approach was developed to decompose the railway system into subregion with similar decisional complexity. Secondly, a non-centralised railway traffic management approach was designed using Alternative Graph conflict resolution and a delay driven communication scheme with a coordinator. The railway traffic management approach was applied to a case study using the proposed hourly traffic density based decomposition approach. This was compared to the same railway traffic management approach but then using the baseline decomposition now used in practice by ProRail. This has shown that the hourly traffic density based decomposition can reduce the computation time by up to 20 % and reduce the total delay with on average 5 %.

From this it can be concluded that the proposed decomposition can distribute decisional complexity evenly across subregions and thereby reduce the decisional complexity and workload of a single subregion, providing more effective and time-efficient rescheduling solutions. This shows great promise for further research and the state of practice.

6.3 Recommendations

The results of this research give insights that could be further investigated or expanded in the future. Also some aspects of the research could be looked into more to investigate the impact of a decomposition approach even more. Because the decomposition approach has proven to be beneficial to the conflict resolution, future research could be very interesting. The recommendations for future research are discussed in this section.

6.3.1 Recommendations following the hourly traffic density based decomposition approach

One of the main focuses of this research has been the development of a hourly traffic density based decomposition approach. The approach has shown promise and therefore it can be interesting to elaborate the approach.
Firstly, an extra step can be added to the decomposition by determining the optimal amount of subregions. In the current approach a fixed amount of subregions is used that is derived from the amount of subregions in the baseline decomposition. However this amount of subregions is not necessarily the optimal amount of subregions. The determining of the optimal amount of subregions could form a step prior to the decomposition. The term 'optimal' can be then chosen to refer to several things.

For example, it could refer to the amount of subregions with which the hourly traffic density can be decomposed most evenly or it can refer to that the subregions' hourly traffic density is as close to the desired complexity as possible.
Secondly, it can be interesting to perform a sensitivity analysis on the decomposition approach. This entails a study of how sensitive the decomposition is to changes in the input. For instance, how does the decomposition change when changes are made in the routes of the trains. Other factors to which sensitivity can be researched are the frequency of train services, the timetable used to define traffic flow, the amount of tracks taken into account at timetabling points and the size of the infrastructural elements used to decompose.
Thirdly, more exploration can be done in terms of criteria. This research has chosen to represent decisional complexity using hourly traffic density as a main criterion, but decisional complexity could also be expressed in other ways. Traffic heterogeneity, the number of potential conflicts, the amount of possible rescheduling actions, connectivity or any other factor can possibly also represent the decisional complexity. This approach has shown that an hourly traffic density based decomposition can have a positive impact on the conflict resolution problem, so in future research this can be elaborated by researching which factors considerably affect the complexity of the conflict resolution problem. These factors can then be used in criteria that could be additional to the existing criteria or maybe even replace them. Furthermore it could be interesting to experiment with the weights of the different criteria, additional or existing, to study the effect that would have on the decomposition and the conflict resolution.
Lastly, this research used a decomposition approach were a region was divided into subregions. This choice was made to allow the comparison to the current state of practice decomposition. However in future research it may be interesting to view other types of decompositions and what their impact is. Examples of other types of decompositions that have been discussed in subsection 2.3.4 are mathematical decomposition, e.g. in Kersbergen et al. (2016) [8]; train based decomposition, e.g. in Luan et al. (2020) [9] or macro-micro decomposition, e.g. in Keita et al. (2020) [50].

6.3.2 Recommendations following the non-centralised railway traffic management approach

The first recommendation for the non-centralised railway traffic management approach concerns the conflict resolution. In the process of implementing the Alternative Graph method, some numerical errors were encountered. This was neither the result of the chosen approach nor were there implementation errors. It is suspected that it was the cause of the cooperation between Python and the commercial solver Gurobi and thus a software problem. Therefore for future research using the Alternative Graph model, it is advised to use different programs or a different combination of programs to potentially prevent this.
Because the focus of this research was on developing a decomposition approach, relatively simple approaches were used for the conflict resolution and coordination. For future research into non-centralised approaches it could be interesting to expand these approaches. The conflict resolution used can only do retiming and reordering measures. This could be expanded to an approach that can also take local rerouting actions such as a change of platform. In D'Ariano et al. (2008) [34] a modification of the Alternative Graph is presented that can incorporate these local rerouting actions. Reordering, retiming and local rerouting are the relevant actions in the case of disturbances. If future research was to look into a non-centralised railway traffic management approach for disruptions, a method would need to be chosen that can also encompass rerouting, short turning and cancelling. This could for example be the methods presented in Meng and Zhou (2011) [37] and Lusby et al. (2013) [5].
Similarly for the used coordination approach, in future research it could be expanded. The boundary constraints are a relatively basic form of coordination and the state of the art has developed several cutting edge methods. The boundary constraints ensure continuity, such that the simulated system can represent the physical system. A next step could be to include consistency in the coordination, as done in Toletti et al. (2020) [57], where the conflict resolution strategy is maintained globally. This means, for instance, that if one agent chooses for train B to precede train A this choice is aligned with and enforced by all other agents. Additionally global feasibility and optimality could be included. Global feasibility/optimality meaning that the local solution found by an agent is feasible/optimal for the whole system. Global feasibility is included in the coordinator graph approach in Corman et al. (2010, 2012, 2014) [46, 49, 53]. Global optimality is included in Kersbergen et al. (2014, 2016) [8, 55] by means of introducing a central cost function that is aimed to be minimised. Considering continuity, consistency, global feasibility and global optimality tackles the main challenges in coordination of a non-centralised approach. However there are still additional steps that could be taken. For example, agents can use prediction to anticipate how the delays in neighbouring subregions, or subregions beyond that, will likely impact their subregions. This incorporated by the rolling horizon framework used in Wang et al. (2014) [44] and Cavone et al. (2020) [38].

6.3.3 Recommendations following the case study

In order to efficiently conduct the case study on the region Roosendaal - Liempde some assumptions and restrictions were made, which in future research might be reconsidered. For example the assumption was made to implement the same minimal headway between all trains. In reality this minimal headway will vary between different individual trains and the order of those trains. In further research, these differences in headway can be calculated for all these combinations of trains and incorporated into the case study. This can be done by determining the critical headway on the shared infrastructure of two trains usin the compression method, amongst others explained in Pachl (2014) [76]. It could also be chosen to calculate the headways in terms of certain categories to reduce the amount of parameters. For example, it could be explored to determine the minimal headways between different train types, e.g. freight trains and intercity trains, instead of all individual trains.

Also, in further research, it may be interesting to expand the test case to analyse the behaviour of both the decomposition approach and the impact it has on the conflict resolution. The expansion of the test case can be interesting on two levels. Firstly it could be interesting to study the decomposition of a larger region. This study has considered four train dispatcher subregions, but this could be expanded to more train dispatcher subregions or even a whole corridor. Secondly, the conflict resolution could be viewed for a longer period of time. This research has viewed the conflict resolution for a 15 minute time window. Viewing a longer time window, for example 30 minutes, a few hours or a whole day, could give more insight on the longer term effects of the decomposition on the conflict resolution. Additionally, this gives the option to incorporate larger initial delays.

Additionally, it can be interesting to incorporate different delay scenarios in the case study. This research has delayed one train at every entry point with the planned headway to the preceding train. In an expansion of the amount of delay scenarios, multiple trains can be delayed at a single entry point and/or the amount of initial delay given to a train can be varied. For example, trains can be delayed with 70 % or 130 % of the planned headway.

6.3.4 Recommendations to ProRail

The thesis research was done in cooperation with ProRail, thus it is of interest to view how they can incorporate this theoretical research into their state of practice. Two main opportunities are identified for ProRail to use the knowledge gained from this research.

First of all, the non-centralised railway traffic management approach in combination with the proposed decomposition can be used to construct a Traffic Management System to advise the train dispatchers in their decision making. The TMS can then provide solutions in cases where the train dispatcher may have difficulty overseeing the situation. Because the proposed decomposition is able to reduce decisional complexity, this opportunity could be used to incorporate additional local constraints into the TMS, such as common connections, local transportation or passenger satisfaction. The TMS can include these constraints in the decision making that a train dispatcher is not always able to take into account. Ultimately, the TMS could support the train dispatcher to make more all-encompassing choices.

As a second usage of the knowledge gained in this thesis, the proposed decomposition approach can form a tool to reflect on the current decomposition of the train dispatcher subregions. The case study has shown that the proposed decomposition results in better delay performance than the baseline decomposition. If this is the case with agents doing the conflict resolution this may also be the case when human train dispatcher are resolving the conflict. Thus the proposed decomposition approach can be used to distribute the decisional complexity more evenly among the train dispatcher subregions, allowing the train dispatcher to solve conflicts in a more effective and efficient manner. It is known that ProRail is considering research into the decomposition of the train dispatcher subregions and the benefits of the proposed decomposition could be a valuable insight. At the very least, the proposed decomposition approach can be a tool that is able to quantify the current decomposition and possible future decompositions.

Concludingly, the knowledge presented in this thesis can be valuable to ProRail both in the field of non-centralised TMS and in the research into train dispatcher subregion decompositions. In other words, the knowledge in this thesis can be valuable in the overall innovative research at ProRail.

Appendices

This part of the thesis contains several appendices to provide more details on the research.

Appendix A describes the inputs to the decomposition of the Roosendaal - Liempde region. Appendix B aims to support the selection of the 15 minute interval within the 24 hour Tuesday timetable with a data analysis. Appendix C shows details on the implementation of the models presented in this thesis. It provides some illustrative algorithms and a reference for the conflict resolution implementation. Appendix D displays the detailed outputs from the conflict resolution in the case study for the various delay scenarios. Lastly, Appendix E contains a scientific paper of the research presented in this thesis.

A. Case study inputs for hourly traffic density based decomposition of the Roosendaal-Liempde region

This section contains the determination of the value of weight factor ζ and also the data input to the decomposition of the Roosendaal-Liempde regions viewed in the case study.

A.1 Determination of the weight factor

The weight factor ζ can take any value between, and including, 0 and 1. For the decomposition of the Roosendaal - Liempde region an analysis is performed to determine which value to use for the weight factor.

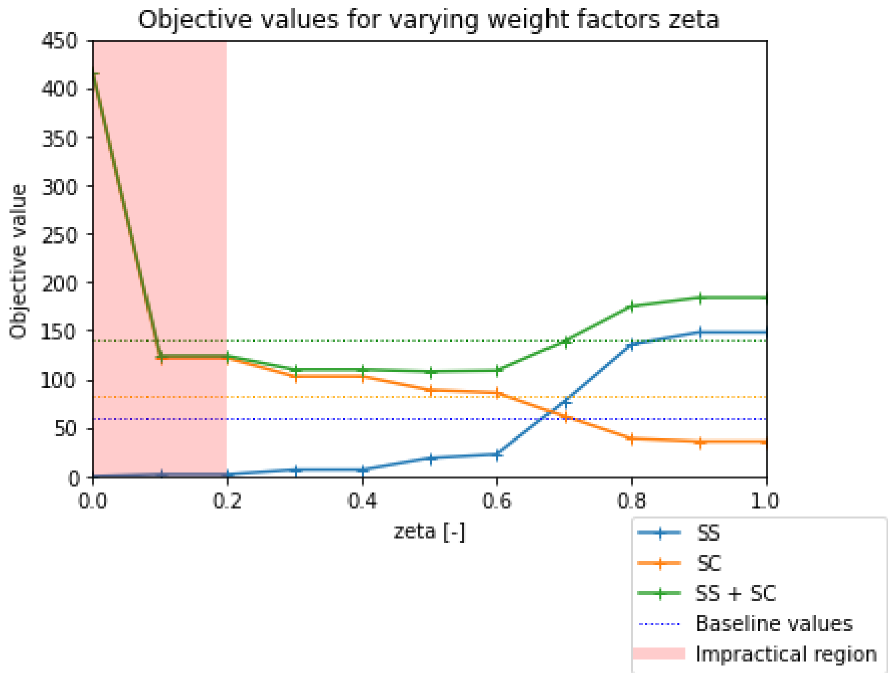


Figure A.1: The objective values for different values of weight factor ζ

The graph in Figure A.1 shows how the objective values change for values of the weight factor. The abbreviation SS stands for the subregion size variance in terms of hourly traffic density and SC stands for the subregion crossings. Values of both these objectives are as defined in Equation 5.1 and Equation 5.2 in subsection 5.3.1. The dotted lines show the value of the respective objective of the baseline decomposition. This value has been calculated in post processing based on the trains dispatcher subregions in practice and thus does not vary with weight factor ζ. The impractical region in red, indicates that for these values of ζ the subregions no longer consist of only neighbouring timetabling points but the timetabling points belonging to one subregion are scattered around the region. This is practically undesirable and thus these values will not be used for the decomposition. Observing the graph shows that the two objectives alternate each other as being the focus of the minimisation. At $\zeta = 0$ the variance in subregion size is reduced to zero, whereas at $\zeta = 1$ the size variance has increased and the subregion crossings are kept to a minimum. In determining the value of ζ to be used, it is desired to keep the total objective $SS + SC$ as low as possible, because for these values of ζ the decomposition will be the best according to the used criteria. The total objective is minimal for ζ values ranging from 0.3 to 0.6 . Within these weight factors, the values of the two individual objectives vary. Although the value of SS increases with increasing ζ, it is still considerably lower than the baseline value. The value of SC on the other hand is actually above the baseline value. It is chosen to use the ζ for which the value of SC as close to the baseline as possible, because then this research is most likely able to improve on the baseline decomposition. Choosing a ζ with a higher amount of subregion crossings, would counteract this. That is why a value of $\zeta = 0.6$ is chosen to use in the decomposition.

A.2 Input data to the decomposition

Table A.1 shows all the timetabling points contained in the region. The table shows their full names, abbrevations and the type of timetabling point. There are two types of stations: the intercity station and the sprinter station. Only sprinter trains stop at the sprinter stations, whereas all passenger trains stop at intercity stations. A freight stop is a timetabling point that is connected to a freight yard, where freight trains can be filled, emptied and/or shunted. At a junction timetabling point, multiple tracks converge to form a smaller amount of tracks, which as a result is a point that is explicity considered in the timetable. The last column in the table shows the hourly traffic density values of the respective timetabling points.
Table A.2 provides an overview of the routes and frequencies inputted to the decomposition model. The routes are recorded by an ordered series of zeros and ones. A one means that the train service travelling that route passes through that respective timetabling point, a zero means it does not. The bottom row shows the frequency of the trains service, i.e. how many trains travel that route per hour.

Timetabling point	**Abbrevation**	**Type**	**Hourly traffic density [trains/track/hour]**
Roosendaal	Rsd	Intercity station	5.9
Oudenbosch	Odb	Sprinter station	4.5
Zevenbergen	Zvb	Sprinter station	4.5
Etten-Leur	Etn	Sprinter station	4.8
Lage Zwaluwe	Zlw	Sprinter station	4.5
Zevenbergschenhoek aansluiting	Zha	Junction	5.5
Breda Prinsenbeek	Bdpb	Sprinter station	6
Breda	Bd	Intercity station	6.1
Gilze Rijen	Gz	Sprinter station	8.3
Tilburg Reeshof	Tbr	Sprinter station	8.5
Tilburg Universiteit	Tbu	Sprinter station	6.8
Tilburg	Tb	Intercity station	12.3
Tilburg Industrie	Tbi	Freight stop	6.5
Oisterwijk	Ot	Sprinter station	6.5
Vught Aansluiting	Vga	Junction	4.7
's Hertogenbosch	Ht	Intercity station	7.8
Vught	Vg	Sprinter station	3.8
Boxtel	Btl	Sprinter station	5.1
Liempde	Lpe	Junction	4.6

Table A.1: Timetabling points in the Roosendaal - Liempde region

Timetabling points	Routes								
	Rsd-Zlw	Rsd-Lpe	Zlw-Lpe	Ht-Lpe	Rsd-Ht	Bdpd-Bd	Bdpd-Lpe	Tbu-Lpe	Zlw-Ht
Rsd	1	1	0	0	1	0	0	0	0
Odb	1	0	0	0	0	0	0	0	0
Zvb	1	0	0	0	0	0	0	0	0
Etn	0	1	0	0	1	0	0	0	0
Zlw	1	0	1	0	0	0	0	0	1
Zha	0	0	1	0	0	1	1	0	1
Bdpb	0	0	1	0	0	1	1	0	1
Bd	0	1	1	0	1	1	1	0	1
Gz	0	1	1	0	1	0	1	0	1
Tbr	0	1	1	0	1	0	1	0	1
Tbu	0	1	1	0	1	0	1	1	1
Tb	0	1	1	0	1	0	1	1	1
Tbi	0	0	0	0	1	0	0	0	1
Ot	0	1	1	0	0	0	1	1	0
Vga	0	0	0	1	1	0	0	0	1
Ht	0	0	0	1	1	0	0	0	1
Vg	0	0	0	1	1	0	0	0	1
Btl	0	1	1	1	0	0	1	1	0
Lpe	0	1	1	1	0	0	1	1	0
Frequency [trains/hour]	8	4	10	15	11	6	4	4	2

Table A.2: Train services and frequencies in the Roosendaal - Liempde region

B. Timetable selection for conflict resolution in the Roosendaal - Liempde case study

For the conflict resolution within the case study, 15 minutes are selected from a basic Tuesday timetable. The figures in this appendix aim to provide some support for the selection of the 20:25-20:40 interval. The main goal of the timetable selection is to show traffic heterogeneity and trains at the Points of Interest for conflict. This is sufficient to use it in conflict resolution to compare the impact of the proposed and baseline decomposition. Figure B.1 shows the train activities in every hour of the 24 hour timetable for the four different train types included in the case study: freight trains, intercity trains, sprinter trains and high speed trains. Here, a train activity is defined as an activity included in the timetable as an arrival, departure or pass through at a timetabling point. A first look at the graph shows that there is significantly less traffic during the night hours, but for the remainder of the day the difference in train activities is minimal. However, the focus of this graph is to give insight in the traffic heterogeneity throughout the timetable. The traffic heterogeneity does not vary largely. An hour is selected where all types of trains are present and somewhat distributed. This hour is from 20:00 till 21:00.

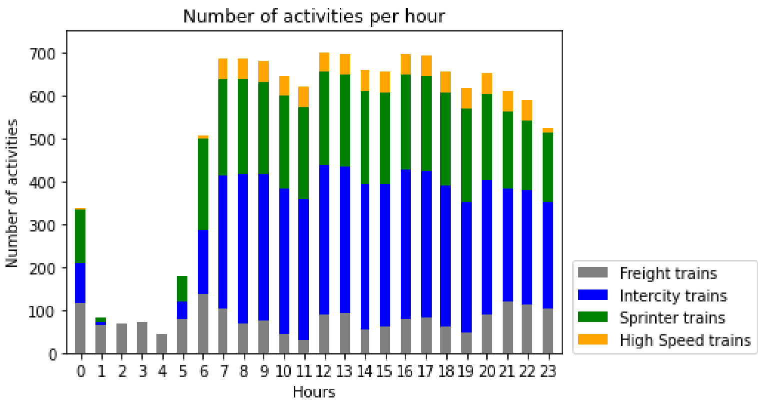


Figure B.1: The number of train activities of different train types in the 24 hour basic Tuesday timetable

Figure B.2 shows the number of trains present at the Points of Interest for conflict during 5 minute intervals in the hour from 20:00 and 21:00, which was selected based on the previous graph. These main Points of Interest are the three timetabling points that contain junctions with converging and crossing tracks: Zevenbergschenhoek aansluiting, Liempde and Vught aansluiting. The most trains are present at these Points of Interest from 20:15-20:20 and 20:30-20:40. Apart from sheer presence, it also important if the trains share infrastructure, which is the starting point of a potential conflict. This was more the case for the 20:30-20:40 interval. Because the trains present at these Points of Interest quickly declines in the interval from 20:40-20:45, the choice was made to select the 15 minute interval from 20:25-20:40 for the test case.

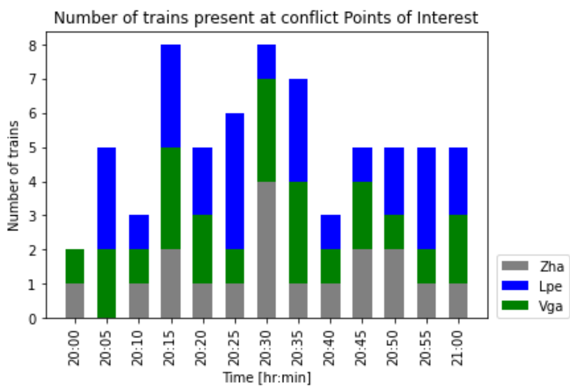


Figure B.2: The number of trains present at three of the Points of Interest for conflict during the interval from 20:00 to 21:00

C. Implementation details of the decomposition model, conflict resolution algorithm and coordination scheme

All three models have been implemented in Python and the optimisation models (the decompisition and conflict resolution model) are solved using Gurobi. This appendix provides details on the implementations.

C.1 Implementation of the decomposition model

Before the decomposition model can be executed as presented in subsection 3.3.3, the input data (see section A.2) is to be manipulated as to fit the mathematical formulation of the decomposition. This is especially the case for parameters E_f and B_f.

Algorithm 1 shows the construction of parameter E_f, which contains the ordered timetabling points that train service f crosses in its route. In the algorithm, *route_matrix* refers to the matrix consisting of zeros and ones, as in Table A.2, which indicates whether a train service f crosses a timetabling point i in its route or not. The vector *timetabling_points* contains the ordered set of all timetabling points. The remaining parameters are the same as described in the mathematical formulation. The parameter B_{fji} describes the route such that the

Algorithm 1 Construction of parameter E_f from input data

```
for f in F do
  E_f[f] ← empty
  for i in E do
    if route_matrix[i, f] · timetabling_points[i] ≠ 0 then
      E_f[f].append(route_matrix[i, f] · timetabling_points[i])
    end if
  end for
end for
```

matrix B_f describes the route matrix for train service f. Every row in the matrix describes a trip from one timetabling point to another in every row and in that row indicates them with -1 and 1 respectively, in the column belonging to those timetabling points. All other values in the row are zero. This is continued to form the whole route of the train service. The implementation of this format was accomplished by using a count function to include one trip in every row and to ensure the values of -1 and 1 are assigned accordingly, as seen in Algorithm 2.

Algorithm 2 Construction of parameter B_f from input data

```
  count ← 0
  B ← empty
for f in F do
    B_f ← zeros(||E_f[f]|| − 1, E)
    i ← 0
    for j in [0,||E_f[f]|| − 1] do
      while count < 1.5 do
      B_f[j, i] ← route_vector[i, f]
      count ← count + route_vector[i, f]
      i ← i + 1
      end while
    count ← 0
    i ← i − 1
    B_f[j, i] ← −1
    end for
B[f] ← B_f
end for
```

C.2 Implementation of the conflict resolution algorithm

The conflict resolution implementation was achieved by using the implementation used in Janssens (2022) [78]. In the respective thesis a moving block version of the Alternative Graph formulation is constructed. For it to be applicable to the fixed block signalling system in the current research, several changes are made. More specifically, several functions and parameters are omitted to return to the original Alternative Graph formulation as described in D'Ariano et al. (2007) [10].

C.3 Implementation of the coordination scheme

The specifics of the algorithm of the coordinator can differ per number of agents and how the subregions of these agents neighbour each other. The functioning of the coordinator is illustrated by showing the coordinator algorithm used in the Roosendaal - Liempde case study in Algorithm 3. The agents are described by the name of their respective subregions: Rsd, Bd, Tb and Ht. The vectors containing the delays of the trains leaving a subregion and entering another, are represented by the names of those subregions. For example, the delays of the trains leaving the Roosendaal region and entering the Breda region are described in vector Rsd_Bd. The algorithm is continued on the next page.

Algorithm 3 The coordinator algorithm

```
  Bd_Rsd, Ht_Tb, Bd_Tb ← 0
  run Rsd(Bd_Rsd, indelay_Rsd)
return Rsd_Bd1

run Tb(Ht_Tb, Bd_Tb, indelay_Tb)
return Tb_Bd1, Tb_Ht1

if Tb_Ht1 > 0 or indelay_Ht > 0 then
    run Ht(Tb_Ht1, indelay_Ht)
    return Ht_Tb1
end if

if Rsd_Bd1 > 0 or Tb_Bd1 > 0 or indelay_Bd > 0 then
    run Bd(Rsd_Bd1, Tb_Bd1, indelay_Zha, indelay_Zlw)
    return Bd_Tb1, Bd_Rsd1
end if

if Bd_Rsd1 > 0 then
    run Rsd(Bd_Rsd1, indelay_Rsd)
    return Rsd_Bd2
end if

if Ht_Tb1 > 0 or Bd_Tb1 > 0 then
    run Tb(Ht_Tb1, Bd_Tb1, indelay_Tb)
    return Tb_Bd2, Tb_Ht2
end if

if Tb_Bd2 ≠ Tb_Bd1 or Rsd_Bd1 ≠ Rsd_Bd2 then
    run Bd(Rsd_Bd2, Tb_Bd2, indelay_Zha, indelay_Zlw)
    return Bd_Tb2, Bd_Rsd2
end if

if Tb_Ht2 ≠ Tb_Ht1 then
    run Ht(Tb_Ht2, indelay_Ht)
    return Ht_Tb2
end if
```

```
if Ht_Tb2 ≠ Ht_Tb1 or Bd_Tb2 ≠ Bd_Tb1 then
   run Tb(Ht_Tb2, Bd_Tb2, indelay_Tb)
   return Tb_Bd3, Tb_Ht3
end if

if Bd_Rsd2 ≠ Bd_Rsd1 then
   run Rsd(Bd_Rsd2, indelay_Rsd)
   return Rsd_Bd3
end if
```

D. Detailed results of the conflict resolution in the Roosendaal - Liempde case study

		Proposed decompostion						**Baseline decomposition**					
Train delayed at entrypoint		**None**	**Zha**	**Ht**	**Tb**	**Rsd**	**Zlw**	**None**	**Zha**	**Ht**	**Tb**	**Rsd**	**Zlw**
Computation time [s]		0.13	0.35	0.29	0.31	0.13	0.26	0.16	0.39	0.43	0.22	0.20	0.26
Interactions		0	0	1	1	0	1	0	1	1	0	0	0
Output delay [s]	**Rsd**	0	0	0	0	236	682	0	0	0	0	236	602
	Bd	0	401	0	127	0	431	0	411	0	0	0	0
	Tb	0	0	169	263	0	0	0	141	223	196	0	0
	Ht	0	0	303	0	0	0	0	0	302	0	0	0
	Region	0	401	223	196	236	592	0	410	276	196	236	602
Maximum output delay [s]		0	141	133	127	124	431	0	141	133	127	124	431
Minimum output delay [s]		0	128	36	669	44	161	0	10	36	69	44	10
Number of delayed trains	**Rsd**	0	0	0	0	3	2	0	0	0	0	3	0
	Bd	0	3	0	1	0	1	0	4	0	0	0	3
	Tb	0	0	2	2	0	0	0	1	3	2	0	0
	Ht	0	0	3	0	0	0	0	0	3	0	0	0
	Region	0	3	3	2	3	2	0	4	4	2	3	3

Table D.1: Detailed results of the scenario without delay and delay scenarios where a train was delayed at one entrypoint

		Proposed decompostion						**Baseline decomposition**					
Train delayed at entrypoints		**Zha Lpe**	**Ht Lpe**	**Zlw Zha**	**Rsd Zlw**	**Rsd Zha**	**Zha Ht**	**Zha Lpe**	**Ht Lpe**	**Zlw Zha**	**Rsd Zlw**	**Rsd Zha**	**Zha Ht**
Computation time [s]		0.261	0.396	0.24	0.253	0.32	0.514	0.438	0.373	0.39	0.272	0.453	0.509
Interactions		1	2	1	1	0	1	1	1	1	0	1	2
Output delay [s]	**Rsd**	0	0	0	919	236	0	0	0	0	236	236	0
	Bd	527	127	682	431	401	401	411		874	602	411	411
	Tb	296	465	703	0	0	169	337	419	141	0	141	364
	Ht	0	303	0	0	0	303	0	302	0	0	0	302
	Region	596	419	864	828	637	623	606	472	874	838	647	687
Maximum output delay [s]		141	133	431	431	141	141	141	133	431	431	141	141
Minimum output delay [s]		69	36	131	44	44	36	10	36	10	10	10	10
Number of delayed trains	**Rsd**	0	0	2	5	3	0	0	0	0	3	3	0
	Bd	4	1	3	1	3	3	4	0	5	3	4	4
	Tb	2	4	0	0	0	2	3	5	1	0	1	4
	Ht	0	3	0	0	0	3	0	3	0	0	0	3
	Region	5	5	4	5	6	6	6	6	5	6	7	8

Table D.2: Detailed results of the delay scenarios where a train was delayed at two entrypoints

		Proposed decompostion					Baseline decomposition				
Train delayed at entrypoints		**Zha Ht Lpe**	**Lpe Zha Zlw**	**Ht Zha Zlw**	**Lpe Zha Rsd**	**Ht Zha Rsd**	**Zha Ht Lpe**	**Lpe Zha Zlw**	**Ht Zha Zlw**	**Lpe Zha Rsd**	**Ht Zha Rsd**
Computation time [s]		0.52	0.35	0.457	0.324	0.51	0.53	0.432	0.524	0.436	0.654
Interactions		2	2	2	1	1	2	1	2	1	2
Output delay [s]	**Rsd**	0	682	682	236	236	0	0	0	236	236
	Bd	527	830	703	527	401	411	874	874	411	411
	Tb	465	296	169	296	169	560	337	364	337	364
	Ht	303	0	303	0	303	302	0	302	0	302
	Region	819	1060	1087	832	860	882	1070	1150	843	923
Maximum output delay [s]		141	431	431	141	141	141	431	431	141	141
Minimum output delay [s]		36	69	36	44	36	10	10	10	10	10
Number of delayed trains	**Rsd**	0	2	2	3	3	0	0	0	3	3
	Bd	4	4	3	4	3	4	5	5	4	4
	Tb	4	2	2	2	2	6	3	4	2	4
	Ht	3	0	3	0	3	3	0	3	0	3
	Region	8	6	7	8	9	10	7	9	9	11

Table D.3: Detailed results of the delay scenarios where a train was delayed at three entrypoints

E. Research Paper

HOURLY TRAFFIC DENSITY BASED DECOMPOSITION IN NON-CENTRALISED RAILWAY TRAFFIC MANAGEMENT

Gitte Hornung, Vasso Reppa, Egidio Quaglietta and Dick Middelkoop

***Abstract* – In Traffic Management Systems for railway traffic management, non-centralised approaches have been receiving an increasing amount of attention. However, one inherent aspect of a non-centralised approach, the decomposition, is not often focused on in research. The contribution of this paper is to develop a decomposition approach that incorporates both aspects of the railway system, the infrastructure and the train traffic, and to provide insight into the impact of a decomposition approach on the non-centralised conflict resolution. A decomposition approach is developed that decomposes a railway network into subregions based on an even distribution of the hourly traffic density among the subregions. A non-centralised railway traffic management approach is constructed using Alternative Graph conflict resolution and delay-driven communication with a coordinator. This approach is applied to the proposed decomposition and a baseline decomposition. It is concluded that the proposed decomposition can reduce the computation time by up to 20 % and reduce the total delay with on average 5 %. Thus it can be said that the proposed decomposition is able to reduce the overall decisional complexity of the conflict resolution problem and improve delay performance. Therefore, the proposed decomposition approach and the provided insights can be valuable to future research.**

I. INTRODUCTION

In the case of conflict in the timetable, action must be taken to resolve this conflict in order for the train operations to continue. In current Dutch practice this task resides with the train dispatchers. Such an action can be changing the timing or order of the trains, changing the route of a train or even cancelling (a part of) the train's route. In the case of conflict it can be become a large puzzle which action(s) will lead to the least delay and passenger discomfort. To decide which action(s) to take the train dispatcher can consult preset action plans, but they will largely rely on their own knowledge and experience. However it can be difficult to oversee the situation when multiple trains get involved and taken actions will have consequences on yet another selection of trains. Besides that it is a difficult and demanding task to manage these conflicts, a train dispatcher will not always be able to take all factors into account when finding a solution.

In these cases a *Traffic Management System (TMS)* can support the train dispatcher. A TMS is an intelligent system that is able to take many factors into account considering the infrastructure and traffic and uses a conflict resolution algorithm to find a solution to the conflict [1]. The intelligence of the TMS combined with the knowledge of the train dispatcher, can let the train dispatcher come to a computer aided decision.

The development of Traffic Management Systems has been a research subject for many years. Initially these approaches were mainly *centralised*, meaning that one single decision system, or *agent*, will find a solution to the whole region. Because the railway traffic management is already a problem with many different factors, finding a decision for a large or complex situation can be challenging and time consuming if done by one single agent [2–4]. That is why research has shown increased interest in *non-centralised* approaches. In these approaches there are multiple agents that are each responsible for their own subsystem of the region, which can be a.o. a selection of trains, a route or a subregion. Using a non-centralised approach has several advantages:

- **Flexibility** The overall railway traffic management problem is broken into smaller problems. This means that within these smaller problems the complexity of the decision is reduced and there is more flexibility in finding a good solution.
- **Scalability** Because the problems are smaller, additional factors can easily be added to the problem. For example, common passenger connections or local transport can be added to be taken into account.
- **Computation time** A reduced decisional complexity also means that the problems are commonly faster to compute. With a reduced computation time the non-centralised approaches are more fit for real-time implementation.

Apart from these advantages, a non-centralised approach also comes with challenges. The subsystems need to be coordinated such that all the sub solutions can form an overall solution. Therefore it is important to address coordination between the agents to find an overall solution that is feasible and maybe even optimal for the whole network.

One of the aspects of a non-centralised railway traffic management system is the *decomposition*: how the railway system is divided into subsystems. Within the field of non-centralised approaches only a few publications have focused on the decomposition approaches, e.g. Lamorgese and Mannino (2015) [5] and Kersbergen et al. (2016) [6]. The railway system consists of two main components, the infrastructure and the train traffic, and within the few papers on decomposition approaches these two are not equally taken into account. As such, only a part of the railway system is taken into account in the decomposition. Moreover, no research could be found that studies what the individual impact of a decomposition approach can be on the performance of the conflict resolution. For example, Luan et al. (2020) [7] studies the presented decomposition approaches in combination with solution approaches, but does not present the impact of solely the decomposition approaches.

Non-centralised approaches have not only been regarded in literature but also in practice. In the current state of practice, the network is divided into regions to form three levels. The national control center views the entire Dutch network, 12 traffic control centers are responsible for their own region and those regions are subdivided into 4 to 12 subregions, or train dispatcher posts. The state of practice of how a region is subdivided into subregions is essentially also a decomposition. This state of practice decomposition has come about historically and organically and has not yet been scientifically founded.

Following from these challenges, the objective of this research is to develop a decomposition approach in non-centralised railway traffic management that incorporates both aspects of the railway system, the infrastructure and the train traffic. Additionally this research aims to provide insight into what the impact of a decomposition approach can be on the non-centralised conflict resolution. This is done by developing a decomposition approach that divides a region into subregions based on hourly traffic density. A non-centralised railway traffic management approach is also developed, by establishing the conflict resolution algorithm and coordination scheme. The proposed decomposition approach is applied to a case study region. The proposed decomposition is then compared with the baseline decomposition used in the current state of practice, by applying the same non-centralised railway traffic management approach to both decompositions. Using a series of delay scenarios, it is analysed what impact the proposed decomposition has on the effectiveness and efficiency of the conflict resolution compared to the baseline.

The paper is structured as follows. Section II discusses the literature related to the current research. Section III presents the proposed decomposition approach. Section IV presents the other two elements of a non-centralised conflict resolution approach: the conflict resolution algorithm and coordination scheme. Section V describes the result obtained from a case study on the region between Roosendaal and Liempde. Lastly, section VI concludes the paper with the conclusions and recommendations.

II. RELATED WORKS

The research field of TMS has received much interest over the years. The publications related to the current research are discussed in this section.

The initial publications on TMS mainly considered a centralised approach. In this research the modelling of the railway traffic management problem plays a large role. D'Ariano et al. present the Alternative Graph method in D'Ariano et al. (2007) [8], which formulates the conflict resolution problem using graph theory. This approach has been used in several other publications, such as D'Ariano et al. (2008) [9] and Corman et al. (2011) [10]. A Mixed Integer Linear Programming (MILP) formulation has been used to describe the conflict resolution problem in Luan et al. (2018a) [4] and Pellegrini et al. (2015) [2]. However the conflict resolution problem can become cumbersome when viewing large or complex situations, resulting in long computation times. To mitigate that, Pellegrini et al. (2015) [2] presents the RECIFE-MILP algorithm to solve the conflict resolution problem, but it remains that centralised approaches often lead to very large problem formulations.

That is why the research has been looking into non-centralised approaches. In Corman et al. (2010) [11], along with other publications ([5, 7, 12–16]), it is shown that using a non-centralised approach in comparison to a centralised approach can lead to reduction in computation time and an improvement in delay reduction. Kersbergen et al. study further implications of using different forms of non-centralisation in [6, 17]. Within the non-centralised approaches three main aspects are discerned: the conflict resolution algorithm, the coordination scheme and the decomposition. Roughly the same conflict resolution algorithms are used as in centralised approaches: the Alternative Graph ([5, 11, 12]) and MILP formulation ([7, 14–16, 18–20]) are common. There has been a variety of approaches for coordination. In [13] Corman et al. present a coordinator graph that formulates the coordination problem using graph theory. Kersbergen et al. use a central cost function in [6, 17]. Another approach is using boundary constraints at the boundaries of the subsystems. A common boundary constraint is the time transition constraint, which ensures continuity at subsystem borders, used in Luan et al. (2018b, 2020) [7, 18] and Toletti et al. (2020) [21].

Paper	Type	Criteria	Result
[5]	MM	Problem scale	Station & Line agents
[6]	MAT	Problem size, Connectivity	Constraints
[19]	GEO	Connectivity, Centrality	Subregions
[7]	GEO	Region size, Connectivity	Subregions
	TRA	Singular train	Trains
	TIN	Occurence estimate	Time intervals
[15]	MM	Problem type	Subproblems

MM: Macro-Micro; MAT: Mathematical; GEO: Geographical; TRA: Train based; TIN: Time interval

Table 1. Decomposition in non-centralised railway traffic management approaches

The last aspect of a non-centralised approach, the decomposition, has been the focus of only a few papers, summarised in Table 1. The decomposition type describes the general approach of the decomposition. The criteria of the decomposition dictate why an element will be assigned to one subsystem or another and the result of the decomposition is what a subsystem contains. For example, Lamorgese and Mannino (2015) [5] applies a Macro-Micro decomposition to a train network. As a criterion the problem scale is used, meaning that if the problem is microscopic or macroscopic dictates to which subproblem it is assigned. The result is an agent for every station and every line between stations, in which the line agents solve the macroscopic problem while the station agents solve the microscopic problem at the stations. In Keita et al. (2020) [15] the problem is decomposed to forms three subproblems, elaborating on the Benders cut. First, the routing problem is solved. Secondly, the optimal rescheduling actions are found. Lastly, the delays that correspond to the decided actions are computed. Kersbergen et al. [6] mathematically decompose the system by making subsystems consisting of constraints. The criteria that they use are: the difference in number of constraints between subproblems and the amount of constraints not in the current subproblem connected to a constraint in the current subproblem, or in other words problem size and connectivity.

Connectivity is used in another sense in the geographical decomposition presented in the research by Chen et al. in [19]. They define nodes within the railway network at every lineside signal. These nodes are then given a degree of connectivity and centrality in the form of two factors: how many trains travel over that node and how many block sections are connected to that node. In the decomposition it is then considered that a region contains only two important nodes, which then gives the nodes an assignment to a certain subregion. The paper in this selection that pays the most attention to the decomposition is Luan et al. (2020) [7]. In this paper three types of decomposition are presented and compared. Firstly they present a geographical decomposition that divides the railway network into subregions considering two criteria: the deviation of the size of a subregion, in terms of amount of block sections in the subregion compared to the mean subregion size and the amount of train services crossing a subregion border. Secondly they present a train based decomposition where every train is viewed as a subsystem. Thirdly, they show a time interval based decomposition, where the rescheduling problem is solved for certain time intervals. Whether an occurrence belongs to a time interval is determined by the estimated occurrence time.

These publications only consider a part of the railway system in their decomposition and do not elaborate on the impact of a decomposition on the conflict resolution. The contribution of this research is to:

- Develop an hourly traffic density based decomposition approach to evenly distribute the decisional complexity
- Give insight into the impact of the decomposition approach on the effectiveness and efficiency of the non-centralised conflict resolution

III. AN HOURLY TRAFFIC DENSITY BASED DECOMPOSITION APPROACH

This research develops a decomposition that results in subregions. One main reason for this choice is that this also how the railway traffic management problem is decomposed in the current state of practice. Having the same end result makes the research comparable to the state of practice and gained insights can more quickly and easily be applied. Another main reason is that this leads to a static decomposition, in the sense that once the problem has been decomposed it will not change [7]. The decomposition approach in the current research takes the form of an optimisation model that assigns parts of the infrastructure to certain subregions, drawing inspiration from the geographical decomposition presented in Luan et al. (2020) [7].

A. Hourly traffic density as a criterion

The aim of developing a decomposition approach is to equally divide decisional complexity among the agents, to improve computation, performance and flexibility. It is on the basis of this aim that the main criterion for the decomposition is defined as the *hourly traffic density* in trains per track per hour. This combines both the infrastructure and the train traffic travelling over it. This describes the decisional complexity in the sense that it shows how likely conflicts are to occur. A high ratio of trains per track per hour means that these trains will not have large headways between them and that a conflict is likely.

Another way of looking at it, is by describing it as the ratio between supply and demand. It describes to what extent the supply of infrastructure can satisfy the demand of the train operating companies. From this view, a high hourly traffic density means that the infrastructure is not fully able to satisfy the demand of the traffic flow. It is calculated by dividing the traffic flow in the area by the amount of tracks the area has. In other words it is the trains travelling through the area every hour divided by the number of tracks. Trains that stop in the area, because it is for example a station, are taken into account with a factor 2, because they require additional scheduling actions and therefore form a complexer part of the railway traffic management problem.

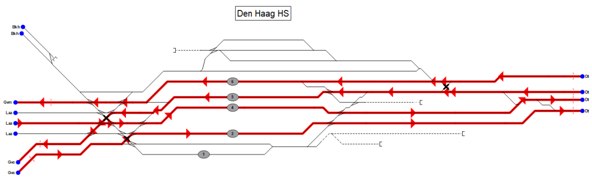


Figure 1. The traffic flow and infrastructure at Den Haag Holland Spoor [22]

How the hourly traffic density is determined is illustrated by means of an example for Den Haag Holland Spoor station. Figure 1 demonstrates the infrastructure at the railway station. The amount of parallel tracks varies throughout the timetabling point: The most left counts 7 tracks, the most right 4 and the middle 10. Understandably, the hourly traffic density will vary accordingly. The choice is made to use the constraining number of tracks, because this is the bottleneck of the timetabling point. In the case of the example this means the 4 tracks on the most right side are used. The traffic in the Dutch network has an hourly cyclic schedule. The station has no additional trains during rush hour and no additional freight trains outside of rush hours. Therefore a regular hour can be used to determine the traffic flow. A total of 28 trains travel through the timetabling point every hour. Of these trains all intercity's and sprinters, 26, make a stop at the station and the 2 freight trains pass through without stopping. The factor 2 for stopping trains is taken into account by multiplying the trains that only pass through with 0.5. This means the hourly traffic, or trains per hour, comes to 27. Dividing this by the 4 constraining tracks gives a final hourly traffic density of 6.75 .

B. Modelling details

The output of the model is that infrastructural elements are assigned to subregions. Contrary to previously presented microscopic decomposition models, this approach will take a macroscopic approach. Theoretically speaking the subregions could be composed of elements any size: block sections, junctions, stations or stations with a corresponding link. As infrastructural elements, this model defines 'dienstregelpunten' (*timetabling points*). These are points in the infrastructure that are recognised to need scheduling actions by network administrator ProRail. So generally the timetabling points are the extraordinary parts of the infrastructure such as stations or large junctions. As these are also the points of interest for railway traffic management, they form a suitable definition of the infrastructural elements in the decomposition.

Besides the hourly traffic density, the decomposition model has several other inputs. The train services within the region are introduced into the model by stating whether a train service crosses an individual timetabling point. Because the timetabling points are ordered in the input, the aggregation of this information leads to the route of the train service.
Another input, is the frequency of those train services. This gives the model insight into which routes are used most and might have a higher importance in the decomposition. This frequency is defined as number of trains per hour.
Lastly, the decomposition is inputted how many subregions the region should be divided into.

C. Model formulation of an hourly traffic density based decomposition approach

The aim of the decomposition model is to assign timetabling points to subregions. This assignment is captured in decision variable μ_i, which has the value of the number of the subregion that timetabling point i is assigned to. This assignment is based on the minimisation of the objective function, which consists of the weighted sum of two criteria: the size variance between the subregions in terms of the hourly traffic density and the number of trains crossing subregion borders per hour. This minimising is enabled by choosing the values of two auxiliary decision variables: γ_{fj} and α_{ri}. Where γ_{fj} represent whether train service f crosses a subregion border at point j and α_{ri} describes whether timetabling point i is assigned to subregion r. The model is constrained by a series of constraints, which serve one of two functions: either defining a relation between the decision variables γ_{fj}, α_{ri} and μ_i, or constraining the assignment problem.

The mathematical formulation can be summarised as below. In this formulation the notations $|x|$ and $||X||$ are used. With $|x|$ it is indicated that the absolute value of x is taken. The notation $||X||$ indicates the cardinality of set X, or in other words the size of set X.

Indices

- f train service f in train service set F
- i timetabling point i in infrastructure set E
- j timetabling point j in train service route set E_f of train f
- r subregion r in subregion set R

Sets

- F set of train services $\{trainservice_1, ...\}$
- E set of timetabling points $\{e_1, ..., i\}$
- R set of subregions $\{1, ..., r\}$

Parameters

- B_{fji} the route matrix for every train service f indicating which timetabling points the train service crosses in its route, with $B_{fji} = 0$ if the timetabling point i is not used in the j^{th} trip in the route of train service f; $B_{fji} = 1$ if train service f leaves timetabling point i in the j^{th} trip in the route and $B_{fji} = -1$ if train service f enters timetabling point i in the $(j+1)^{th}$ trip in the route $B_{fji} \in \{-1, 0, 1\}$, [-]
- E_{fj} the timetabling point i used as the j^{th} timetabling point in the route of train service f $E_{fj} \in E$, [-]
- t_i the hourly traffic density of timetabling point i $t_i \in \mathbb{R}^+$, [trains/track/hour]
- ζ the weight factor in the objective function $\zeta \in (0,1) \subset \mathbb{R}$
- ρ_f the frequency of train service f $\rho_f \in \mathbb{R}^+$, [trains/hour]

Decision variables

- α_{ri} the assignment variable of timetabling point i to subregion r: $\alpha_{ri} = 1$ if timetabling point i belongs to subregion r, else $\alpha_{ri} = 0$
- γ_{fj} if two consecutive timetabling points j and $j+1$ in the route of train service f belong to different subregions: $\gamma_{fj} = 1$ if E_{fj} and $E_{f(j+1)}$ belong to a different region, else $\gamma_{fj} = 0$
- μ_i the region from set R that timetabling point i belongs to

Objective function

$$min\ [\zeta \cdot \sum_{f \in F} \left(\sum_{j \in \{1, ||E_f||-1\}} (\gamma_{fj}) \cdot \rho_f \right) + (1-\zeta) \cdot \sum_{r \in R} \left| \sum_{i \in E} (\alpha_{ri} \cdot t_i) - \frac{\sum_{i \in E} t_i}{||R||} \right|] \tag{1}$$

Subject to:

$$\frac{\left|\sum_{i \in E} (B_{fji} \cdot \mu_i)\right|}{||R|| - 1} \leq \gamma_{fj} \quad \forall j \in \{1, ||E_f||-1\} \forall f \in F \tag{2}$$

$$\alpha_{ri} \leq 1 - \frac{|\mu_i - r|}{||R||} \qquad \forall r \in R, i \in E \tag{3}$$

$$\sum_{i \in E} \alpha_{ri} \geq 1 \qquad \forall r \in R \tag{4}$$

$$\sum_{r \in R} \alpha_{ri} = 1 \qquad \forall i \in E \tag{5}$$

The objective function, Equation 1, consists of two elements. The first element is to minimise the amount of subregions train service f crosses when travelling over all trips j in its route. The second element is to minimise the difference between the hourly traffic density over all timetabling points i in subregion r and the average hourly traffic density per subregion. The objective is to minimise the weighted sum of both elements.
Equation 2 defines the value of binary variable γ_{fj} for all train services f and all timetabling points j in the route of train service f. If the timetabling point j and $j+1$ in the route of train service f are in the same subregion the term $\left|\sum_{i \in E} (B_{fji} \cdot \mu_i)\right|$ will be zero, because the positive and negative values of the multiplication cancel each other out in the summation. Therefore the entire fraction will be zero and thus so will γ_{fj}. In contrast when timetabling point j and $j+1$ belong to a different subregion, the term will become a positive, nonzero value which will give the fraction a value higher than 0 but less than or equal to 1. Therefore γ_{fj} will be 1.
Equation 3 defines the value of binary variable α_{ri} for all subregions r and all timetabling points i. If a specific timetabling point i belongs to a specific subregion r then $\mu_i = r$, as a consequence $|\mu_i - r| = 0$.

Therefore the fraction will also be equal to zero and thus α_{ri} will be equal to 1. Contrary, if a specific timetabling point i does belong to a specific subregion r, the fraction will have a positive, nonzero value not larger than 1. This will ensure α_{ri} equals zero.
Equation 4 states that every subregion r should contain at least one timetabling point. In other words, this ensures that there cannot be empty subregions.
The last constraint, Equation 5, dictates that the sum of α_{ri} over all subregion should be one. This entails that all timetabling points are assigned to exactly one subregion, so that no timetabling point is left unassigned.

IV. A NON-CENTRALISED RAILWAY TRAFFIC MANAGEMENT MODEL

All aspects of the non-centralised conflict resolution model refer to the agents of which it is built up. The decomposition determines which parts of the network should be overseen by an agent. The coordination will ensure communication between these agents and the conflict resolution is what is happening inside the agent. These two latter aspects make up the non-centralised railway traffic management model.

A. A railway traffic management approach for conflict resolution

This research chooses to use the Alternative Graph formulation as presented in D'Ariano et al. (2007) [8] as a conflict resolution approach. It is a fast and efficient model, which are characteristics that are beneficial for the real-time implementation envisioned for Traffic Management Systems. Additionally, the model has a compact formulation. This means that even when non-centralised railway traffic management approaches view relatively large networks, the model will remain manageable.

The Alternative Graph method models the conflict resolution using graph theory. The nodes represent operations of a certain train at a certain time, the arcs represent the succession of events and the arc weight represents the minimal time between events. Besides fixed arcs the Alternative Graph model also contains alternative arcs. These arcs show the precedence between two trains and an example is demonstrated in Figure 2. The arcs come in alternative arc pairs and in the solution only one of those arcs can be chosen. By changing the precedence of trains and/or the time of certain operations, the Alternative Graph finds a solution to the conflict problem. This solution will be chosen based on how the maximum secondary delay can be minimised. More specifically, the goal of Alternative Graph model is to: *choose a selection S of alternative arcs that has no positive length cycle and minimises the length of the longest path from 0 to n,* $l^s(0,n)$ [1].

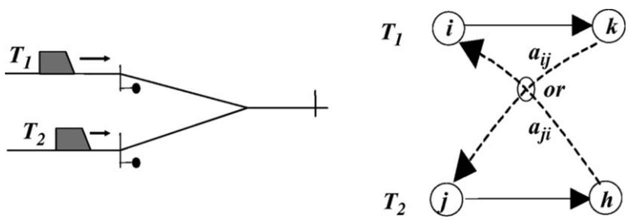


Figure 2. Fragment of an Alternative Graph model at a converging junction with two trains, as in [8]

In this paper it is chosen to reformulate the Alternative Graph model as presented in D'Ariano et al. (2007) [8] to more accurately represent the MILP problem as which it is implemented. The mathematical content remains identical, but Equation 8 is reformulated and Equation 9 is added.

The main objective is to minimise the maximum consecutive delay, which is represented in the equation $t_n - t_0$, meaning the time stamp of sink node n minus the time stamp of source node 0., see Equation 6. This objective can be reached by influencing the two decision variables t_i and z_{ij}. The decision variable t_i signifies the starting time of the operation at node i, e.g. entering a block section, a departure or an arrival. The decision variable z_{ij} is used to show which arc out of the alternative arc pair is selected. The indices i and j indicate that the variable refers to alternative arc (i, j) from node i to node j , which is part of alternative arc pair $((i, j), (h, k))$ in the set of alternative arc pairs D. The decision variable z_{ij} influences the precedence of trains and by choosing that precedence has an influence on the variable t_i, such that this can then influence the value of the sink node n.

The mathematical formulation of the Alternative Graph model in this research is notated as follows:

Index

i,... for node i in node set N

Sets

N set of nodes $\{0,i,...,n\}$
A set of all arcs $\{(i, j),...\}$
F set of fixed arcs $\{(i, j),...\}$
D set of alternative arc pairs $\{((i, j), (h, k)),...\}$

Parameter

f_{ij} the arc weight of arc (i, j) from node i to node j $f_{ij} \in \mathbb{R}^+ \; \forall (i, j) \in A$, [s]

Decision variables

t_i the starting time of the operation at node i [s]
z_{ij} the decision variable if alternative arc (i, j) is selected: $z_{ij} = 1$ if alternative arc is used, if not $z_{ij} = 0$ [-]

Objective Function

$$min[t_n - t_0] \quad (6)$$

Subject to:

$$t_i - t_j \geq f_{ij} \quad \forall (i, j) \in F \quad (7)$$

$$z_{ij} \cdot (t_j - t_i) + z_{hk} \cdot (t_k - t_h) \leq z_{ij} \cdot f_{ij} + z_{hk} \cdot f_{hk} \quad \forall ((i, j), (h, k)) \in D \quad (8)$$

$$\sum_{(l,m)\in\{(i,j),(h,k)\}} z_{lm} = 1 \quad \forall ((i, j), (h, k)) \in D \quad (9)$$

The objective of the model is to minimise the maximum secondary delay as in Equation 6. This value is represented in $t_n - t_0$, where t_0 is the overall starting time which is often equal to zero. Therefore the objective function can often be reduced to $min\, t_n$.
Equation 7 constrains the minimal technical time f_{ij} that should be between two consecutive actions connected with a fixed arc. For example, it can represent the technical running time: how fast a train can technically get from one point to another without taking into account buffer times.
Equation 8 has a similar goal as Equation 7. In this case, the constraint concerns the alternative arcs such that the technical time represents the minimal headway between two trains. An extra consideration for the alternative arcs is that they come in pairs. Multiplying with the respective variable z_{ij} or z_{hk}, ensures that the only non-zero terms in the equation enforce the minimal technical running time only for the used alternative arc.
Equation 9 dictates that exactly one of the alternative arcs in an alternative arc pair is used. This is done by saying that the sum over the variable z_{lm} for both the alternative arcs in the pair should be equal to one, such that exactly one value of z_{lm} can be equal to one.

B. A coordination approach for non-centralised railway traffic management

In this research it is chosen to use boundary constraints as a coordination method. This approach is used and mentioned in several publications, [7, 18, 21]. These are coupling constraints that couple the subregions by constraining a certain characteristic of trains crossing a subregion border, for example the time of crossing, should be the same in both subregions to ensure continuity. This research proposes an alternative to the time transition constraint: a delay transition constraint. This entails that the delay of a train is coordinated between agents to be the same at both ends of a subregion border.

$$d_{final,tr} = d_{entrance,tq} \quad \forall t \in T_{rq}, r \in R, q \in Q_r \qquad (10)$$

The delay transition constrain in Equation 10 dictates that the delay with which train t leaves subregion r, i.e. $d_{final,tr}$, is equal to the delay with which that same train t will enter the neighbouring subregion q, i.e. $d_{entrance,tq}$. Using a boundary constraint in this formulation, opens up the possibility of communicating only in the case of delay.

Communicating only in the case of a delay, can be categorised as event-driven communication, in this case the event is a delay. To enforce it a coordinator is added as an entity. The coordinator is placed in a higher layer than the conflict resolution agents, which allows it to oversee the global conflict resolution task. It has three main tasks:

1. Communicating initial delay to the relevant agent(s)
2. Assigning the conflict resolution task to the agents, by overseeing in which order the subregions are to be solved and if they are to be solved at all.
3. Prompting communication between agents, or not. An agent will only need to if a train leaving its subregion is delay and then only to the relevant agent. This determined by the coordinator.

As is reflected in these tasks, the coordinator has continuous communication with all agents. The coordinator gives the agent a solving or communication task and the agent returns the result of the conflict resolution of the subregion to the coordinator. Shown schematically, the system as whole can be drawn as in Figure 3. Here $d_{in,r}$ signifies the initial delay for subregion r and d_{tr} stands for the delay of all trains t leaving subregion r.

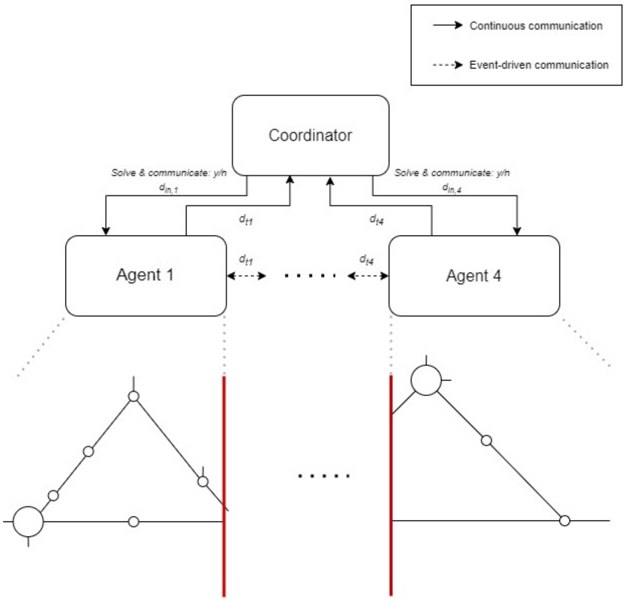


Figure 3. Schematic representation of the coordination scheme

Combining the Alternative Graph conflict resolution approach and the delay driven coordination scheme gives a non-centralised railway traffic management approach. This approach can be used to test the impact of the decomposition on the conflict resolution.

V. CASE STUDY AND RESULTS

The hourly traffic density based decomposition is validated by applying it to a case study within the Dutch network. The proposed decomposition is compared to the baseline decomposition, the train dispatcher subregions currently used at ProRail. This comparison is made both on the level of the decompositions as the impact they have on the conflict resolution problem.

A. Region Roosendaal - Liempde: A case study

The test case is chosen as the network between Roosendaal and Lage Zwaluwe, and Den Bosch and Liempde, see Figure 4, which is part of the IJssellijn corridor and freight route, 'de Brabant route'. It is chosen to let the outline of the test case region to be defined by the current train dispatcher subregions. This test case contains some interesting aspects. First of all, at this point in the Dutch network many different routes come together, which makes it interesting for the decomposition approach. Additionally, the region has heterogeneous traffic in the sense that besides the intercity, sprinter and high speed passenger trains also freight travels travel through the regions. The varying characteristics of the traffic result in interesting factors in the conflict resolution and the decomposition.

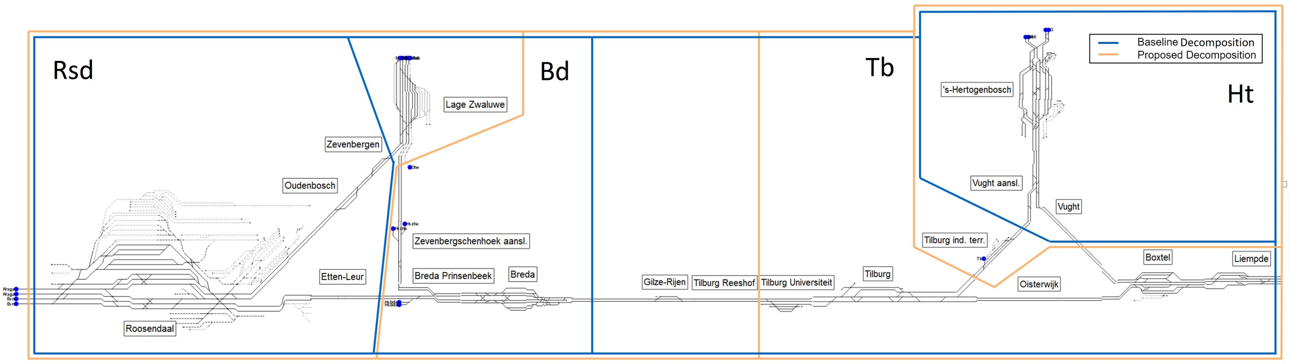


Figure 4. The two decompositions, edited from [22]
Rsd: Roosendaal; Bd: Breda; Tb: Tilburg; Ht: Den Bosch

Moreover the triangular formation of the infrastructure means that it contains many different Points of Interest for conflict, such as converging tracks and crossovers.

Part of the test case is also the timetable and delay scenarios used. For the timetable a 15 minute interval from 20:25 to 20:45 is used of a basic timetable for Tuesday in 2022 by ProRail [22]. The Dutch timetable consists of an hourly repeating pattern. So a consciously chosen quarter of an hour can give insights valuable for the whole day. The selection for the schedule is based on the fact that it is as susceptible to conflict as possible, or in other words when the traffic flow is as busy and complex as possible. As a result, the conflict resolution problem is the most complex and it is expected to increase the visibility of differences between the effects of both decompositions.

One delayed train		**Two delayed trains**	
Entrypoint	Delay [s]	Entrypoints	Delay [s]
Rsd	203	Rsd, Zlw	733
Zlw	530	Zlw, Zha	705
Zha	175	Rsd, Zha	378
Ht	165	Zha, Ht	340
Lpe	303	Zha, Lpe	478
		Ht, Lpe	468

Three delayed trains	
Entrypoints	Delay [s]
Rsd, Zha, Ht	543
Rsd, Zha, Lpe	681
Zlw, Zha, Ht	870
Zlw, Zha, Lpe	1008
Zha, Ht, Lpe	643

Table 2. The delay scenarios categorised by the amount intially delayed trains with the entry point(s) of the delayed train(s) and the total intial delay
Rsd: Roosendaal; Zlw: Lage Zwaluwe; Zha: Zevenbergschenhoek aansluiting; Ht: Den Bosch; Lpe: Liempde

To create delay scenarios, five different initial delay types are chosen. Each delay type is the delaying of one train upon entering the region at one of the five entry points, that is also early on in the 15 minute time frame. This way the delay has time to propagate but the algorithm also has time to resolve it. The five entrypoints of the region are: Roosendaal, Lage Zwaluwe, Zevenbergschenhoek aansluiting, Den Bosch and Liempde.

The size of the delay is set to the planned headway the train has to the train in front of it. This way two trains are planning to be at the same place at the same time, unavoidably leading to a conflict. In total, 17 delay scenarios are used, where each delay scenario consists of one or multiple delay types. The delay scenarios are summarised in Table 2. An abbreviation refers to the entry point where a train is given an initial delay. Besides these delay scenarios, the scenario where there is no delay is also studied. This is when all trains travel according to schedule.

With the timetable and delay scenarios, the train traffic can be simulated. This gives the tools needed to test the impact of the decomposition on the conflict resolution

B. Decomposing the region Roosendaal-Liempde based on hourly traffic density

Using these inputs, a decomposition was found by the decomposition approach presented in section III. In Figure 4, the orange line shows the border of the subregions in the proposed decomposition and the blue lines show the border of the baseline decomposition. The baseline decomposition refers to how the train dispatcher subregions are currently defined. The subregions are referred to according to the intercity station it contains or the abbreviation of that station. This means the four subregions are: Roosendaal or Rsd; Breda or Bd; Tilburg or Tb and Den Bosch or Ht. This name refers to the subregion in the respective decomposition.

Looking only at the proposed decomposition a few observations can be made. All subregions contain 4 to 5 timetabling points. Also, all subregions contain exactly one intercity station. One large difference between the decompositions can be seen in the size of the Tilburg subregion. In the baseline this subregion is quite large, reaching from Gilze Rijen to Tilburg Industrie and Liempde and containing 8 timetabling points. Using the hourly traffic density, Tilburg is identified as the timetabling point with the highest value, being two times larger than the average value. As a result the decomposition model chooses to include less timetabling points in the subregion containing Tilburg, to come to a more even distribution of hourly traffic density over the subregions. Another difference in subregion size can be seen in the Roosendaal subregion, where the proposed decomposition includes Lage Zwaluwe contrary to the baseline decomposition. This is again a result of the hourly traffic density criterion.

Roosendaal and the surrounding timetabling points, Etten-Leur, Oudenbosch and Zevenbergen, have a relatively small hourly traffic density. So to evenly distribute the hourly traffic density over the subregions, the proposed decomposition approach chooses to also include Lage Zwaluwe in that subregion. Overall, at first glance, the subregions seem to be more evenly distributed in terms of hourly traffic density. To further study this, the subregions are compared quantitatively in the form of various graphs.

C. The impact of a decomposition on the non-centralised conflict resolution problem

Both decompositions are subjected to the railway traffic management approach presented in section IV. To gain insight into the impact of the different decompositions on the non-centralised conflict resolution problem, several KPIs are compared in the graphs below.

The computation time of the conflict resolution problem is used to measure the distribution of decisional complexity over the subregions. How long the computer takes for calculation is related to how complex the problem is to resolve. Therefore faster computation times suggest a more equal distribution of decisional complexity. The comparison in computation time is made in terms of how many trains are initially delayed. In the previous section, it was suspected that the size of the conflict resolution problem is more evenly distributed across the subregions when using the proposed decomposition. As a result it is expected that the decisional complexity is also more equally distributed among the subregions using the proposed decomposition, which would result in a lower computation time.
In Figure 5, the relative computation time is plotted against the number of initially delayed trains. The relative computation time entails the computation time of the conflict resolution using the proposed decomposition divided by the computation time of the conflict resolution using the baseline decomposition. Each of the dots can be translated back to one of the delay scenarios shown in Table 2. Between these dots a trendline is fitted, which can visualise the general trend in the data. In the absolute data, it is observed that using the non-centralised railway traffic management approach leads to a computation time lower than 1 second using either decomposition.

This can be considered fast computation and to be suitable for real-time implementation as a digital aid for train dispatchers. Secondly, in the graph it can be observed that the computation time of the conflict resolution using the proposed decomposition is generally lower than when using the baseline decomposition. Moreover, when more trains are delayed, so the conflict resolution problem becomes more complex, all relative computation times are lower than 100 %. The trendline indicates that the relative computation time decreases to *an average computation time reduction of 20 %*. Meaning that the proposed decomposition allows significantly faster computation of the conflict resolution, especially for more complex formulations. This behaviour can be especially beneficial in complexer problem formulations, such as larger networks, more complex delay situations or additionally incorporated local constraints.

A delay performance indicator that is studied, is the relative total output delay of the conflict resolution using the proposed decomposition compared to the baseline decomposition. The output delay is the delay a train has when it leaves the simulated traffic, which can be for two reasons: the train leaves the region or the end of the simulated 15 minute time period is reached. The total output delay is the output delay summed over all trains. The relative total output delay is defined as the total output delay of the conflict resolution using the proposed decomposition divided by the total output delay using the baseline decomposition. Making this comparison gives insight in the impact of the decomposition on the absorption of delay. Resulting from the formulation of the relative output delay, a value lower than 100 % means that there is less output delay when using the proposed decomposition.
Figure 6 plots the relative total output delay against the number of initially delayed trains. The measurements of the delay scenarios are represented by the dots and the size of the dot signifies how many delay scenarios return that value. The line visually represents the trend in the data. From this plot two conclusions can be drawn. Firstly, all values of the delay reduction are equal to or smaller than 100 %, which means when using the proposed decomposition the conflict resolution solution resulted in the same if not less output delay than when using the baseline decomposition. The proposed decomposition allows the conflict resolution to reduce the output delay up to 19 %, compared to when using the baseline decomposition.

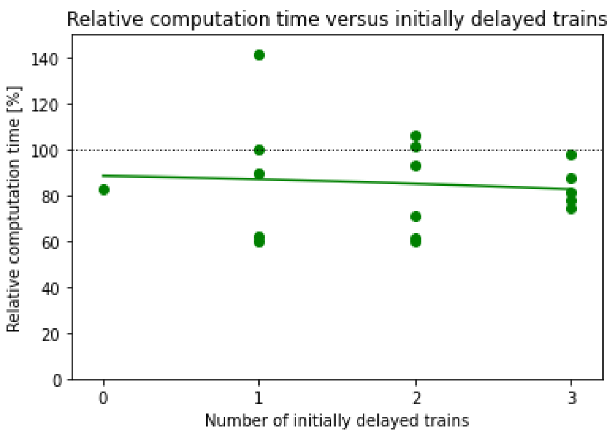


Figure 5. The relative computation time of the conflict resolution problem versus the number of initially delayed trains

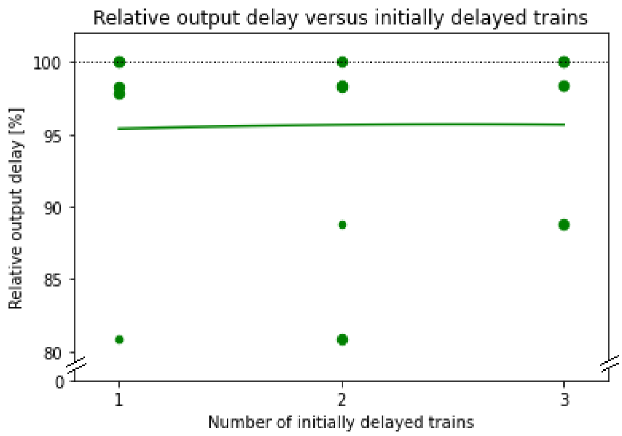


Figure 6. Relative total output delay of the conflict resolution versus the number of initially delayed trains

The last measurement is the number of delayed trains, i.e. the number of trains that have an output delay larger than zero. Subtracting the number of delayed trains in the conflict resolution using the proposed decomposition from those using the baseline decomposition, gives the difference in number of delayed trains. A negative value for difference in number of delayed trains means that the initial delay has had a lasting impact on less trains. Also this measurement is insightful to study over how many trains the total delay is distributed. In Figure 7 the difference in number of delayed trains is against how many trains are delayed in the initial delay scenario. Every dot signifies a measurement from a specific delay scenario and the size of the dot signifies for how many delay scenarios that value occurs. In the plot it can be seen that the proposed decomposition leads to the conflict resolution delaying consistently less trains than the baseline decomposition, decreasing the number of delayed trains with on average 17 %. Moreover this trend seems to strengthen as the complexity of the delay scenario increases. Meaning that for increasingly complex initial delay scenarios the proposed decomposition allows for a decreasing amount of delayed trains. This is especially observed going from the delay scenarios with one delayed train to the delayed scenarios with two delayed trains.

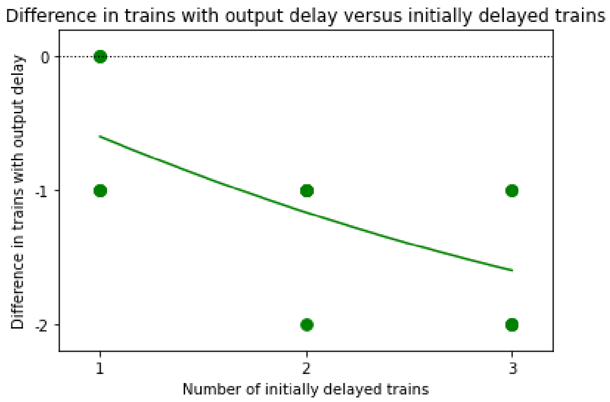


Figure 7. Difference in number of trains with an output delay versus the number of initially delayed trains

Overall, several conclusions can be drawn. The trend in the computation time shows that the proposed decomposition allows up to *20 % faster computation* of the conflict resolution problem. This suggests that the proposed decomposition is able to more equally distribute the decisional complexity of the conflict resolution problem over the subregions, than the baseline decomposition. Thus using the size variance in hourly traffic density as a criterion in decomposition forms a way to reduce the overall decisional complexity of the conflict resolution problem. The delay reduction shows that the proposed decomposition allows the conflict resolution to mitigate more delay than when using the baseline decomposition, leading to *on average 5% less total output delay*. Additionally, the proposed decomposition is able to allow the conflict resolution to *reduce the number of delayed trains with 17 %*. Combining these statements shows that using the proposed decomposition allows the conflict resolution to absorb delays better, especially for the smaller delays.

VI. CONCLUSIONS AND RECOMMENDATIONS

In this research a traffic flow complexity based decomposition approach was developed to decompose the railway system into subregion with similar decisional complexity. Additionally, a non-centralised railway traffic management approach was designed using Alternative Graph conflict resolution and a delay driven communication scheme with a coordinator. The railway traffic management approach was applied to a case study using the proposed traffic complexity based decomposition approach. This was compared to the same railway traffic management approach but then using the baseline decomposition now used in practice by ProRail. This has shown that the traffic flow complexity based decomposition can reduce the computation time by up to 20 % and reduce the total delay with on average 5 %. From this it can be concluded that the proposed decomposition can distribute decisional complexity evenly across subregions and thereby reduce the decisional complexity and workload of a single subregion, providing more effective and time-efficient rescheduling solutions. This shows great promise for further research and the state of practice.

For further research there are three main recommendations. Firstly, it would be interesting in future research to include the amount of subregions the region is to be divided into in the optimisation, as to determine the optimal number of subregions for that specific region. Secondly, it is recommended that future research looks into expansions and variations of decomposition criteria to explore different ways that the complexity of the conflict resolution problem can be distributed over subregions. Lastly, it is advised that in future research the decomposition is tested with more advanced forms of the non-centralised railway traffic management approach.

References

[1] A. D'Ariano, F. Corman and D. Pacciarelli, "Rescheduling," in *Railway Timetabling & Operations*, I. Hansen and J. Pachl, Eds., 2nd ed., 2014.

[2] P. Pellegrini, G. Marlière, R. Pesenti and J. Rodriguez, "RECIFE-MILP: An Effective MILP-Based Heuristic for the Real-Time Railway Traffic Management Problem," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2609–2619, Oct. 2015.

[3] R. M. Lusby, J. Larsen, M. Ehrgott and D. M. Ryan, "A set packing inspired method for real-time junction train routing," *Computers and Operations Research*, vol. 40, no. 3, pp. 713–724, Mar. 2013.

[4] X. Luan, Y. Wang, B. De Schutter, L. Meng, G. Lodewijks and F. Corman, "Integration of real-time traffic management and train control for rail networks - Part 1: Optimization problems and solution approaches," *Transportation Research Part B: Methodological*, vol. 115, pp. 41–71, Sep. 2018.

[5] L. Lamorgese and C. Mannino, "An exact decomposition approach for the real-time train dispatching problem," *Operations Research*, vol. 63, no. 1, pp. 48–64, Jan. 2015.

[6] B. Kersbergen, T. van den Boom and B. De Schutter, "Distributed model predictive control for railway traffic management," *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 462–489, Jul. 2016.

[7] X. Luan, B. De Schutter, L. Meng and F. Corman, "Decomposition and distributed optimization of real-time traffic management for large-scale railway networks," *Transportation Research Part B: Methodological*, vol. 141, pp. 72–97, Nov. 2020.

[8] A. D'Ariano, M. Pranzo and I. A. Hansen, "Conflict resolution and train speed coordination for solving real-time timetable perturbations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 208–222, Jun. 2007.

[9] A. D'Ariano, F. Corman, D. Pacciarelli and M. Pranzo, "Reordering and local rerouting strategies to manage train traffic in real time," *Transportation Science*, vol. 42, no. 4, pp. 405–419, 2008.

[10] F. Corman, A. D'Ariano, I. A. Hansen and D. Pacciarelli, "Optimal multi-class rescheduling of railway traffic," *Journal of Rail Transport Planning and Management*, vol. 1, no. 1, pp. 14–24, 2011.

[11] F. Corman, A. D'Ariano, D. Pacciarelli and M. Pranzo, "Centralized versus distributed systems to reschedule trains in two dispatching areas," *Public Transport*, vol. 2, no. 3, pp. 219–247, Aug. 2010.

[12] F. Corman, A. D'Ariano, D. Pacciarelli and M. Pranzo, "Optimal inter-area coordination of train rescheduling decisions," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 1, pp. 71–88, 2012.

[13] F. Corman, A. D'Ariano, D. Pacciarelli and M. Pranzo, "Dispatching and coordination in multi-area railway traffic management," *Computers and Operations Research*, vol. 44, pp. 146–160, Apr. 2014.

[14] G. Cavone, T. Van Den Boom, L. Blenkers, M. Dotoli, C. Seatzu and B. De Schutter, "An MPC-Based Rescheduling Algorithm for Disruptions and Disturbances in Large-Scale Railway Networks," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 1, pp. 99–112, Jan. 2020.

[15] K. Keita, P. Pellegrini and J. Rodriguez, "A three-step Benders decomposition for the real-time Railway Traffic Management Problem," *Journal of Rail Transport Planning and Management*, vol. 13, May 2020.

[16] H. Zhang, S. Li, Y. Wang, L. Yang and Z. Gao, "Collaborative real-time optimization strategy for train rescheduling and track emergency maintenance of high-speed railway: A Lagrangian relaxation-based decomposition algorithm," *Omega: The International Journal of Management Science*, vol. 102, p. 102 371, Jul. 2021.

[17] B. Kersbergen, T. Van Den Boom and B. De Schutter, "Distributed model predictive control for rescheduling of railway traffic," in *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, Institute of Electrical and Electronics Engineers Inc., Nov. 2014, pp. 2732–2737.

[18] X. Luan, B. De Schutter, T. van den Boom, F. Corman and G. Lodewijks, "Distributed optimization for real-time railway traffic management," in *15th IFAC Symposium on Control in Transportation Systems*, vol. 51, Elsevier B.V., Jan. 2018, pp. 106–111.

[19] Z. Chen, R. Liu, W. Hu and T. Peng, "A Distributed Rescheduling Approach to Railway Network with Fine Granularity in case of Emergencies," in *Proceedings of the 32nd Chinese Control and Decision Conference*, 2020, pp. 314–319.

[20] Y. Wang, S. Zhu, S. Li, L. Yang and B. De Schutter, "Hierarchical Model Predictive Control for on-Line High-Speed Railway Delay Management and Train Control in a Dynamic Operations Environment," *IEEE Transactions on Control Systems Technology*, 2022.

[21] A. Toletti, M. Laumanns and U. Weidmann, "Coordinated railway traffic rescheduling with the Resource Conflict Graph model," *Journal of Rail Transport Planning and Management*, vol. 15, Sep. 2020.

[22] InfraAtlas, *Infrastructure and Timetable: Dataset 20211212-008, viewed in InfraMonitor*, 2022.

Bibliography

[1] ProRail, *prorail.nl/over-ons: Over ons*, 2021.

[2] Centraal Bureau voor Statistiek, *longreads.cbs.nl/nederland-in-cijfers-2020: Nederland in Cijfers*, 2020.

[3] Rijksoverheid: Ministerie van Infrastructuur en Waterstaat, *Toekomstbeeld Openbaar Vervoer (TBOV) 2040*, 2019.

[4] P. Pellegrini, G. Marlière, R. Pesenti, and J. Rodriguez, "RECIFE-MILP: An Effective MILP-Based Heuristic for the Real-Time Railway Traffic Management Problem," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2609–2619, Oct. 2015.

[5] R. M. Lusby, J. Larsen, M. Ehrgott, and D. M. Ryan, "A set packing inspired method for real-time junction train routing," *Computers and Operations Research*, vol. 40, no. 3, pp. 713–724, Mar. 2013.

[6] X. Luan, Y. Wang, B. De Schutter, L. Meng, G. Lodewijks, and F. Corman, "Integration of real-time traffic management and train control for rail networks - Part 1: Optimization problems and solution approaches," *Transportation Research Part B: Methodological*, vol. 115, pp. 41–71, Sep. 2018.

[7] L. Lamorgese and C. Mannino, "An exact decomposition approach for the real-time train dispatching problem," *Operations Research*, vol. 63, no. 1, pp. 48–64, Jan. 2015.

[8] B. Kersbergen, T. van den Boom, and B. De Schutter, "Distributed model predictive control for railway traffic management," *Transportation Research Part C: Emerging Technologies*, vol. 68, pp. 462–489, Jul. 2016.

[9] X. Luan, B. De Schutter, L. Meng, and F. Corman, "Decomposition and distributed optimization of real-time traffic management for large-scale railway networks," *Transportation Research Part B: Methodological*, vol. 141, pp. 72–97, Nov. 2020.

[10] A. D'Ariano, M. Pranzo, and I. A. Hansen, "Conflict resolution and train speed coordination for solving real-time timetable perturbations," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 208–222, Jun. 2007.

[11] A. D'Ariano, F. Corman, and D. Pacciarelli, "Rescheduling," in *Railway Timetabling & Operations*, I. Hansen and J. Pachl, Eds., 2nd ed., 2014.

[12] "Interview with a train dispatcher," *Verkeersleider post Utrecht 25-04-2022*, 2022.

[13] NOS, *ProRail: treinverkeer mogelijk in problemen door thuiszittende verkeersleiders*, Oct. 2020.

[14] Nederlandse Spoorwegen, *Spoorkaart 2022*, 2022.

[15] X. Luan, B. De Schutter, T. van den Boom, F. Corman, and G. Lodewijks, "Distributed optimization for real-time railway traffic management," in *15th IFAC Symposium on Control in Transportation Systems*, vol. 51, Elsevier B.V., Jan. 2018, pp. 106–111.

[16] ProRail, *www.youtube.com/watch?v=vadfRAhE4Oc&t=125s: Treinbeveiliging met het Central Safety System (CSS) van ERTMS: de uitleg*, 2022.

[17] E. Quaglietta, *Introduction, Lecture Notes Course Railway Traffic Management, Delft University of Technology*, 2020.

[18] R. Negenborn and H. Hellendoorn, "Intelligence in Transportation Infrastructures via Model-Based Predictive Control," in *Intelligent Infrastructures*, R. Negenborn, Z. Lukszo, and H. Hellendoorn, Eds., vol. 42, Dordrecht: Springer, 2010, pp. 3–24.

[19] S. Khayyam, F. Ponci, J. Goikoetxea, V. Recagno, V. Bagliano, and A. Monti, "Railway Energy Management System: Centralized-Decentralized Automation Architecture," *IEEE Transactions on Smart Grid*, vol. 7, no. 2, pp. 1164–1175, Mar. 2016.

[20] Q. Gu, T. Tang, F. Cao, H. R. Karimi, and Y. Song, "Peak power demand and energy consumption reduction strategies for trains under moving block signalling system," *Mathematical Problems in Engineering*, pp. 36–47, 2013.

[21] L. Jiang, Z. Bie, T. Long, H. Xie, and Y. Xiao, "Distributed energy management of integrated electricity-thermal systems for high-speed railway raction grids and stations," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 3, pp. 541–554, May 2021.

[22] W. Fang, S. Yang, and X. Yao, "A Survey on Problem Models and Solution Approaches to Rescheduling in Railway Networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 6, pp. 2997–3016, Dec. 2015.

[23] V. Cacchiani *et al.*, *An overview of recovery models and algorithms for real-time railway rescheduling*, 2014.

[24] ProRail, *prorail.nl/over-ons/wat-doet-prorail/coordinatie-treinverkeer/personeelstekort-treinverkeersleiding: Personeelstekort treinverkeersleiding*, 2021.

[25] A. Nandram, *NS-medewerkers staken: 'Normaal gesproken gebruiken 1,3 miljoen reizigers per dag de trein'*, Aug. 2022.

[26] Movares and ProRail, "Telefoonnummers van Verkeersleidinggebieden in Nederland," 2020.

[27] E. Philipsen and R. Oostveen, "Interview," *De Inkpot, Utrecht 05-06-2022*, 2022.

[28] ProRail, *www.prorail.nl/over-ons/wat-doet-prorail/coordinatie-treinverkeer: Coördinatie van het treinverkeer*, 2020.

[29] A. Bruinsma, *Het TreinAfhandelingsDocument (TAD)*, 2021.

[30] P. Kecman, F. Corman, A. D'Ariano, and R. M. Goverde, "Rescheduling models for railway traffic management in large-scale networks," *Public Transport*, vol. 5, no. 1-2, pp. 95–123, Sep. 2013.

[31] J. Törnquist and J. A. Persson, "N-tracked railway traffic re-scheduling during disturbances," *Transportation Research Part B: Methodological*, vol. 41, no. 3, pp. 342–362, 2007.

[32] F. Corman, A. D'Ariano, A. D. Marra, D. Pacciarelli, and M. Samà, "Integrating train scheduling and delay management in real-time railway traffic control," *Transportation Research Part E: Logistics and Transportation Review*, vol. 105, pp. 213–239, Sep. 2017.

[33] J. Rodriguez, "A constraint programming model for real-time train scheduling at junctions," *Transportation Research Part B: Methodological*, vol. 41, no. 2, pp. 231–245, 2007.

[34] A. D'Ariano, F. Corman, D. Pacciarelli, and M. Pranzo, "Reordering and local rerouting strategies to manage train traffic in real time," *Transportation Science*, vol. 42, no. 4, pp. 405–419, 2008.

[35] G. Caimi, M. Fuchsberger, M. Laumanns, and M. Lüthi, "A model predictive control approach for discrete-time rescheduling in complex central railway station areas," *Computers and Operations Research*, vol. 39, no. 11, pp. 2578–2593, Nov. 2012.

[36] F. Corman, A. D'Ariano, I. A. Hansen, and D. Pacciarelli, "Optimal multi-class rescheduling of railway traffic," *Journal of Rail Transport Planning and Management*, vol. 1, no. 1, pp. 14–24, 2011.

[37] L. Meng and X. Zhou, "Robust single-track train dispatching model under a dynamic and stochastic environment: A scenario-based rolling horizon solution approach," *Transportation Research Part B: Methodological*, vol. 45, no. 7, pp. 1080–1102, 2011.

[38] G. Cavone, T. Van Den Boom, L. Blenkers, M. Dotoli, C. Seatzu, and B. De Schutter, "An MPC-Based Rescheduling Algorithm for Disruptions and Disturbances in Large-Scale Railway Networks," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 1, pp. 99–112, Jan. 2020.

[39] International Union of Railways (UIC), *uic.org/com/enews/nr/362/article/rail-topo-model-and-railml-r: Rail Topo Model and railML*, 2013.

[40] A. Mascis and D. Pacciarelli, "Discrete Optimization Job-shop scheduling with blocking and no-wait constraints," *European Journal of Operational Research*, vol. 143, pp. 498–517, 2002.

[41] T. Yamada and R. Nakano, "Ch 7: Job-shop scheduling," in *Genetic algorithms in engineering systems*, A. Zalzala and P. Fleming, Eds., The Institution of Electrical Engineers, 1997, pp. 134–160.

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

[43] P. Pellegrini, R. Pesenti, and J. Rodriguez, "Efficient train re-routing and rescheduling: Valid inequalities and reformulation of RECIFE-MILP," *Transportation Research Part B: Methodological*, vol. 120, pp. 33–48, Feb. 2019.

[44] Y. Wang, B. De Schutter, T. J. Van Den Boom, and B. Ning, "Optimal trajectory planning for trains under fixed and moving signaling systems using mixed integer linear programming," *Control Engineering Practice*, vol. 22, no. 1, pp. 44–56, Jan. 2014.

[45] F. Corman, "Interactions and equilibrium between rescheduling train traffic and routing passengers in microscopic delay management: A game theoretical study," *Transportation Science*, vol. 54, no. 3, pp. 785–822, Jun. 2020.

[46] F. Corman, A. D'Ariano, D. Pacciarelli, and M. Pranzo, "Optimal inter-area coordination of train rescheduling decisions," *Transportation Research Part E: Logistics and Transportation Review*, vol. 48, no. 1, pp. 71–88, 2012.

[47] S. Narayanaswami and N. Rangaraj, "A MAS architecture for dynamic, realtime rescheduling and learning applied to railway transportation," *Expert Systems with Applications*, vol. 42, no. 5, pp. 2638–2656, Apr. 2015.

[48] Y. Guo and C. Zhang, "Real-time railway transit management based on multi-agent system (MAS)," in *Journal of Physics: Conference Series*, vol. 1828, IOP Publishing Ltd, Mar. 2021.

[49] F. Corman, A. D'Ariano, D. Pacciarelli, and M. Pranzo, "Dispatching and coordination in multi-area railway traffic management," *Computers and Operations Research*, vol. 44, pp. 146–160, Apr. 2014.

[50] K. Keita, P. Pellegrini, and J. Rodriguez, "A three-step Benders decomposition for the real-time Railway Traffic Management Problem," *Journal of Rail Transport Planning and Management*, vol. 13, May 2020.

[51] H. Zhang, S. Li, Y. Wang, L. Yang, and Z. Gao, "Collaborative real-time optimization strategy for train rescheduling and track emergency maintenance of high-speed railway: A Lagrangian relaxation-based decomposition algorithm," *Omega: The International Journal of Management Science*, vol. 102, p. 102 371, Jul. 2021.

[52] S. K. Sinha, S. Salsingikar, and S. SenGupta, "An iterative bi-level hierarchical approach for train scheduling," *Journal of Rail Transport Planning and Management*, vol. 6, no. 3, pp. 183–199, Dec. 2016.

[53] F. Corman, A. D'Ariano, D. Pacciarelli, and M. Pranzo, "Centralized versus distributed systems to reschedule trains in two dispatching areas," *Public Transport*, vol. 2, no. 3, pp. 219–247, Aug. 2010.

[54] Y. Wang, S. Zhu, S. Li, L. Yang, and B. De Schutter, "Hierarchical Model Predictive Control for on-Line High-Speed Railway Delay Management and Train Control in a Dynamic Operations Environment," *IEEE Transactions on Control Systems Technology*, 2022.

[55] B. Kersbergen, T. Van Den Boom, and B. De Schutter, "Distributed model predictive control for rescheduling of railway traffic," in *2014 17th IEEE International Conference on Intelligent Transportation Systems, ITSC 2014*, Institute of Electrical and Electronics Engineers Inc., Nov. 2014, pp. 2732–2737.

[56] C. Strotmann, "Railway scheduling problems and their decomposition," Tech. Rep., 2007.

[57] A. Toletti, M. Laumanns, and U. Weidmann, "Coordinated railway traffic rescheduling with the Resource Conflict Graph model," *Journal of Rail Transport Planning and Management*, vol. 15, Sep. 2020.

[58] M. Faris, A. Núñez, S. Zhou, and B. De Schutter, "Distributed Optimization for Railway Track Maintenance Operations Planning," in *21st International Conference on Intelligent Transportation Systems*, 2018, pp. 1194–1201.

[59] Z. Chen, R. Liu, W. Hu, and T. Peng, "A Distributed Rescheduling Approach to Railway Network with Fine Granularity in case of Emergencies," in *Proceedings of the 32nd Chinese Control and Decision Conference*, 2020, pp. 314–319.

[60] Y. Guo, C. Zhang, C. Wu, and S. Lu, "Multiagent System–Based Near-Real-Time Trajectory and Microscopic Timetable Optimization for Rail Transit Network," *Journal of Transportation Engineering, Part A: Systems*, vol. 147, no. 2, p. 04 020 153, Feb. 2021.

[61] H. Hassanabadi, B. Moaveni, and M. Karimi, "A comprehensive distributed architecture for railway traffic control using multi-agent systems," *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 229, no. 2, pp. 109–124, Feb. 2015.

[62] H. Jaakkola and B. Thalheim, "Architecture-driven modelling methodologies," in *Proceedings of the 2011 conference on Information Modelling and Knowledge Bases XXII*, 2011, p. 98.

[63] P. Su and B. Park, "Auction-based highway reservation system an agent-based simulation study," *Transportation Research Part C: Emerging Technologies*, vol. 60, pp. 211–226, Nov. 2015.

[64] C. Yong, M. Ullrich, and L. Jiajian, "Decentralized, autonomous train dispatching using swarm intelligence in railway operations and control," in *Proceedings of the 7th International Conference Railway Operations Modelling and Analysis*, 2017, pp. 521–540.

[65] E. Marcelli and P. Pellegrini, "Literature Review Toward Decentralized Railway Traffic Management," *IEEE Intelligent Transportation Systems Magazine*, vol. 13, no. 3, 18p, Jan. 2020.

[66] T. Letia, M. Hulea, and R. Miron, "Distributed scheduling for real-time railway traffic control," in *Proceedings of the International Multiconference on Computer Science and Information Technology, IMCSIT 2008*, vol. 3, 2008, pp. 679–685.

[67] T. Anwar, C. Liu, H. L. Vu, and C. Leckie, "Partitioning road networks using density peak graphs: Efficiency vs. accuracy," *Information Systems*, vol. 64, pp. 22–40, Mar. 2017.

[68] X. Lin and J. Xu, "Road network partitioning method based on canopy-kmeans clustering algorithm," *Archives of Transport*, vol. 54, no. 2, pp. 95–106, 2020.

[69] L. Meng and X. Zhou, "Simultaneous train rerouting and rescheduling on an N-track network: A model reformulation with network-based cumulative flow variables," *Transportation Research Part B: Methodological*, vol. 67, pp. 208–234, 2014.

[70] InfraAtlas, *Infrastructure and Timetable: Dataset 20211212-008, viewed in InfraMonitor*, 2022.

[71] A. G. Williams *et al.*, "Radon as a tracer of atmospheric influences on traffic-related air pollution in a small inland city," *Stockholm Uni Press*, vol. 68, no. 1, p. 30 967, 2016.

[72] A. Sayegh, J. E. Tate, and K. Ropkins, "Understanding how roadside concentrations of NOx are influenced by the background levels, traffic density, and meteorological conditions using Boosted Regression Trees," *Atmospheric Environment*, vol. 127, pp. 163–175, Feb. 2016.

[73] Y. Son *et al.*, "Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters," *Science of The Total Environment*, vol. 639, pp. 40–48, Oct. 2018.

[74] L. Sørensen, K. B. Lindberg, I. Sartori, and I. Andresen, "Residential electric vehicle charging datasets from apartment buildings," *Data in Brief*, vol. 36, pp. 105–116, Jun. 2021.

[75] E. Quaglietta, *Real-time rescheduling, Lecture Notes Course Railway Traffic Management, Delft University of Technology*, 2020.

[76] J. Pachl, "Timetable Design Principles," in *Railway Timetabling & Operations*, I. A. Hansen and J. Pachl, Eds., 2nd ed., 2014.

[77] J. Koning, "Comparing the performance ERTMS level 2 fixed block and ERTMS level 3 moving block signalling systems usign simulation techniques," in *Computers in Railways*, vol. 8, WIT Press, 2002, pp. 43–52.

[78] M. L. Janssens, *Multi machine approaches for conflict resolution under moving block signalling*, 2022.