Adding flexibility to the timetable in real-time railway traffic management

Runsheng Zhou

TU Delft

Adding flexibility to the timetable in real-time railway traffic management

by

Runsheng Zhou

to obtain the degree of Master of Science in Civil Engineering at the Delft University of Technology,

Student number:	5451590	
Date:	November 30, 2023	
Thesis committee:	Prof. dr. R. M. P. Goverde,	TU Delft, chair
	Ir. Z. Wang,	TU Delft, daily supervisor
	Dr. ir. G. Correia,	TU Delft, second examiner
	Ir. E. P. Philipsen,	ProRail, external supervisor

Delft University of Technology

Preface

Dear reader,

Sitting at home in Beijing, I cannot help but think back to my lively times at Delft University of Technology in the Netherlands. This acknowledgment goes to everyone who made this academic journey possible and meaningful.

First and foremost, my heartfelt appreciation goes to all committee members. You sparked my interest in learning and patiently guided me through my research. Your tireless tutoring was the key factor for my successful thesis project. Special thanks to my daily supervisor, Ziyulong, a tutor and friend. Whenever I felt hard to find the balance between life and study, he was always there to guide me through. In certain tough times, I would ask you questions during non-working hours, and you were always willing to sacrifice your time to solve my problems.

To my family members in China, even though you were far away from the Netherlands during the entire journey, your constant support kept me going. A special mention to my girlfriend, Yucen – your warm presence during the intense days and nights of study motivated me. Heartfelt thanks to my parents for their strong financial and emotional support. The prospect of reuniting with you fuels my dedication to the thesis.

To my friends, no matter where you are in the world, your friendship has been my anchor. To my friends in the Netherlands, our gatherings after tough working days were a great stress-buster. I deeply miss the evenings filled with laughter. To friends beyond the Netherlands, our online chats and videos were a joy, thanks for sharing your lives and happiness.

My thesis dives into the world of railways. Over nine months, I learned a lot about the Dutch railway system, both in academics and industry. Due to my great passion for this thesis project, even though the entire process was not easy, I never regretted it.

Next, I will enter the next stage of my life, I cannot wait to start my career in China. For some reason, I may not continue to explore the railway field, but I firmly believe that this thesis is not in vain. During this journey, I learned not only academic and practical knowledge related to railways, but also methodology for discovering, understanding, exploring, and solving problems. These will have a profound impact on my future career, regardless of the exact field.

Writing this now, night falls quietly, reminding me of the nights working on my thesis in the Netherlands. There is more I want to say, but let me wrap it up. Huge thanks to everyone who played a part in this thesis project. Wishing you all good health, ongoing success, and the fulfillment of your dreams.

Runsheng Zhou
Rotterdam, November 28, 2023

Executive summary

With the imperative of transitioning toward sustainable transportation modes, several projects and strategies have been suggested to support the development of European railway transport. With the strong backing and advocacy from the EU for railway transport, it is crucial to focus on innovative and efficient technologies to maintain service quality. A vital consideration in railway operations is uncertainty, encompassing both external factors (such as weather conditions, passenger numbers, and passenger behavior) and internal factors (like infrastructure and vehicle equipment reliability, and personnel behavior). In everyday train operations, uncertainties can lead to perturbations in the timetable, necessitating the rescheduling of trains and the restoration of timetable feasibility through traffic management systems. Failure to address these perturbations properly may give rise to two types of track occupation conflicts: disturbances and disruptions.

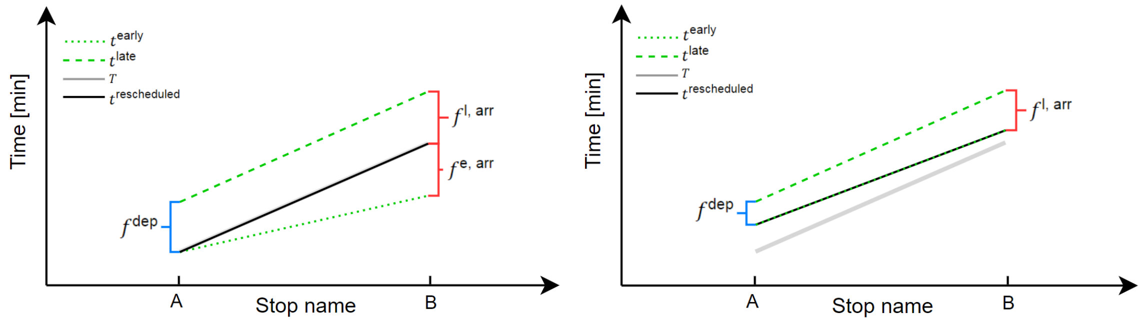
At the tactical level, timetable flexibility is defined as the freedom to choose the event times from a periodic timetable. In the case of traffic uncertainties, timetable flexibility can be used to adjust the departures and arrivals of trains, thus preventing conflicts from occurring. Some attempts have been made to explore the concept of timetable flexibility; however, a universally accepted definition has yet to emerge. Robustness is preventive, which assigns more buffer times to increase timetable robustness in compromise with the capacity of the infrastructure. Resilience is proactive and reactive since a resilient timetable helps to withstand, absorb, accommodate, and recover from disturbances and disruptions. Timetable flexibility can be considered as a means to achieve resilience, which is preventive and proactive. Flexibility is preventive similar to robustness, while it assigns time slots to the arrival and departure events of different trains at different stations to deal with real-time uncertainties. Meanwhile, it is also proactive since it responds to conflicts that are predicted to be happening in real traffic. In this research, we are not changing the timetable, indicating that flexibility is added on top of the rescheduling plan or the rigid timetable. Timetable flexibility is defined as the ability of a timetable to be easily modified to withstand small disturbances and absorb delays, as well as to offer a larger solution space in the application of dispatching measures (retiming, reordering, rerouting) to solve larger disturbances without changing the given (re)scheduled timetable. In the current research, retiming and reordering are under consideration.

In accordance with the official performance indicator of train punctuality reported by ProRail, the punctuality of passenger trains is gauged by calculating the percentage of arrivals in which the deviation between the rigid arrival time and the actual arrival time falls below the 3-minute punctuality threshold. Three types of flexibility are allowed in our research, namely, late departure flexibility, early arrival flexibility, and late arrival flexibility. The maximum allowable departure flexibility and late arrival flexibility in event timings are closely associated with the punctuality threshold. Two illustrative graph representations of timetable flexibility are presented in Figure 1.

In both examples, a train departs from station A and arrives at station B. The grey solid line indicates the rigid train path, and the black solid line indicates the train path of the rescheduling plan. The green dashed line indicates the flexible time with late departures and late arrivals, and the green dotted line represents the flexible time with early arrivals. The punctuality threshold regulates the boundaries of departure flexibility and late arrival flexibility. It indicates that the difference between the green dashed line and the grey line at the two ends cannot exceed the punctuality threshold. In our research, we aim to add timetable flexibility to the rescheduling plan, thus, providing train dispatchers and signalers with more solution spaces in finding a new Real-Time Traffic Plan (RTTP) when encountering traffic uncertainties.

The main objective of the current research is to add and optimize timetable flexibility in traffic disturbance management for passenger railway transport. Therefore, the main research question and research questions are constructed as follows.

Main research question:



(a) Timetable flexibility without initial delays

(b) Timetable flexibility with initial delays

Figure 1: Illustrative graph representations of timetable flexibility

How to introduce and optimize flexible event times to the rescheduling plan in railway traffic management?

Research questions:

1. What is the state-of-the-art on real-time railway traffic management, and what is the state-of-the-practice in the Netherlands?
2. How can timetable flexibility be defined and modeled in the conflict detection and resolution process to output an optimal rescheduling plan that efficiently addresses traffic disturbances?
3. How can the performance of the model be analyzed?
4. How can the performance of timetable flexibility be evaluated?

Based on the literature review, mainstream rescheduling models utilized in railway traffic disturbance management encompass Mixed Integer Linear Programming (MILP), Integer Programming (IP), and the Alternative graph (AG) model. Additionally, Constraint Programming (CP), knowledge-based, and learning-based rescheduling methods are also employed. However, the incorporation of flexibility into the timetable remains relatively unexplored within the railway field. Efforts have been made to introduce flexibility at both tactical and operational planning levels, but a consensus on the definition of timetable flexibility is lacking, and there is a dearth of follow-up studies in this domain.

In terms of the state-of-the-practice, ProRail manages railway traffic disturbance. The primary objective is to adapt plans in response to traffic disturbances. This process involves five stages: traffic state monitoring, traffic state prediction, conflict detection, conflict resolution, and plan updating. Rescheduling models come into play during conflict resolution to mitigate potential conflicts.

As our contribution, a framework for real-time railway operations is constructed, which is shown in Figure 2. Timetable flexibility is added in the Conflict Detection and Resolution (CDR) module. In case of newly detected delays, a rescheduling plan is generated in which flexibility is maximized. Otherwise, timetable flexibility is added on top of the original plan, which can be the rigid timetable or the last RTTP. The rescheduling plan is determined whether to be accepted by train dispatchers and signalers in the rescheduling decision model. Once accepted, a new RTTP will be generated and fed to train operations and traffic control. Otherwise, the last RTTP is maintained. Traffic control is about executing the RTTP generated in traffic management, which guarantees the safe implementation of the plan. Train operations are about the efficient running of individual trains. Real traffic is a combination of traffic control and train operation.

Adding flexibility to the rescheduling plan is realized by an AG-based MILP model. This model is represented as a graph with nodes and arcs. Each node is associated with a time instant, indicating the starting of an operation, where the operation can be a train entering a block section or station. An arc represents the relation between two nodes. All arcs are directed arcs in an AG formulation, thus, the order between two events is inherently indicated. There are three types of timetable flexibility, namely departure flexibility, early arrival flexibility, and late arrival flexibility. Departure flexibility is represented as the time difference between the latest departure time and the rescheduled departure time. Early

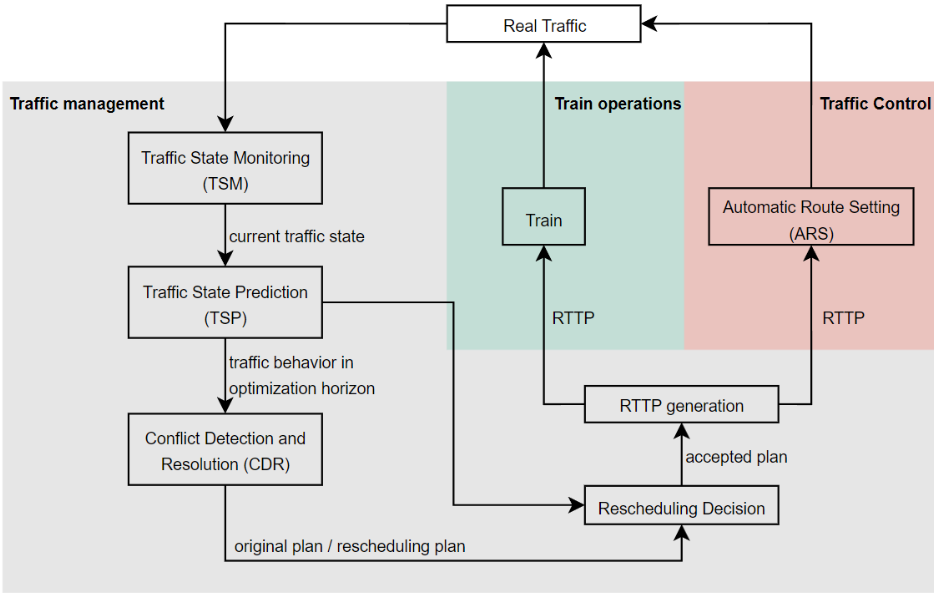


Figure 2: Framework for real-time railway operations

arrival flexibility is the time difference between the rescheduled arrival time and the earliest arrival time. Late arrival flexibility is the time difference between the latest arrival time and the rescheduled arrival time. ProRail officially assesses train punctuality for passenger trains using a performance indicator that focuses on the percentage of arrivals. A train is considered punctual when the difference between the rigid arrival time and the actual arrival time is less than a punctuality threshold of 3 minutes. The concept of the punctuality threshold also applies to regulate the values of late departure flexibility and late arrival flexibility. Notably, early arrival flexibility is not bound by this threshold, permitting trains to arrive significantly ahead of schedule. Operational constraints should be respected by the times to ensure the feasibility of the rescheduling plan and the flexible plans.

A toy network, which is a simplified and smaller version of a real-life railway network, is designed to verify the model. Based on the results, it can be observed that the control measure of retiming and reordering is functioning well. When no initial delays are experienced by trains, timetable flexibility is added on top of the rigid timetable. When trains are delayed, flexibility is added on top of the rescheduling plan that has the least deviation compared to the rigid one.

Following model verification, a case study was conducted in the network of four major train stations, namely, ’s-Hertogenbosch (Ht), Tilburg (Tb), Breda (Bd), and Eindhoven (Ehv). This network is a small part of the Dutch railway network, and an illustrative overview is indicated by the red border in Figure 3. Trains running from Ehv to Bd, from Ht to Bd, and from Ehv to Ht are incorporated in the case study. A passenger transfer between trains from different directions is incorporated at station Tb. In addition to the four major stations, there are intermediate stops along the route, including Eindhoven Philips Strijp (Ehs), Best (Bet), Btl, Oisterwijk (Ot), Tilburg Reeshof (Tbr), Tilburg Universiteit (Tbu), Gilze Rijen (Gz), and Vught (Vg). In this case study area, all trains except the freight train have a period of half an hour, to make sure that a full cycle is involved, the first trains of the second periods are also incorporated. The case study incorporated a total of 13 trains, comprising 9 intercity trains (IC), 3 sprinter trains (SPR), and 1 freight train (GO). To investigate the relationship between input delays and output flexibility, 10 sets of initial delays were randomly generated, with the delays following a Weibull distribution, indicating 10 random cases. In this case study, the model comprises 3280 continuous variables, 1099 binary variables, and 22484 constraints. It was found that our model is able to provide solutions in an average computation time of 1.12 s, proving the feasibility of the model in real-time railway traffic management.

In order to investigate the influencing factors of timetable flexibility, one illustrative application and two sensitivity analyses were conducted. In the illustrative applications of the model to various inputs, initial delays from 0 to 10 minutes with an interval of 100s are added to a single train, and initial delays of all other trains are fixed to 0 s. By comparing the train paths and optimization results of different delay scenarios, it was found that timetable flexibility basically decreases as the initial delay increases.

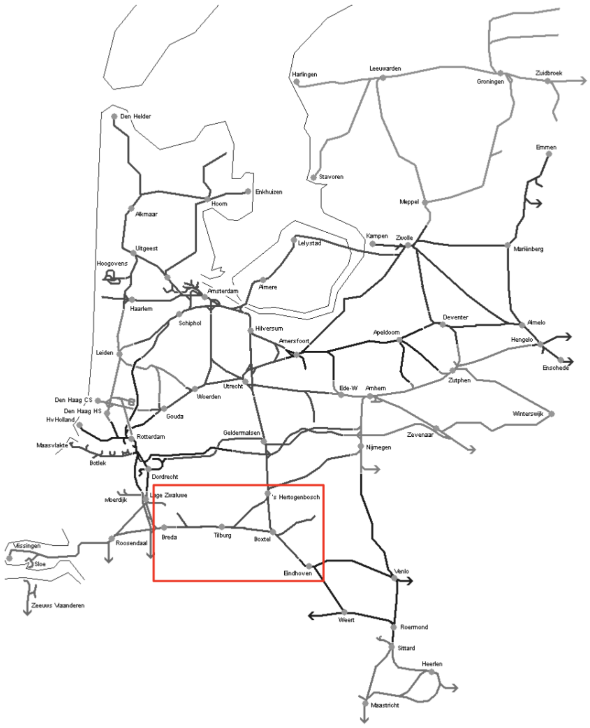


Figure 3: An overview of the selected dispatching area[1]

The reason is that the rescheduled train paths are closer to each other to recover the initial delays, leading to fewer solution spaces to add flexibility. However, there were also situations where flexibility increased with the increasing initial delays. These are because of the interactions between trains. In these scenarios, the delay of a train leads to more buffer times for the former train to arrive and depart later than the rescheduled time. Thus, the late arrival flexibility and departure flexibility of the former train increases.

In the objective function of our model, a weight of departure flexibility is introduced, which indicates the preference between departure flexibility and arrival flexibility. In the first sensitivity analysis, the relationship between the weight and the distribution of timetable flexibility was investigated. The weights equal 1.1, 1.0, and 0.9 are tested for the case without initial delays and the 10 random cases. It was found that a trade-off only exists between departure flexibility and early arrival flexibility, and late arrival flexibility remains the same while changing the weight. When the weight is greater than 1, departure flexibility is maximized. When the weight is less than one, early arrival flexibility reaches its maximum. When the departure flexibility of a train at a station is reduced, the late arrival flexibility is also diminished to ensure that the dwell time requirement is met. Simultaneously, though the early arrival flexibility of the following train increases, the overall timetable flexibility decreases, leading to reduced objective value. Therefore, in order to maximize the collective timetable flexibility of trains within the selected dispatching area, it is essential to maintain the late arrival flexibility at a consistent level when modifying the weighting of departure flexibility.

In the second sensitivity analysis, we explored the effects of adjusting the specific value of the punctuality threshold to understand its impact on the flexibility that can be integrated into the rescheduling plan within the case study network. Our findings revealed that as the punctuality threshold value increases, the flexibility allocated to each stop also increases. This phenomenon occurs because, with a higher punctuality threshold, trains are allowed to have much later arrival and departure, as long as they adhere to other operational constraints. Furthermore, we introduced a concept termed flexibility increment, which represents the difference in flexibility per stop between adjacent tested punctuality thresholds. This indicator helps illustrate how flexibility increases as the punctuality threshold gradually rises. The results indicate that the flexibility increment tends to decrease as the punctuality threshold

[1] http://www.sporenplan.nl/

exceeds the range from $0.5p$ to $3p$, p is the value of the punctuality threshold that is currently used by ProRail. This decrease suggests that flexibility may not continually increase at all stops, as it might have already reached its maximum potential. It is important to note that in our case study, not all trains are subject to headway constraints since only a subset of trains is considered. When the punctuality threshold exceeds $3p$, late arrival flexibility and departure flexibility are introduced for these non-regulated trains, resulting in constant flexibility increment.

ProRail's current practice involves generating an RTTP with precise departure and arrival times, which can be challenging for trains to adhere to due to real-time traffic uncertainties. When these uncertainties arise, it necessitates the use of computationally expensive algorithms to create a new RTTP. To address this issue, the concept of timetable flexibility is introduced, allowing for the generation of a rescheduling plan with event time slots. This approach offers more flexibility to train dispatchers and signalers, enabling them to select new arrival and departure times from event time slots to adapt to the uncertainties.

Timetable flexibility is advantageous in various situations. Departure flexibility becomes crucial during peak hours when platforms are crowded, leading to longer dwell times than initially scheduled. Late departures are then required to accommodate passengers and minimize inconvenience. Arrival flexibility, on the other hand, is valuable in scenarios such as bad weather, which can cause tracks to become slippery and affect estimated traction and braking force. Consequently, trains may arrive earlier or later than planned. Additionally, complex train interactions in congested areas can result in deviations from rigid arrival times. In all these situations, the incorporation of early and late arrival flexibility helps mitigate the impact of uncertainties.

The choice between early arrival flexibility and departure flexibility depends on the specific scenarios. In high passenger volume situations or peak hours, late departure flexibility is preferred to ensure smooth operations. In contrast, in extreme weather conditions or congested dispatching areas, early arrival flexibility is preferred over departure flexibility.

It is important to note that the introduction of timetable flexibility does not directly affect railway travelers, as they continue to rely on the published timetable. However, the indirect benefits for passengers are significant. Most delays are induced by the mishandling of traffic uncertainties. Timetable flexibility allows for the timely resolution of uncertainties by selecting a new RTTP from the rescheduling plan with event time slots. This, in turn, reduces delays throughout the network, resulting in improved punctuality and passenger satisfaction.

Contents

1

Introduction

In this chapter, essential context regarding the current research is provided. Section 1.1 discusses the change in demand for railway transport due to the COVID-19 crisis. It demonstrates the importance of developing railway transport and handling railway traffic uncertainties, such as weather conditions, passenger behavior, etc. In Section 1.2, the research problem is identified. This is done by presenting the railway planning levels and delving into the specific focus of this study. Moreover, we expound upon some relevant concepts that have been thoroughly studied in the existing literature. Furthermore, this section provides the motivation behind the research, while also offering a comprehensive definition of timetable flexibility. Section 1.3 illustrates the objective of the current research and how the identified problems will be addressed. Section 1.4 presents the outline of the current research.

1.1. Trends in the railway sector

The railway service can be classified into railway passenger transport and railway freight transport. The demand for passenger transport and freight transport can be expressed by passenger kilometers and tonne kilometers, respectively. The annual statistics for both indicators from 2015 to 2021 are shown in Figure 1.1. Due to the COVID-19 crisis, many European countries reported a significant decrease in passenger-kilometers between 2019 and 2020 in passenger transport for main railway undertakings. In the same period, a decline can also be observed in tonne-kilometers in freight transport for main railway undertakings. However, the demand for freight transport in 2021 exceeded the value in 2019 and almost recovered to the peak level since 2015, while the demand for passenger transport in 2021 did not recover too much from the pandemic.

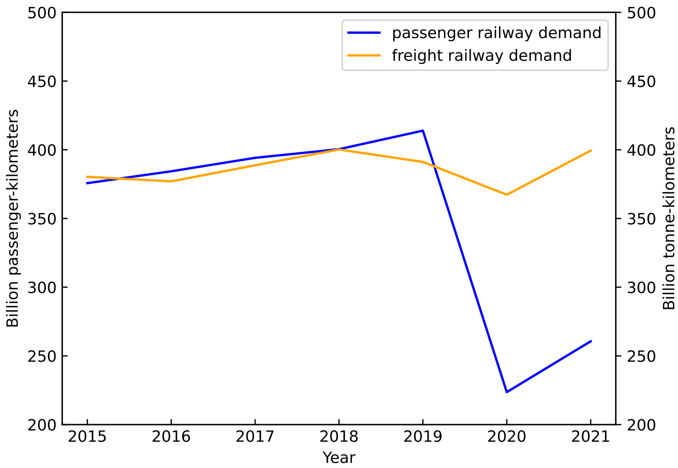


Figure 1.1: Railway transport for main undertakings in Europe from 2015-2021(Eurostat, 2022)

Furthermore, the railway passenger demand may not fully recover to the state before the crisis due to the prevalence of hybrid work. 20-25% of workers in advanced economies and about 10% in emerging economies could work from home three to five days a week, which indicates a decline in commuter and business trips (Lund et al., 2021).

Although the total demand for railway transport may not be able to return to the previous levels, railway transport is deemed as one of the most sustainable and safest modes of transport (Ott et al., 2021). In terms of sustainability, Ritchie (2020) found that walking and cycling are nearly always the lowest carbon way to travel over short to medium distances, and traveling by train is the most sustainable option compared to traveling by car or airplane over medium to long distances. Regarding the safety aspect, the rail appears as the safest mode of land transport in the EU, with the fatality rate for passengers gradually approaching that for aircraft on-board passengers(European Union Agency for Railways, 2022).

Under the imperative of transitioning towards sustainable transportation modes, various projects and strategies have been proposed to support the advancement of European railway transport. Europe's Rail Joint Undertaking (EU-RAIL) envisions creating a high-capacity, integrated European railway network by eliminating interoperability barriers, providing comprehensive integration solutions, and accelerating the implementation of innovations. Additionally, EU-RAIL advocates digitalization and automation to enhance cost-effectiveness in the railway industry, boost capacity, and reinforce flexibility and reliability (European Union, 2021). The European Green Deal, on the other hand, aims to transform the EU into a modern, resource-efficient, and competitive economy, with the ambitious target of achieving net-zero greenhouse gas emissions by 2050. From a transport perspective, this initiative aims to improve the well-being and health of citizens and future generations by prioritizing public transportation (European Commission, 2021). Recognizing the significance of rail as a sustainable, smart, and safe mode of transport, the European Commission declared 2021 as the European Year of Rails, emphasizing the importance of rail in the journey towards a greener future (European Commission, 2022). The commission has also set key milestones to guide the transition towards sustainability and innovation, such as doubling high-speed rail traffic across Europe by 2030 and doubling rail freight traffic by 2050 (European Commission, 2020).

With the EU's strong support and promotion of railway transport, attention must be paid to innovative and efficient technologies to uphold service levels. One of the crucial aspects to consider in railway operations is uncertainty, which encompasses both external factors (e.g., weather conditions, passenger numbers, passenger behavior) and internal factors (e.g., infrastructure and vehicle equipment reliability, personnel behavior) (Kecman et al., 2013). This dynamic and stochastic nature of railway transport necessitates precise estimation of future train states, especially when confronted with disturbances and disruptions. While extensive research has focused on online static train rescheduling, accounting for the dynamic parts remains a critical area that requires further exploration (Corman and Meng, 2014). In daily train operations, uncertainties can lead to timetable perturbations, necessitating train rescheduling and restoration of timetable feasibility through traffic management systems (D'Ariano et al., 2007b). Failure to address perturbations properly may result in two types of track conflicts, namely, disturbances and disruptions. Disturbances are relatively small perturbations in railway operations that can be managed by adjusting only the timetable, while disruptions are relatively large incidents that require the rolling stock and crew schedule to be modified (Cacchiani et al., 2014). Both disturbances and disruptions result in primary delays compared to the planned train schedule, with these primary delays propagating as secondary delays to other trains (Corman and Meng, 2014), thereby significantly impacting operations and passengers (Liu et al., 2021). Consequently, effectively dealing with uncertainty and promptly resolving railway traffic disturbances and disruptions to maintain high railway operation performance represents a compelling and worthwhile research area.

1.2. Problem identification

In the railway context, a hierarchical decision-making structure is generally adopted to deal with the planning problem, which includes three levels comprising strategic level, tactical level, and operational level (Lusby et al., 2011). In addition, real-time railway operations comprise three stages, including traffic management, traffic control, and train operations. Considering the trends in the railway sector, solutions are proposed to improve the efficiency of the railway system to mitigate the influence of train conflicts. In transportation systems, Morlok and Chang (2004) define flexibility as the ability of a system to adapt to external changes while maintaining satisfactory system performance. Flexibility is increasingly desired in transportation systems in order to accommodate changing demands and traffic patterns, however, in the railway context, only a few researchers have tried to add flexibility in different planning levels and operation stages, and there is no consensus among studies on the definition of timetable flexibility.

In order to increase the chance of finding feasible train paths, Caimi et al. (2011) investigated the possibility of adding flexibility in the timetabling stage of the tactical planning level. They define timetable flexibility as the freedom to choose the event times from a periodic timetable. They proposed a Flexible Periodic Event Scheduling Problem (FPESP) model on the basis of the conventional Periodic Event Scheduling Problem (PESP), in which a flexible timetable with event time slots is generated rather than exact event times, which provides larger solution space for the allocation of railway tracks (Caimi et al., 2011).

Apart from adding flexibility at the tactical planning level, the Train Path Envelope (TPE) can be considered as adding flexibility to train runs in the stage of train operations. A TPE is a sequence of time windows in which trains can run in an energy-efficient way without hindering the following trains (Quaglietta et al., 2016). Therefore, trains are allowed to run flexibly within a given range taking into account the situation at the moment.

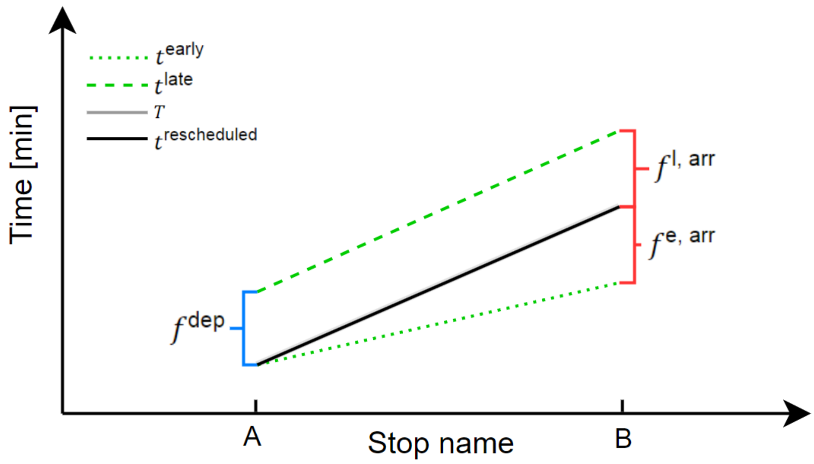
D'Ariano et al. (2008b) assessed the flexible timetable in traffic management. In their definition of a flexible timetable, a set of feasible platform tracks for each train and for each station, the time windows of arrival and departure times, and a provisional order of trains at overtakes and junctions are presented. They found that flexible timetables are preferable to rigid ones since flexibility offers more freedom to solve conflicts. Therefore, timetable flexibility is a promising concept to improve train punctuality by efficiently resolving train delays and accommodating traffic uncertainties.

Based on the definition of flexible timetable proposed by D'Ariano et al. (2008b), the flexible timetable that is used in our research can be identified. In the current research, we do not mean to generate a newly published timetable for passengers but aim to generate a new rescheduling plan to train dispatchers and signalers, in which we provide them with larger solution spaces to handle real-time traffic uncertainties.

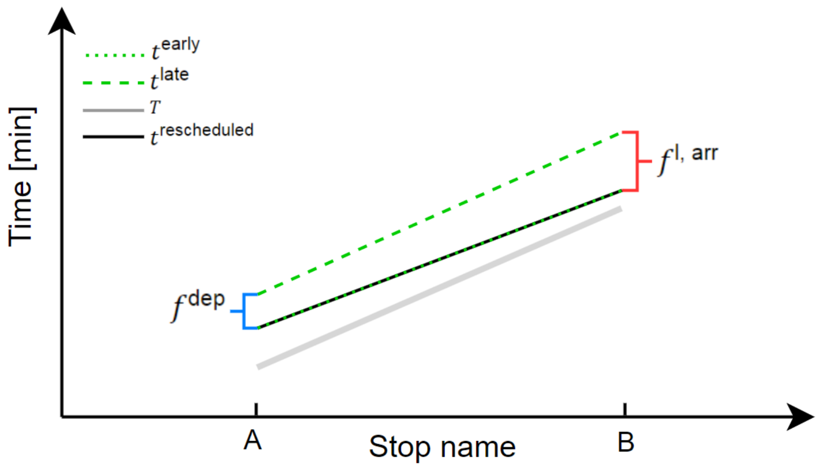
Two concepts, namely, robustness and resilience, were often mentioned in traffic management and have been widely studied. Robustness is preventive, which assigns more buffer times to increase timetable robustness in compromise with the capacity of the infrastructure. Resilience is both proactive and reactive since a resilient timetable helps to withstand, absorb, accommodate, and recover from disturbances and disruptions (Goverde and Hansen, 2013). According to these two concepts, timetable flexibility can be defined as the ability of a timetable to be easily modified to withstand small disturbances and absorb delays, as well as to offer a larger solution space in the application of dispatching measures (retiming, reordering, rerouting) to solve larger disturbances without changing the given (re)scheduled timetable. Timetable flexibility can be considered as a means to achieve resilience, which is preventive and proactive. Flexibility is preventive similar to robustness, while it assigns time slots to the arrival and departure events of different trains at different stations to deal with real-time uncertainties. Meanwhile, it is also proactive since it responds to conflicts that are predicted to be happening in real traffic.

A rigid timetable is a timetable published to passengers, while a rescheduling plan has the least deviation from the rigid timetable. On top of the rescheduling plan, we aim to add and maximize flexibility if the punctuality threshold allows. According to the performance indicator of train punctuality that is officially reported by ProRail, train punctuality for passenger trains is measured by the percentage of arrivals where the difference between the rigid arrival time and the realized arrival time is less than the punctuality threshold of 3 minutes. Regarding departure events, early departures compared to rigid departure times are strictly forbidden since this will prevent passengers who are following the rigid published timetable from catching the train. Therefore, there are three types of timetable flexibility, namely, departure flexibility, early arrival flexibility, and late arrival flexibility. Since punctuality is only a measure of lateness, the concept of the punctuality threshold is extended to late departure flexibility and late arrival flexibility. It means that very early arrival times at stations are allowed.

In order to provide a more straightforward explanation of timetable flexibility, two illustrative examples are presented in Figure 1.2. A and B are two stations, the grey solid line indicates the rigid train path, where the start of this line is the rigid departure time at station A and the end of the line is the rigid arrival time at station B. The black solid line indicates the rescheduled train path, where the two ends are the rescheduled arrival time and departure time of the rescheduling plan, respectively. The green dashed line indicates the flexible time with late departure and late arrival, and the green dotted line represents the flexible time with early arrivals. In Figure 1.2a, no initial delays are added, thus, the rescheduled train path is the same as the rigid train path, and the black line and the grey line overlap. The punctuality threshold regulates the boundaries of departure flexibility and late arrival flexibility. Thus, The difference between the green dashed line and the grey line at the two ends cannot exceed the punctuality threshold. Consequently, departure flexibility is calculated as the difference between the latest departure time and the rescheduled departure time. Late arrival flexibility is calculated as the difference between the latest arrival time and the rescheduled arrival time. Early flexibility is the difference between the rescheduled arrival time and the earliest arrival time. According to Figure 1.2b, some initial delays are assigned, thus, the black line moves above the grey line. Since the rescheduled departure and arrival times in the rescheduling plan are the earliest possible times, early arrival flexibility cannot be added anymore, leading to the overlapped green dotted line and black line. Regarding late departures and late arrivals, they can still be added, only if the difference between the green dashed line and the grey line at the two ends is lower than the punctuality threshold. In general, we aim to add timetable flexibility to the rescheduling plan, and the scenario without initial delays is a special case, in which the rescheduling plan is the same as the rigid timetable. The area between the green dashed line and the green dotted line is the possible solution space that we aim to provide to train dispatchers and signalers in finding a Real-Time Traffic Plan (RTTP) when encountering traffic uncertainties.



(a) Timetable flexibility without initial delays



(b) Timetable flexibility with initial delays

Figure 1.2: Illustrative graph representations of timetable flexibility

1.3. Objectives and research questions

In the current research, we do not change the published timetable for passengers but aim to generate a new working plan to train dispatchers and signalers, in which we provide them with larger solution spaces to handle real-time traffic uncertainties. Given the potential and the definition of timetable flexibility, the overall objective of the current research is to add and optimize timetable flexibility in traffic disturbance management for railway transport. In addition, we aim to investigate the impacts of timetable flexibility on the rescheduling plan determined in the Conflict Detection and Resolution (CDR) process in traffic management. From a scientific point of view, the objective is to construct a model that generates a rescheduling plan for train dispatchers and signalers, in which timetable flexibility is maximized, which leads to an optimal objective value under given initial delays. In the current practice of the railway sector, CDR generates a rescheduling plan, if the plan is accepted by the train dispatchers and signalers, an RTTP will be generated and provided to trains and the route setting. However, the RTTP with time points can be hard for trains to follow due to real-time traffic uncertainties. By adding timetable flexibility, a rescheduling plan with departure and arrival time slots can be obtained, providing train dispatchers and signalers with larger solution spaces in determining the RTTP. When encountering traffic uncertainties, a new set of departure and arrival times can be easily generated based on the rescheduling plan, while not conducting the computationally expensive solution algorithms to obtain a new rescheduling plan. Therefore, the practical objective of the current research is to improve the robustness of the RTTP.

Based on the research objectives, the main research question is constructed as follows.

Main research question:

How to introduce and optimize flexible event times to the rescheduling plan in railway traffic management?

In order to guide the investigation of adding flexibility to the rescheduling plan, the main research question is divided into four sub-research questions containing the theoretical and practical aspects of the research project.

Sub-research questions:

1. What is the state-of-the-art on real-time railway traffic management, and what is the state-of-the-practice in the Netherlands?
2. How can timetable flexibility be defined and modeled in the CDR process to output an optimal rescheduling plan that efficiently addresses traffic disturbances?
3. How can the performance of the model be analyzed?
4. How can the performance of timetable flexibility be evaluated?

1.4. Outline of the thesis

The remainder of this thesis report is organized as follows.

In Chapter 2, the state-of-the-practice of the railway sector in the Netherlands is presented according to the relevant literature and an interview with an advisor from ProRail, the railway infrastructure manager in the Netherlands. Besides, the state-of-the-art related to railway traffic rescheduling and flexibility in railways is investigated. In addition, the research gaps are identified.

In Chapter 3, a new framework for railway operations is constructed. Besides, the core of the current research is presented, which is constructing a MILP model to add flexibility to the timetable while addressing traffic disturbances. According to the results of the toy network, the model is verified.

Chapter 4 presents the model demonstration. Given the results of the real-life case study, the performance of the model is tested. Besides, three analyses are conducted to investigate the performance of timetable flexibility and the influencing factors.

In Chapter 5, the conclusions of the current research are drawn by answering the main research question and the sub-research questions. In addition, the practical implications and some recommendations for future research are presented.

2

Literature Review

In this chapter, a literature review is conducted, including the state-of-the-practice in the railway sector in the Netherlands and the state-of-the-art regarding railway traffic rescheduling, as well as flexibility in railways. Regarding railway traffic rescheduling, the literature is reviewed based on the adopted models. Regarding flexibility in railways, the literature is reviewed based on the application stages, namely, tactical planning, operational planning, and train operations. According to the literature review, the research gaps are identified.

2.1. State-of-the-Practice

The current practices in the railway sector are discussed in this section, including the hierarchical structure of the railway planning process, as well as the stages within each planning level. Furthermore, the current practice in ProRail is discussed.

2.1.1. Planning process in railways

In the railway context, railway companies face a planning problem that consists of several consecutive planning stages. A hierarchical decision-making approach is generally adopted to address this problem (Lusby et al., 2011). Three decision-making levels are included, namely, the strategic level, the tactical level, and the operational level, which are shown in Figure 2.1.

Strategic level

In the strategic level of the planning process, two tasks are accommodated consecutively, namely, network planning and line planning. Strategic problems are driven by the estimation for the long-term traffic demand (Goossens et al., 2004). A railway network can be represented as a graph composed of nodes and edges. The nodes in this graph correspond to stations or junctions where trains interact, while the edges represent track sections that connect two nodes. In addition, the main concern of the network planning stage is the construction and/or modification of the existing infrastructure according to the estimated demand (Lusby et al., 2011). Both network design and demand data play critical roles as inputs during the line planning stage. This stage involves the determination of train lines, including their routes, stops, and frequencies (Goossens et al., 2004). The aim of line planning is to maintain efficient train connections while pursuing specific objectives, such as minimizing operational costs or maximizing passenger satisfaction (Lusby et al., 2011).

Tactical level

Given a complete network and line plan, resources are allocated over infrastructure. Three tasks are included in the tactical planning level, namely, train timetabling and routing, rolling stock scheduling, and crew scheduling.

The train timetabling problem aims at determining the periodic or aperiodic timetables for a set of trains that do not violate track capacities and satisfies some operational constraints (Caprara et al.,

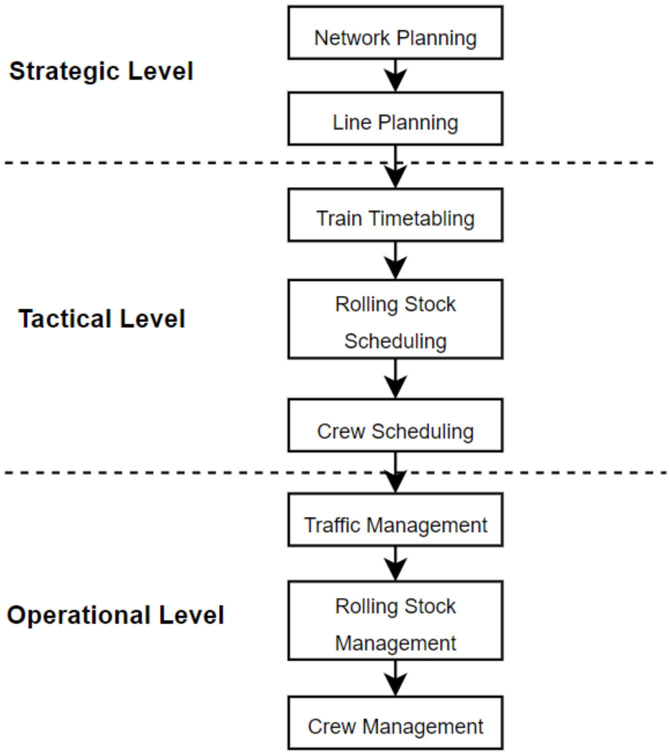


Figure 2.1: Planning process in railways

2002). Route planning and event time determination are crucial components of the timetabling process in railway tactical planning. These aspects focus on defining the specific routes that trains will follow and determining the exact departure and arrival times at various stations, junctions, and bridges along those routes (Kroon et al., 2009). Route planning involves identifying the optimal paths for trains to follow between their origin and destination stations. There are often multiple possible routes that a train can take to reach its destination. The route planning process considers factors such as track availability, capacity, train priority, and operational constraints. After the routing plan is determined, the departure and arrival times of trains at stations are determined, which are so-called events. Besides, some passing event times should also be determined such as junctions and bridges (Kroon et al., 2009). The train running times indicate the duration between a train's departure event and its subsequent arrival event. Dwell times at stations represent the duration between a train's arrival and departure events at the same station. Headway time is typically defined as the time between two train departures or passages. Minimum headway times indicate the shortest time allowed between consecutive train movements on the same track (Lusby et al., 2011). It can be defined between pairs of arrival times and departure times, as well as between arrival and departure times. In order to make the designed timetable more robust, running time supplements are added to the running times and dwell times, and buffer times are added to headway times based on experience and expert options (Kroon et al., 2009). By combining route planning and event time determination, the timetabling process in railway tactical planning creates a well-structured and optimized timetable for train services. Many objectives can be used when constructing the timetable, for instance, maximizing passenger satisfaction, maximizing timetable robustness, minimizing the operational cost, etc.

Rolling stock scheduling allocates an appropriate amount of rolling stock units from available fleets to accomplish trips in the given timetable, as well as arrange rolling stock circulations. Rolling stock circulations comprises circulations over one train line, several train lines, and complex circulations with intermediate combining and splitting. In the Dutch railway context, three conflicting objectives are considered in rolling stock scheduling by the main railway undertaking (Nederlandse Spoorwegen, NS), namely, service, efficiency, and robustness. Regarding service consideration, it ensures that as many passengers as possible can be offered a seat; efficiency aims at minimizing the amount of rolling stock and the number of rolling stock kilometers while fulfilling the expected passenger demand; in the robustness aspects, shunting movements are reduced and a line-based rolling stock circulation is adopted by NS to maintain the robustness of rolling stock schedules (Kroon et al., 2009). Line-based rolling stock circulation refers to a method of managing and deploying rolling stock in a railway network based on specific train lines. In this approach, each train line is assigned a dedicated set of rolling stock that is tailored to the specific demands and characteristics of that line.

A train driver and a number of conductors are required to operate a train, and the number of conductors is decided according to the rolling stock composition of the train (Kroon et al., 2009). The crew scheduling process involves creating optimized work schedules, known as rosters, for train crew members. In crew scheduling, a task means that a crew member of the right type and with the right qualifications must be present at the corresponding train. A duty is a sequence of consecutive tasks to be carried out by a single crew member. The rosters describe the allocation of the duties to the crew members, and they are generated by an individual rostering system in many companies (Abbink et al., 2018). Three objectives are also considered in crew scheduling by NS that includes efficiency, acceptability, and robustness. Efficiency indicates that the total number of duties is as small as possible; acceptability can be guaranteed by labor rules and company agreement; with respect to the robustness of the crew schedule, it can be affected by several elements, for instance, the slack time and relief locations (Ravichandran, 2013). In addition, novel objectives are being developed considering the trade-off between fairness and attractiveness in crew rostering (Breugem et al., 2022).

Operational level

In general, operational planning in railways can be thought of as the distribution of resources and facilities to satisfy the demand of various trains based on a predetermined timetable. Operational planning problems are defined to be those that occur on a day-to-day basis when references determined at the tactical level have to be adjusted due to unforeseen disturbances (Lusby et al., 2011), which means that timetable, rolling stock schedules, and crew schedules can be modified in the real-time operational planning. Since the aim of the current research is to investigate timetable flexibility and its consequences, the real-time modification of rolling stock schedules and crew schedules will not be discussed.

In daily operations, traffic perturbations that occur in real-time may hinder the realization of the timetable determined at the tactical level. If the perturbations cannot be absorbed by the running time supplement and buffer times added to the timetable during the timetabling stage, primary and secondary delays will be induced that will cause adverse impacts on railway operations and passenger satisfaction. In order to minimize the negative effects of unexpected incidents, online railway traffic rescheduling approaches have been introduced into the field to make up for the inability of offline timetables in dealing with real-time uncertainties. Corman and Meng (2014) reviewed papers related to online railway rescheduling and classified them into two categories, namely, static rescheduling (open-loop) and dynamic rescheduling (closed-loop). A static rescheduling process indicates that rescheduling approaches are performed only once with full information available, while in the dynamic rescheduling process, real-time information is provided to produce an up-to-date timetable.

However, apart from modifying the tactical plans, the safe execution of the updated plan is also a crucial component in railway operations. In order to discuss the operational practice in detail, the ON-TIME (Optimal Networks for Train Integration Management across Europe) framework for the real-time management of railway traffic perturbations proposed by Quaglietta et al. (2016) is used as a reference, which is shown in Figure 2.2. In the framework, a web-service event-dispatcher is designed as an interface to support the data communication between different modules, and a microscopic simulator is adopted to reproduce the real traffic.

The concept of dynamic management of railway traffic perturbations has been widely understood as a set of proactive actions with the aim of minimizing the consequences of actual delays, which comprises the supervision of traffic state, detection of deviations from the original timetable, resolution of conflicts affecting the network performance, etc (Kecman et al., 2013). These correspond to the modules illustrated in the ON-TIME framework, including Traffic State Monitoring (TSM), Traffic State Prediction (TSP), CDR, Connection Conflict Detection and Resolution (CCDR), Human-Machine Interface (HMI), Automatic Route Setting (ARS), Train Path Envelope Computation (TPEC), and Driver Advisory System (DAS).

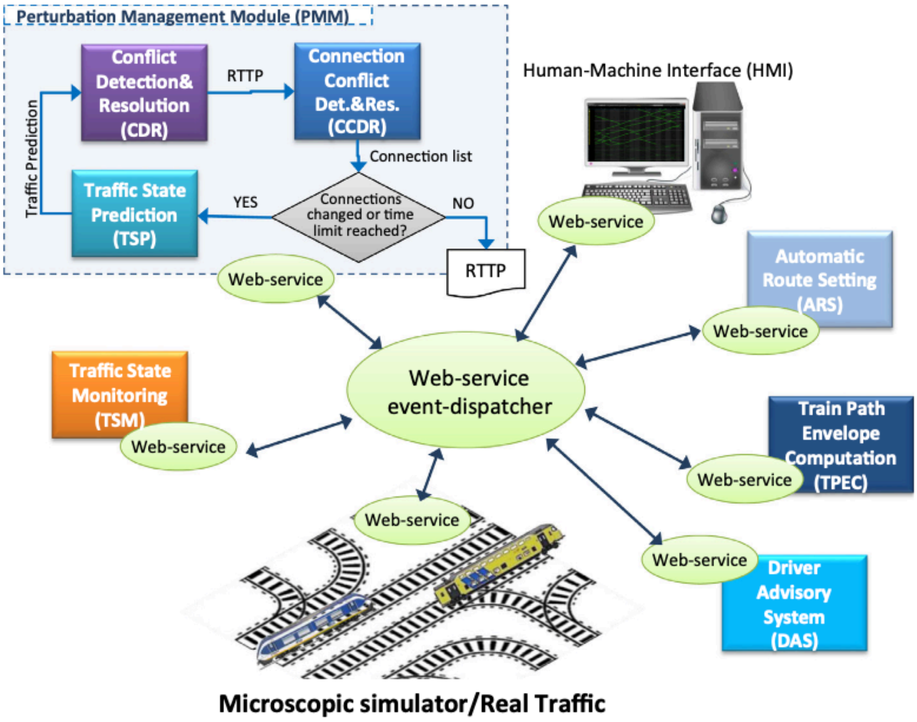


Figure 2.2: The ON-TIME framework for the real-time management of railway traffic perturbations (Quaglietta et al., 2016)

The TSM module collects data from the microscopic simulator/real traffic to estimate the current traffic state, the collected data comprise signal state, track section occupation/release, train position, and train speed (Quaglietta et al., 2016).

The TSP module, CDR module, and CCDR module form a closed-loop Perturbation Management Module (PMM), which aims to predict the traffic within the Optimization Horizon (OH) and generate updated RTTP to ensure track conflict-free and connection conflict-free train operations (Quaglietta et al., 2016). Based on the current traffic state, the TSP module forecasts the time-distance and speed-distance trajectories of trains within OH, and then the updated traffic prediction is transferred to the CDR module, in which the Conflict Detection detects the potential track conflicts within OH and the Conflict Resolution generates RTTP that comprises a set of control measures (retiming, reordering, rerouting, canceling trains, etc.). In order to minimize the impact of delays on passengers, especially those requiring transfers between trains, the CCDR module detects connection conflicts and determines whether to keep the connection or not according to the RTTP provided by the CDR. If there are no connection conflicts detected or to be dropped, the RTTP will be ready for the other modules, otherwise, the TSP generates updated traffic predictions and the CDR computes new RTTPs until no changes in connections.

The HMI module is a visualization of the detailed network topology, including platforms, switches, signals, etc, in which the RTTP or the original timetable, as well as the estimated traffic state until the current time are also presented (Quaglietta et al., 2016). Given the updated RTTP provided by the PMM, dispatchers determine whether to accept the control measures in the RTTP via the HMI.

After the eventual RTTP is decided by dispatchers, the ARS module generates the route setting plan and implements the commands in the microscopic simulator. The time to conduct the ARS is worth discussing since a late implementation that leads to trains meeting restricted signals needs to be prevented while it better be late enough to provide flexibility to change the routing plan (Quaglietta et al., 2016). In the meanwhile, the TPEC module is conducted based on the eventual RTTP. TPE comprises feasible time and speed intervals, which optimize train driving without impacting capacity (Jaekel and Albrecht, 2013). In this framework, the output of the TPEC module is time windows and/or time targets of trains that can be exploited for every track detection section to minimize energy consumption while avoiding track conflicts.

Given that TPE is computed and the route setting plan is implemented, the DAS is used in the ON-TIME framework, which enables the driver to control the train following an energy-optimal trajectory within the TPE. Apart from the DAS, manual driving is usually adopted nowadays, and the Automatic Train Operation (ATO) is mostly under development in mainline railway (Yin et al., 2017).

2.1.2. Current practice in the Netherlands

In the Netherlands, the railway sector is organized by the Infrastructure Manager (IM) and the Railway Undertakings (RUs). ProRail is the only IM and it owns the visible infrastructure, for instance, track, switches, and signals, as well as some invisible infrastructure, such as traffic control systems and underground cables. In addition, ProRail is responsible for the construction, maintenance, and management of the Dutch rail network, including all relevant facilities such as tunnels, overpasses, overhead power lines, signals, switches, and stations, on behalf of the national government. ProRail also allocates capacity on the rail network and is responsible for rail traffic control. The RUs are responsible for the rolling stock and the train crew, thereinto, NS is the largest passenger RU and operates on the mainline railway network. Apart from NS, there are other regional and private operators that run specific train services in the Netherlands, including Arriva, Syntus, and Connexxion. ProRail and NS are jointly responsible for the stations, which means that ProRail builds the stations and NS carries out daily maintenance on behalf of ProRail.

In order to show the organization of the IM and RUs, a framework for the logical domains in Dutch railways is presented in Figure 2.3. According to the framework, a four-stage organization can be observed, namely, planning, operational management, operational control, and physical operations. The planning stage corresponds to the tactical planning stages mentioned in subsection 2.1.1. The RUs are responsible for timetable design, while the IM oversees capacity allocation, and both processes are interconnected. During the planning phase, the RUs submit their timetable proposals to ProRail, detailing their desired train paths and service plans, while ProRail ensures the allocation of capacity to different RUs. Once the timetable is determined, rolling stocks and crews can be scheduled accordingly.

In daily railway operations, traffic perturbations occurring in real-time may compromise the predetermined timetable and plans, thus, operational conduct should be carried out. In the operational management stage, the responsibility of the IM is called traffic management, which aims to keep the predetermined plans for the use of infrastructure up to date. While, for the RU, it is called rolling stock management and train crew management whose objective is to update the deployment of rolling stock and train crew according to real-time traffic information. The RU can only update the rolling stock and train crew schedules after a feasible timetable is generated by the IM.

The up-to-date plans are executed in the operational control stage. For the IM, the responsibility is to set routes for the movement of trains, distract infrastructure for engineering works, and operate the movable bridges. Next is the physical operation, in which the switches, signals, and movable bridges are operated by the IM. While for the RU, the objective of both stages is to run the trains based on the coordination of the plans determined in traffic management and the physical infrastructure operation.

The determination and execution of the timetable and routing plan, as well as the infrastructure operations, are the job of the IM. The current situation in ProRail is presented in Figure 2.4. In capacity management, an integrated plan comprising a timetable and routing plan is generated by planners using DONNA, which is the planning system for railway capacity allocation. Then, the complete plan is distributed to traffic management as the reference for daily operations. The original plan is kept when no variations occur in the real traffic, however, the distributed plan is not always feasible due to traffic perturbations, hence, a plan update is necessary. Currently, a hierarchical structure is adopted in traffic management, including national traffic management and regional traffic management. The plan is kept up to date on a regional level by regional dispatchers (Decentrale Verkeersleider) via VOS (Verkeersleiding Ondersteunend Systeem), and the corresponding routing plan is actualized by signallers (Treindienstleider) via PRL (Procesleidings system). In case of larger disturbances and disruptions that affect more than one region, the national dispatchers (Verkeersleider and Planner CMBO) take the decision on timetable updates. Afterward, the plans are executed under the coordination of three systems, namely, PRL, ASTRIS (Aansturing en Statusmeldingen RailInfraStructuur), and TROTS (TRain Observation and Tracking System). Signallers use PRL to operate signals and switches and set routes for trains. ASTRIS is a highly secured system that ensures the safe operation of signals and switches. TROTS monitors the state of the trains in real time. The physical operations are conducted under the regulation of the signaling system, which includes interlocking systems and the ATB (Automatische

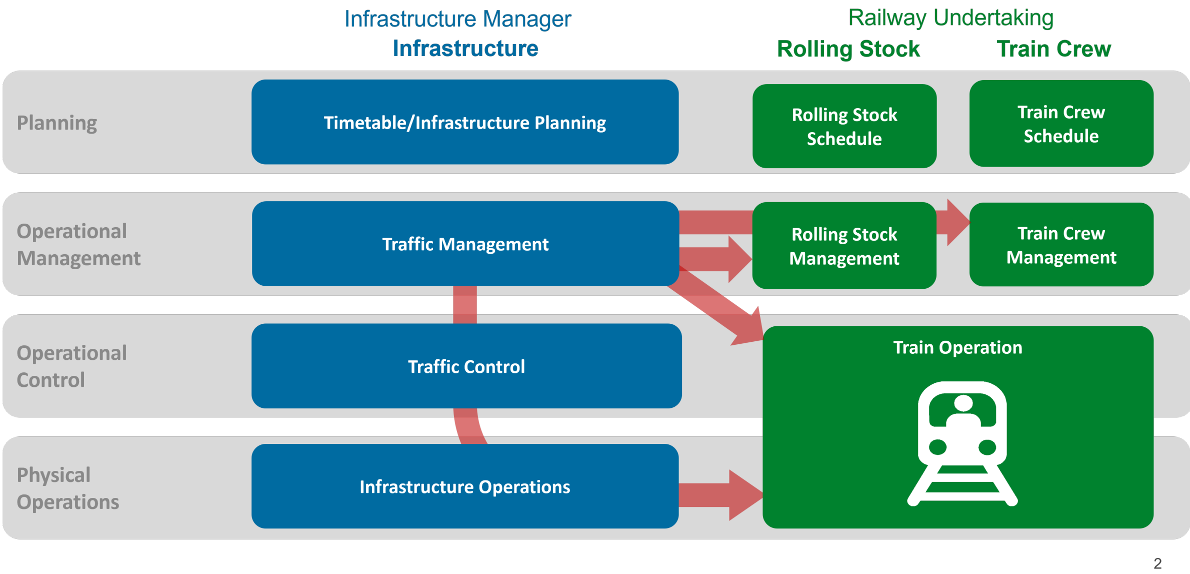


Figure 2.3: Logical domains in Dutch railways (E. P. Philipsen, personal communication, April 18, 2023)

TreinBeïnvloeding), which is the automatic train protection system in the Netherlands.

Traffic management comprises delay management and disruption management. Since we aim to add and optimize timetable flexibility in traffic disturbance management, and delay management aligns closely with the concept of disturbance management. The logical functions of delay management are shown in Figure 2.5.

In delay management, five stages are included, which have a low level of automation. In traffic state monitoring, the current traffic state is estimated based on the traffic data collected by the track detection sections, which means that delays occurring in real traffic can be reflected and observed. In traffic state prediction and conflict detection stages, future train trajectories can be predicted in the form of a time-distance diagram. By inserting the delays observed in the previous stage, potential conflicts can be visually identified in accordance with a track occupation graph. After identifying the conflicts, conflict resolution and plan optimization should be performed, in which resolution is determined according to human decisions that are made based on predefined solutions as much as possible. Then, the feasibility of the resolution is checked by colleagues if necessary. In the last stage, The adjustments may be made just in the routing plan by signallers for minor delays. While for larger delays, the timetable is adjusted by train dispatchers using VOS, and followed by routing plan adjustments by signallers.

2.2. State-of-the-art

The literature in relation to adding flexibility to the timetable in traffic management will be discussed thematically in this section. Disturbances are relatively small perturbations in railway operations that can be managed by adjusting only the timetable, while disruptions are relatively large incidents that require the rolling stock and crew schedule to be modified (Cacchiani et al., 2014). Since the aim of the flexible timetable is to better address traffic disturbances and handle real-time traffic uncertainty, disruption management is not within the scope of the current research, and only literature regarding traffic disturbance or delay management is taken into consideration. In subsection 2.2.1, the literature on the railway traffic rescheduling process under traffic disturbances is discussed, in which various models and algorithms are developed to solve the rescheduling problem. In addition, an overview of the discussed literature is also provided. In subsection 2.2.2, the concept of flexibility is discussed according to its definition in other contexts. Besides, literature regarding adding flexibility to the timetable in the railway context is reviewed. In addition, the concept of TPE is discussed, which aims to add flexibility in train operations rather than in timetabling stage or traffic management. Regarding railway traffic rescheduling, there is also research on the solution algorithms. However, the current research focuses on mathematical modeling. Therefore, solution algorithms are not reviewed.

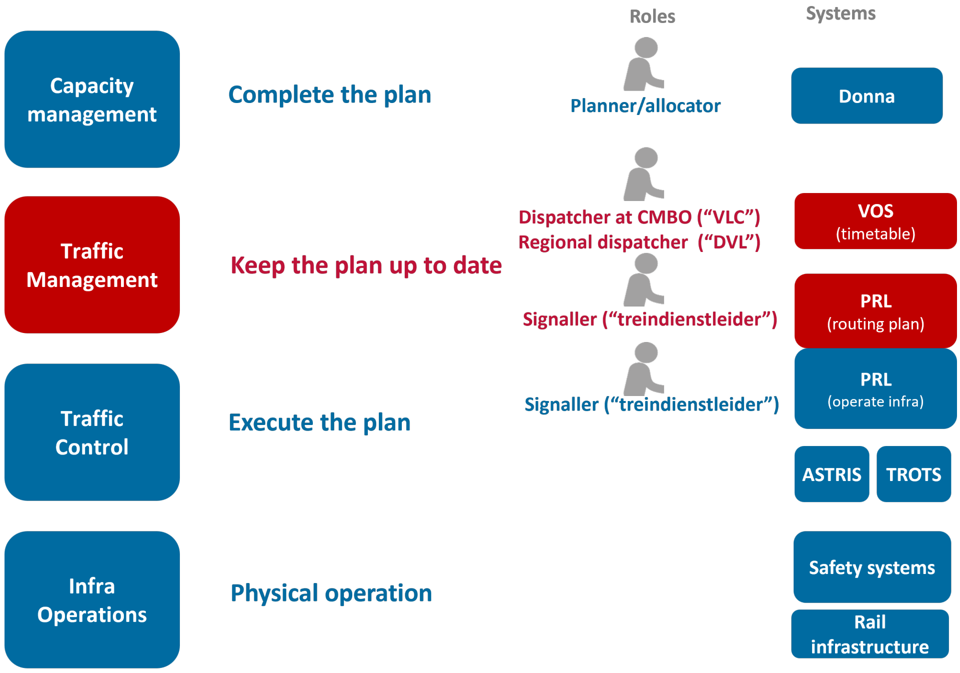


Figure 2.4: Current situation in ProRail (E. P. Philipsen, personal communication, April 21, 2023)

2.2.1. Railway traffic rescheduling models

In the railway sector, timetables that are predetermined at the tactical planning level are designed to be robust and resilient to traffic perturbations. However, it is impossible to develop perfect plans to guard against all types of perturbations, thus, railway traffic rescheduling approaches have been defined that modify the plans in order to address disturbances and disruptions that occur in real traffic.

Corman and Meng (2014) review approaches on online railway traffic rescheduling problems and exhibit the dynamic and stochastic aspects of the problem. Online traffic rescheduling falls in the operational level of the planning process in railways, in which up-to-date plans are generated during operations in a short computation time of a few seconds/minutes. They sum up the control measures used by dispatchers and classify them into time, speed, order, local route, global route, and service. The time aspect indicates shifting in planned times at reference points, which can be achieved by trains going slower or faster than planned; regarding speed, train dispatchers provide adjusted time targets to drivers to avoid conflicts or save energy; changing train orders at shared infrastructure elements is seen as a critical measure under limited infrastructure capacity; the objective of the local route aspect is to change a default route for an alternative one, such as changing platforms at stations; while in the global route aspect, different routing plans that skip stations and pass through different stations are computed to fulfill the service intention; regarding the service aspect, the often used measures are canceling trains, short turning, canceling or adding stops (Corman and Meng, 2014). They also divide the rescheduling approaches into open-loop approaches and closed-loop approaches. The former approach generates rescheduling solutions only once based on full knowledge of events happening far in the future, while in the latter approach, optimal rescheduling solutions are generated iteratively based on deviations or time steps.

Meanwhile, Cacchiani et al. (2014) present an overview regarding recovery models under traffic disturbances. They distinguish Train Timetable Rescheduling (TTR) approaches by the level of detail considered in the railway system, known as microscopic and macroscopic. At the macroscopic level, stations and tracks are represented by nodes and arcs of a graph, while block sections and signals are not taken into consideration. At the microscopic level, blocking time models and the underlying data are used to derive the running and headway times. Since various models are developed to solve the rescheduling problem, papers regarding railway traffic rescheduling under disturbances are reviewed in a methodological structure.

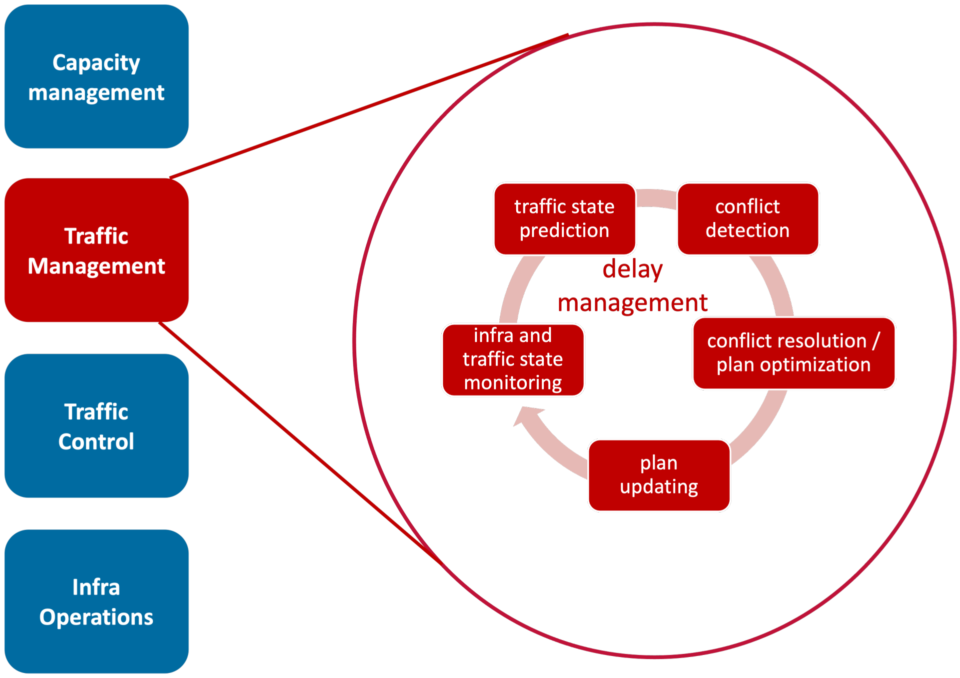


Figure 2.5: Logical functions in traffic management (E. P. Philipsen, personal communication, January 24, 2023)

Mixed Integer Linear Programming (MILP)

Törnquist and Persson (2007) discuss the propagation of disturbances and the control actions to minimize the consequences. They propose a MILP model to formulate the rescheduling problem for a railway network composed of a mix of single- to multiple-track segments. In the MILP model, continuous variables represent the occurring times, as well as the delays of events. Besides, binary variables are introduced to represent whether a track will be used by an activity, as well as the train orders. The objective functions of the model are minimizing the total final delay of the traffic, and minimizing the total final cost associated with delays. Since applying the full model requires a long computation time, four strategies are evaluated which allow different degrees of track swaps and order changes. In order to analyze the influence of disturbance characteristics and time horizons, they conduct a case study on the southern part of the Swedish railway network, taking into account delays from 5 to 40 minutes, and time horizons of 30, 60, and 90 minutes.

Dotoli et al. (2014) propose a Decision Support System (DSS) architecture for real-time management of railway networks. The DSS adopts a MILP formulation to address the rescheduling problem in a mixed-tracked network under disturbances, based on the model for n-tracked networks presented by Törnquist and Persson (2007). Dotoli et al. (2014) emphasize the importance of choosing an appropriate time horizon, since too small a time horizon limits the efficiency of the rescheduling process, while too long a time horizon induces long computation times. The objective functions used in this paper are minimizing the cost of total delay for all events within the time horizon and minimizing total delay. Afterward, a heuristic algorithm is applied to obtain near-optimal solutions, so that train dispatchers can determine the new timetable and communicate with drivers efficiently. The case study conducted on a regional railway network in southern Italy considers three time horizons from 30 to 50, and the results show that the DSS outperforms the model proposed by Törnquist and Persson (2007) in all cases.

Pellegrini et al. (2012) study the real-time railway traffic management after an unexpected event perturbs the operations. They propose a MILP formulation to solve the problem, which models the infrastructure in terms of track-circuits to achieve a fine realistic representation of the control area. This formulation considers all train routing alternatives and all rescheduling alternatives for trains along these routes, with the objective of minimizing the maximum consecutive delay. Continuous variables are the time in which a train enters a track-circuit and the delay assigned to a train in a track-circuit, while, binary variables indicate whether a train uses a route, as well as the order between two trains using a track-circuit. Several constraints are introduced to the model, namely, time concerning constraints, constraints for managing delays, constraints due to the change of rolling stock configuration, and capacity constraints. An experiment is conducted in a control area including the Lille Flandres Station (France), and the results demonstrate the good performance of the formulation in terms of computation time.

Cavone et al. (2017) investigate the self-learning decision making procedure for robust real-time train rescheduling in case of disturbances, based on a Data Envelopment Analysis (DEA) method. It can be used for aperiodic timetables of mixed-tracked networks. The overall procedure consists of three steps. First, an optimal timetable is derived using a MILP. The objective is to achieve a balance between minimizing train delays and maximizing timetable robustness across an appropriate time horizon. Next, a merging procedure is used to join the optimal timetable with the nominal one, followed by identifying and solving arising conflicts iteratively using a real-time heuristic procedure. In the final step, a cross-efficiency fuzzy DEA method is utilized to predict the results and effectiveness of alternative control actions. This step plays a critical role in updating an external database, where the most suitable solution is stored for potential use in future instances of similar disturbances. The self-learning capability of this procedure improves the quality of the rescheduling at each reapplication of the method. An experiment is conducted in southern Italy, and the results show that the procedure is effective for both railway companies and passengers.

Zhang et al. (2021) study the real-time rescheduling problem based on updated information for a single-track high-speed railway system with traffic in one direction. They construct a MILP model, where the decision variables are arrival times, departure times, arrival orders, departure orders, and dwelling plans. The objective is to minimize the deviation from the original timetable, which is expressed by four components, including positive arrival delay, negative arrival delay, departure delay, and penalty for changing the dwelling plan. Several constraints are added to the model including dwell, running, headway, overtaking, timetable, and station capacity constraints. The authors also develop a scenario-based chance-constrained model predictive control (SC-MPC) algorithm to solve the MILP model constructed for the real-time rescheduling problem. The MPC algorithm adopts a strategy of rolling optimization that repeatedly optimizes the timetable computation in real time, which reduces the computation time as well as improves the timetable robustness. The case study on the Beijing-Shanghai high-speed railway line shows that the proposed model effectively solves disturbances and the MPC algorithm increases the robustness of the output.

Integer Programming (IP)

Schöbel (2007) investigate the delay management problem which specifically aims at determining whether passenger transfers between trains should be kept or dropped, with the objective of minimizing the sum of all delays over all customers. They represent the delay management problem as an activity-on-arc project network and propose a path-based and an activity-based IP model, with the objective of minimizing the sum of all delays over all customers. An event-activity network is adopted, in which nodes denote arrival and departure events, and arcs denote the waiting and driving activities of trains and changing activities of passengers. In the path-based model, binary variables for each path are introduced to determine whether to keep all connections on paths or not. In the activity-based model, binary variables for each changing activity are used to describe if train connections are missed or maintained. Based on these, a cubic activity-based model is proposed, in which two additional variables are introduced. The additional binary variable indicates whether an activity is reached on a path without any missed connection before or not. An integer variable indicates the number of customers who use an activity. The authors proposed a never-meet property to indicate whether a case is appropriate for a linearization. They simplified the cubic model to an Integer Linear Programming (ILP) when the property holds. Constraints are added to guarantee that delays can be transferred from the start of an activity to the end so that they can be absorbed by the slack times. Besides, additional constraints are added to satisfy the minimum transfer time for passengers. The outputs of those models are timetables that minimize the total passenger delays.

On the basis of the IP formulation of the delay management problem proposed by Schöbel (2007), Schachtebeck and Schöbel (2008) add priority decisions to the IP model to make it capable of dealing with the capacitated case. The limited capacity of the railway tracks indicates that the order between two or more trains that use the same piece of infrastructure should also be determined.

Dollevoet et al. (2012) investigate the delay management problem with rerouting of passengers.

Since passengers will not wait for a cycle time to take their original routes when they miss a connection, the authors take into consideration that passengers will adjust their routes. They develop a delay management model in the formulation of IP based on an event-activity network, with the objective function of minimizing the weighted sum of realized arrival times, which is equivalent to minimizing the weighted sum of delays. Since the IP model proposed by Dollevoet et al. (2012) is hard to solve under large-scale real-world cases, Dollevoet and Huisman (2014) develop several heuristic approaches, including dispatching rules in practice, classical delay management model without passenger rerouting, and iterative model parameters updating. The experimental study based on real-time data from Dutch railways demonstrates that the iterative heuristic can solve complex instances in a short computation time.

Caimi et al. (2012) investigate the computer-aided systems to deal with the increasing challenges in railway traffic management of complex station areas. The authors propose a dispatching assistant based on an MPC framework, and the closed-loop discrete-time system adopts an IP formulation to model the rescheduling problem. Two binary variables are introduced, one indicates whether a blocking-stairway is assigned to a train, and the other one indicates whether the connection between two trains is maintained. The objective function of the model is maximizing customer satisfaction, with several constraints taken into consideration, such as clique constraints, connection constraints, and platform-related constraints. The robustness of the rescheduled timetable can be ensured by the temporal scopes, in which three parameters are introduced, namely, the rescheduling frequency interval, the fixing horizon, and the planning time horizon. The rescheduling frequency interval represents the frequency of the train rescheduling process; due to safety considerations, the routes of trains that are between the current time and the fixing horizon cannot be modified; the planning time horizon defines how far into the future trains will be considered in the current rescheduling process. The results of the case study demonstrate the potential of this approach compared to the human dispatching process, especially when rerouting is considered.

Alternative Graph (AG) Model

D'Ariano et al. (2007a) study the train scheduling problem, which aims to generate a new conflict-free timetable, with minimal deviation from the original timetable. They consider the problem as a huge Job Shop Problem (JSP) with no-store constraints, and model the problem with an AG formulation, with the objective function of minimizing the maximum consecutive delay for all trains at all visited stations. In an AG representation, a set of nodes, fixed arcs, and pairs of alternative arcs are introduced. Thereinto, nodes represent operations and are connected by directed arcs. Each arc represents a precedence relation between two nodes. A fixed arc represents the running time of a train through a block section, which corresponds to the processing time of the previous node. Since only one train can be assigned to a block section during operation at a time (Pachl, 2014), pairs of alternative arcs are used to model the possible precedence relations between two trains. The length of an alternative arc is the minimum time headway between the associated trains. Finally, the AG model assigns start times to operations, ensuring that all fixed precedence relations and exactly one of each pair of the alternative precedence relations are satisfied with no positive length cycles. In addition, they develop a branch and bound algorithm with implication rules to speed up the computation. The authors take into consideration the precedence and meeting constraints. The precedence constraint indicates that a train must wait for the arrival of the former train before departing, in case the former train carries the crew or necessary rolling stocks, and the meeting precedence indicates that a minimum dwell time should be met, during which two given trains must be together at a given station for exchanging passengers. The computational experiments carried out on a bottleneck area of the Dutch railway network indicate that the proposed algorithm generates near-optimal solutions in a short computation time.

In order to investigate the capability of conflict resolution and train speed coordination for solving real-time timetable perturbations, D'Ariano et al. (2007b) adopt a detailed AG model for the train dispatching problem, with the objective function of minimizing the maximum consecutive delay, and they mainly focus on the retiming and reordering control measures. The dispatching system comprises two modules, generating an updated conflict-free timetable, and ensuring a minimum headway between trains while keeping acceptable speed profiles. In the former module, conflicts between trains are effectively detected and solved, and a traffic plan is generated. In the second module, the blocking time model is used to check whether the headway requirements are met, and speed coordination issues are considered among consecutive trains. The interactive procedure over two modules generates a feasible speed profile for each train over the time horizon. Ultimately, a conflict-free timetable that obeys the signaling and safety constraints can be obtained. The performance of the automated dispatching system is tested on the Schiphol railway network, and better solutions are obtained compared to human dispatching.

D'Ariano et al. (2008a) describe the implementation of a real-time traffic management system ROMA (Railway traffic Optimization by Means of Alternative graphs). The authors adopt the AG formulation and propose a branch and bound algorithm for train reordering and rerouting. This dispatching system helps the human dispatcher as an effective support tool to improve train punctuality under the influence of delays and disturbances.

Corman et al. (2009) study the railway rescheduling in complex interlocking areas, in which the passage of trains through short track sections is constrained by operational rules. In this paper, the network is presented at a microscopic level. The authors adopt two AG formulations, one is based on block sections and another is based on the aggregation of block sections into station routes, with the objective of minimizing the maximum consecutive delay at each relevant point of the studied area. An aggregated block is a sequence of consecutive block sections that a train may traverse on after each other. The results of the computational study show that the aggregated representation with non-sequence-dependent setup times is a good approximation of sectional-release route locking operations, which leads to small extra buffer times.

Based on (D'Ariano et al., 2008a) and (Corman et al., 2009), in order to tackle the increased complexity of busy stations with multiple conflicting paths and high service frequency, Corman et al. (2011) investigate enhancements to the train dispatching model as well as solution algorithms. The authors also adopt aggregated blocks with the goal of reducing the number of decision variables and constraints. The main focus of this literature is changing dwell times, train orders, and train routes in the real-time dispatching process. To test the performance of the dispatching tool in terms of solution quality and computation time, a case study is conducted based on the accepted statistical distribution of train delays for Utrecht Central Station.

Corman et al. (2012) study the railway timetable rescheduling problem when disturbances perturb the timetable feasibility in the view of the train operating companies and passengers. They adopt an AG model to restore the schedule feasibility, and the objective is to simultaneously minimize the consecutive delays between trains and maximize the total value of satisfied connections. Regarding the constraint, a headway constraint based on the blocking time theory and a passenger connection constraint is introduced. In addition, two heuristic algorithms are proposed to compute the Pareto front of non-dominated schedules. The results of the computational experiment based on real-world data indicate that a compromise solution between delay minimization and connection satisfaction is crucial for advanced performance management.

Corman et al. (2014) study the timetable rescheduling problem when daily planned operations are perturbed by disturbances, in which control actions should be taken by train dispatchers to maintain the feasibility of operations and limit the delay propagation. The authors adopt the dispatching support tool ROMA and evaluate the robustness of the rescheduled plans towards potential disturbances. Regarding the timetable, two timetables are compared, namely, regular timetables and shuttle timetables. In the shuttle timetable, trains arrive at a station and turn around to start a new service in the opposite direction. The robustness of the timetables is studied as the impact of stochastic input on the key performance indicators related to trains and passengers. The computational experiment results indicate that shuttle services may be profitable when high priority needs to be assigned to some services with the drawbacks of high resources usage, management cost, and dispatcher workload.

Mazzarello and Ottaviani (2007) introduce an advanced Traffic Management System (TMS), in which two core modules are the CDR and the Speed Profile Generator (SPG). In the CDR system, an AG formulation is adopted to model real-time train scheduling and routing, since it is powerful when response time is a critical factor in evaluating the performance of an approach. The authors introduce some new constraints, namely minimum speed constraint, passing constraint, out-of-order constraint, and precedence constraint. The minimum speed constraint regulates that trains must travel at a speed higher than the minimum speed within a block section. The passing constraint indicates that a train can only pass through a node after a given time. When a block section is unavailable in a given time interval, the out-of-order constraint regulates that a train should arrive or exit a block section after the end time or before the start time of the unavailable periods, respectively. The precedence constraint directly selects one from each pair of alternative arcs at certain block sections. The authors exploit the rerouting capabilities of the network while limiting the computation time. Besides, they propose a post-processing strategy to ensure the feasibility of the plan. Based on the rescheduled plan, the main output of the SPG module is an advisory speed sent to each train whenever a change of speed is required. The pilot carried out on the Dutch railway routes indicates that advanced TMS can improve train punctuality and achieve energy efficiency, by preventing conflicts and optimally guiding train runs.

Kecman et al. (2013) dedicated to the development of macroscopic rescheduling models for railway traffic management. The authors extend the AG model to a macroscopic scale by aggregating multiple block sections into open track segments and platform tracks into timetable points. They introduce four macroscopic models to investigate the impact of the level of detail and the number of operation constraints on the performance of the rescheduling model in terms of minimization of consecutive delay and computation time. Case studies carried out on the corridor Utrecht-Den Bosch and the Dutch national railway network demonstrate that large instances can be solved with the most complex macroscopic model within reasonable computation time.

Other Models

Chiu et al. (2002) construct the train rescheduling problem as a Constraint Satisfaction Problem (CSP), in which the solution is searched from a large search space under various constraints. In the CSP model, a finite set of possible values is assigned to each variable, which is called the domain, and a finite set of constraints is constructed to regulate the possible combination of values that variables can take. Two optimality criteria are defined to measure the solution quality, including minimum-delay optimal and minimum-change optimal compared to the original timetable. The authors introduce four types of constraints, namely, scheduling constraints, stopover-maintenance constraints, modification constraints, and forward-labeling constraints. The scheduling constraints are set according to six scheduling rules, namely the speed rule, the station occupancy rule, the station entry rule, the station exit rule, the line time rule, and the stopover rule. The speed constraint limits the speed of trains traveling between two stations. The station occupancy constraint sets the time interval between two trains occupying a track. The station entry and exit constraint set time intervals between two trains' entrance to and departure from a station using one line, respectively. When two trains travel in the same direction on a line, the line time constraint forbids overtakes. When two trains run in the opposite direction on a line, it ensures that the line must be unoccupied for at least a certain time interval. The stopover constraint sets the minimum time duration that a train stays in a station. The stopover-maintenance constraints enforce that every train stays in the station no less than its original waiting time. The modification constraints enforce the modifications to original timetables to stay fixed during rescheduling. The forward-labeling constraints ensure that arrival and departure times at stations can only be delayed. In addition, two heuristics are proposed to achieve the two different optimal solutions, and the performance is demonstrated by an experiment using real-life data.

Fay (2000) investigates a dispatching support system for use in railway operation control systems, in which expert knowledge is contained in fuzzy rules of the 'IF-THEN' type. The author describes a Fuzzy Petri Net (FPN) notion that combines the graphical power of Petri Nets and the capabilities of Fuzzy Sets to model rule-based expert knowledge in a decision support system. The fuzziness exists since the conditions that have to be fulfilled for the application of a certain rule can only be specified imprecisely by the experts, and the conditions are not available precisely as well. The author rates the dispatching actions according to their utility functions. The utility function involves the delay of all affected trains weighted per rank of trains, the summed delay for all passengers, and the costs of additional resources. The case study shows that a knowledge-based expert system can improve traffic quality and reduce operation costs.

Nowadays, learning-based approaches are a research hotspot in solving traffic management problems, which can be considered as an extension to the knowledge-based approaches. Since the currently used rescheduling approaches are limited due to the strong reliance on predefined knowledge, Šemrov et al. (2016) present a train rescheduling method based on reinforcement learning, more precisely a Q-learning algorithm. The basic components of the Q-learning algorithm are environment, agent, states, actions, and reward. The learning agent aims to learn a policy that effectively reduces the total delay resulting from an initial disturbance. By actively exploring the environment, the agent interprets the received reinforcement signal. Through accumulated experience, the agent strategically chooses actions aimed at maximizing the sum of rewards. The empirical results indicate that the Q-learning algorithm generates superior or comparable rescheduling solutions compared to those of non-learning-based rescheduling methods.

Zhu et al. (2020) propose a reinforcement learning-based timetable rescheduling method and also adopt the Q-learning algorithm. This method entails offline learning of timetable rescheduling techniques, enabling instantaneous application of optimal dispatching decisions according to the current state of the railway environment. The computational experiment shows that the solution quality and convergence are affected by the state representation of the railway environment. However, it is verified that this method generates superior solutions within limited training episodes.

Apart from reinforcement-based methods, Huang et al. (2023) propose a data-driven train model that exploits past realization data to provide decision support for traffic control. This model better involves the understandability of the practitioners compared to prescriptive techniques, such as mathematical programming and heuristics. Meanwhile, they adopt decision graphs to identify which control action leads to the best solution, in terms of reduction of delays, based on the past performance of the same action in similar conditions. The authors mainly focus on determining the order of two consecutive trains in merging stations, and the computational results indicate that the presented model can identify solutions that lead to the largest delay reduction.

2.2.2. Flexibility in railways

In the dictionary of business and management, the explanation of flexibility is the ability to adapt an operating system to respond to changes in the environment (Jonathan, 2009). In transportation systems, Morlok and Chang (2004) define flexibility as the ability of a system to adapt to external changes while maintaining satisfactory system performance. Flexibility is increasingly desired in transportation systems in order to accommodate changing demands and traffic patterns, however, in the railway context, only a few researchers investigate different levels where flexibility can be introduced.

Tactical planning

Caimi et al. (2011) address the problem of generating conflict-free periodic train timetables for large railway networks. The timetable is constructed at the tactical planning level, based on train service intentions, such as train lines, stop stations, and interconnection possibilities. the Periodic Event Scheduling Problem (PESP) is the classical model for generating periodic railway timetables. At a macroscopic level, a periodic railway timetable comprises departure and arrival times at stations for all trains running in the network within a period. Each departure and arrival of a train at a station is called an event. The dependencies between events are modeled as constraints, for instance, trip time constraints, headway constraints, dwell time constraints, and connection constraints. The authors propose a Flexible Periodic Event Scheduling Problem (FPESP) model, which is an extension of the PESP, to generate periodic timetables with event slots. In FPESP, the decision variables are lower bounds of the events and the corresponding flexibility. Four objective functions are proposed including minimizing a weighted sum of the passenger-relevant times, maximizing a weighted sum of flexibility, simultaneously minimizing travel time, and maximizing flexibility, etc. In addition, they develop a FLEXBOX model, in which a flexible timetable is determined at the macroscopic level, and a feasible routing plan and platform assignment is determined at the microscopic level. The results of the case study indicate that the FLEXBOX model increases the total flexibility of the timetable, and a flexible timetable provides larger solution space for microscopic solutions, which leads to better solution quality. Besides, the FPESP model only moderately increases the computation time compared to conventional PESP.

Operational planning

Under the trend of managing congested areas by planning less in the off-line phase and solving train conflicts in real time, D'Ariano et al. (2008b) investigate the concept of flexible timetable to improve the timetable robustness. In this literature, they focus on determining time windows of arrival/departure times, a partial order of trains at overtakes and junctions in off-time timetable planning, and the exact train arrival/departure times at critical points, as well as the exact order of train at overtakes and junctions are determined in real-time traffic management. The authors adopt the alternative graph formulation to model the Conflict Resolution Problem (CRP), with the objective function of minimizing the maximum consecutive delay. In addition, they develop three greedy heuristics and a branch and bound algorithm to solve the model, which are tested on a bottleneck area of the Dutch railway network. In the computational experiment to analyze different CRP algorithms, 5 values of flexibility ranging from 0 to 120 seconds are assigned to each of the 60 perturbation schemes, and the results show that the branch and bound algorithm outperforms other algorithms in terms of maximum total delays and average delays, and increasing flexibility helps decrease the delays. In another experiment that compare flexible timetables with different buffer time distribution, 7 values of flexibility ranging from 0 to 180 seconds are assigned to each of the 60 perturbation schemes, and the results indicate that flexible timetables are preferable compared to the rigid ones and flexibility offers more freedom to solve conflicts and to increase punctuality without decreasing the throughput.

Train operations

The concept of train path envelope is first proposed by ON-TIME (2013), which is used to manage and optimize train paths in railway networks, thus improving capacity and efficiency of railway networks. The TPE comprises feasible time and speed intervals, and it defines the maximum and minimum allowed speeds for a train on a specific segment of track, based on the geometry and signaling characteristics of the track, as well as the operational requirements of the train. Within the range provide by the TPE, train driving can be optimized without impact on capacity (Jaekel and Albrecht, 2013), thus, the TPE can be considered as adding flexibility in train trajectories. The railway infrastructure can be regarded as a series of block sections, and the blocking time theory is important for ensuring the safety and reliability of railway operations, as it helps to prevent train collisions and other incidents that can disrupt train services and cause delays (Pachl, 2014). For a given train, the TPE is computed by taking into account the buffer times with adjacent trains, which requires the computation of the blocking times of all trains (Quaglietta et al., 2016).

In order to integrate the Conflict Resolution System (CRS) and the DAS, Jaekel and Albrecht (2013) present the concept of the TPE, and refers to the sequence of target time windows for consecutive locations along a train run. This literature focuses on how target windows for DAS can be computed based on the results of CRS. The authors adopt a non-linear constraint programming for constructing the TPE, and the computation of the TPE is achieved following two steps, the first is generating a drivable path that minimizes the energy consumption, and the second is allocating the available buffer times to the train paths. The computational experiment indicates the feasibility and potential of the model in computing the TPE.

Wang and Goverde (2016) study the train trajectory optimization problem, which computes a feasible train speed profile, taking into account the operational constraints and signaling constraints. This paper formulates the real-time traffic plan as a set of TPEs, and the authors adopt a multiple-phase optimal control model, which is able to capture varying gradients and speed limits, as well as constraints provided by the TPE. The TPE consists of target points and target windows, which are defined by triples of position, time, and speed information. In case of disturbances, trains may recover from delays and get back to the planned timetable by regulating their speed within the range provided by the TPE.

Albrecht and Dasigi (2016) present the overall approach of the ON-TIME project and the approaches to real-time perturbation management and real-time train speed regulation. Besides, Quaglietta et al. (2016) introduce a framework for the automatic real-time management of railway traffic. In real-time traffic management, the TPEC module is introduced between the computation of the RTTP and its implementation to train operations, which aims to identify energy-efficient train speed profiles compatible with the RTTP, by investigating the buffer times that can be exploited. A TPE must be computed for every train in the area and any available RTTP, and it serves as the input to all computations related to train speed control.

Wang et al. (2023) investigate the conflict-free train path planning in Automatic Train Operation (ATO) over the European Train Control System (ETCS). They propose a Train Path Slot model in the formulation of a linear programming model, which considers three driving strategies, namely, Energy-Efficient Train Control, Reduced Maximum Speed, and Minimum Time Train Control. They optimise the location of Timing Points (TPs) and their associated time targets or windows, which are eventually assembled in a TPE. This research bridges the IM and RUs by providing time targets or feasible time windows in the form of TP constraints at critical locations.

2.3. Research gaps

In the literature regarding railway traffic rescheduling under disturbances, different research scopes are considered by the authors, such as the level of detail and case study on corridor/network, in addition, various models, objective functions, control measures, and constraints are adopted. In order to sum up the papers regarding railway traffic disturbance management, a literature overview is presented in Table 2.1.

According to the literature review, only a limited amount of literature is found dealing with adding flexibility in railways. TPE can be considered as adding flexibility in the perspective of individual train. However, adding flexibility to the timetable is at the operational planning level, which is from the perspective of the traffic in the network.

The research gaps can be identified as follows. First, among the literature discussing timetable flexibility, there is no consensus on its definition. Timetable robustness is the ability of a timetable to withstand design errors, parameter variations, and changing operational conditions. Timetable resilience is the flexibility of a timetable to prevent or reduce secondary delays using dispatching measures. However, most literature investigating timetable resilience focuses on traffic disruption management rather than traffic disturbance management. Therefore, we define timetable flexibility as the ability to be easily modified to withstand small disturbances and absorb delays, as well as to offer a larger solution space in the application of control measures to solve larger disturbances without changing the given (re)scheduled timetable.

Second, the FPESP model is designed for tactical planning, which investigates the allocation and trade-off between capacity and flexibility. The PESP model is often designed for long-term planning and optimization, which can involve complex mathematical formulations and computationally intensive algorithms. Therefore, the computation time of the FPESP model is too long for real-time rescheduling. Besides, PESP is well-suited for scenarios where events repeat at regular intervals. However, real-time railway rescheduling involves efficient responses to dynamic and unpredictable disturbances, such as varying passenger volumes, different driver behavior, and changing weather conditions. This means that the rescheduled plans are not necessarily periodic. In a nutshell, the FPESP model is not suitable in the current research.

By comparison, the AG model is a widely used model to generate rescheduling solutions within tight time constraints, which aims to handle real-time traffic disturbances. In the AG formulation constructed by D'Ariano et al. (2008b), they make the scheduled departure times earlier than the rigid departure times and publish a new timetable to passengers, by which flexibility is actually an early departure compared to the rigid timetable. However, this is unfair to passengers who stick to the rigid published timetable since the rescheduled departure times are actually early departures. Besides, flexibility is a parameter rather than a decision variable, which means that the values of flexibility are assigned to test the impacts of different algorithms on flexibility, as well as the impacts of flexibility on different levels of buffer time. Therefore, they are not aiming at optimization timetable flexibility. In addition, in the computational experiment, the maximum arrival times are fixed to the rigid arrival times, thus, flexibility is added only to the departure events. Furthermore, the experiments are conducted on the railway corridor in the Schipol dispatching area, thus, connection constraints that indicate the transfer requirements of passengers between trains at stations are not taken into account.

Table 2.1: An overview of literature with common railway rescheduling models used

Authors(s)	Model	Objective function (s)	Control measures	Connection constraint	Level(s)	Case study
Tornquist and Persson	MILP	minimize total final delays /minimize total cost of final delays	retiming, reordering	x	macroscopic	network
Dotoli et al.	MILP	minimize the cost of total delays /minimize total delays	retiming, reordering		macroscopic	corridor
Pellegrini et al.	MILP	minimize the maximum secondary delays	retiming, reordering, rerouting	x	macroscopic	network
Cavone et al.	MILP	simultaneously minimize train delays and maximize timetable robustness	retiming, reordering		macroscopic	corridor
Zhang et al.	MILP	minimize the deviation from the planned timetable	retiming, reordering		macroscopic	corridor
Schobel	IP	minimize total passenger delays	retiming	x	macroscopic	network
Schachtebeck and Schobel	IP	minimize total passenger delays	retiming, reordering	x	macroscopic	network
Dollevoet et al.	IP	minimize total passenger delays	retiming, reordering	x	macroscopic	network
Caimi et al.	IP	maximize passenger satisfaction	retiming, rerouting	x	microscopic	network
D' Ariano et al.	AG	minimize the maximum secondary delays	retiming, reordering		microscopic	corridor
D' Ariano et al.	AG	minimize the maximum secondary delays	retiming, reordering		microscopic	corridor
D' Ariano et al.	AG	minimize the maximum secondary delays	retiming, reordering		microscopic	corridor
D' Ariano et al.	AG	minimize the maximum secondary delays	retiming, reordering, rerouting	x	microscopic	corridor
Corman et al.	AG	minimize the maximum secondary delays	retiming, reordering		microscopic	corridor
Corman et al.	AG	minimize consecutive delays	retiming, reordering, rerouting	x	microscopic	corridor
Corman et al.	AG	simultaneously minimize consecutive delays and maximize total value of satisfied connections	retiming, reordering	x	microscopic	corridor
Corman et al.	AG	minimize average total delays /minimize passenger travel times	retiming, reordering	x	microscopic	network
Mazzarello and Ottaviani	AG	minimize the maximum secondary delays	retiming, reordering, rerouting	x	microscopic	corridor
Kecman et al.	AG	minimize the maximum secondary delays	retiming, reordering	x	macroscopic	corridor
Chiu et al.	CSP	minimize the maximum secondary delays /minimize the number of modified station visits	retiming		macroscopic	corridor
Fay	knowledge-based	minimize a utility function (train delays, passenger delays, additional resource costs)	retiming, reordering, rerouting	x	microscopic	corridor
Šemrov et al.	reinforcement learning-based	minimizing total train delays	retiming, reordering		microscopic	corridor
Zhu et al.	reinforcement learning-based	minimizing total train delays	retiming, reordering		microscopic	corridor
Huang et al.	data-driven	minimizing total train delays	retiming, reordering		microscopic	corridor
Current research	AG	minimize the deviation from the rigid timetable and maximizing total timetable flexibility	retiming, reordering	x	microscopic	network

3

Methodology

In the current chapter, the methodology for constructing and solving the mathematical model, as well as analyzing the results, are described. In Section 3.1, a framework for real-time railway operations is proposed based on the state-of-the-practice in Dutch railways, and detailed discussion is made on the components within the framework, after which the exact positioning of the current research is identified. In Section 3.2, the scope of our research is stated and justified. In Section 3.3, the pros and cons of the model that will be used in this research are discussed, and some actions and scopes are identified to limit the impact of model flaws. In Section 3.4, the principle of the AG model is described and an illustrative example is presented. Besides, the MILP model and its application in modeling the train rescheduling problem are discussed. In Section 3.5, the formulation of the AG-based MILP model is presented, including the indices, sets, parameters, decision variables, objective function, and constraints. In addition, explanations of the objective function and constraints are also provided. In Section 3.6, the general model inputs are described. Besides, the Gurobi Optimizer is introduced, as well as the algorithms and techniques used in this commercial solver. Section 3.7, the toy network layout, and the trains running on the toy network are introduced. In order to verify the model, two illustrative examples including the scenarios with and without initial delays are presented.

3.1. Railway operations framework

According to the discussion on the ON-TIME framework for real-time management of railway traffic perturbations and the interview on the current practice in ProRail, a framework for real-time railway operations is proposed, which is shown in Figure 3.1. The modules form a closed-loop structure together with real traffic and can be distinguished into three stages, namely, traffic management, traffic control, and train operations. The functions of the modules and their interactions are explained below on the basis of the stage at which they are positioned. Our main focus is the CDR process in traffic management, which belongs to the operational planning instead of the execution part.

Traffic management

The main purpose of traffic management is to generate the RTTP, which guides the implementation of modules in traffic control and train operations. Traffic management comprises five modules, namely TSM, TSP, CDR, Rescheduling Decision, and RTTP Generation.

The TSM module is the interface between real traffic and traffic management, it monitors the movement of trains, track conditions, and signaling systems so that the current traffic state can be estimated and delays that occur in real-time can be observed. The current traffic state is then transferred to the TSP module, in which the future conditions and behavior of railway traffic within the optimization horizon are forecast.

By comparing the future traffic state with the rigid timetable, potential track conflicts can be identified and resolved in the CDR module. If no conflicts are detected, the original plan is maintained, which can be the rigid timetable or the last RTTP. Otherwise, a new rescheduling plan, which aims to solve the identified conflicts and address all possible new conflicts while optimizing certain objectives, is

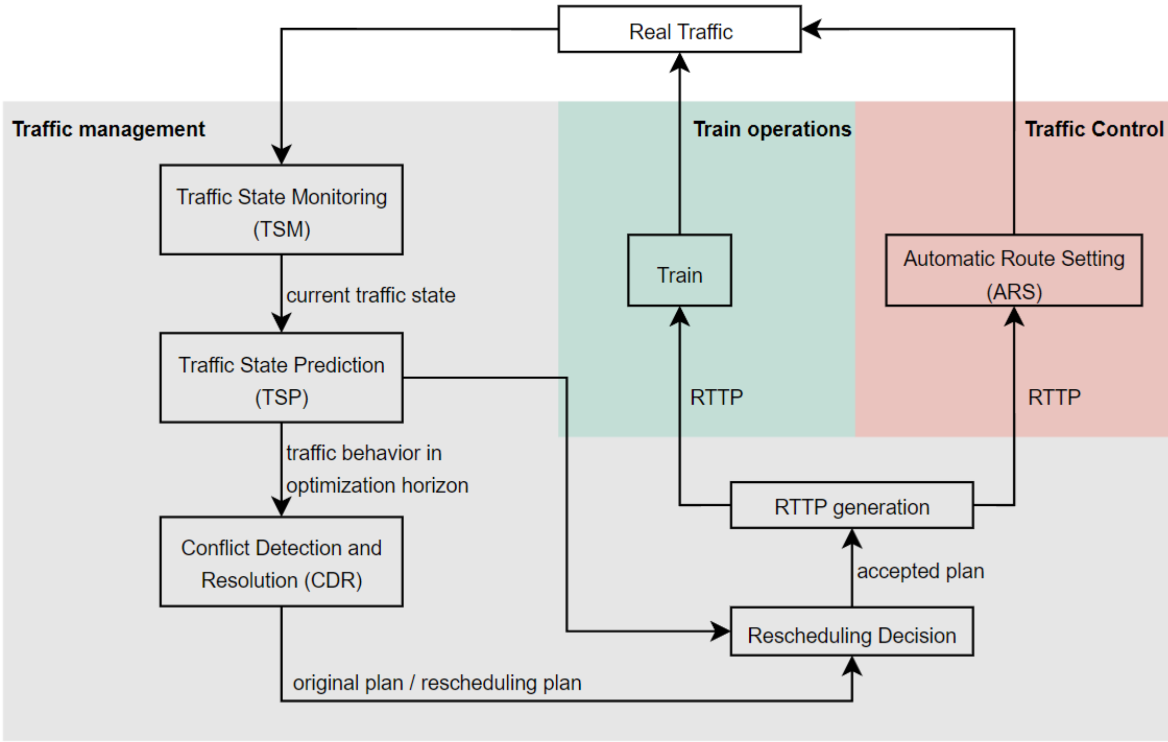


Figure 3.1: Framework for real-time railway operations

generated and provided to the rescheduling decision module. In the current practice of ProRail, the rescheduling plan is a plan that has the least deviation from the rigid timetable, which comprises exact event times. By adding flexibility, the rescheduling plan includes event time slots.

In the Rescheduling Decision module, the rescheduling plan is compared with traffic prediction, after which train dispatchers and signalers determine whether or not to accept the rescheduling plan. In the RTTP Generation module, if the rescheduling plan is accepted, a new RTTP will be generated accordingly. Otherwise, the last RTTP is maintained. In the current practice, the RTTP is the same as the rescheduling plan if accepted. However, when flexibility is added on top of the rescheduling plan, a conflict-free RTTP can be obtained by picking departure and arrival times from the event time slots.

Traffic control and train operations

Traffic control is all about executing the RTTP generated in traffic management, which guarantees the safe implementation of the plan. The function of the ARS module is to automate the process of setting train routes on a railway network. Train operations are about the efficient running of individual trains. The determined RTTP is fed to the train. Real traffic is a combination of traffic control and train operation.

3.2. Assumptions justification

In the current research, we assume that the NS'54/ATB signaling and train protection system is used. With respect to dispatching measures, only retiming and reordering are considered, while rerouting is not within our assumption.

The Netherlands has been progressively implementing ETCS as part of its commitment to the European Rail Traffic Management System (ERTMS). ETCS Level 1 and Level 2 have been rolled out on various rail lines to improve interoperability across European rail networks. NS'54/ATB is the legacy signaling and train protection system, which has been in use in the Netherlands for several decades and has undergone various versions and upgrades. These systems aim to enhance safety, efficiency, and interoperability in rail operations.

The transition from NS'54/ATB to ETCS aligns with long-term industry trends and future-proof railway systems. It enables seamless and integrated train operations across Europe. Standardizing on ETCS allows for a consistent train control system, reducing the need for multiple signaling systems and promoting harmonization within the European railway network. Besides, ETCS offers advanced safety features and functionalities that can enhance overall railway safety. In addition, ETCS can contribute to increased capacity on rail networks through improved signaling and communication. The system enables more efficient train operations, optimized speed profiles, and better utilization of track infrastructure. In the case study of this research, NS'54/ATB data are used. However, the method can be generically applied to ETCS.

Regarding the CDR, three dispatching measures are usually used in railway traffic disturbance management, namely, retiming, reordering, and rerouting. Retiming involves adjusting the departure and arrival times of trains to create a feasible timetable. It can help alleviate conflicts and reduce delays without significant route changes. Retiming can be implemented relatively quickly and is suitable for managing disturbances. Reordering involves rearranging the sequence of trains within a timetable to minimize delays and optimize resource allocation. Rerouting refers to the process of redirecting trains from their originally planned routes to alternative routes. Railway networks are intricate systems with numerous tracks, junctions, and stations. Finding optimal rerouting solutions that consider all possible alternatives is computationally intensive. Besides, incorporating local or global rerouting can have significant impacts on the result of our model. Other dispatching measures, namely canceling trains, short-turning, and canceling or adding stops, are often used when disruptions occur, thus, it is unnecessary to use these measures in traffic disturbance management.

By comparison, retiming and reordering are relatively practical measures while rerouting is not a primary choice due to its complexities. In this research, we just aim to showcase how flexibility can be added to timetable events, therefore, rerouting is not in our assumption. However, the model can be generalized to any rescheduling measure that can be realized in an AG model.

3.3. Model justification

In the literature regarding railway traffic rescheduling, the conflict resolution problem is usually modeled as a mathematical optimization problem. Regarding adding flexibility to the timetable, FPESP is a model proposed to generate flexible time slots for departure and arrival times instead of exact times. It is adopted at the tactical planning level, and its output is a periodic timetable published to passengers. However, our main focus is real-time railway traffic rescheduling at the operational planning level, which means the solution is not necessarily periodic. In addition, FPESP is a macroscopic model, which means a simplified track topology is used and a microscopic model needs to be constructed to test the macroscopic solution. Therefore, FPESP is not suitable for the current research.

Considering our research objective and scope, We plan to develop a rescheduling model based on an AG (Mascis and Pacciarelli, 2002) but reformulated as a MILP. The possible benefits of an AG-based MILP model are presented as follows.

- **Clear representation**: The railway rescheduling problem involves numerous interconnected elements, including trains, block sections, stations, and operational constraints. The AG-based model offers an intuitive and clear representation that captures these elements and their relationships. By representing the railway network as a graph, where nodes represent the start of processes and arcs represent the relationship between the processes, the model can comprehensively and effectively represent the topology of the network and the traffic within.

- **Microscopic representation**: The AG representation naturally models the railway traffic rescheduling problem at a microscopic level, which allows for a detailed and accurate representation of the railway system. It captures individual train movements, infrastructure layout, signal systems, and other operational constraints more precisely. Therefore, it can identify opportunities for optimizing resource allocation and reducing conflicts.

- **Objective variety**: Rescheduling decisions often involve trade-offs between various conflicting objectives, such as minimizing delays and maximizing passenger satisfaction. In an AG model, the objective is minimizing the maximum consecutive delays. However, a MILP formulation can incorporate multiple objectives by defining them as optimization criteria with corresponding weights, which allows decision-makers to analyze and explore the trade-offs involved and generate rescheduling solutions that align with their desire.

However, the AG-based MILP model also has some drawbacks in its application, which are shown below.

- **Computation complexity**: Due to the microscopic representation, the AG-based MILP model can become computationally extensive in large-scale railway networks. Therefore, the time required to solve the model may become impractical in real-time applications.
- **Limited scalability**: Developing an AG-based MILP model for railway rescheduling requires careful consideration of the intricacies and constraints. Constructing an accurate and comprehensive graph representation and model formulation that adequately represents the various aspects of the railway system can be challenging, not to mention extremely large railway networks.
- **Inherent simplifications**: The AG-based MILP model represents a simplified abstraction of the complex reality of railway rescheduling. Certain assumptions or simplifications, such as fixed travel times, static capacities, or idealized resource availability, may be necessary for model tractability. However, these simplifications can introduce deviations from the real operational conditions, potentially impacting the model's accuracy and real-world applicability.

For the computation complexity, it is a common problem of MILP models. Thus, using the AG formulation will not impose significant impacts on the computation time compared to other MILP formulations. Besides, the AG-based MILP model becomes more efficient when rerouting is out of scope. Regarding scalability, a toy network can be designed to verify the model and a relatively small network can be used to conduct a case study. In addition, there are attempts that try to connect different regions of a large network. Thus, the scalability problem can be efficiently solved using other techniques. Regarding the deviations from real-life operations, the microscopic topology used in an AG has already limited the occurrence of deviations as much as possible.

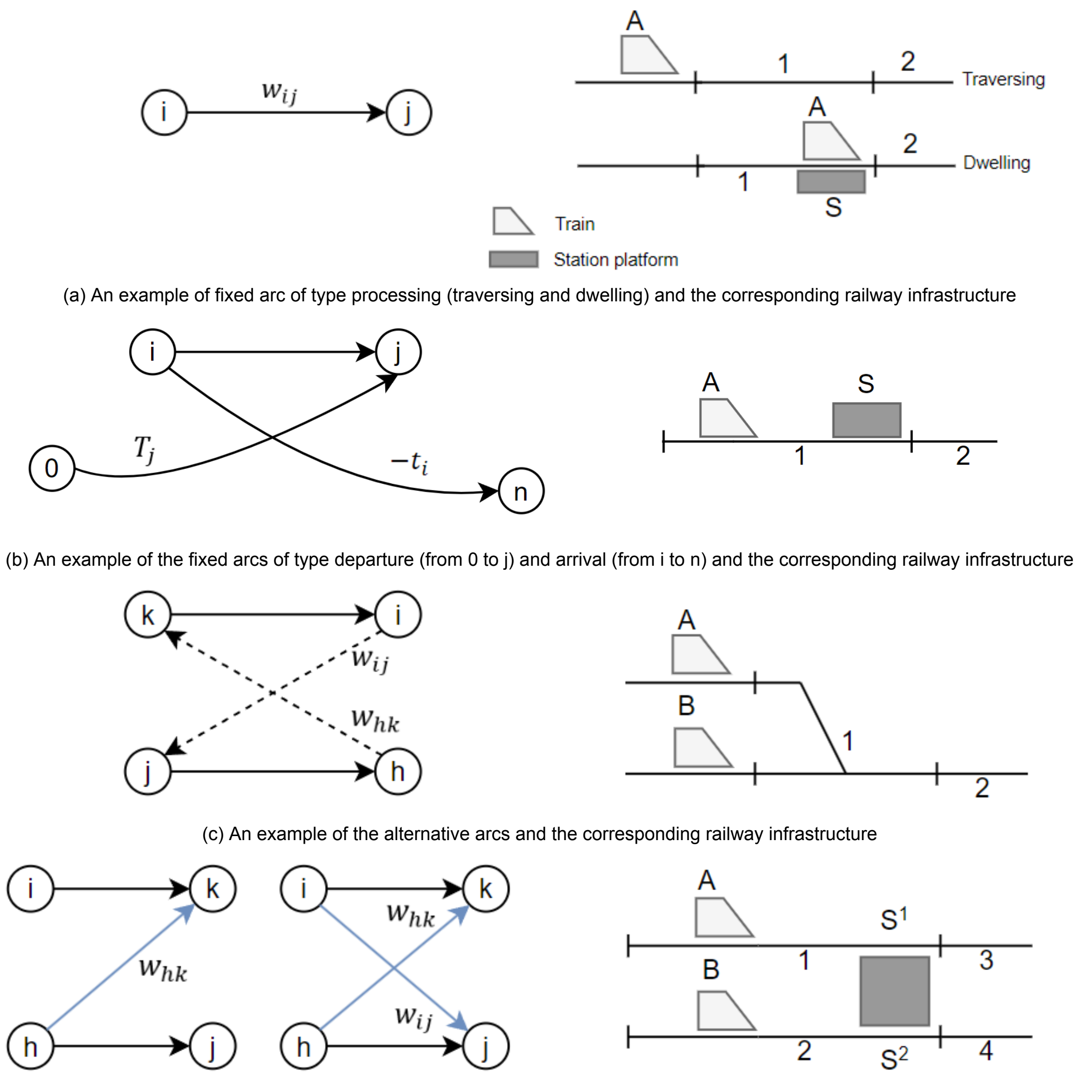
3.4. Alternative Graph-based Mixed Integer Linear Programming model

3.4.1. Principle of Alternative Graph formulation

In the AG formulation, the model is represented as a graph with nodes and arcs. Each node is associated with a time instant, indicating the starting of an operation, and the operation can be a train entering a block section or station. An arc represents the relation between two nodes. All arcs are directed arcs in an AG formulation, thus, the order between two events is inherently indicated. Basically, three types of arcs are included, namely, fixed arcs, alternative arcs, and connection arcs. Thereinto, fixed arcs can be distinguished into processing arcs, timetabling arcs, and arcs that represent entering and leaving the dispatching area. In order to present the involved arcs in a clear manner, a summary of the arcs used in an AG model is shown in Figure 3.2.

In 3.2a, a fixed arc and the corresponding railway infrastructure are presented. A processing arc implies the traversing of a train through a block section, or the dwelling of a train at the station. In the first case, train A traverses block section 1, and then, enters block section 2. Node i indicates train A entering block section 1, and node j indicates train A entering block section 2. w_{ij} is the weight of the processing arc, indicating the minimum running time of train A in block section 1. If node i represents the arrival at a station, w_{ij} is weighted by the minimum dwell time at the station.

In order to incorporate the rigid timetable in an AG, as well as to model trains entering and leaving the dispatching area, dummy node 0 and node n are added. Node 0 is called the start node, which represents the entrance of trains into the dispatching area, at which the starting time is often set to 0. Node n is called the end node, which is used to model the maximum consecutive delays over all trains when they leave the dispatching area. In Figure 3.2b, a pair of timetabling arcs and the corresponding railway infrastructure are presented. A pair of timetabling arcs comprise a departure arc and an arrival arc. In this example, node i is an arrival node, which indicates a train arrives at a station. Besides, node j indicates a train departs from the station, which is equivalent to entering the block section after the station. T represents the rigid time of a node and t is the time instant of a node. The departure arc is an arc from node 0 to a departure node, which is weighted by the rigid departure time T_j from the station. It means that trains cannot depart earlier than the rigid departure times. The arrival arc is an arc from an arrival node to node n, which is weighted by $-t_i$.



(a) An example of fixed arc of type processing (traversing and dwelling) and the corresponding railway infrastructure

(b) An example of the fixed arcs of type departure (from 0 to j) and arrival (from i to n) and the corresponding railway infrastructure

(c) An example of the alternative arcs and the corresponding railway infrastructure

(d) An example of the connection arcs and the corresponding railway infrastructure

Figure 3.2: A summary of the arcs in an AG

Apart from processing arcs and timetabling arcs, there are arcs that represent entering and leaving the dispatching area. Entry arcs are arcs from node 0 to the first nodes of each train and exit arcs are arcs from the last nodes of each train to node n.

An example of two trains approaching a junction and the alternative arcs are shown in Figure 3.2c. Alternative arcs are used to represent possible train orders at shared resources. In this example, train A and train B simultaneously require the same junction block section 1. In the AG representation, nodes k and i indicate train A entering block sections 1 and 2, respectively. Nodes j and h indicate train B entering block sections 1 and 2, respectively. The intersecting dashed arcs are called a pair of alternative arcs, and only one in the pair can be chosen, indicating the precedence relationship between two trains entering the shared resource. Choosing the alternative arc from node i to node j implies that train B uses block section 1 after train A while choosing the arc from node h to node k implies the opposite sequence. The weight of an alternative arc ensures that the following train can only enter the shared resource at a certain time after the preceding train has fully entered the next block section. w_{ij} represents the minimum required time between train A exiting and train B entering block section 1. w_{hk} represents the minimum required time between train B exiting and train A entering block section 1.

Connection arcs are used to model passenger transfers at stations. Passenger transfer in railway refers to the process of transferring passengers from one train to another within a railway station, and it typically occurs when passengers have to change trains to reach their desired destination. Basically, there are two types of passenger transfers, namely asymmetric and symmetric transfers. Asymmetric transfers indicate transfers from one train to another and symmetric transfers indicate transfers between two trains in both directions. In Figure 3.2d, two simple visualizations are presented.

In this visualization, node i and node j indicate train A and train B enter different platforms in a station, respectively, and nodes h and k imply the departure of two trains from the station. The blue-directed arcs connecting the arrival of a train and the departure of another train are considered connection arcs, and the weight of the connection arc indicates the minimum required transfer time between two trains at the station. If the connection holds, it means that train A can only leave the station w_{jh} time interval after the arrival of train B at the station, meanwhile, train B can only leave the station w_{ik} time interval after the arrival of train A at the station. The difference between connection arcs and alternative arcs is that connection arcs can be kept or discarded simultaneously. ProRail has a document named TAD (Treindienst Afhandelings Document), in which the conditions that designed passenger transfers have to be discarded are specified. In practice, the passenger transfer is designed between two trains. A feeder train is a train that brings passengers to the transfer station. A waiting train is a train that waits for the passengers on the feeder train at the transfer station. The determination of whether to maintain or discard a passenger transfer is based on two standards. For a passenger transfer, there is an arrival delay measurement point at a specific station located upstream. If the arrival delay of the feeder train exceeds the threshold, the transfer has to be discarded. Besides, there is also the latest departure time requirement. If the departure time of the waiting train from the transfer station exceeds the threshold, the transfer has to be discarded. In other words, the designed passenger transfer can only be maintained when both thresholds are respected.

In order to incorporate three types of timetable flexibility into our model, namely departure flexibility, early arrival flexibility, and late arrival flexibility, some improvements are made on top of the conventional AG. An illustrative representation of three types of timetable flexibility is presented in Figure 3.3. In this illustrative example, nodes i and j represent the arrival and departure node of a train, respectively, and w_{ij} is the dwell at this stop. The grey nodes represent the rescheduled times, the red nodes represent the latest flexible times, and the blue nodes represent the earliest flexible times. According to Figure 3.3, departure flexibility is indicated by the arc from the rescheduled departure node to the latest flexible departure node. The arc from the earliest departure node to the rescheduled departure node is weighted 0, indicating that trains are never allowed to depart earlier. The early arrival flexibility is indicated by the arc from the earliest arrival node to the rescheduled arrival node, and the late arrival flexibility is indicated by the arc from the rescheduled arrival node to the latest arrival node. Not only the nodes representing the rescheduled times, but the nodes representing the earliest and the latest flexible times should also respect the operational constraints. In addition, the alternative arcs and the connection arcs should be constructed between each pair of the three types of nodes.

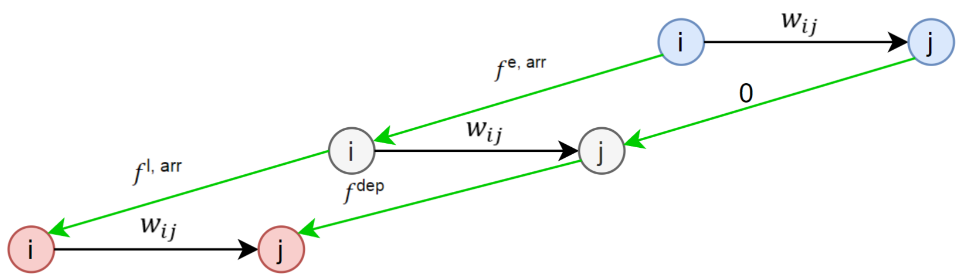


Figure 3.3: An illustrative representation of three types of timetable flexibility

3.4.2. Illustrative example of alternative graph

In order to provide a comprehensive discussion on the AG model with timetable flexibility, as well as the interactions between nodes and arcs, a full example of infrastructure layout and the corresponding AG representation is presented in Figure 3.4. In this example, we simplified the departure event as the entry to the successor block section of a station.

A complete AG consists of a quadruple $G = (N, F, A, C)$, where N is a set of nodes, F is a set of fixed arcs, A is a set of pairs of alternative arcs, and C is a set of connection arcs. They represent the railway network topology, operations, and all operational requirements. In this illustrative example, two

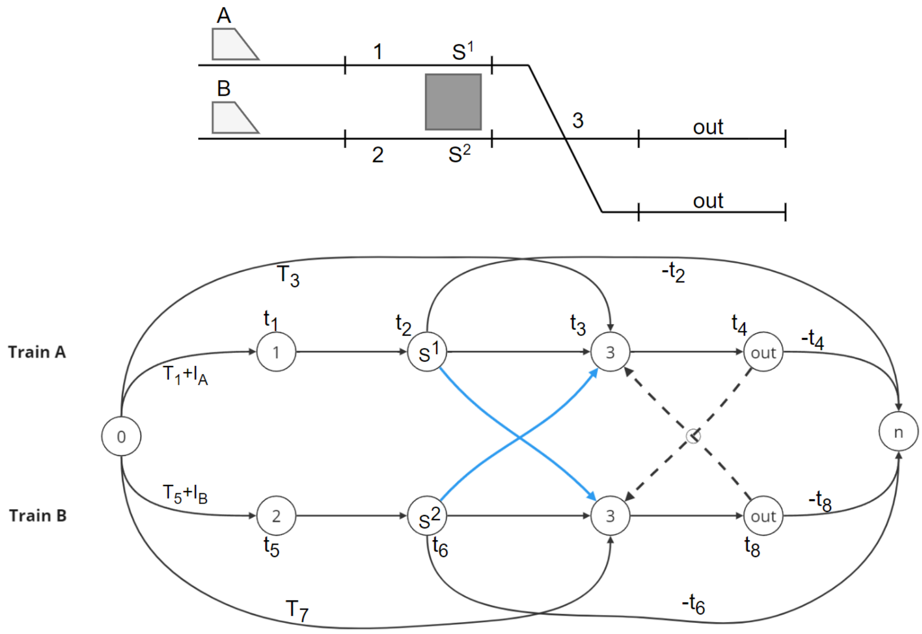


Figure 3.4: An example of infrastructure layout and alternative graph representation with all arcs

trains are scheduled in a small dispatching area consisting of 5 block sections and 1 station. Trains A and B run on block sections 1 and 2, respectively. Then train A stop at platform 1 and train B stop at platform 2 in the station. Upon departure, they share a diamond crossing located at block section 3. The block sections named out indicate that trains leave the dispatching area. In the AG representation, fixed arcs are indicated as black solid lines, alternative arcs are indicated by black dashed lines, and connection arcs are indicated by blue solid lines.

To simplify the representation, we assign an ID to each node representing the arrival at a block section or station. t represents the time of a node, in this illustrative example, t is a set of times including the rescheduled times $t_i^{\text{rescheduled}}$, the earliest flexible times t_i^{early}, and the latest flexible times t_i^{late}. I represents the initial delay when a train enters the dispatching area, and T represents the rigid time at a node.

Before running the model, the weights of some arcs should be updated incorporating parameters or decision variables. For an entry arc, it is weighted by $T + I$, which is the sum of the rigid entry time and initial delay of the corresponding train. It indicates the exact time that a train enters the dispatching area. For a departure arc, it is weighted by T, indicating that trains are not allowed to leave the station earlier than the rigid departure time. With respect to the arrival arcs, their weights differ when obtaining different types of times. When finding the rescheduled arrival times $t_i^{\text{rescheduled}}$, the weights of the arrival arcs are $-t_i^{\text{rescheduled}}$. Likewise, the weights of the arrival arcs are $-t_i^{\text{early}}$ when seeking for t_i^{early} and the weights of the arrival arcs are $-t_i^{\text{late}}$ when seeking for t_i^{late}. For an exit arc, it is weighed by $-t_i^{\text{rescheduled}}$. Since flexibility is not added at exit nodes, thus, the exit time of t_i^{early} and t_i^{late} should equal to $t_i^{\text{rescheduled}}$.

Based on the principle of the AG model, all three types of times are fixed to 0 s at node 0 and node n. In addition, a set of flexibility pairs S is constructed to pair the departure from a station and the arrival at the following station of the same train. According to the operational data of the case study area, there can be arrival nodes that have no preceding departure nodes, as well as departure nodes that have no following arrival nodes. For arrivals and departures that cannot form a departure-arrival pair, timetable flexibility is set to 0 s, meaning flexibility is not added to those nodes.

In the optimization process, the connection arcs can be determined whether to maintain or discard. By choosing one alternative arc from the pair, the train order at the shared resource is determined. In this example, choosing the arc from node 4 to node 7 indicates that train A precedes train B while choosing the arc from node 3 to node 8 indicates the other precedence relationship. Ultimately, a set of arcs is obtained containing the fixed arcs, the alternative arcs obtained by selecting one from each pair, and the connection arcs that are maintained. Simultaneously, the rescheduled times, the earliest flexible times, and the latest flexible times are generated, leading to the optimal objective value.

3.4.3. Mixed Integer Linear Programming

MILP is a mathematical optimization technique that can be considered as an extension to linear programming (IP), in which some of the decision variables are constrained to take integer values, unlike in linear programming where variables are continuous, which makes MILP suitable for solving optimization problems with discrete and continuous decision variables. The objective of MILP is to find the optimal values for the decision variables that minimize or maximize a linear objective function while satisfying a set of linear constraints. Various solvers, such as Gurobi and CPLEX optimizer, and various algorithms, such as branch and bound or cutting plane methods, are available to efficiently solve MILP problems and find the optimal solution.

In the context of railway traffic rescheduling, MILP can be utilized to optimize the allocation of resources and decision-making processes. In the current research, the rescheduling measures involve retiming, reordering, and passenger transfer connection decisions in response to traffic disturbances. The AG model can be formulated as MILP, which allows for more general objective functions, for instance, minimizing total delays, minimizing the cost of total delays, and minimizing the drift from the planned timetable, rather than minimizing the maximum consecutive delays. The model constraints can capture the operational rules and limitations of the railway network, such as track capacities, signaling restrictions, and connection requirements.

3.5. Model formulation

In this section, we aim to present the model formulation that generates the rescheduling plan based on given initial delays. On top of the rescheduling plan, timetable flexibility is added and maximized based on operational constraints.

According to the principle of the AG and incorporating the addition components, the mathematical notations and formulas that are used in the AG-based MILP model are given below. The symbols and descriptions of the indices, sets, and parameters are presented in Table 3.1. Besides, the symbols and descriptions of the decision variables are presented in Table 3.2. In addition, the objective function and constraints are presented, followed by a detailed explanation.

Objective function

$$minimize \sum_{i \in N_a \cup N_d \cup N_e} e^{\text{rescheduled}} * t_i^{\text{rescheduled}} - \sum_{(i,j) \in S} (e^{\text{f, dep}} * f_i^{\text{dep}} + f_j^{\text{e, arr}} + f_j^{\text{l, arr}}) \tag{3.1}$$

Constraints

$$t_j^{\omega} \geq t_i^{\omega} + w_{ij}^{\text{fixed}} \qquad \forall \omega \in \Omega, (i,j) \in F \tag{3.2}$$

$$t_j^{\delta_1} \geq t_i^{\delta_2} + w_{ij}^{\text{alternative}} - M * \alpha_{((i,j),(h,k))} \qquad \forall (\delta_1, \delta_2) \in \Delta, ((i,j),(h,k)) \in A \tag{3.3}$$

$$t_k^{\delta_1} \geq t_h^{\delta_2} + w_{hk}^{\text{alternative}} - M * (1 - \alpha_{((i,j),(h,k))}) \qquad \forall (\delta_1, \delta_2) \in \Delta, ((i,j),(h,k)) \in A \tag{3.4}$$

$$t_j^{\text{early}} \geq t_i^{\text{late}} + w_{ij}^{\text{connection}} - M * \beta_{ij} \qquad \forall (i,j) \in C \tag{3.5}$$

$$t_i^{\text{late}} - T_i - u_i^{\text{delay}} \leq M * \gamma_i^{\text{delay}} \qquad \forall i \in N_t \tag{3.6}$$

$$t_i^{\text{late}} - T_i - u_i^{\text{delay}} \geq M * (\gamma_i^{\text{delay}} - 1) \qquad \forall i \in N_t \tag{3.7}$$

$$t_j^{\text{early}} - u_j^{\text{departure}} \leq M * \gamma_j^{\text{departure}} \qquad \forall j \in N_d, (i,j) \in C \tag{3.8}$$

$$t_j^{\text{early}} - u_j^{\text{departure}} \geq M * (\gamma_j^{\text{departure}} - 1) \qquad \forall i \in N_d, (i,j) \in C \tag{3.9}$$

Table 3.1: Notations for indices, sets, and parameters

Symbol	Description (unit)
Sets	
$G = (N, F, A, C)$	graph
Θ	set of trains θ
N	set of nodes i, j, h, k
F	set of fixed arcs (i, j)
A	set of alternative arcs $((i, j), (h, k))$
C	set of connection arcs (i, j)
S	set of flexibility pairs (i, j)
$N_a \subset N$	set of arrival nodes
$N_d \subset N$	set of departure nodes
$N_e \subset N$	set of exit nodes
$N_t \subset N$	set of transfer delay measurement nodes
Ω	set of superscripts ω for trip time constraints $\{t^{\text{rescheduled}}, t^{\text{early arrival}}, t^{\text{late arrival}}\}$
Δ	set of pairs of superscripts (δ_1, δ_2) for headway time constraints $\{(t^{\text{rescheduled}}, t^{\text{rescheduled}}), (t^{\text{early}}, t^{\text{early}}), (t^{\text{late}}, t^{\text{late}}), (t^{\text{rescheduled}}, t^{\text{early}}), (t^{\text{rescheduled}}, t^{\text{late}}), (t^{\text{early}}, t^{\text{late}}), (t^{\text{late}}, t^{\text{early}}), (t^{\text{early}}, t^{\text{rescheduled}}), (t^{\text{late}}, t^{\text{rescheduled}})\}$
Indices	
i, j, h, k	indices for nodes
θ	index for trains
Parameters	
w_{ij}^{fixed}	arc weight of fixed arc (i, j) (s)
$w_{ij}^{\text{alternative}}, w_{hk}^{\text{alternative}}$	arc weight of alternative arcs (i, j) and (h, k) (s)
$w_{ij}^{\text{connection}}$	arc weight of connection arcs (i, j) (s)
T_i	the rigid time at node i (s)
I_θ	initial delay of train θ entering the network (s)
u_i^{delay}	the largest acceptable delay at the delay measurement node i of a transfer (s)
$u_i^{\text{departure}}$	the latest departure time of the waiting train at the departure node i of a transfer (s)
p	punctuality threshold (s)
$e^{\text{rescheduled}}$	weight of rescheduled times
$e^{\text{f, dep}}$	weight of departure flexibility
M	a sufficient large value

Table 3.2: Notations for decision variables

Symbol	Description
$t_i^{\text{rescheduled}}$	continuous; rescheduled time at node i (s)
t_i^{early}	continuous; earliest flexible time at node i (s)
t_i^{late}	continuous; latest flexible time at node i (s)
$f_i^{\text{e, arr}}$	continuous; early arrival flexibility at arrival node i (s)
$f_i^{\text{l, arr}}$	continuous; late arrival flexibility at arrival node i (s)
f_i^{dep}	continuous; departure flexibility at departure node i (s)
$\alpha_{((i,j),(h,k))}$	binary; the choice from the alternative pair (equals 0, if arc (i,j) is selected; equals 1, if arc (h,k) is selected)
β_{ij}	binary; the choice to maintain or discard the connection arc (equals 0, if the arc is maintained; equals 1, if the arc is discarded)
γ_i^{delay}	binary; whether the largest acceptable delay of a transfer is exceeded (equals 0, if not exceeded; equals 1, if exceeded)
$\gamma_i^{\text{departure}}$	binary; whether the latest departure time of a transfer is exceeded (equals 0, if not exceeded; equals 1, if exceeded)
$\upsilon_i^{\text{arrival}}$	binary; whether the difference between the rescheduled arrival time and the rigid arrival time exceeds the p (equals 0, if exceeded; equals 1, if not exceeded)
$\upsilon_i^{\text{departure}}$	binary; whether the difference between the rescheduled departure time and the rigid departure time exceeds p (equals 0, if exceeded; equals 1, if not exceeded)

$$\beta_{ij} \geq \gamma_h^{\text{delay}} \qquad \forall h \in N_t, (i,j) \in C \tag{3.10}$$

$$\beta_{ij} \geq \gamma_j^{\text{departure}} \qquad \forall (i,j) \in C \tag{3.11}$$

$$\beta_{ij} \leq \gamma_j^{\text{departure}} + \gamma_h^{\text{delay}} \qquad \forall h \in N_t, (i,j) \in C \tag{3.12}$$

$$t_i^{\text{rescheduled}} \geq T_i \qquad \forall i \in N_a \cup N_d \cup N_e \tag{3.13}$$

$$t_i^{\text{rescheduled}} - T_i - p \leq M * (1 - \upsilon_i^{\text{departure}}) \qquad \forall i \in N_d \tag{3.14}$$

$$t_i^{\text{rescheduled}} - T_i - p \geq -M * \upsilon_i^{\text{departure}} \qquad \forall i \in N_d \tag{3.15}$$

$$t_i^{\text{late}} - T_i - (1 - \upsilon_i^{\text{departure}}) * M \leq p \qquad \forall i \in N_d \tag{3.16}$$

$$f_i^{\text{dep}} \leq t_i^{\text{late}} - t_i^{\text{rescheduled}} \qquad \forall i \in N_d \tag{3.17}$$

$$f_i^{\text{dep}} \leq p * \upsilon_i^{\text{departure}} \qquad \forall i \in N_d \tag{3.18}$$

$$t_i^{\text{rescheduled}} - T_i - p \leq M * (1 - \upsilon_i^{\text{arrival}}) \qquad \forall i \in N_a \tag{3.19}$$

$$t_i^{\text{rescheduled}} - T_i - p \geq -M * \upsilon_i^{\text{arrival}} \qquad \forall i \in N_a \tag{3.20}$$

$$f_i^{\text{e, arr}} \leq t_i^{\text{rescheduled}} - t_i^{\text{early}} \qquad \forall i \in N_a \tag{3.21}$$

$$t_i^{\text{late}} - T_i - (1 - \upsilon_i^{\text{arrival}}) * M \leq p \qquad \forall i \in N_a \tag{3.22}$$

$$f_i^{\text{l, arr}} \leq t_i^{\text{late}} - t_i^{\text{rescheduled}} \qquad \forall i \in N_a \tag{3.23}$$

$$f_i^{\text{l, arr}} \leq p * v_i^{\text{arrival}} \qquad \forall i \in N_a \tag{3.24}$$

$$f_i^{\text{dep}} \geq 0 \qquad \forall i \in N_d \tag{3.25}$$

$$f_i^{\text{e, arr}} \geq 0 \qquad \forall i \in N_a \tag{3.26}$$

$$f_i^{\text{l, arr}} \geq 0 \qquad \forall i \in N_a \tag{3.27}$$

According to the objective function (3.1), by minimizing the rescheduled times at arrival, departure, and exit nodes, we aim to obtain the rescheduling plan that has the smallest deviation from the rigid timetable. The minimization of the rescheduled times at intermediate nodes is not considered since the rigid times at these nodes are not given. The rescheduled plan can be considered as the baseline to add timetable flexibility. On top of the rescheduling plan, timetable flexibility is added, and the objective is to maximize the total timetable flexibility. $e^{\text{rescheduled}}$ is the weight of the rescheduled times at arrival, departure, and exit nodes. In this research, it is set to a sufficient large number indicating that our primary aim is to obtain the rescheduling plan, and then add flexibility. $e^{\text{f,dep}}$ is the weight of departure flexibility. $e^{\text{f,dep}}$ equals 1 indicates that there is no preference on the distribution of flexibility. $e^{\text{f,dep}}$ greater than 1 means more departure flexibility is preferred, and $e^{\text{f,dep}}$ less than 1 indicates that more arrival flexibility is preferred.

The retiming aspect of the model can be represented by the rescheduled times at nodes, which need to satisfy the trip time constraint (3.2). w_{ij}^{fixed} indicates the minimum running time through a block section or the dwell time at a stop, which can be obtained from the operational data. According to Constraint (3.2), the arrival time at a block section should be greater than or equal to the arrival time at the former block section plus the minimum running time at the former block section. For an arrival node and a succeeding departure node, the weight implies the minimum dwell time at the station. Constraint (3.2) should be applied to all event times attributed to a node, including the rescheduled time, the earliest flexible time, and the latest flexible time. In order to simplify the formulation regarding trip time constraints, a set Ω is constructed, in which a superscript ω is used to represent each type of event times.

Constraint (3.3) and Constraint (3.4) are headway constraints that regulate the minimum required time between two trains. $w_{ij}^{\text{alternative}}$ is the weight of alternative arcs, which is calculated as the minimum headway time between trains minus the block running time of the preceding train. Two preceding possibilities are provided, and only one of the alternative pairs can be chosen. When alternative arc (i, j) is chosen, $\alpha_{((i,j),(h,k))}$ equals 0, which means Constraint (3.4) is relaxed, and only Constraint (3.3) takes effect. When alternative arc (h, k) is chosen, $\alpha_{((i,j),(h,k))}$ equals 1, which means Constraint (3.3) is relaxed, and only Constraint (3.4) takes effect. To ensure that the results of the rescheduled times, the earliest flexible times, and the latest flexible times are all feasible, and no conflicts occur between each pair of train paths, the headway constraints should be respected between rescheduled times, between flexible times, as well as between rescheduled times and flexible times. In order to simplify the formulation related to headway constraints, a set Δ is constructed, in which a superscript (δ_1, δ_2) is used to represent the types of event times that should respect the headway time constraints.

Constraint (3.5) is the connection constraint, which ensures that if a passenger transfer is maintained, a minimum transfer time should be satisfied between the arrival time of a train and the departure time of the waiting train of a designed passenger transfer. Since timetable flexibility is added to arrival and departure events, we only need to ensure that the minimum passenger transfer time can be realized between the latest arrival time and the earliest departure time. Therefore, any arrival and departure times within the time windows respect the minimum transfer time when the passenger transfer is maintained. When the connection arc (i, j) is maintained, β_{ij} equals 0 and the constraint takes effect. Otherwise, the constraint is relaxed.

Two parameters u_i^{delay} and $u_i^{\text{departure}}$ are introduced, representing the arrival delay threshold at the measurement point of and the departure time threshold at a passenger transfer, respectively. Constraints (3.6) and (3.7) are used to model the logic of the delay threshold at the measurement point.

When the threshold is violated, γ_i^{delay} has to be 1, otherwise, it equals 0. Constraints (3.8) and (3.9) are used to model the logic of the departure threshold at the transfer station. When the threshold is violated, γ_i^{delay} has to be 1, otherwise, it equals 0. By constructing Constraints (3.10), (3.11), and (3.12), the passenger transfer is discarded when either threshold is exceeded, and it can only be maintained when two thresholds are respected.

Since the times in the rescheduling plan should have the smallest deviation from the rigid times, we need to ensure that the rigid times are used when no initial delays are experienced by trains. Besides, trains should arrive at stations and leave the network as early as possible when they are delayed. Constraint (3.13) is added to ensure that trains cannot arrive at stations, depart from stations, and leave the network earlier than the rigid times. The earliest possible times at arrival nodes and exit nodes can be obtained based on the trip time constraint. As our primary objective is to minimize the rescheduled times, the minimum arrival, departure, and exit times can be obtained.

In order to obtain the values of timetable flexibility, the conditions that flexibility can be added should be specified. Based on the punctuality threshold, flexibility can only be added when the difference between the rescheduled times and the rigid times do not exceed the threshold. Regarding departure flexibility, Constraints (3.14) and (3.15) are used to determine whether the threshold is exceeded for departure events. Constraint (3.16) is added to ensure that the difference between the latest departure time and the rigid departure time is less than or equal to the punctuality threshold. According to the definition of departure flexibility, Constraint (3.17) is constructed, indicating that departure flexibility is calculated as the difference between the latest flexible departure time and the rescheduled departure time. In addition, Constraints (3.18) and (3.25) are added to regulate that departure flexibility can only be added when the punctuality threshold is not exceeded, that is $\upsilon_i^{\text{departure}}$ equals 1. Otherwise, departure flexibility equals 0, and the latest flexible departure times equal the rescheduled departure times.

Regarding arrival flexibility, it can only be added when the difference between the rescheduled arrival times in the rescheduling plan and the rigid times do not exceed the threshold, which is represented by Constraint (3.19) and (3.20). With respect to late arrival flexibility, Constraint (3.22) is added to ensure that the difference between the latest flexible arrival time and the rescheduled arrival time is less than or equal to the punctuality threshold. According to Constraint (3.23), late arrival flexibility is calculated as the difference between the latest flexible arrival time and the rescheduled arrival time. Constraints (3.27) and (3.24) are added to regulate that early arrival flexibility can only be added when the punctuality threshold is not exceeded, that is $\upsilon_i^{\text{arrival}}$ equals 1. Otherwise, late arrival flexibility equals 0, and the latest flexible arrival times equal the rescheduled arrival times.

In our model, early arrival flexibility is defined as the difference between the rescheduled arrival time and the earliest flexible arrival time, which is represented by Constraint (3.21). Since punctuality is only defined for lateness, the upper bound of early arrival flexibility is not limited, which means very early arrivals are also allowed. Constraint (3.26) is added to make sure positive early arrival flexibility is generated, thus, ensuring early arrival times are always smaller than or equal to the rescheduled arrival times.

3.6. Model implementation

3.6.1. General input

In order to fulfill the realization of the AG-based MILP model, 6 types of input are required, including the rigid timetable, processing times, headway times, connection times, initial delays, and other inputs, which are explained below.

- **Rigid timetable**: In the rigid timetable, rigid times are arrival, departure, and passing times at timetable points. Regarding the departure, the lower bounds for the rescheduled departure times are specified to be the rigid departure times, indicating that trains cannot depart earlier than the rigid times.
- **Processing times**: Processing times include block running times and dwell times. For block sections, the processing time is the minimum block running time, where the trains run at the maximum allowed speed. For stations, the processing time is the minimum dwell time, indicating the minimum required times that trains should stop at the station. The weight of fixed arcs comprises the minimum running and dwell times.

- **Headway times**: According to the block time theory (Pachl, 2014), headway time consists of the clearing, release, and block running times of the preceding train, and the setup, sight, reaction, and approach times of the following train. In our model, block running times are represented by the processing times. Therefore, the exact value that is used as the weight of an alternative arc is the minimum headway time between trains minus the block running time of the first train.
- **Connection times**: Connection times indicate the time between a train arriving at the station and another train leaving the station. A minimum passenger transfer time should be provided as the weights for the connection arcs. For cross-platform transfers, the minimum required time is 2 minutes. For long-platform transfers, the minimum required time is 3 minutes. For one-platform-further transfers, the minimum required time is 4 minutes. For cross-station transfers, the minimum required time is 5 minutes.
- **Initial delays**: The initial delays are used to represent traffic disturbances, and they should be added to the trains when they enter the dispatching area.
- **Other inputs**: Other inputs comprise the value of punctuality threshold p, delay threshold of a transfer u_i^{delay}, and departure threshold of a transfer $u_i^{\text{departure}}$. The value of p is provided according to current practice in railway traffic management, which is 3 minutes. The values of u_i^{delay} and $u_i^{\text{departure}}$ can be obtained from the TAD of ProRail.

3.6.2. Commercial solver: Gurobi optimizer

Gurobi Optimizer (Gurobi Optimization, LLC, 2023) is a powerful and widely used mathematical optimization software that provides efficient solvers for a wide range of optimization problems. In this research, we employ Python as our programming language, which serves as the interface for interacting with Gurobi. Gurobi Optimizer employs a variety of algorithms and techniques to solve MILP problems efficiently, which are discussed as follows.

- **Branch and Bound (B&B)**: Gurobi's MILP solver utilizes a Branch and Bound algorithm. It systematically explores the solution space by branching on integer variables and bounding the objective function. The branching process splits the problem into subproblems, forming a tree structure, and the bounds of the objective function guide the search towards the optimal solution.
- **LP Relaxation**: Gurobi starts by solving the Linear Programming (LP) relaxation of the MILP problem, where the integer constraints are relaxed, and the problem is treated as a continuous linear programming problem. This LP relaxation provides an initial bound on the objective function and helps in guiding the branching decisions during the Branch and Bound process.
- **Cut Generation**: Gurobi employs various cutting-plane algorithms to strengthen the LP relaxation and tighten the bounds. These algorithms generate valid inequalities (cuts) that are added to the LP relaxation to eliminate infeasible or suboptimal solutions.
- **Heuristics**: Gurobi incorporates several heuristic algorithms to quickly find good-quality feasible solutions and improve the search process. These heuristics help in generating promising initial solutions and exploring the solution space effectively.
- **Presolve Techniques**: Gurobi's presolve phase simplifies the problem formulation by exploiting the problem structure and eliminating redundant constraints or variables. Presolve techniques reduce the problem size and complexity, leading to faster solution times.
- **Parallel Processing**: Gurobi is designed to utilize multiple threads and parallel processing to speed up the solution process for MILP problems. It distributes the computational workload across multiple cores or machines, enabling efficient exploration of the solution space.

3.7. Model verification

A toy network refers to a simplified and smaller version of a real-life railway network. Before diving into the complexities of a large-scale network, a toy network can help verify the basic principles and the feasibility of the model formulation. Besides, in the programming process, bugs and errors may emerge, and using a toy network with a simplified problem setting can make it easier to track and fix any errors. In the toy network, we simplified the departure event as the entry to the successor block section of a station. Regarding the times in the rigid timetable, entry times of trains entering the network, exit times of trains leaving the network, and the arrival and departure times at stations should be provided.

3.7.1. Toy network description

The layout of the toy network is presented in Figure 3.5. The toy network consists of 3 stations and 16 block sections, of which 13 out of 16 are running block sections, and the other three are block sections named out, representing that trains leave the dispatching area. 3 trains are designed to run on the toy network, the routes of trains A, B, and C are indicated by blue, red, and green lines, respectively. Train A passes through block sections 1, 2, 3, 4, 5, 6, 9, and out1, and stops at station S_1 and platform 2 of station S_2. The different platforms in the station are denoted using superscripts, for instance, S_2^2 means platform 2 of station S_2. Train B passes through block sections 1, 2, 3, 11, 12, 13, and out2, and stops at station S_1 and platform 2 of station S_3, denoted as S_3^2. Train C traverses through block sections 7, 8, 9, 10, 13, and out3, and stops at platform 1 of station S_2 and platform 1 of station S_3, denoted as S_2^1 and S_3^1, respectively. The designed passenger transfer is from train C to train B at station S_3. Train A and Train B approach the switch located in block section 1, train A and train C approach the diamond crossing in Block Section 7 and run towards two different directions, and train B and train C approach the diamond crossing in block section 10 and run towards two different directions, those block sections are locations where the train sequence can be changed.

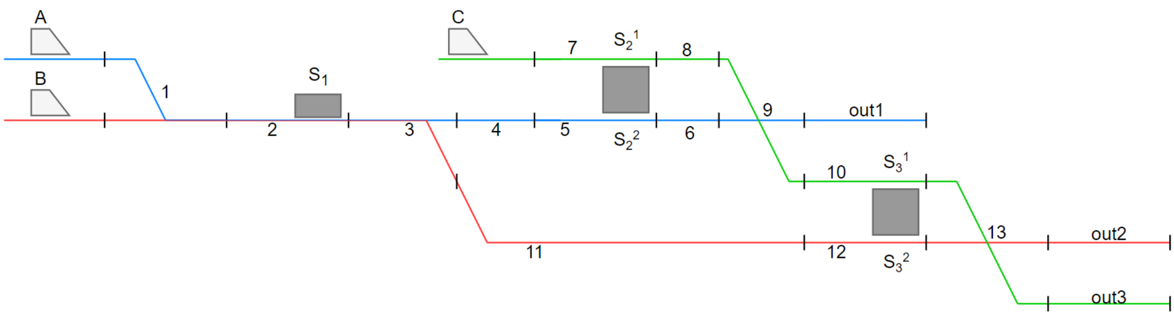


Figure 3.5: Layout of the toy network

3.7.2. Illustrative examples of toy network

In order to run the model, a rigid timetable should be given and some fictitious values of parameters should be given as model inputs. The weights of fixed arcs, alternative arcs, and connection arcs are the minimum running/dwell times, minimum headway time between trains minus the block running time of the preceding train, and minimum required transfer times, respectively. So, we first assign values to the arcs, and the weights of the arcs are presented in the AG representation of the toy network, which is shown in Figure 3.6. In our toy network, the minimum blocking running times are set to 10 s or 12 s, the minimum dwell times are set to 20 s, the weights of alternative arcs are set to 15 s, and the minimum transfer time at station S_3 is set to 20 s.

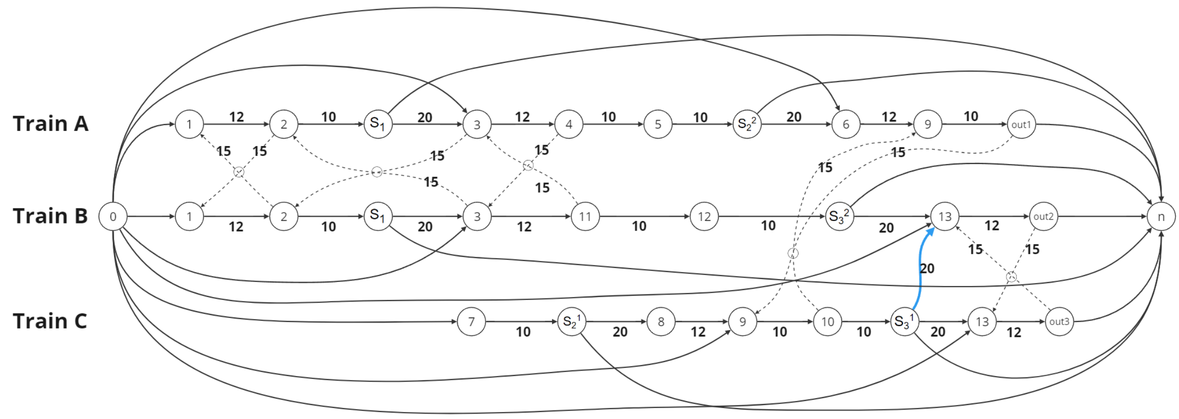


Figure 3.6: The AG representation of the toy network

Based on the minimum processing, headway, and connection times, the rigid departure and arrival times can be generated considering the running time supplements and buffer times. The rigid timetable of the toy network is presented in Table 3.3.

In the illustrative examples, the weight of rescheduled times $e^{\text{rescheduled}}$ is set to 100, indicating that flexibility is added on top of the rescheduling plan. The weight of departure flexibility $e^{\text{f, dep}}$ is set to 1, indicating no preference is imposed on departure or arrival flexibility. Besides, the punctuality threshold p is set to 10 s, and the sufficient large value M is set to 9999. Regarding the passenger transfer from feeder train C to waiting train B at station S_3, the delay measurement node is the arrival node of train C at station S_2. The largest acceptable delay u_i^{delay} of the feeder train at the measurement point is set to 85 s. For the latest departure time of train B at station S_3, $u_i^{\text{departure}}$ is set to 225 s.

Table 3.3: Rigid timetable of the toy network

	Rigid timetable		
Block section	A	B	C
1	*	*	*
2	*	*	*
S1	35	100	*
3	65	130	*
4	*	*	*
5	*	*	*
7	*	*	*
S21	*	*	65
S22	105	*	*
6	135	*	*
8	*	*	95
9	*	*	*
10	*	*	*
11	*	*	*
12	*	*	*
S31	*	*	140
S32	*	175	*
13	*	205	170
out	*	*	*

Illustrative example 1 without initial delays

In the first illustrative example of the toy network, no initial delays are assigned to trains when they enter the dispatching area. After running the model in the Gurobi optimizer, the optimal results of decision variables can be obtained. The AG representation of the optimization results without initial delays is presented in Figure 3.7. It can be observed that train A precedes train B at block sections 1, 2, and 3, train C precedes train A when approaching the diamond crossing located at block section 9, and train C precedes train B when approaching the diamond crossing located at block section 13. Regarding the connection arc, it is maintained since both thresholds are respected.

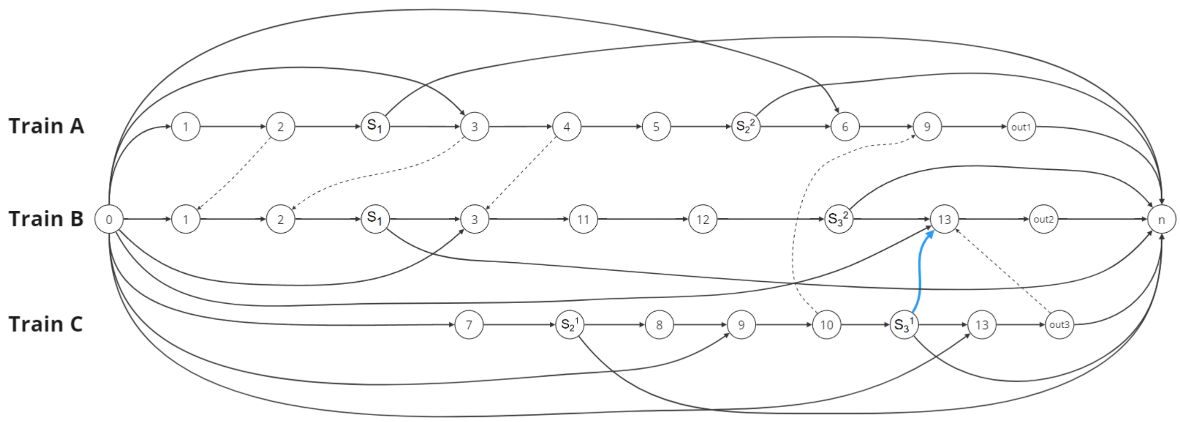


Figure 3.7: The AG representation of the optimization results of the example without initial delays

A comparison of the rescheduling plan, early flexible plan, and late flexible plan is shown in Table 3.4. It can be observed that the rescheduled times are the same as the corresponding values in the rigid timetable, which means that the model finds the rescheduling plan that has no deviation from the rigid timetable. The times in the flexible plans are the flexible times generated by the optimization model. In the early flexible plan, early arrival flexibility is added, and in the late flexible plan, late arrival flexibility and departure flexibility are added. In order to explicitly show the timetable flexibility, the results are illustrative in a time-distance diagram, which is shown in Figure 3.8.

Table 3.4: A comparison of the rescheduling plan, early flexible plan, and late flexible plan of illustrative example 1

	Rescheduling plan			Early flexible plan			Late flexible plan		
Block section	A	B	C	A	B	C	A	B	C
1	*	*	*	*	*	*	*	*	*
2	*	*	*	*	*	*	*	*	*
S_1	35	100	*	35	100	*	45	100	*
3	65	130	*	65	130	*	75	140	*
4	*	*	*	*	*	*	*	*	*
5	*	*	*	*	*	*	*	*	*
7	*	*	*	*	*	*	*	*	*
S_2^1	*	*	65	*	*	65	*	*	65
S_2^2	105	*	*	97	*	*	115	*	*
6	135	*	*	135	*	*	135	*	*
8	*	*	95	*	*	95	*	*	105
9	*	*	*	*	*	*	*	*	*
10	*	*	*	*	*	*	*	*	*
11	*	*	*	*	*	*	*	*	*
12	*	*	*	*	*	*	*	*	*
S_3^1	*	*	140	*	*	130	*	*	150
S_3^1	*	175	*	*	165	*	*	185	*
13	*	205	170	*	205	170	*	205	170
out	*	*	*	*	*	*	*	*	*

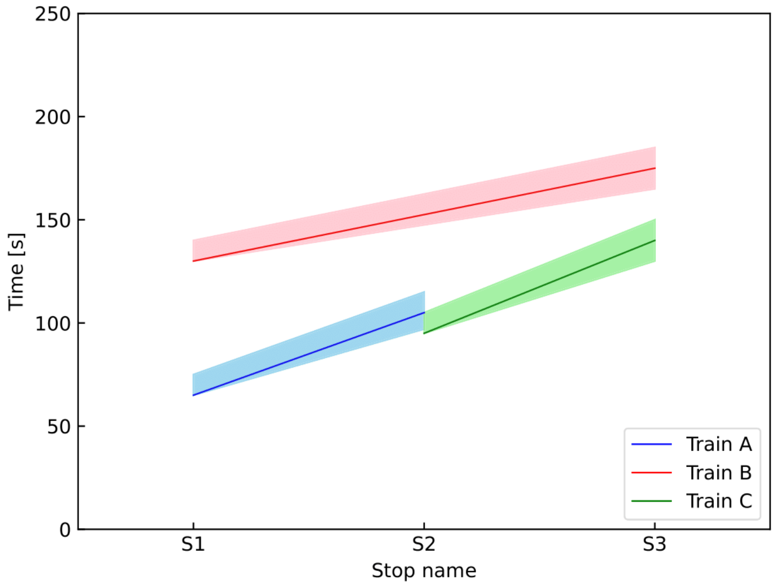


Figure 3.8: The train paths of three trains without initial delays

In Figure 3.8, flexibility added to the departure and arrival events is represented by the colored thick train paths. Train A stops at S_1 and S_2, however, flexibility is only added to its departure from station 1 and its arrival at station 2. The reason is that information outside the dispatching area is unknown, thus, flexibility is added to the arrival events that have upstream departures, as well as to the departure events that have downstream arrivals. The same is true for trains B and C. Regarding the train path of train A which is shown in blue, departure flexibility is added at station 1, and both early and late arrival flexibility are added to station 2. By looking at the numerical results shown in Table 3.4, the departure flexibility is 10 s, which can be calculated as the time difference between the latest flexible time and the rescheduled time of train A entering block section 3. Regarding the arrival flexibility of train A at S_2, the early arrival flexibility is 8 s, which is the time difference between the rescheduled arrival time and the earliest flexible time of train A arriving at S_2. The late arrival flexibility is 10 s, which is the time difference between the latest flexible time and the rescheduled time of train A arriving at S_2. The flexibility of train B at S_1 and S_3, as well the flexibility of train C at S_2 and S_3 can be calculated in the same way. With respect to train B, the departure flexibility at S_1, the early arrival flexibility at S_2, and the late arrival flexibility at S_2 are 10 s. For train C, the departure flexibility at S_2, the early arrival flexibility at S_3, and the late arrival flexibility at S_3 are 10 s.

Illustrative example 2 with initial delays

After making sure that the rescheduling plan is the same as the rigid timetable when no initial delays are added, the feasibility of the model under delayed scenarios should also be verified. In other words, it should be ensured that timetable flexibility is actually added to the rescheduling plan that has the least deviation from the rigid timetable. In the meantime, the punctuality threshold should also be respected.

In the second illustrative example, 80 s initial delays are assigned to train A, 10 s initial delays are assigned to train B, and 40 s initial delays are assigned to train C. The AG representation of the optimization results when initial delays are assigned is presented in Figure 3.9. It is obvious that the order between train A and train B traversing block sections 1, 2, and 3 changes compared to the illustrative example 1. Regarding the passenger transfer, the arrival delay of train C at S_2 is 100 s, which exceeds the delay threshold of 85 s. Therefore, the train connection of the passenger transfer from train C to train B at S_3 is discarded.

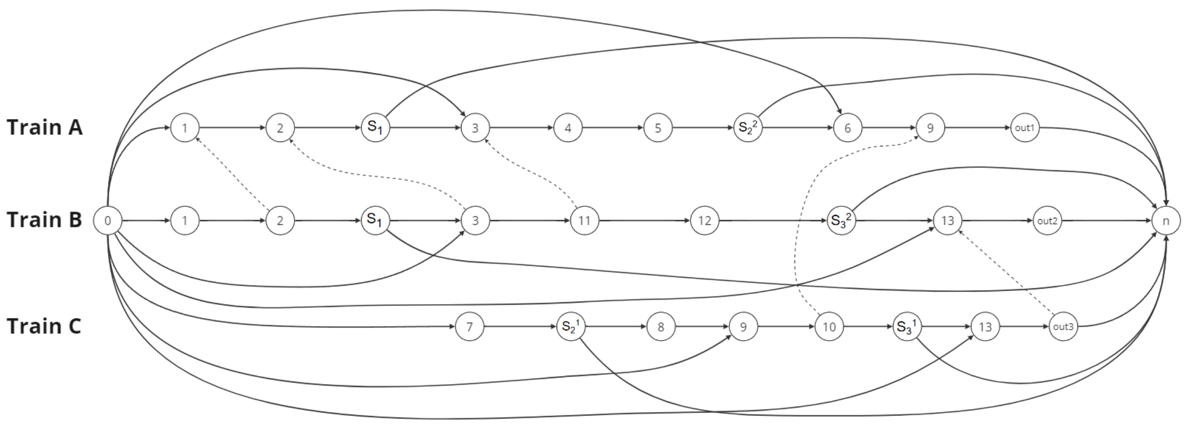


Figure 3.9: The AG representation of the optimization results of the example with initial delays

A comparison of the rescheduling plan, early flexible plan, and late flexible plan when trains experience some initial delays is shown in Table 3.5. Since trains are delayed when entering the dispatching area, trains should arrive and depart as early as possible to recover those delays. Therefore, the arrival and departure times at stations, as well as the exit times in the rescheduling plan should be the ones with the least deviation from the rigid plan. In order to present timetable flexibility more clearly, a time-distance diagram is shown in Figure 3.10, in which three train paths are presented.

Table 3.5: A comparison of the rescheduling plan, early flexible plan, and late flexible plan of illustrative example 2

	Rescheduling plan			Early flexible plan			Late flexible plan		
Block section	A	B	C	A	B	C	A	B	C
1	*	*	*	*	*	*	*	*	*
2	*	*	*	*	*	*	*	*	*
S_1	155	100	*	155	100	*	155	100	*
3	175	130	*	175	130	*	175	130	*
4	*	*	*	*	*	*	*	*	*
5	*	*	*	*	*	*	*	*	*
7	*	*	*	*	*	*	*	*	*
S21	*	*	100	*	*	100	*	*	100
S22	207	*	*	207	*	*	207	*	*
6	227	*	*	227	*	*	227	*	*
8	*	*	120	*	*	120	*	*	120
9	*	*	*	*	*	*	*	*	*
10	*	*	*	*	*	*	*	*	*
11	*	*	*	*	*	*	*	*	*
12	*	*	*	*	*	*	*	*	*
S31	*	*	152	*	*	152	*	*	152
S32	*	175	*	*	165	*	*	185	*
13	*	205	172	*	205	172	*	205	172
out	*	*	*	*	*	*	*	*	*

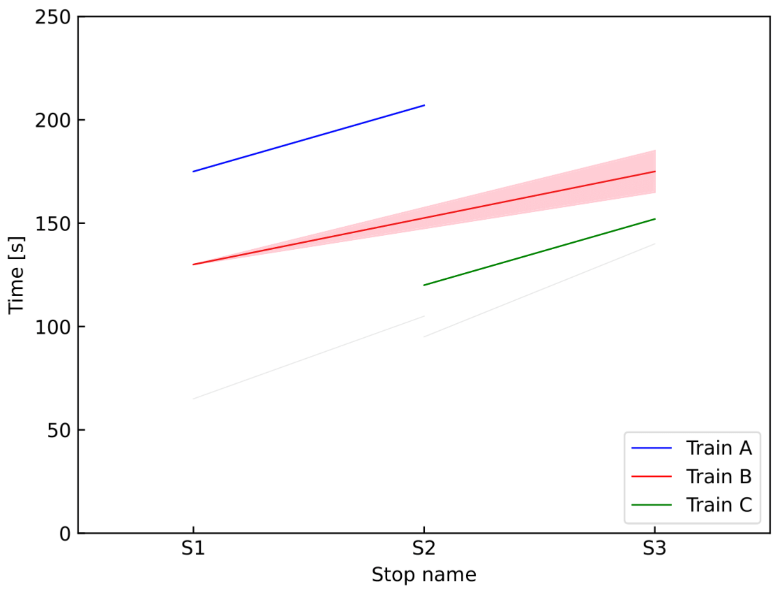


Figure 3.10: The train paths of three trains with initial delays

According to Figure 3.10, it can be observed that no flexibility is added to train A and train C. Regarding train A, by comparing the rescheduling plan to the rigid timetable, it can be seen that the difference between rescheduled departure time and rigid departure time at S_1 exceed the punctuality threshold of 10 s. It means that no departure flexibility can be added. Likewise, the difference between rescheduled arrival time and rigid arrival time at S_2 also exceeds 10 s, indicating that no arrival flexibility can be added. The same is true for train C. With respect to train B, the thick train path indicates that no departure flexibility is added at S_1, while early arrival flexibility and late arrival flexibility is added. Since train B precedes train A in block sections 1, 2, and 3, the headway time requirement between two trains should be respected. Due to the initial delays of 80 s experienced by train A, the rescheduled arrival time of train A at S_1 is 155 s, and the rescheduled time of train A entering block section 2 is 145 s due to the 10-second trip time. Considering the 15-second weight of the alternative arc, train B can depart from S_1 at 130 s at the latest. Therefore, no departure flexibility can be added. For the arrival of train B at S_3, it is not affected by other trains, thus, the early arrival flexibility is 10 s, which is calculated as rescheduled arrival time minus the earliest flexible arrival time at S_3. Likewise, the late arrival flexibility is also 10 s, which is calculated as the latest flexible arrival time minus the rescheduled arrival time at S_3.

According to the two illustrative examples on the toy network, it can be ensured that the objective function and constraints are functioning well, thus, proving the feasibility of the model.

4

Case study

In this chapter In section 4.1, an introduction to the real case study area and the incorporated trains is provided. Section 4.2 presents the preparations that are necessary to run the model, namely data set construction, data correction, and parameter settings. In section 4.3, several sets of initial delays are generated, according to which our model is applied to the real-life network. In section 4.4, the relationship between initial delays and timetable flexibility is investigated. Interpretations are provided based on comparisons of train paths, as well as the numerical results. In section 4.5, the impacts of departure flexibility weight on timetable flexibility are explored. In addition, the interactions between different types of flexibility are also studied. In section 4.6, the relationship between the punctuality threshold and timetable flexibility is investigated.

4.1. Case study description

The selected dispatching area for the case study is located in the southern part of the Dutch railway network. The case study area includes stations ’s-Hertogenhosch (Ht), Tilburg (Tb), Breda (Bd), and Eindhoven (Ehv) as the major nodes in the network. In Figure 4.1, a schematical view of the track layout of the selected dispatching area is presented. It can be observed that lines from different directions meet at stations Ht, Tb, and Ehv, thus, passenger transfers are incorporated in those stations. There are converging tacks, diverging tracks, and crossing tracks in the selected dispatching area, indicating that more interesting conflict scenarios can be obtained in such a network compared to a railway corridor, in which trains run on parallel tracks or the same track. In addition, the selected area also contains traffic heterogeneity. Three types of trains traverse through the railway infrastructure, namely intercity (IC), sprinter (SPR), and freight train (GO). Apart from the four major stations, there are also stops in between, namely, Eindhoven Philips Strijp (Ehs), Best (Bet), Btl, Oisterwijd (Ot), Tilburg Reeshof (Tbr), Tilburg Universiteit (Tbu), Gilze Rijen (Gz), and Vught (Vg). Since the stop names in Figure 4.1 can be a bit out of date, the stop Tilburg West is the same station as Tbu.

In order to clearly illustrate the train information, an overview of the selected trains is presented in Table 4.1. IC trains are a type of long-distance train service in the Netherlands. These trains connect major cities and regions within the country. SPR trains are a type of regional train service in the Netherlands. Unlike IC trains, SPR trains make frequent stops, serving not only major cities but also smaller towns and suburban areas. GO stands for freight train. According to the train data provided by ProRail, specific IDs are assigned to trains. However, those IDs are not continuous since only a small set of trains is selected. Those IDs will be mentioned in the discussion of the train paths. In the current research, we mainly focus on the investigation of adding flexibility rather than the application of the model to complicated real-life cases. Besides, AG is not able to model the interaction between trains running in different directions. Therefore, train services in three directions are selected, namely, from Ehv to Bd, from Ht to Bd, and from Ehv to Ht. All trains except the freight train have a period of half an hour, to make sure that a full cycle is involved, the first trains of the second periods are also incorporated. They are indicated by the serial number, for example, train 800-H-2 runs in the next period of train 800-H-1, and train 3600-T-2 runs in the next period of train 3600-T-1. Since no data are given for the train that runs in the next period of train 1900-H-1, train 1900-H-2 is not incorporated in

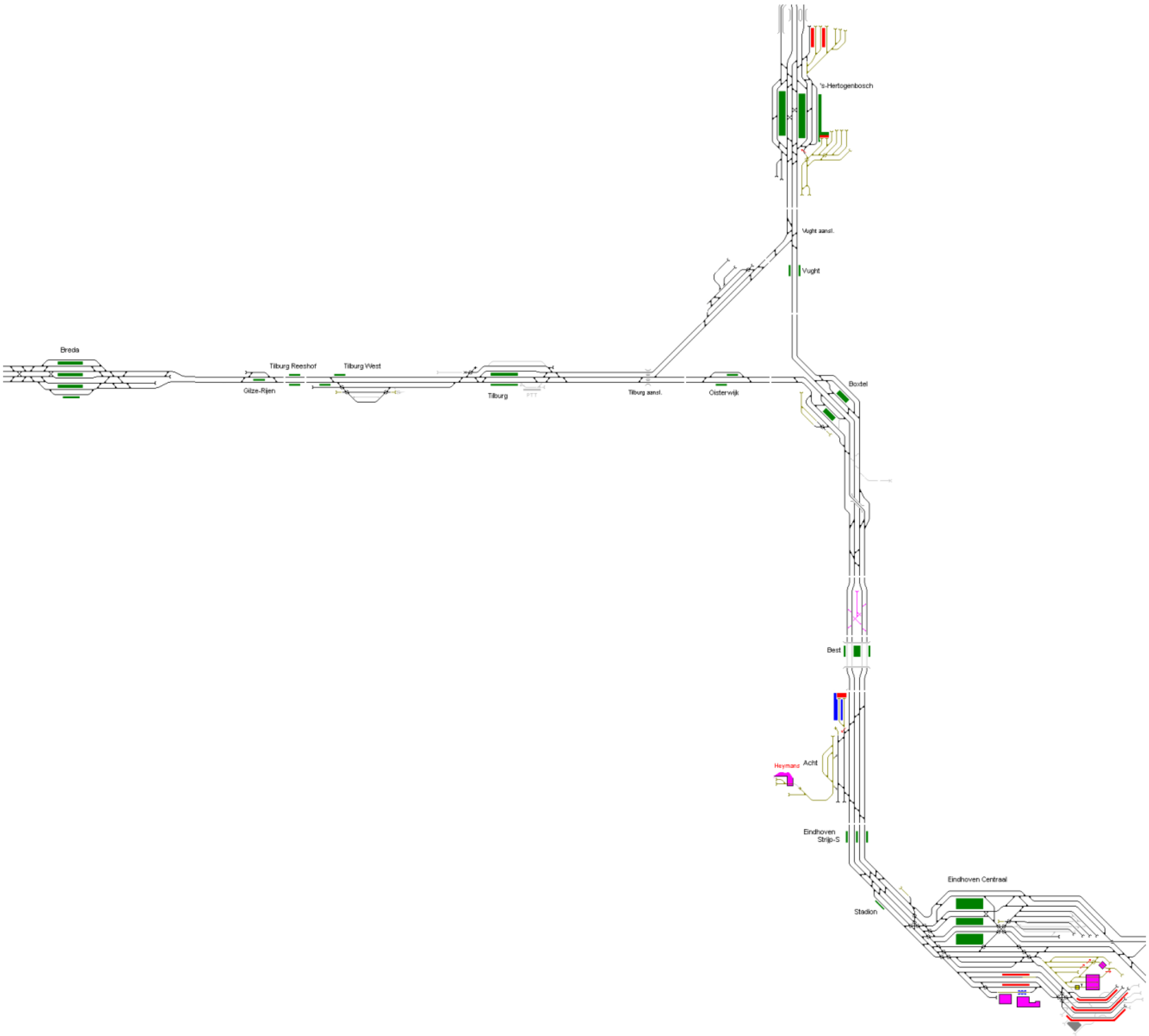


Figure 4.1: Track layout of the dispatching area[1]

the case study.

According to the microscopic track layout and the operational data, the tracks that are used for each train are described as follows. At station Ehv, trains 5, 15, 17, and 70 dwell at platform 4, trains 1 and 64 dwell at platform 5, and trains 6, 61, and 66 dwell at platform 6. In the corridor from Ehv to Btl, trains 6, 15, and 66 run on the same tracks, and trains 1, 5, 17, 61, 64, and 70 share the same tracks. Before approaching station Btl, some trains diverge to different tracks. Trains 1, 5, 6, and 15 enter the corridor from Btl to Tb, and the other trains run towards Ht. At station Ht, trains 34, 43, and 79 dwell at platform 7b. Since train 18 is a freight train, it has no scheduled stop and passes platform 8. After leaving the platform area, they merge at a switch and run on the same tracks in the corridor from Ht to Tb. At station Tb, trains 34, 43, and 79 from Ht dwell at platform 3, and the freight train passes platform 3. Besides, trains from Ehv dwell at platform 2. After leaving station Tb, trains 1, 5, and 6 diverge with train 15 since it is running towards its terminal station Tbu. The other trains merge and run on the same tracks in the corridor from Tb to Bd.

The rigid timetable is the actual timetable used by ProRail, NS, and other operators. The rigid timetables for trains that run from Ehv to Bd, from Ht to Bd, and from Ehv to Ht are presented in Table B.1, Table B.2, and Table B.3, respectively. For major stations, namely Ehv, Tb, Ht, and Bd, the rigid arrival and departure times are given. While for all other short stops, only the rigid departure times

[1] `http://www.sporenplan.nl/`

Table 4.1: An overview of the selected trains in the dispatching area

Train ID	Train serie	Direction	Serial number	Train stops	Train type
1	1100	H	1	Ehv-Tb-Bd	IC
5	1900	H	1	Ehv-Tb-Bd	IC
6	21400	H	1	Ehv-Tb-Bd	IC
15	6400	H	1	Ehv-Ehs-Bet-Btl-Ot-Tb-Tbu	SPR
17	800	H	1	Ehv-Ht	IC
18	AABE1	H	1	-	GO
34	3600	T	1	Ht-Tb-Bd	IC
43	6600	T	1	Ht-Tb-Tbu-Tbr-Gz-Bd	SPR
61	3500	H	2	Ehv-Ht	IC
64	3900	H	2	Ehv-Ht	IC
66	4400	H	2	Ehv-Ehs-Bet-Btl-Vg-Ht	SPR
70	800	H	2	Ehv-Ht	IC
79	3600	T	2	Ht-Tb-Bd	IC

are given. Regarding the freight train, no stops are planned in the case study area, which means that no departure or arrival times are given. Based on the TAD and the selected trains, a passenger transfer from train 6600-T-1 to train 6400-H-1 at station Tb can be obtained. According to the rigid timetables, it can be observed that train 6600-T-1 arrives at Tb at :48:00 and train 6400-H-1 leaves Tb at :53:30. :48:00 indicates the 48 minutes of each hour, the same is true for :53:30. It means that a transfer time of five and a half minutes is designed for passengers.

4.2. Model preparation

For the model to run smoothly, some preparations need to be made on the model inputs. These processes include data filtering, data combination, data correction, and parameter determination.

Data set construction

The data used in the case study is generated by two tools, namely, FRISO (Flexible Rail Infrastructure Simulation model of Operations) and ROBERTO. FRISO is a microscopic railway simulation tool that generates running times, track occupation, and signaling aspects, based on databases including infrastructure, timetable, and train routes. The outputs of FRISO are then fed to ROBERTO, in which the headway times between trains approaching shared infrastructure can be obtained. However, the outputs of both FRISO and ROBERTO are not tailored for the AG-based MILP model, and there can be redundant or missing data. Therefore, some adjustments have to be made.

In the first step, the desired points need to be filtered out from a large amount of data. Regarding the track section boundary (dutch: SPOORTAKSTUKBEGRENZER), points with type timetable point track (dutch: DRGLPT_SPOOR), signal (dutch: SEIN), or switch (dutch: WISSEL) are extracted. Regarding the timetable point track, there can be station tracks for train stops and station tracks for passing trains. The tracks for passing trains are not extracted from the original data since the passing can be represented by the entry and exit signals of a station track. Concerning the switches, they are extracted since headway requirements on some switches are provided by ROBERTO.

In the second step, useful information about the selected points is kept to simplify the data structure. The maintained information will be used to construct data frames of nodes, arcs, and trains. This information includes train name, track section boundary name, signal direction, dispatching area code, distance from the starting point, time instant of a train passing the track section boundary with the front end, and the minimum headway between a pair of trains at signal or switch.

Since there can be signals of a direction and opposite direction along the track, the corresponding signals of the running direction of the trains should be extracted. A block section is a part of the railway track between two signals, and the time of a train traversing the block section is called the trip time. However, the trip time of a train cannot be directly obtained. Therefore, they are calculated by subtracting the front-end passage time of a train at the entry signal of a block section from the front-end passage time of the train at the exit signal of the block section.

For the dwell time at a stop, the minimum dwell time at a major station is 60 s and the minimum dwell time at a short stop is 36 s. Since only the rigid departure times are given at short stops, the rigid arrival times are manually added, which are the rigid departure times minus the minimum dwell time of 36 s. For the freight, although no rigid departure and arrival times are given, the entry time and the exit time from the dispatching should be calculated based on the rigid passing time through a station.

In our model, the exact weight of a processing arc is the minimum trip time of a train over the block section. The exact value that is used as the weight of an alternative arc is the minimum headway time between trains minus the block running time of the first train. Regarding the passenger transfer time requirement, it is known that the transfer from train 6600-T-1 to train 6400-H-1 is a cross-platform transfer according to the TAD and station layout. Thus, the minimum required transfer time is 2 minutes. In this case, the weights of fixed arcs that represent dwelling and running, alternative arcs, and the connection arc can be obtained.

Data correction

After running the model for the first time, some inherent delays were found when no initial delays were added to trains. It means that the rigid timetable is not consistent with the data generated by FRISO and ROBERTO. To remove the inherent delays, the time instances in the rigid timetable are changed according to rescheduled times generated by the model. The corrected rigid timetables of trains running from Ehv to Bd, from Ht to Bd, and from Ehv to Ht are presented in Table C.1, Table C.2, and Table C.3, respectively.

In addition, the given trip times from the original data are the scheduled trip times. However, minimum trip times are required as the weights of fixed arcs in the model, which can be obtained by removing the running time supplements from actual trip times. In the case study, the minimum trip times are assumed 93% of the scheduled running times.

Regarding the minimum headway times, some headway times are missing due to the scope of the case study area. For instance, the entry of a train into the dispatching area can be represented by the departure from a station. Thus, there is no information about the signal before the station. In this case, trains can depart from the same platform at the station at the same time, which is impossible in reality. Besides, some headway times between trains passing a shared switch are also missing. To correct the data, headway times are manually added between pairs of train departures, and between pairs of trains passing shared switch as alternative arcs.

Parameter settings

After data processing, the last step before running the model is to assign values to all parameters. According to the TAD, the delay measurement point of train 6600-T-1 is station Ht and the largest accepted arrival delay u_i^{delay} is 7 minutes. It means that if the feeder train arrives at Ht more than 7 minutes later than the rigid arrival time, the passenger transfer has to be discarded. Besides, the latest departure time of the waiting train $u_i^{\text{departure}}$ at the transfer station Tb is 50:00, and 50:00 is the time in an hourly timetable. It indicates that if the transfer can only be maintained with the departure of train 6400-H-1 from Tb later than 50 minutes of each period, the transfer also has to be discarded. The weight of rescheduled time $e^{\text{rescheduled}}$ in the objective function is set to 100, indicating that the primary aim is to obtain the rescheduled times that have the least deviation from the rigid times. Then, flexibility is added and maximized on top of this rescheduling plan. The weight of departure flexibility $e^{\text{f, dep}}$ is set to 1 for now. It means that there is no particular preference between departure and arrival flexibility. The punctuality threshold p is set to 3 minutes based on the performance indicator in practice. M is set to 9999, which is sufficient in the current case study.

4.3. Model demonstration

To apply the model to the real-life railway network, some cases with different combinations of initial delays are tested. According to Yuan (2006), the statistical contribution of train delays fits the Weibull distribution. Yuan (2006) performed a statistical analysis of train operations based on train traffic data recorded at station Den Haag Holland Spoor, the Netherlands, and nearly 10000 trains were recorded in September 1999. The shape parameter k and the scale parameter λ are fine-tuned for different train series and tested using the K-S test. Two types of delays were investigated, namely, arrival delay and departure delay. Regarding departure delays, the average value of k is 1.2 and the average value of λ is 76 for IC trains. The average values of k and λ are 0.9 and 59 for SPR trains, respectively. With respect to arrival delays, the average value of k is 1.1 and the average value of λ is 97 for IC trains. The average value of k is 0.8 and the average value of λ is 102 for SPR trains. The average values of k and λ are used for all train series.

According to the data inputs of our model, the nodes that represent trains entering the dispatching area can be arrival nodes or departure nodes. The entry nodes of trains with id 1, 5, and 6 are departure nodes, and the entry nodes of trains with id 15, 17, 34, 43, 61, 64, 66, 70, and 79 are arrival nodes. Regarding train 18, no initial delays are added since the entry node of this train is neither a departure nor an arrival node.

After determining the shape and scale parameters, we let the model randomly generate the integer values of arrival and departure delays based on the Weibull distribution. In order to explore the potential relationship between input delays and output flexibility, 10 random cases are generated, which are indicated by R1 to R10. The computation time of the optimization process is queried. The initial delay inputs and the CPU time of each random case and the case without any initial delays are presented in Table 4.2. According to the model outputs, the average value of initial delay per train generated based on the Weibull distribution is 47.89 s, and the average CPU time of the optimization process is 1.12 s.

Table 4.2: Inputs of initial delays and the computation times for random cases and the case without initial delays

Train id	Rigid	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
1	0	76	135	113	113	138	138	108	121	63	30
5	0	108	53	99	56	64	58	72	77	71	324
6	0	33	68	135	68	64	98	135	104	35	122
15	0	54	1	26	47	27	12	12	7	66	78
17	0	64	12	18	61	18	34	45	48	46	76
18	0	0	0	0	0	0	0	0	0	0	0
34	0	5	67	39	29	80	58	10	58	67	45
43	0	53	18	19	49	70	18	54	37	27	40
61	0	71	21	51	68	79	75	46	40	71	75
64	0	28	15	8	22	13	9	16	51	5	53
66	0	58	26	11	44	14	55	29	18	27	9
70	0	8	11	31	21	45	3	46	74	6	24
79	0	26	80	69	5	4	29	39	17	48	58
Initial delay per train [s]	0.00	44.92	39.00	47.62	44.85	47.38	45.15	47.08	50.15	40.92	71.85
CPU time [s]	1.18	1.44	1.34	1.03	1.04	0.97	1.08	1.06	1.05	1.02	1.10

After optimization, the maximum timetable flexibility that can be added to each train of each random case can be obtained, which is presented in Table 4.3. At a stop, three types of timetable flexibility are added. To measure timetable flexibility, flexibility per stop is used as a performance indicator, which is calculated as the total flexibility added to trains over 45 stops in the network. By comparing the input delay per train with the output flexibility per stop for all cases shown in Table 4.2 and Table 4.3, a general trend can be observed that flexibility decreases as initial delays increase. However, there can be exceptions. By comparing random case 2 with random case 4, as the initial delay per train increases from 39.00 s to 44.85 s, the corresponding flexibility per stop increases from 222.82 s to 223.18 s. Regarding the flexibility added to each train, no intuitive relationship can be found with the value of initial delay per train. It means that the flexibility of each train can increase, decrease, or remain the same when the initial delay added to the train changes. Besides, the timetable flexibility of a train can even change when the initial delays remain the same. Those situations are possibly due to the interaction among trains, which needs further investigation.

To explicitly present the flexibility added to the timetable events, the flexible train path plot is generated. In a flexible train path plot, the colored solid lines indicate the rescheduled train paths of different types of trains. IC train paths are in blue, SPR train paths are in red, and GO train paths are in black.

Table 4.3: Results of timetable flexibility for random cases

Train id	Rigid	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
1	562.65	502.86	502.86	538.73	502.86	502.86	502.86	468.59	530.73	588.73	532.65
5	699.33	516.52	593.33	525.52	587.33	571.33	583.33	589.60	547.52	557.33	300.52
6	545.77	512.77	477.77	410.77	477.77	481.77	447.77	410.77	441.77	510.77	423.77
15	2515.65	2515.65	2515.65	2515.65	2515.65	2515.65	2515.65	2515.65	2515.65	2515.65	2515.65
17	497.52	436.33	497.52	454.33	497.52	497.52	497.52	497.52	495.33	497.52	326.41
18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
34	785.24	785.24	785.24	785.24	785.24	785.24	785.24	785.24	785.24	785.24	785.24
43	773.53	773.53	773.53	773.53	773.53	773.53	773.53	773.53	773.53	773.53	773.53
61	408.93	392.73	333.73	319.85	355.73	330.73	330.73	360.73	319.85	319.85	408.93
64	465.63	465.63	465.63	465.63	465.63	465.63	465.63	465.63	465.63	465.63	465.63
66	2119.56	2119.56	2051.12	2119.56	2119.56	2119.56	2119.56	2119.56	2119.56	2119.56	2119.56
70	429.07	429.07	497.52	429.07	429.07	429.07	429.07	429.07	429.07	429.07	429.07
79	533.18	533.18	533.18	533.18	533.18	533.18	533.18	533.18	533.18	533.18	533.18
Flexibility per stop [s]	229.69	221.85	222.82	219.36	223.18	222.36	221.87	221.09	221.27	224.36	213.65

The train paths in the rescheduling plan are plotted by connecting the rescheduled departure times from stations and the rescheduled arrival times at the following stations. Regarding timetable flexibility, they are represented by colored areas, which are plotted based on the earliest and the latest flexible arrival and departure times. The rigid train paths are indicated by grey solid lines. Regarding the trains that diverge from or merge into the plotting corridor, the times at the diverging and merging points are obtained using extrapolation based on the times at the last and the following stops. However, there can be overlaps between the thick train paths of two adjacent trains. The reason is that two trains are running on parallel tracks.

In Figure 4.2, the train paths with flexibility without any initial delays are presented. Two trains are indicated in Figure 4.2. Train 5 is not the fifth train in the train path plot, however, it indicates train 1900-H-1, whose ID is 5 according to the operational data. Regarding the departure events, it can be observed that only late departure flexibility is added. For arrival events, both early arrival flexibility and late arrival flexibility are added. In the case that no delays are experienced by trains, the timetable flexibility per stop reaches its maximum, which is 229.69 s. To showcase some changes in departure flexibility, early arrival flexibility, and late arrival flexibility, the train paths of two random cases are generated and presented as follows.

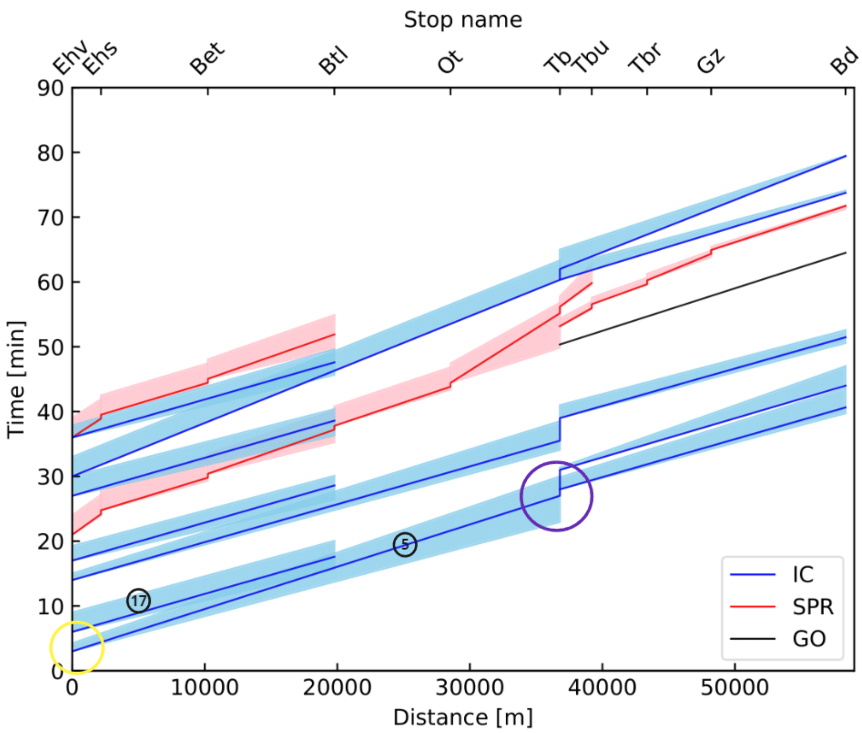


Figure 4.2: Train paths from Ehv to Bd without initial delays

By comparing the areas indicated yellow circles in Figure 4.2 and Figure 4.3, it can be observed that the departure flexibility of train 1 at station Ehv decreases. According to the numerical results of the optimization model, the departure flexibility that is added to train 1 is 74.81 s when there are no initial delays. However, it decreases to 0 s when the train experiences an initial delay of 99 s.

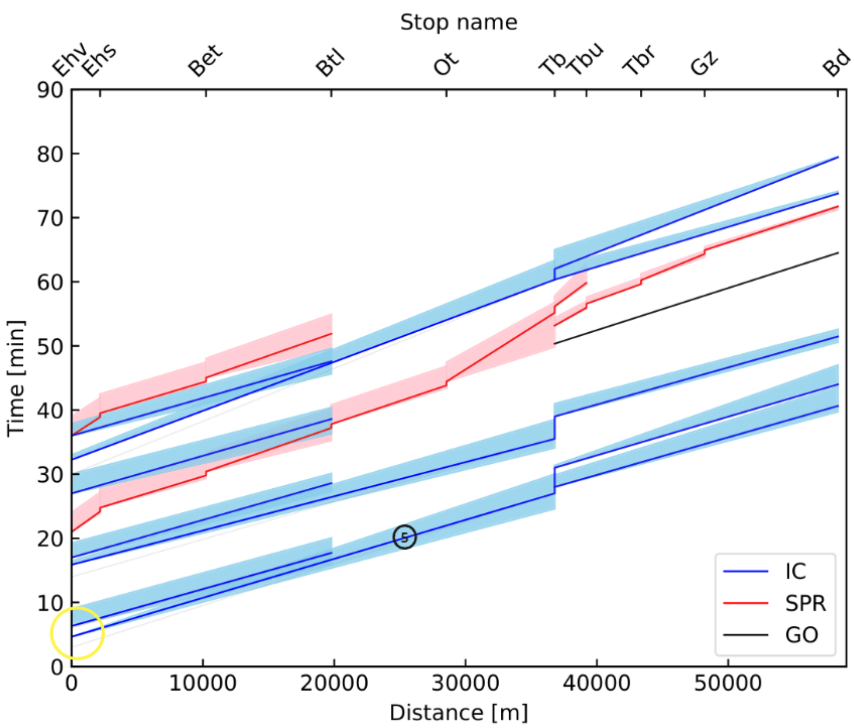


Figure 4.3: Train paths from Ehv to Bd of case R3

By comparing the areas indicated purple circles in Figure 4.2 and Figure 4.4, it can be observed that both the early and late arrival flexibility of train 1 at station Tb decrease. Regarding the early arrival flexibility, it is 242.69 s when no trains experience initial delays. However, it drops to 0 s when an initial delay of 324 s is added to train 1. With respect to the late arrival flexibility, it reaches the punctuality threshold of 180 s when there are no initial delays, and it decreases to 98.69 s after train 1 experiences an initial delay of 324 s. Besides, the order between trains 5 and 17 changes, indicating that our model is capable of reordering in real-life cases.

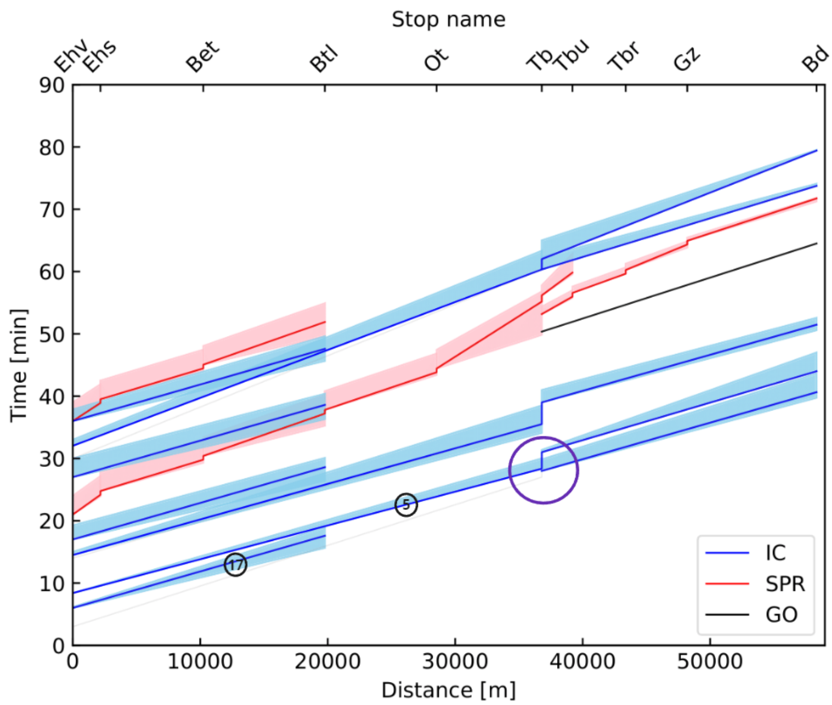


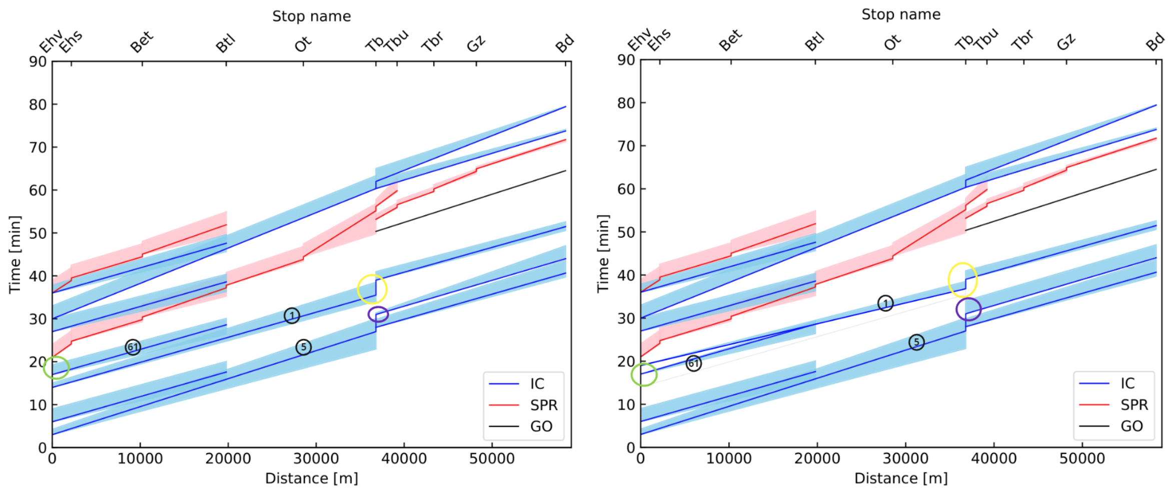
Figure 4.4: Train paths from Ehv to Bd of case R10

4.4. Relationship between initial delays and timetable flexibility

In the model demonstration, the assignment of timetable flexibility can have various changes with the changing initial delays. Those changes are possibly due to the interactions between trains. In this subsection, illustrative applications of the model to various inputs are presented. We aim to investigate the relationship between initial delays and each type of flexibility, such that the reasons for the changes can be revealed. In this analysis, $e^{\text{rescheduled}}$ is set to 100 and $e^{\text{f, dep}}$ is set to 1. In order to investigate how timetable flexibility changes with the different inputs of initial delays, initial delays from 0 to 10 minutes with an interval of 100s are added to a single train, and initial delays of all other trains are fixed to 0 s. To reveal the interactions between trains and the changes in departure flexibility, early arrival flexibility, and late arrival flexibility as initial delays change, we present and discuss some comparisons between the train paths. These comparisons involve different initial delays and are presented as follows.

A comparison between flexible train paths without initial delay and with 300 s initial delays assigned to train 3 is presented in Figure 4.5. Figure 4.5a presents the flexible train paths without initial delays, in which the rescheduled train paths are the same as the rigid ones. Flexible train paths with 300 s initial delays added to train 3 are presented in Figure 4.5b. Some typical areas that flexibility changes are highlighted by colored circles, which are discussed below.

By comparing the areas marked by green circles in both Figure 4.5a and Figure 4.5b, it can be easily observed that the rescheduled departure time of train 1 at Ehv increases. Due to the initial delay of 300 s added to train 1, the order between trains 1 and 61 passing a switch between Ehv and Ehs even changes. Besides, it can also be found that the departure flexibility of both trains 1 and 61 at Ehv decreases. According to the numerical results, the departure flexibility of train 1 at Ehv decreases from 59.79 s to 0 s. The reason is that train 1 is largely delayed, thus, the difference between the



(a) Flexible train paths from Ehv to Bd with 0 s initial delay added to trains in the network

(b) Flexible train paths from Ehv to Bd with 300 s initial delays added to train 1

Figure 4.5: Comparison between flexible train paths from Ehv to Bd without initial delays and with 300 s initial delays on train 1

rescheduled and the rigid departure times exceeded the punctuality threshold of 180 s. As a result, no flexibility can be added. Regarding train 61, its departure flexibility at Ehv is 134.67 s when no initial delay is added to train 1. When the order between trains 1 and 61 changes, train 61 should leave station Ehv and pass the shared switch as early as possible, such that the rescheduled times of train 1 have the least deviation from the rigid times. As a consequence, the departure flexibility of train 61 at Ehv is 0 s.

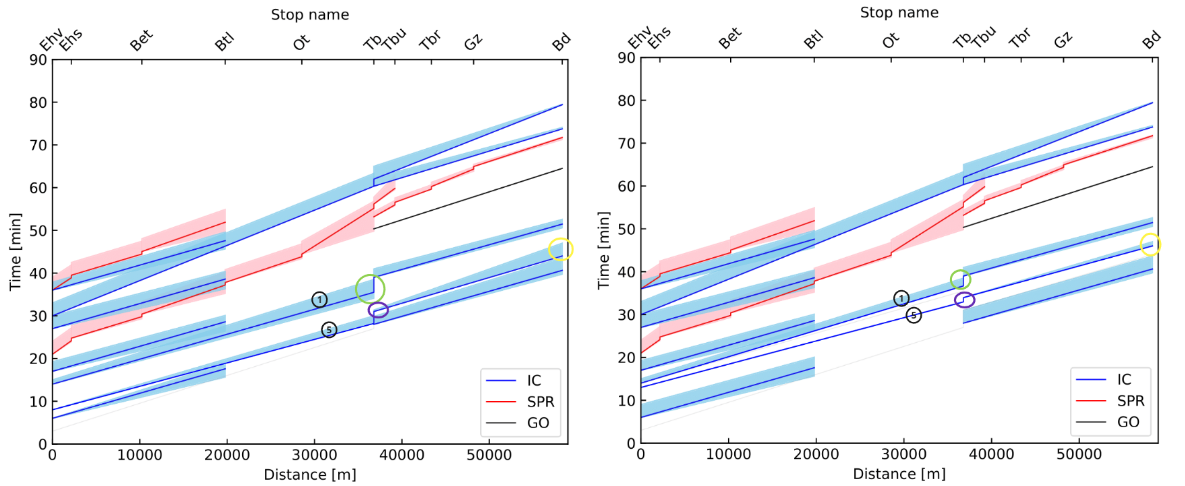
Regarding the areas indicated by yellow circles in Figure 4.5a and Figure 4.5b, it is obvious that both the arrival and departure flexibility of train 1 at Tb decrease. The initial delay of 300 s added to train 1 propagates to station Tb, which can be reflected by the fact that the rescheduled arrival time increases compared to the rigid one. When there are no initial delays, the arrival flexibility of train 1 at Tb is 180 s, and it decreases to 101.14 s when a 300 s delay is added to train 1. Since the arrival flexibility already already reached its maximum, as the rescheduled arrival time increases, the difference between the latest flexible time and the rescheduled time decreases, leading to the decrease in late arrival flexibility of train 1 at Tb. Regarding early arrival flexibility, the rescheduled arrival time of train 1 at Tb is already the earliest, meaning that the train cannot arrive earlier. Therefore, no early arrival flexibility can be added.

According to the areas highlighted by purple circles in Figure 4.5a and Figure 4.5b, the departure flexibility of train 5 at Tb increases. It is due to the interaction between trains 1 and 5 around Tb. As the initial delay of train 1 at Ehv propagates to Tb, the earliest arrival time of train 1 at Tb increases, leading to more buffer times between trains 1 and 5. Therefore, the departure flexibility increases from 21.83 s to 180 s.

There are some other conditions that also lead to similar flexibility changes. A comparison between flexible train paths with 300 s and 600 s initial delays added to train 5 is presented in Figure 4.6. Figure 4.6a presents the flexible train paths from Ehv to Bd with 300 s initial delays, and flexible train paths with 600 s initial delays on train 5 are presented in Figure 4.6b.

Regarding the areas surrounded by purple circles, the departure flexibility of train 5 at Tb decreases. According to the numerical results, it decreases from 21.83 s to 0 s. Due to the 600 s initial delays, the rescheduled departure time of train 5 at Tb is largely affected compared to the case scenario with 300 s initial delays. Therefore, the difference between the rescheduled and the rigid departure time of train 5 from Tb exceeded the punctuality threshold. As a consequence, no departure flexibility can be added.

By comparing the areas indicated by green circles, it can be found that both early and late arrival flexibility decreases. Due to the delay of train 5, train 1 is affected, leading to the late arrival at Tb. According to the numerical results, the early arrival flexibility of train 1 at Tb decreases from 84.42 s to 0 s, since the rescheduled arrival time is the earliest possible arrival time. Regarding the late arrival flexibility of train 1 at Tb, it decreases from 180 s to 108.94 s. The reason is that the rescheduled arrival



(a) Flexible train paths from Ehv to Bd with 300 s initial delays added to train 5

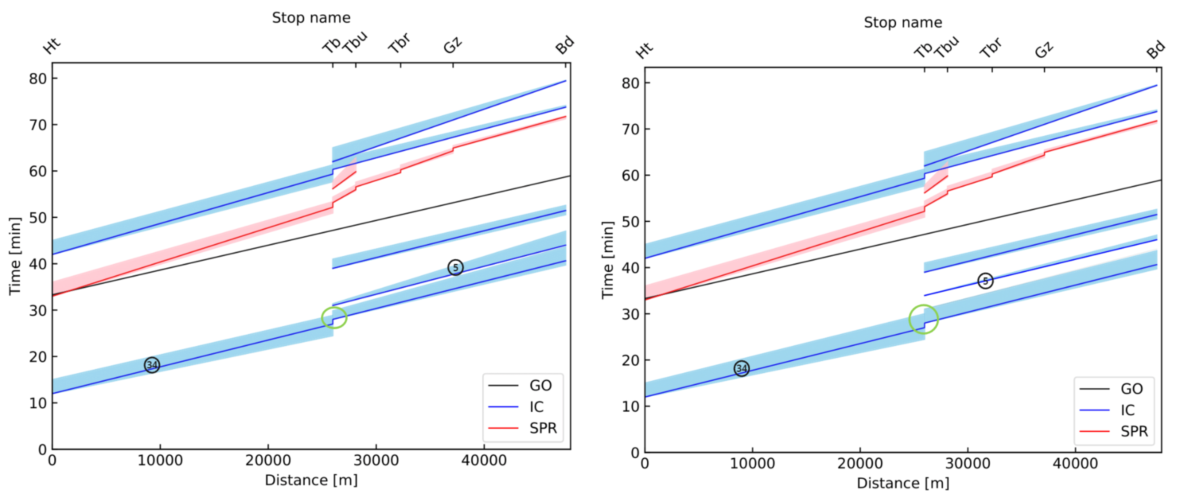
(b) Flexible train paths from Ehv to Bd with 600 s initial delays added to train 5

Figure 4.6: Comparison between flexible train paths from Ehv to Bd with 300 s and 600 s initial delays on train 5

increases due to the delay of train 5, considering the headway constraints, train 5 has to arrive later than the rigid time, leading to a decrease in late arrival flexibility.

Regarding the areas marked by yellow circles, it is obvious that the late arrival flexibility of train 5 at Bd decreases. The model results show that the late arrival flexibility decreases from 180 s to 58.79 s. When train 5 experiences an initial delay of 300 s, the delay is not propagated to Bd, thus, the late arrival flexibility equals the punctuality threshold. When the initial delay increases to 600 s, the rescheduled arrival time of train 5 at Bd increases, leading to a decrease in late arrival flexibility.

Considering the same initial delays, we present the train paths of the corridor between Ht and Bd in Figure 4.7. Figure 4.7a presents the flexible train paths from Ht to Bd with 300 s initial delays, and flexible train paths with 600 s initial delays on train 5 are presented in Figure 4.7b. Regarding the areas indicated by green circles, it can be observed that both the late arrival flexibility and departure flexibility of train 34 at Tb increase. The reason is that the rescheduled departure time of train 5 increases due to the increasing initial delays. Therefore, the buffer times between trains 5 and 34 increase, leaving more space for train 34 to arrive and depart from Tb later than the rigid times. According to the model output, both the late arrival flexibility and the departure flexibility of train 34 at Tb increase from 110.37 s to 180 s.

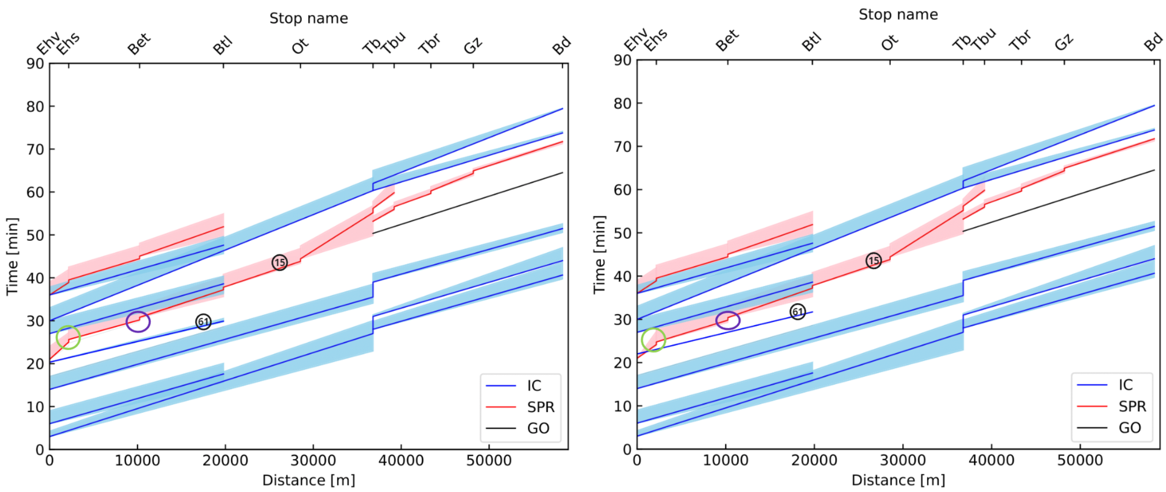


(a) Flexible train paths from Ht to Bd with 300 s initial delays added to train 5

(b) Flexible train paths from Ht to Bd with 600 s initial delays added to train 5

Figure 4.7: Comparison between flexible train paths from Ht to Bd with 300 s and 600 s initial delays on train 5

A comparison between flexible train paths from Ehv to Bd with 500 s and 600 s initial delays added to train 61 is presented in Figure 4.8. Figure 4.8a presents the flexible train paths from Ehv to Bd with 500 s initial delays, and flexible train paths with 600 s initial delays on train 61 are presented in Figure 4.8b.



(a) Flexible train paths from Ehv to Bd with 500 s initial delays added to train 61

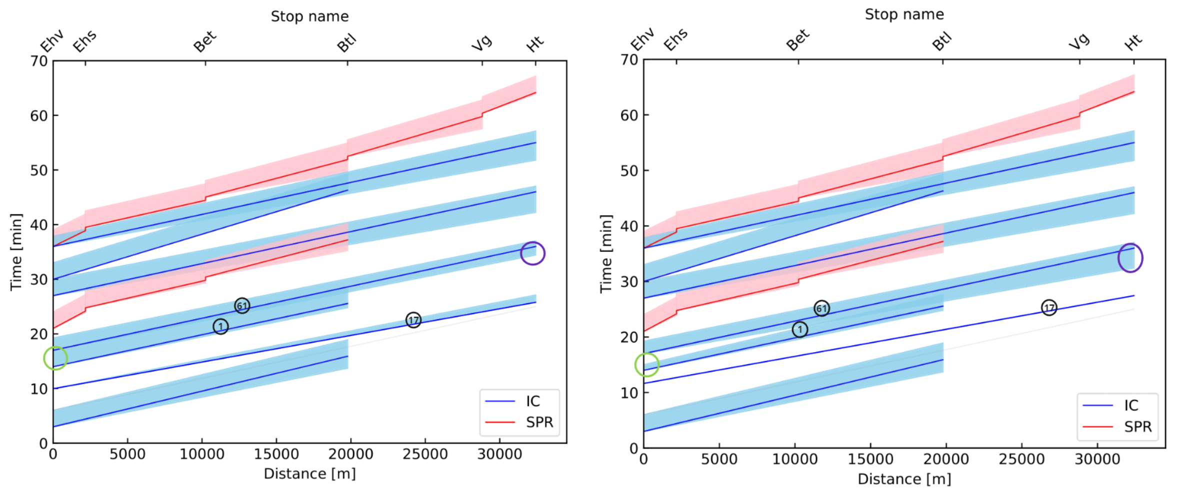
(b) Flexible train paths from Ehv to Bd with 600 s initial delays added to train 61

Figure 4.8: Comparison between flexible train paths from Ehv to Bd with 500 s and 600 s initial delays on train 61

By looking at the train paths of trains 61 and 15, it can be observed that the order between the two trains changes as the initial delay added to train 61 increases from 500 s to 600 s. Regarding the areas highlighted by green circles in Figure 4.8a and Figure 4.8b, it can be found that the late arrival flexibility of train 15 at Ehs increases with the increasing initial delays on train 61. In the corridor between Ehv and Btl, trains 15 and 61 depart from different platforms in Ehv and approach a switch between Ehv and Ehs. After passing the shared switch, the two trains run on parallel tracks. When a 500 s initial delay is added to train 61, train 61 precedes train 15 at the shared infrastructure, and the minimum headway time passing the switch should be respected between the two trains. Therefore, train 15 can only arrive at Ehs later than the rigid arrival time. When an initial delay of 600 s is added to train 61, train 15 passes the shared switch first, indicating that the rescheduled arrival time of train 15 at Ehs equals the rigid one. Thus, the difference between the latest arrival time and the rescheduled arrival time increases in Figure 4.8a compared to that in Figure 4.8b. According to the numerical results, the late arrival flexibility of train 15 at Ehs increases from 128.57 s to 180 s, when the initial delay added to train 61 increases from 500 s to 600 s.

With respect to the areas surrounded by purple circles, it is obvious that the early arrival flexibility of train 15 increases as the initial delay on train 61 increases from 500 s to 600 s. In Figure 4.8a, the earliest possible arrival time of train 15 at Bet is later than the rigid one due to the delay of train 61, as well as the passing sequence between the two trains through the shared switch. However, in Figure 4.8b, it is not necessary for train 15 to wait for train 61 to pass the shared switch first. It indicates that train 15 can arrive at Bet as early as possible. Therefore, the early arrival flexibility of train 15 at Bet increases from 0 s to 28.45 s.

A comparison between flexible train paths from Ehv to Ht with 400 s and 500 s initial delays added to train 17 is presented in Figure 4.9. Figure 4.9a presents the flexible train paths from Ehv to Bd with 400 s initial delays, and flexible train paths with 500 s initial delays on train 17 are presented in Figure 4.9b. Regarding the areas indicated by green circles, it can be observed that the departure flexibility of train 1 at Ehv decreases, however, the corresponding rescheduled departure time remains the same. According to the area marked by purple circles, the early arrival flexibility of train 61 increases, while the corresponding rescheduled arrival time at Ht remains the same. These changes should not occur since trains 1 and 61 are not influenced by the increasing initial delay on train 17. Regarding these changes, they are possibly due to the setting of the weight of departure flexibility. In this text, we set $e^{\mathrm{f,\,dep}}$ to 1, indicating that there is no particular preference for the distribution of timetable flexibility. Thus, the model just helps find the results that lead to maximum total timetable flexibility on top of the rescheduling plan, and the exact distribution between departure and arrival flexibility is not constrained.



(a) Flexible train paths from Ehv to Ht with 400 s initial delays added to train 17

(b) Flexible train paths from Ehv to Ht with 500 s initial delays added to train 17

Figure 4.9: Comparison between flexible train paths from Ehv to Ht with 400 s and 500 s initial delays on train 17

Based on the results of cases in the model demonstration, as well as the investigation of the relationship between initial delays and timetable flexibility. It can be concluded that timetable flexibility basically decreases with the increasing amount of initial delays added to trains in the network. The reason is that our model finds the rescheduled times that have the least deviation from the rigid times. Thus, the rescheduled train paths are closer to each other compared to the rigid train paths, such that the deviation can be minimized, leading to fewer solution spaces to add timetable flexibility on top of the rescheduling plan. In addition, the solution spaces of late arrival flexibility and departure flexibility can only be found within the punctuality threshold. It means that late arrival and departure flexibility decreases when a timetable event is delayed in most cases. In addition to the decreases in timetable flexibility with the increasing initial delays, special cases can be found, in which flexibility increases when more initial delays are added. They are mainly due to the interaction between trains, for instance, the delay of a train might lead to more buffer times for the former train to arrive and depart later than the rigid timetable.

4.5. Relationship between departure flexibility weight and timetable flexibility

In the analysis exploring the relationship between initial delays and timetable flexibility, special scenarios were found, in which timetable flexibility is randomly distributed. In this sensitivity analysis, we aim to adjust the value of departure flexibility weight $e^{\text{f, dep}}$, such that investigate the impacts of $e^{\text{f, dep}}$ on the results of timetable flexibility, as well as the trade-off between different types of timetable flexibility.

In this analysis, the weight of rescheduled times $e^{\text{rescheduled}}$ is still set to 100, indicating that timetable flexibility is added on top of the rescheduling plan. Regarding $e^{\text{f, dep}}$, the values of 1.1, 1.0, and 0.9 are tested for the case without initial delays, as well as the random cases that are generated based on Weibull distribution in section 4.3.

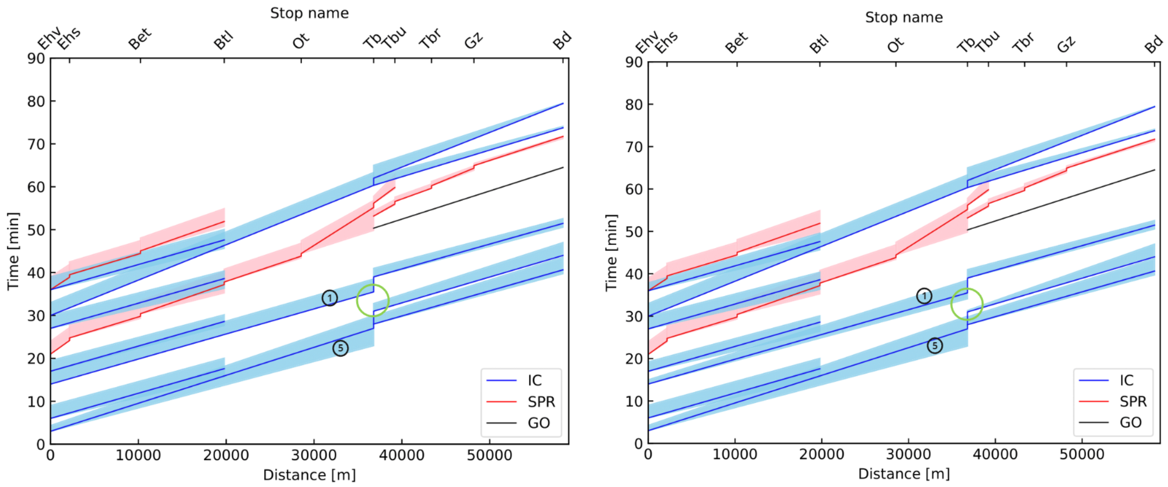
The results of each type of flexibility for all test cases are presented in Table 4.4. It can be found that, for each test case, departure flexibility reaches its maximum when $e^{\text{f, dep}}$ is 1.1. Regarding arrival flexibility, early arrival flexibility is maximized when $e^{\text{f, dep}}$ equals 0.9. However, late arrival flexibility remains the same as the weight of departure flexibility changes in each case. For the results when $e^{\text{f, dep}}$ is set to 1.0, early arrival flexibility is maximized in the rigid case and random cases R1, R4, R5, R6, and R10. For all other test cases, the results lead to the maximum total timetable flexibility, however, neither early arrival flexibility nor late arrival flexibility is maximized.

In order to present and discuss the trade-off between different types of flexibility, the train paths of the rigid case are generated, in which no initial delays are experienced by trains. A comparison

Table 4.4: Results of each type of flexibility with $e^{\text{f, dep}}$ equals 1.1, 1.0, and 0.9 for the case without initial delays and all random cases

		$e^{\text{f, dep}}$		
Test cases	Flexibility types	1.1	1.0	0.9
Rigid	Early arrival [s]	2196.04	2488.20	2488.20
	Late arrival [s]	3718.78	3718.78	3718.78
	Departure [s]	4421.26	4129.10	4129.10
R1	Early arrival [s]	2060.04	2330.80	2330.80
	Late arrival [s]	3718.78	3718.78	3718.78
	Departure [s]	4204.26	3933.49	3933.49
R2	Early arrival [s]	2143.04	2291.54	2359.99
	Late arrival [s]	3718.78	3718.78	3718.78
	Departure [s]	4165.26	4016.75	3948.30
R3	Early arrival [s]	2078.04	2275.92	2311.80
	Late arrival [s]	3718.78	3718.78	3718.78
	Departure [s]	4074.26	3876.37	3840.49
R4	Early arrival [s]	2140.04	2378.99	2378.99
	Late arrival [s]	3718.78	3718.78	3718.78
	Departure [s]	4184.26	3945.30	3945.30
R5	Early arrival [s]	2132.04	2345.99	2345.99
	Late arrival [s]	3718.78	3718.78	3718.78
	Departure [s]	4155.26	3941.30	3941.30
R6	Early arrival [s]	2138.04	2351.99	2351.99
	Late arrival [s]	3718.78	3718.78	3718.78
	Departure [s]	4127.26	3913.30	3913.30
R7	Early arrival [s]	2124.04	2333.72	2367.99
	Late arrival [s]	3718.78	3718.78	3718.78
	Departure [s]	4106.26	3896.57	3862.30
R8	Early arrival [s]	2119.04	2319.92	2347.80
	Late arrival [s]	3718.78	3718.78	3718.78
	Departure [s]	4119.26	3918.37	3890.49
R9	Early arrival [s]	2125.04	2328.11	2413.99
	Late arrival [s]	3718.78	3718.78	3718.78
	Departure [s]	4252.26	4049.18	3963.30
R10	Early arrival [s]	1958.53	2245.51	2245.51
	Late arrival [s]	3637.47	3637.47	3637.47
	Departure [s]	4018.15	3731.18	3731.18

between flexible train paths from Ehv to Bd when $e^{\text{f, dep}}$ equals 1.1 and 0.9 is presented in Figure 4.10. Figure 4.10a presents the flexible train paths with $e^{\text{f, dep}}$ equals 1.1, and flexible train paths with $e^{\text{f, dep}}$ equals 0.9 is shown in Figure 4.10b.



(a) Flexible train paths from Ehv to Bd without initial delay with $e^{\text{f, dep}}$ is 1.1

(b) Flexible train paths from Ehv to Bd without initial delay with $e^{\text{f, dep}}$ is 0.9

Figure 4.10: Comparison between flexible train paths from Ehv to Bd when $e^{\text{f, dep}}$ equals 1.1 and 0.9

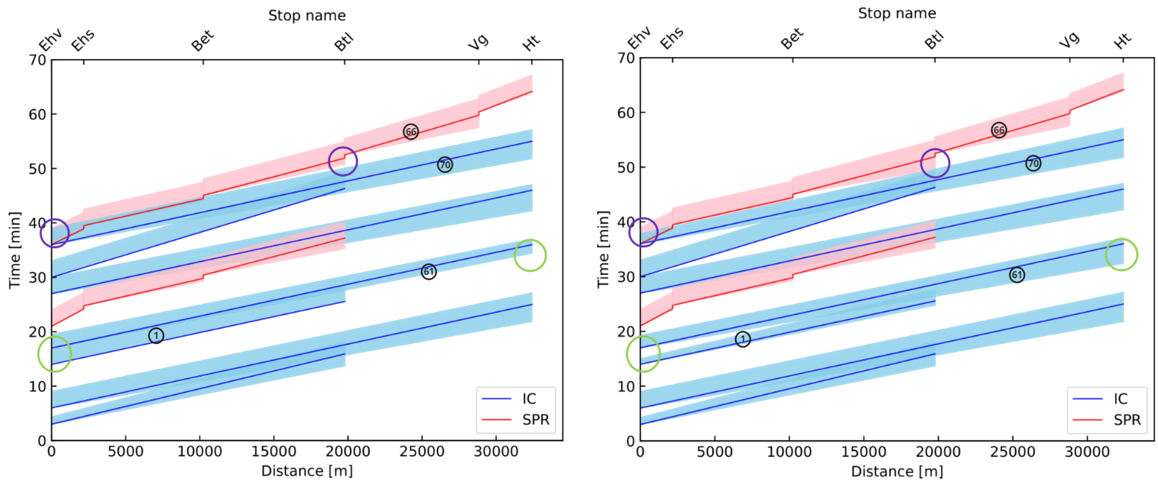
By comparing the areas marked by green circles, it can be observed that the departure flexibility of train 5 at Tb decreases, and the early arrival flexibility of train 1 at Tb increases when $e^{\text{f, dep}}$ changes from 1.1 to 0.9. Trains 1 and 5 run on the same track in the corridor between Ehv and Tb, and they stop at the same platform in Tb. The decrease in the departure flexibility of train 5 at Tb indicates that more buffer times are left for train 1, leading to as early arrival of train 1 at Tb as possible. According to the numerical results, the early arrival flexibility of train 1 at Tb increases from 0 s to 84.24 s as $e^{\text{f, dep}}$ changes from 1.1 to 0.9, and the departure flexibility of train 5 at Tb decreases from 106.25 s to 21.83 s. When arrival flexibility is preferred over departure flexibility, there is still 21.83 s departure flexibility. This is because the maximum early arrival flexibility of train 1 is 84.24 s, and it cannot arrive any earlier. Therefore, some buffer times are left for train 5 to have a late departure.

Apart from the scenario where a trade-off in different types of flexibility exists between two adjacent trains at the same station, there are some other scenarios worth exploring. A comparison between flexible train paths from Ehv to Ht when $e^{\text{f, dep}}$ equals 1.1 and 0.9 is presented in Figure 4.11. Figure 4.11a presents the flexible train paths with $e^{\text{f, dep}}$ equals 1.1, and flexible train paths with $e^{\text{f, dep}}$ equals 0.9 is shown in Figure 4.11b.

Regarding the areas surrounded by purple circles, it is obvious that the departure flexibility of train 70 at Ehv decreases, while the early arrival flexibility of train 66 at Btl increases as $e^{\text{f, dep}}$ changes from 1.1 to 0.9. Trains 70 and 66 run parallel in the corridor between Ehv and Bet, and they merge into the same track at a switch upstream to Btl, then share the same block sections until Ht. As train 70 departs earlier from Ehv, the rescheduled times of train 70 passing signals also decrease, resulting in more buffer times left for train 66 to pass the shared infrastructure earlier. Therefore, the arrival time of train 66 at Btl can be much earlier. According to the numerical results, the maximum departure flexibility of train 70 at Ehv is 180 s, and the maximum early arrival flexibility of train 66 at Btl is 136.20 s.

Some changes similar to the areas in purple circles can also be observed in the areas highlighted by green circles, in which the departure flexibility of train 1 at Ehv decreases, while the early arrival flexibility of train 61 at Ht increases as $e^{\text{f, dep}}$ changes from 1.1 to 0.9. The reason is still the interactions between the two trains traversing through shared block sections between stops. Based on the model outputs, the maximum departure flexibility that can be added to train 1 at Ehv is 180 s, and the maximum early arrival flexibility of train 61 at Ht is 213.60 s.

Based on this sensitivity analysis, it can be concluded that the trade-off only exists between early arrival flexibility and late arrival flexibility. In a station, the platform track is protected by an entry signal and an exit signal, a minimum time should be respected between the leaving of the preceding train from the block section and the entry of the following train into the block section. Regarding the preceding



(a) Flexible train paths from Ehv to Ht without initial delay with $e^{\mathrm{f,\,dep}}$ is 1.1
(b) Flexible train paths from Ehv to Ht without initial delay with $e^{\mathrm{f,\,dep}}$ is 0.9

Figure 4.11: Comparison between flexible train paths from Ehv to ht when $e^{\mathrm{f,\,dep}}$ equals 1.1 and 0.9

train, the arrival and departure at the station are closely related since the minimum dwell time should be respected. If the departure flexibility of a train at a station decreases, the late arrival flexibility decreases correspondingly to meet the dwell time requirement. Although the early arrival flexibility of the following train increases, the overall timetable flexibility decreases, which contradicts our objective function. Therefore, to maximize the overall timetable flexibility added to trains in the selected dispatching area, the late arrival flexibility remains the same when changing the weight of departure flexibility. When $e^{\mathrm{f,\,dep}}$ is greater than 1, it indicates that departure flexibility is preferred over arrival flexibility, leading to the maximum amount of departure flexibility in compromise of the decreasing early arrival flexibility. When $e^{\mathrm{f,\,dep}}$ is less than 1, flexibility is added to arrival events, resulting in the maximum early arrival flexibility and the minimum departure flexibility.

4.6. Relationship between punctuality threshold and timetable flexibility

After knowing the relationship between initial delays and timetable flexibility, as well as the relationship between the weight of departure flexibility and timetable flexibility. We also want to investigate the impacts of the punctuality threshold on the amount of flexibility that can be added to the trains in the selected network. In the current practice of ProRail, the punctuality threshold p is set to 180 s. It indicates that the difference between late arrival time and rigid arrival time, as well as the difference between late departure time and rigid departure time, cannot exceed the threshold. In this sensitivity analysis, we aim to gradually increase the value of the punctuality threshold, and then investigate what is the maximum time that can be added to the trains as flexibility in the selected dispatching area.

Different values of the punctuality threshold are tested, namely, from $0.5p$ to $5p$ with an increment of $0.5p$, and the results are presented in Table 4.5. It can be found that flexibility per stop increases with the increasing punctuality threshold. The reason is that as the value of the punctuality threshold increases, trains are allowed to arrive and depart from stops much later, as long as the headway requirements are respected. Besides, flexibility increment is also introduced, which is calculated as the difference between the flexibility per stop of adjacent tested values of the punctuality threshold. It can be observed that the flexibility increment decreases as the punctuality threshold increases. However, the flexibility increment remains 22 s when the punctuality threshold is greater than $3p$.

Table 4.5: Results of flexibility per stop with different punctuality thresholds and the flexibility increment

Punctuality threshold [s]	**0.5p**	**p**	**1.5p**	**2p**	**2.5p**	**3p**	**3.5p**	**4p**	**4.5p**	**5p**
Flexibility per stop [s]	160.75	229.69	287.66	332.28	368.52	397.44	419.44	441.44	463.44	485.44
Flexibility increment [s]	-	68.94	57.97	44.62	36.24	28.92	22.00	22.00	22.00	22.00

To explicitly present the changes in flexibility per stop with the punctuality threshold, the results are illustrated in a line chart, which is shown in Figure 4.12. When the punctuality threshold is larger than $3p$, the line becomes a straight line. The reason is that the distribution of timetable flexibility is regulated by headway requirements, trip time requirements, and the punctuality threshold simultaneously. As the punctuality threshold increases, timetable flexibility does not necessarily increase since flexibility may have reached the maximum at some stops. In our case study, only a part of the Dutch railway network is selected, and not all trains are incorporated. Therefore, the late arrival flexibility and departure flexibility of the last trains running in each corridor are not regulated by headway requirements. As a consequence, the late arrival flexibility and departure flexibility of those trains equals the punctuality threshold, leading to unchanged flexibility increments when the punctuality threshold is larger than $3p$.

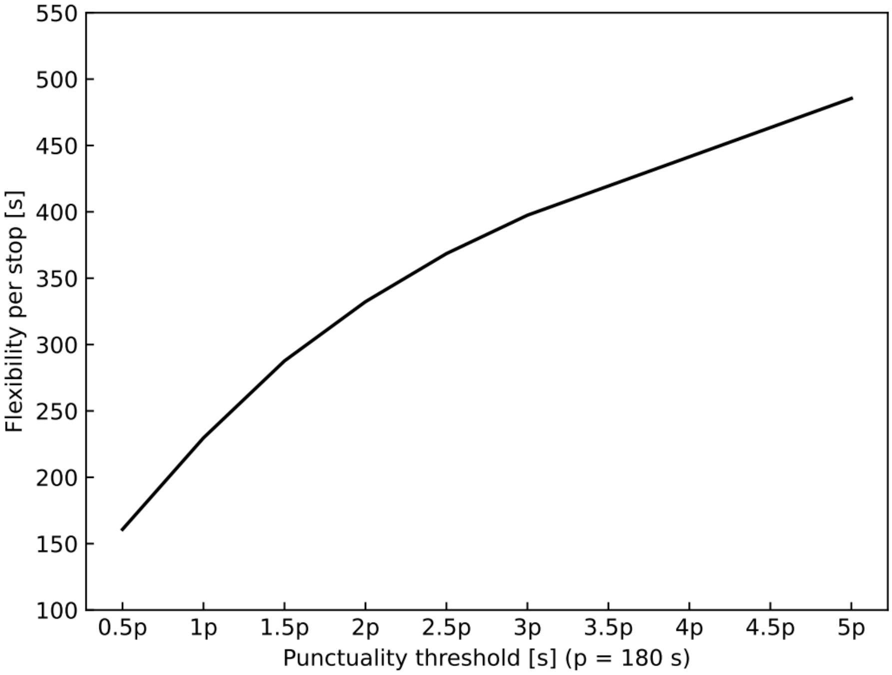


Figure 4.12: Change in flexibility per stop with increasing punctuality threshold

To conclude, the increasing rate of flexibility decreases when the punctuality threshold increases from $0.5p$ to $3p$. When the punctuality threshold equals $3p$, the buffer times between trains are fully distributed as timetable flexibility. As the punctuality threshold further increases, the flexibility increment remains the same since flexibility is added to the last trains running in each corridor.

5
Conclusions and recommendations

The current research into adding timetable flexibility to the Real-Time Traffic Plan (RTTP) in railway traffic management has been carried out to investigate the performance of timetable flexibility. An Alternative Graph (AG)-based Mixed Integer Linear Programming (MILP) model was constructed to minimize the total delay in real-time railway traffic management while maximizing timetable flexibility. Both model verification and model validation are performed by means of a toy network and a real-life case study on a part of the Dutch railway network around Eindhoven-Tilburg-'S-Hertogenbosch area. Conclusions can be drawn by answering all research questions and the main research question. Besides, the implementations of our research on the current practice of ProRail are discussed. In addition, recommendations for future research are provided.

5.1. Answer to the sub-research questions

In this section, four research questions are answered based on the findings from the corresponding chapter.

1. What is the state-of-the-art on real-time railway traffic management, and what is the state-of-the-practice in the Netherlands?

   According to the literature review, the mainstream rescheduling models that are used in railway traffic disturbance management include Mixed Integer Linear Programming (MILP), Integer Programming (IP), and the AG model. Besides, some other models are also used, such as Constraint Programming (CP), knowledge-based, and learning-based rescheduling methods. Among them, the learning-based approaches have been a hot research topic in recent years.

   With respect to adding flexibility to the timetable, it was studied at the tactical planning level, where a Flexible Periodic Event Scheduling Problem (FPESP) model was proposed. Attempts were also made to add flexibility at the operational planning level. Train Path Envelope belongs to train operations, which is the downstream of traffic management. It can be considered as adding flexibility to individual train. Based on the previous studies, the research gap is that there is no consensus on the definition of timetable flexibility, and no study aims to add and maximize timetable flexibility on top of the rescheduling plan in real-time railway traffic management.

   Regarding the state-of-the-practice, railway traffic disturbance management is the work of ProRail. The main aim is to keep the plan up to date considering the traffic perturbations. In disturbance management, five stages are included, namely, traffic state monitoring, traffic state prediction, conflict detection, conflict resolution, and plan updating. The traffic rescheduling models are used in conflict resolution to solve potential conflicts. In the current practice, RTTP with exact times is generated and provided to train dispatchers and signalers, then sent to trains and Automatic Route Setting (ARS), respectively. However, time points are hard for train drivers to follow due to real-time traffic uncertainties. Thus, a new RTTP should be generated by conducting the computationally expensive algorithms to handle traffic disturbances. By adding flexibility, a rescheduling plan with event time slots is provided to train dispatchers, according to which they can determine departure and arrival times to handle traffic uncertainties.

2. How can timetable flexibility be defined and modeled in the CDR process to output an optimal rescheduling plan that efficiently addresses traffic disturbances and real-time uncertainties?

   The definition of timetable flexibility can be given based on two other similar concepts, namely, timetable robustness and timetable resilience. Timetable flexibility is a means to achieve resilience. It is defined as the ability of a timetable to be easily modified to withstand small disturbances and absorb delays, as well as offer a larger solution space in the application of dispatching measures (retiming, reordering, rerouting) to solve larger disturbances without changing the given (re)scheduled timetable.

   In order to model timetable flexibility concisely, the exact stage that flexibility is added, as well as its connection to other stages should be specified. Thus, a framework for real-time railway operations is constructed. Besides, the AG model based on a MILP formulation is selected and constructed, in which retiming and reordering control measures are considered. In our model, three types of timetable flexibility are incorporated, namely early arrival flexibility, late arrival flexibility, and departure flexibility. Early arrival flexibility is the time difference between the rescheduled arrival time and the earliest flexible arrival time. Late arrival flexibility is the difference between the latest flexible arrival time and the rescheduled arrival time. Departure flexibility is defined as the difference between the latest flexible departure time and the rescheduled departure time. According to the performance indicator of train punctuality that is officially reported by ProRail, train punctuality for passenger trains is measured by the percentage of arrivals where the difference between the rigid arrival time and the realized arrival time is less than the punctuality threshold of 3 minutes. Since punctuality is a measure of lateness, the concept of the punctuality threshold is extended to late departure flexibility and late arrival flexibility. The values of late arrival flexibility and departure flexibility are regulated by the punctuality threshold, while early arrival flexibility is not constrained by the threshold, which means very early arrival times at stations are allowed.

3. How can the performance of the model be evaluated?

   The model was verified by means of a toy network, which refers to a simplified and smaller version of a real-life railway network. The results of the toy case indicate that the control measure of retiming and reordering was functioning well in our model. Besides, when no initial delays are added to trains, timetable flexibility is added on top of the rigid timetable. When trains experience initial delays, flexibility is added to the rescheduling plan that has the least deviation from the rigid timetable. After the model verification, a case study is conducted in the area around four major stations, namely, 's-Hertogenhosch (Ht), Tilburg (Tb), Breda (Bd), and Eindhoven (Ehv). Passenger transfer between trains from different directions is designed at station Tb. Apart from the four major stations, there are also stops in between, namely, Eindhoven Philips Strijp (Ehs), Best (Bet), Btl, Oisterwijd (Ot), Tilburg Reeshof (Tbr), Tilburg Universiteit (Tbu), Gilze Rijen (Gz), and Vught (Vg). 13 trains are incorporated in our case study, including 9 intercity trains (IC), 3 sprinter trains (SPR), and 1 freight train (GO). To investigate the relationship between input delays and output flexibility, 10 sets of initial delays are randomly generated based on Weibull distribution. The values of the shape parameter k and the scale parameter λ are specified for departure delays of IC, departure delays of SPR, arrival delays of IC, and arrival delays of SPR, respectively. In this case study, our model includes 3280 continuous variables, 1099 binary variables, and 22484 constraints. After applying our model to the random cases, it can be obtained that the average computation time of the optimization process is 1.12 s.

4. How can the performance of timetable flexibility be analyzed?

   In the model demonstration, the case without initial delays, as well as 10 random cases that are generated based on Weibull distribution are tested. To measure timetable flexibility, flexibility per stop is used as a performance indicator, which is calculated as the total flexibility added to trains over 45 stops in the network. According to the results, a general trend can be found that flexibility per stop decreases with increasing initial delays per train, however, there are exceptions. Regarding the flexibility added to each train, no intuitive relationship can be found with the initial delays added to the corresponding train.

In order to investigate the influencing factors of timetable flexibility, one illustrative application and two sensitivity analyses were conducted. In order to explore the impacts of initial delays on each type of flexibility, initial delays from 0 to 10 minutes with an interval of 100s are added to a single train, and initial delays of all other trains are fixed to 0 s. Based on the results, it can be concluded that timetable flexibility basically decreases as the initial delay increases. It is because the rescheduled train paths are closer to each other to recover the initial delays, leading to fewer solution spaces to add flexibility on top of the rescheduling plan. However, exceptions can be found, in which flexibility increases with the increasing initial delays. The reason is the interaction between trains. In these scenarios, the delay of a train leads to more buffer times for the former train to arrive and depart later than the rescheduled time.

Apart from the changes in timetable flexibility due to the changing initial delays and the interaction between trains, some dynamic changes in departure and arrival flexibility were also found. It is because $e^{\text{f, dep}}$ is set to 1, indicating that there is no preference for the distribution of timetable flexibility. In the first sensitivity analysis, $e^{\text{f, dep}}$ equals 1.1 and 0.9 are also tested for the case without initial delays and the 10 random cases. It was found that a trade-off only exists between departure flexibility and early arrival flexibility, and late arrival flexibility remains the same while changing the value of $e^{\text{f, dep}}$. When $e^{\text{f, dep}}$ is greater than 1, departure flexibility reaches its maximum. When $e^{\text{f, dep}}$ is less than 1, early arrival flexibility reaches the maximum. If the departure flexibility of a train at a station decreases, the late arrival flexibility decreases correspondingly to meet the dwell time requirement. Although the early arrival flexibility of the following train increases, the overall timetable flexibility decreases, which results in a lower objective value. Therefore, to maximize the overall timetable flexibility added to trains in the selected dispatching area, the late arrival flexibility remains the same when changing the weight of departure flexibility.

In the second sensitivity analysis, we change the exact value of the punctuality threshold, such that investigate the impacts of the punctuality threshold on the amount of flexibility that can be added to the trains in the case study network. According to the results, flexibility per stop increases with the increasing value of the punctuality threshold. It is because trains are allowed to arrive and depart from stops much later as the punctuality threshold increases, as long as other operational constraints are met. In addition, a flexibility increment is introduced, which is the difference between the flexibility per stop of adjacent tested punctuality threshold. It reflects the increasing rate of flexibility as the punctuality threshold gradually increases. Based on the results, the flexibility increment decreases with the punctuality threshold increases from $0.5p$ to $3p$, since flexibility does not necessarily increase since it may have reached the maximum at some stops. In our case study, there are some trains that are not regulated by headway constraints since not all trains are incorporated. When the punctuality threshold is greater than $3p$, the late arrival flexibility and departure flexibility are added to those trains.

5.2. Answer to the main research question

In this section, the main research question is answered based on the entire research.

- *How to introduce and optimize flexible event times to the rescheduling plan in railway traffic management?*

  Firstly, the definition of timetable flexibility was specified, as well as its connection with the punctuality threshold from ProRail. Then, the range of timetable flexibility and the conditions of its application were identified. There are three types of timetable flexibility, namely, early arrival flexibility, late arrival flexibility, and departure flexibility. Thereinto, only late arrival flexibility and departure flexibility should respect the punctuality threshold since punctuality is a measurement of lateness. There is no limit on early arrival flexibility, indicating very early arrivals are allowed in our model.

  Then, an AG-based MILP model is proposed based on the definitions regarding timetable flexibility. The primary objective of our model is to find the rescheduling plan that has the least deviation compared to the rigid timetable. On top of the rescheduling plan, the model seeks the maximum timetable flexibility that can be added to timetable events based on the operational constraints, as well as the punctuality threshold regulating the value of each type of timetable flexibility.

The model is verified by means of a toy network, and a model demonstration is presented for a part of the Dutch railway network with 12 stations and 13 trains. In the case study, our model includes 3280 continuous variables, 1099 binary variables, and 22484 constraints, and the average computation time of the optimization process is 1.12 s.

In order to have a deeper exploration of timetable flexibility, one illustrative application and two sensitivity analyses are conducted. In the illustrative application of the model to various inputs, the impacts of initial delays on timetable flexibility are investigated. In the first sensitivity analysis, the relationship between the weight of departure flexibility and timetable flexibility is explored. In the second sensitivity analysis, the impacts of the punctuality threshold on the amount of flexibility that can be added to the trains in the selected network are tested by gradually changing the exact value of the punctuality threshold.

5.3. Practical implications

In the current practice of ProRail, an RTTP with exact departure and arrival time points is generated in traffic management, and is provided to train dispatchers and signalers. However, those time points can be hard for trains to follow due to real-time traffic uncertainties. In this case, the computationally expensive solution algorithms should be conducted again to generate a new RTTP. By incorporating timetable flexibility, a rescheduling plan with event time slots is generated. Thus, more solution spaces can be provided with train dispatchers and signalers to determine the RTTP. As a consequence, the RTTP can be more robust since a new set of arrival and departure times can be chosen from event time slots to handle real-time uncertainties. Timetable flexibility is preferred in several scenarios, which are discussed below.

During peak hours, it can be crowded at the platforms of major stations, leading to longer dwell time required than scheduled. Thus, late departures are required, such that accommodate passengers and minimize the inconvenience they face. This kind of traffic uncertainty can be handled by introducing departure flexibility. Regarding arrival flexibility, it is useful in other scenarios. When it comes to bad weather, the tracks can be slippery, leading to inconsistency between the estimated and realized traction and braking force. Therefore, trains can arrive earlier or later than planned. Besides, train interactions can be complicated in congested areas. Thus, the exact arrival times can deviate from the rigid one. In those scenarios, the influence of the uncertainties can be mitigated by incorporating early and late arrival flexibility.

With respect to the distribution of early arrival flexibility and departure flexibility, it can also be connected to the scenarios mentioned above. During peak hours or other scenarios where passenger volume is high, late departures might be needed. Therefore, late departure flexibility is preferred in compromise of early arrival flexibility. When it comes to extreme weather conditions or congested dispatching areas, early arrival flexibility is preferred over departure flexibility.

Railway travelers will not be directly impacted by the introduction of timetable flexibility because what they depend on remains the published timetable. Nonetheless, the benefits of timetable flexibility for passengers are implicit. Most delays are induced by untimely and inappropriate handling of traffic uncertainties. By incorporating timetable flexibility, some uncertainties can be directly solved by picking a new RTTP from the rescheduling plan with event time slots. As a result, uncertainties will not lead to delays that propagate throughout the network, resulting in fewer delays experienced by railway travelers. This, in turn, enhances punctuality and passenger satisfaction.

5.4. Future research

The current research shows great promise in investigating timetable flexibility, and the results of this research provide insights that could be further explored or expanded upon in the future.

In terms of planning levels, our research centers on operational planning, which involves the continuous generation of rescheduling plans with timetable flexibility in real time. Nevertheless, this work is also applicable at the tactical planning level, where timetables with event time slots are generated in advance.

Regarding control measures, only retiming and reordering are considered in the current research. It would be interesting to also incorporate other control measures, such as rerouting, which may lead to more timetable flexibility added to trains, thus, further improving the robustness of the RTTP.

The concept of adding flexibility to the timetable can be combined with the learning-based approaches. In our model, the trade-off between early arrival flexibility and departure flexibility is controlled by the weight of departure flexibility. It means that only extreme scenarios can be obtained, namely, scenarios with maximum early arrival flexibility and maximum departure flexibility. However, the best distributions of each disturbance scenario remain known. Using learning-based methods, the values of each type of flexibility can be directly related to real conditions, leading to further improvements in the rescheduling plan for train dispatchers and signalers to determine the RTTP.

A

Input data frame

In order to efficiently manage the inputs of the model, as well as establish connections between nodes and arcs, three data frames are constructed, namely, the node data frame, the arc data frame, and the train data frame, which are discussed as follows.

Nodes

For all nodes in an alternative graph formulation, they are linked to unique IDs. For each ID, information including the node name, node type, and rigid time is provided. An example of the node data frame is shown in Table A.1. Regarding the node name, they are presented in the form of 'Train name_Timetable points', while for the dummy nodes 0 and n, they are expressed as 'N_0' and 'N_n', respectively. The nodes are divided into five types, namely entry, exit, arrival, departure, and other. Entry nodes represent the entry of trains into the dispatching area, exit nodes represent the exit of trains from the dispatching area, arrival nodes represent the arrival of trains at stations, departure nodes represent the departure of trains from stations, and other nodes are all other nodes including node 0, intermediate nodes between arrival and departure nodes, and node n. Regarding the rigid times of the entry, exit, arrival, and departure nodes, they can be obtained from the rigid timetable or the operational data.

Table A.1: An example of node data frame

Node ID	Name	Type	Rigid time (s)
0	N_0	normal	-
1	A_1	entry	T_1
2	A_S1	arrival	T_2
3	A_2	departure	T_3
4	A_2	normal	-
5	A_2	exit	T_5
6	N_n	normal	-
…	…	…	…

Arcs

Arcs are also linked to unique IDs, and five attributes are linked to each arc, including the name, source, destination, weight, and type of the arcs. An example of an arc data frame is shown in Table A.2. For each arc, the arc name is expressed as a pair of node names, and the source and destination are the starting and ending nodes of an arc, respectively, which are expressed as the corresponding node IDs. There are three types of arcs, namely fixed arc, alternative arc, and connection arc. The weights can be either directly obtained from the operational data or expressed by decision variables.

Table A.2: An example of arc data frame

Arc ID	Name	Source	Destination	Weight (s)	Type
0	(A_0,A_1)	0	1	w_0	fixed
1	(A_2,B_1)	2	3	w_1	alternative
2	(A_S2_1,B_3)	4	5	w_2	connection
…	…	…	…	…	…

Trains

Trains are linked to unique IDs, and 3 attributes are linked to each train, including train name, passing arcs, and initial delay. An example of the train data frame is presented in Table A.3. The passing arcs are expressed as a series of arc names that a train travels through. The value of the initial delay can be assigned according to the scenario.

Table A.3: An example of train data frame

Train ID	Name	Arcs	Initial Delay (s)
0	A	(A_0,A_1),(A_1,A_2), ..., (A_n-1,A_n)	I_0
1	B	(B_0,B_1),(B_1,B_2),..., (B_n-1,B_n)	I_1
…	…	…	…

B

Original rigid timetable

Table B.1: Rigid timetable from Ehv to Bd

	Ehv		Ehs		Bet		Btl		Ot		Tb		Tbu		Bd	
Train name	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D
1900-H-1		:03:00									:27:00	:31:00			:44:00	:50:00
6400-H-1	:13:00	:21:00		:24:42		:30:24		:37:48		:44:24	:50:30	:53:30	:58:00			
1100-H-1		:14:00									:35:30	:39:00			:51:00	:53:00
21400-H-1		:30:00									:59:00	:02:00			:18:00	:20:00

Table B.2: Rigid timetable from Ht to Bd

	Ht		Tb		Tbu		Tbr		Gz		Bd	
Train name	A	D	A	D	A	D	A	D	A	D	A	D
3600-T-1	:07:00	:12:00	:27:00	:28:00							:40:00	:45:00
6600-T-1	:28:00	:33:00	:48:00	:49:00		:52:06		:56:06		:00:30	:07:00	:09:00
AABE1-H-1												
3600-T-2	:37:00	:42:00	:57:00	:58:00							:10:00	:15:00

Table B.3: Rigid timetable from Ehv to Ht

	Ehv		Ehs		Bet		Btl		Vg		Ht	
Train name	A	D	A	D	A	D	A	D	A	D	A	D
800-H-1	:02:18	:06:00									:25:00	:28:00
3500-H-2	:11:00	:17:00									:36:00	:38:00
3900-H-2	:21:00	:27:00									:46:00	:48:00
4400-H-2	:27:30	:36:00		:39:06		:45:00		:52:30		:58:36	:03:00	
800-H-2	:32:18	:36:00									:55:00	:58:00

C
Corrected rigid timetable

Table C.1: The corrected rigid timetable from Ehv to Bd

	Ehv		Ehs		Bet		Btl		Ot		Tb		Tbu		Bd	
Train name	A	D	A	D	A	D	A	D	A	D	A	D	A	D	A	D
1900-H-1		:03:00									:27:00	:31:00			:44:00	:50:00
6400-H-1	:13:00	:21:00	:24:10	:24:46	:29:48	:30:24	:37:12	:37:48	:43:50	:44:26	:55:10	:56:10	:59:50			
1100-H-1		:14:00									:35:30	:39:00			:51:28	:53:28
21400-H-1		:30:00									:00:20	:02:00			:19:25	:20:25

Table C.2: The corrected rigid timetable from Ht to Bd

	Ht		Tb		Tbu		Tbr		Gz		Bd	
Train name	A	D	A	D	A	D	A	D	A	D	A	D
3600-T-1	:07:00	:12:00	:27:00	:28:00							:40:38	:45:00
6600-T-1	:28:00	:33:00	:52:10	:53:10	:55:58	:56:34	:59:39	:00:15	:04:20	:04:56	:11:44	:12:44
AABE1-H-1												
3600-T-2	:37:00	:42:00	:59:20	:00:20							:13:45	:15:00

Table C.3: The corrected rigid timetable from Ehv to Ht

	Ehv		Ehs		Bet		Btl		Vg		Ht	
Train name	A	D	A	D	A	D	A	D	A	D	A	D
800-H-1	:02:18	:06:00									:25:00	:28:00
3500-H-2	:11:00	:17:00									:36:00	:38:00
3900-H-2	:21:00	:27:00									:46:00	:48:00
4400-H-2	:27:30	:36:00	:38:54	:39:30	:44:27	:45:03	:51:54	:52:30	:59:47	:00:23	:04:10	
800-H-2	:32:18	:36:00									:55:00	:58:00

Bibliography

Abbink, E., Huisman, D., Kroon, L., 2018. Railway crew management, in: Handbook of Optimization in the Railway Industry. Springer. chapter 11, pp. 243–264.

Albrecht, T., Dasigi, M., 2016. ON-TIME: A framework for integrated railway network operation management. Traffic Management 3, 167–181.

Breugem, T., Dollevoet, T., Huisman, D., 2022. Is equality always desirable? Analyzing the trade-off between fairness and attractiveness in crew rostering. Management Science 68, 2619–2641.

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

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

Caimi, G., Fuchsberger, M., Laumanns, M., Schüpbach, K., 2011. Periodic railway timetabling with event flexibility. Networks 57, 3–18.

Caprara, A., Fischetti, M., Toth, P., 2002. Modeling and solving the train timetabling problem. Operations research 50, 851–861.

Cavone, G., Dotoli, M., Epicoco, N., Seatzu, C., 2017. A decision making procedure for robust train rescheduling based on mixed integer linear programming and data envelopment analysis. Applied Mathematical Modelling 52, 255–273.

Chiu, C., Chou, C., Lee, J.H.M., Leung, H.f., Leung, Y., 2002. A constraint-based interactive train rescheduling tool. Constraints 7, 167–198.

Corman, F., D'Ariano, A., Hansen, I.A., 2014. Evaluating disturbance robustness of railway schedules. Journal of Intelligent Transportation Systems 18, 106–120.

Corman, F., D'Ariano, A., Pacciarelli, D., Pranzo, M., 2012. Bi-objective conflict detection and resolution in railway traffic management. Transportation Research Part C: Emerging Technologies 20, 79–94.

Corman, F., D'Ariano, A., Pranzo, M., Hansen, I.A., 2011. Effectiveness of dynamic reordering and rerouting of trains in a complicated and densely occupied station area. Transportation Planning and Technology 34, 341–362.

Corman, F., Goverde, R.M.P., D'Ariano, A., 2009. Rescheduling dense train traffic over complex station interlocking areas. Robust and online large-scale optimization: models and techniques for transportation systems , 369–386.

Corman, F., Meng, L., 2014. A review of online dynamic models and algorithms for railway traffic management. IEEE Transactions on Intelligent Transportation Systems 16, 1274–1284.

D'Ariano, A., Corman, F., Pacciarelli, D., Pranzo, M., 2008a. Reordering and local rerouting strategies to manage train traffic in real time. Transportation science 42, 405–419.

D'Ariano, A., Pacciarelli, D., Pranzo, M., 2007a. A branch and bound algorithm for scheduling trains in a railway network. European journal of operational research 183, 643–657.

D'Ariano, A., Pacciarelli, D., Pranzo, M., 2008b. Assessment of flexible timetables in real-time traffic management of a railway bottleneck. Transportation Research Part C: Emerging Technologies 16, 232–245.

D'Ariano, A., Pranzo, M., Hansen, I.A., 2007b. Conflict resolution and train speed coordination for solving real-time timetable perturbations. IEEE Transactions on intelligent transportation systems 8, 208–222.

Dollevoet, T., Huisman, D., 2014. Fast heuristics for delay management with passenger rerouting. Public Transport 6, 67–84.

Dollevoet, T., Huisman, D., Schmidt, M., Schöbel, A., 2012. Delay management with rerouting of passengers. Transportation science 46, 74–89.

Dotoli, M., Epicoco, N., Falagario, M., Turchiano, B., Cavone, G., Convertini, A., 2014. A decision support system for real-time rescheduling of railways, in: 2014 European Control Conference (ECC), IEEE. pp. 696–701.

European Commission, 2020. Sustainable and Smart Mobility Strategy. URL: https://transport.ec.europa.eu/transport-themes/mobility-strategy_en.

European Commission, 2021. A European Green Deal: Striving to be the first climate-neutral continent. URL: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal_en.

European Commission, 2022. The journey begins – 2021 is the European Year of Rail! URL: https://ec.europa.eu/commission/presscorner/detail/en/IP_20_2528.

European Union, 2021. Europe's Rail Joint Undertaking. URL: https://european-union.europa.eu/institutions-law-budget/institutions-and-bodies/institutions-and-bodies-profiles/europes-rail-joint-undertaking_en.

European Union Agency for Railways, 2022. Railway Safety and Interoperability: the 2022 Report. URL: https://www.era.europa.eu/content/railway-safety-and-interoperability-2022-report_en.

Eurostat, 2022. Railway passenger transport statistics - quarterly and annual data. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Railway_passenger_transport_statistics_-_quarterly_and_annual_data&oldid=507441.

Fay, A., 2000. A fuzzy knowledge-based system for railway traffic control. Engineering Applications of Artificial Intelligence 13, 719–729.

Goossens, J.W., Van Hoesel, S., Kroon, L., 2004. A branch-and-cut approach for solving railway line-planning problems. Transportation Science 38, 379–393.

Goverde, R.M.P., Hansen, I.A., 2013. Performance indicators for railway timetables, in: 2013 IEEE International Conference on intelligent rail transportation proceedings, IEEE. pp. 301–306.

Gurobi Optimization, LLC, 2023. Gurobi Optimizer Reference Manual. URL: https://www.gurobi.com.

Huang, P., Li, Z., Zhu, Y., Wen, C., Corman, F., 2023. Train traffic control in merging stations: A data-driven approach. Transportation Research Part C: Emerging Technologies 152, 104155.

Jaekel, B., Albrecht, T., 2013. Interfacing conflict resolution and driver advisory systems in railway operations, in: Proceedings of the 3rd International Conference on Models and Technologies for Intelligent Transportation Systems 2013, pp. 333–343.

Jonathan, L., 2009. flexibility. in a dictionary of business and management. URL: https://www.oxfordreference.com/view/10.1093/acref/9780199234899.001.0001/acref-9780199234899-e-2590.

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

Kroon, L., Huisman, D., Abbink, E., Fioole, P.J., Fischetti, M., Maróti, G., Schrijver, A., Steenbeek, A., Ybema, R., 2009. The new Dutch timetable: The OR revolution. Interfaces 39, 6–17.

Liu, T., Ma, Z., Koutsopoulos, H.N., 2021. Unplanned disruption analysis in urban railway systems using smart card data. Urban Rail Transit 7, 177–190.

Lund, S., Madgavkar, A., Manyika, J., Smit, S., Ellingrud, K., Meaney, M., Robinson, O., 2021. The future of work after COVID-19. McKinsey Global Institute 18.

Lusby, R.M., Larsen, J., Ehrgott, M., Ryan, D., 2011. Railway track allocation: models and methods. OR spectrum 33, 843–883.

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

Mazzarello, M., Ottaviani, E., 2007. A traffic management system for real-time traffic optimisation in railways. Transportation Research Part B: Methodological 41, 246–274.

Morlok, E.K., Chang, D.J., 2004. Measuring capacity flexibility of a transportation system. Transportation Research Part A: Policy and Practice 38, 405–420.

ON-TIME, 2013. Deliverable D4.1: Functional and technical requirements specification for perturbation management .

Ott, A., Sandri, N., Speksnijder, I., Wolfs, K., 2021. Safe, smart, and green: Boosting European passenger rail's modal share. URL: `https://www.mckinsey.com/industries/travel-logistics-and-infrastructure/our-insights/safe-smart-and-green-boosting-european-passenger-rails-modal-share#/`.

Pachl, J., 2014. Timetable Design Principles, in: Hansen, I.A., Pachl, J. (Eds.), Railway Timetabling & Operations: Analysis, Modelling, Optimisation, Simulation, Performance Evaluation. Eurailpress, Hamburg, pp. 13 – 46.

Pellegrini, P., Marlière, G., Rodriguez, J., 2012. Real time railway traffic management modeling track-circuits, in: ATOMOS 2012, 12th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems, pp. 23–34.

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

Ravichandran, H., 2013. Evaluating the robustness of crew schedules for rail transit systems. Ph.D. thesis. Massachusetts Institute of Technology.

Ritchie, H., 2020. Which form of transport has the smallest carbon footprint. Our World in Data URL: `https://ourworldindata.org/travel-carbon-footprint`.

Schachtebeck, M., Schöbel, A., 2008. IP-based techniques for delay management with priority decisions, in: 8th Workshop on Algorithmic Approaches for Transportation Modeling, Optimization, and Systems (ATMOS'08), Schloss Dagstuhl-Leibniz-Zentrum für Informatik.

Schöbel, A., 2007. Integer programming approaches for solving the delay management problem, in: Algorithmic Methods for Railway Optimization: International Dagstuhl Workshop, Dagstuhl Castle, Germany, June 20-25, 2004, 4th International Workshop, ATMOS 2004, Bergen, Norway, September 16-17, 2004, Revised Selected Papers, Springer. pp. 145–170.

Šemrov, D., Marsetič, R., Žura, M., Todorovski, L., Srdic, A., 2016. Reinforcement learning approach for train rescheduling on a single-track railway. Transportation Research Part B: Methodological 86, 250–267.

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

Wang, P., Goverde, R.M.P., 2016. Multiple-phase train trajectory optimization with signalling and operational constraints. Transportation Research Part C: Emerging Technologies 69, 255–275.

Wang, Z., Quaglietta, E., Bartholomeus, M., Cunillera, A., Goverde, R.M.P., 2023. Conflict-free train path planning using ato timing points, in: RailBelgrade 2023: 10th International Conference on Railway Operations Modelling and Analysis.

Yin, J., Tang, T., Yang, L., Xun, J., Huang, Y., Gao, Z., 2017. Research and development of automatic train operation for railway transportation systems: A survey. Transportation Research Part C: Emerging Technologies 85, 548–572.

Yuan, J., 2006. Stochastic modelling of train delays and delay propagation in stations. volume 2006. Eburon Uitgeverij BV.

Zhang, H., Li, S., Wang, Y., Wang, Y., Yang, L., 2021. Real-time optimization strategy for single-track high-speed train rescheduling with disturbance uncertainties: A scenario-based chance-constrained model predictive control approach. Computers & Operations Research 127, 105135.

Zhu, Y., Wang, H., Goverde, R.M.P., 2020. Reinforcement learning in railway timetable rescheduling, in: 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), IEEE. pp. 1–6.