Queueing-Based Methods for Timetable-Independent Performance Analysis of Railway Nodes

Von der Fakultät für Bauingenieurwesen der Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

Tamme Paul Emunds

Berichter: Univ.-Prof. Dr.-Ing. Nils Nießen
Univ.-Prof. Dr. rer. nat. Britta Peis

Tag der mündlichen Prüfung: 30. Juni 2026

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

Abstract

Railway transportation is increasingly recognized as a cornerstone for sustainable mobility of passengers and goods. However, existing railway networks are confronted with growing capacity constraints, particularly within node areas such as junctions and stations. In practice, many infrastructure managers rely on timetable-based performance indicators to assess operational quality, which require the assumption of fixed train sequences or timetables in early planning phases. Given that railway infrastructure is designed to last several decades, planning processes must account for uncertainties in future demand and regulatory requirements. Therefore, it is crucial to establish performance metrics that evaluate the infrastructure capacity independently of predetermined timetables.

This dissertation introduces three novel queueing-based methods for the timetable-independent performance analysis of railway nodes and stations. The first contribution is a framework that models the capacity allocation process in railway route nodes using Continuous-Time Markov Chains, enabling the application of probabilistic model-checking to derive performance parameters. Furthermore, two extensions are developed: One incorporates phase-type distributions to explicitly model non-exponential arrival and service processes, enhancing modeling accuracy. The second extension presents, for the first time, a queueing-based method to analyze the performance of an entire railway station within a combined multi-channel system.

All methods decompose railway nodes by infrastructure routes, facilitating broad applicability without the need for detailed timetable data or extensive expert knowledge. The approaches are validated through case studies, demonstrating their utility for infrastructure managers in assessing capacity constraints and planning optimal layouts for junctions and stations. Given that the presented framework obtains performance parameters on explicitly formulated Continuous-Time Markov Chains, it allows for efficient iterations over input parameters and supports model validation, offering substantial improvements over existing simulation and queueing-based approaches.

The results highlight the practicality of the proposed methods in real-world scenarios. By enabling timetable-independent performance evaluation, these methods provide infrastructure managers with powerful tools to support the robust development of railway networks in early planning stages.

Kurzfassung

Der Schienenverkehr ist eine zentrale Säule der nachhaltigen Mobilität von Personen und Gütern. Gerade in den Knotenbereichen des Eisenbahnnetzes treten jedoch zunehmend Kapazitätsengpässe auf. In der Praxis nutzen viele Infrastrukturbetreiber fahrplanbasierte Leistungskennzahlen zur Bewertung der Infrastrukturqualität, was voraussetzt, dass bereits in frühen Planungsphasen feste Fahrpläne angenommen werden. Da die Infrastruktur für Nutzungsdauern von mehreren Jahrzehnten ausgelegt ist, müssen die Planungsprozesse jedoch auch Unsicherheiten hinsichtlich der zukünftigen Nachfrage und regulatorischer Anforderungen berücksichtigen. Daher ist es notwendig, Leistungskennzahlen zu etablieren, mit denen sich die Infrastrukturkapazität unabhängig von vorgegebenen Fahrplänen bewerten lässt.

Diese Dissertation stellt drei neuartige bedientheoretische Methoden zur fahrplanunabhängigen Leistungsanalyse von Abzweigen und Bahnhöfen vor. Zunächst wird eine Methode entwickelt, die den Kapazitätszuweisungsprozess in Eisenbahnknoten mithilfe zeitkontinuierlicher Markov-Ketten modelliert und probabilistische Model-Checking-Verfahren zur Ableitung von Leistungskennzahlen einsetzt. Darauf aufbauend werden zwei Erweiterungen präsentiert: Zum einen wird die Integration von Phasenverteilungen beschrieben, um nicht-exponentielle Ankunfts- und Bedienprozesse explizit zu modellieren und dadurch die Modellgenauigkeit zu erhöhen. Zum anderen wird erstmals eine bedientheoretische Methode zur Analyse der Leistungsfähigkeit eines gesamten Bahnhofs auf Basis eines einzigen kombinierten Mehrkanalsystems entwickelt.

Alle Methoden ermöglichen eine breite Anwendbarkeit über die Zerlegung der betrachteten Knoten nach Fahrwegen. Durch Fallbeispiele wird demonstriert, wie Infrastrukturbetreiber die entwickelten Ansätze zur Engpassdetektion und Infrastrukturplanung einsetzen können. Da die vorgestellten Methoden Kennzahlen mit explizit formulierten Markov-Ketten gewinnen, können Eingabeparameter effektiv iteriert, und zusätzliche Kennzahlen zur Modellvalidierung aufwandsarm integriert werden.

Die Ergebnisse unterstreichen die Praxistauglichkeit der entwickelten Methoden. Durch die fahrplanunabhängige Leistungsbewertung bieten die vorgeschlagenen Ansätze Infrastrukturbetreibern effektive Werkzeuge zur Unterstützung der robusten Entwicklung von Bahnnetzen in frühen Planungsphasen.

Acknowledgements

At this point, I would like to express my sincere gratitude to the people who accompanied and supported me during my dissertation.

First and foremost, I am deeply grateful to my supervisor, Nils Nießen, for his trust and for striking a remarkable balance between granting me the freedom to pursue my research and providing sound academic advice whenever needed. His unwavering support and thoughtful feedback were essential to the success of this work.

I am also grateful to my second examiner, Britta Peis, for the straightforward collaboration and her interest in this work. I greatly appreciated the opportunity to work with her and to benefit from her mathematical perspectives.

I wish to thank my colleagues at the institute for creating an inspiring working environment and for the many valuable academic exchanges. The countless insightful conversations and shared experiences, both within and beyond the professional context, have shaped my time at the institute. In particular, I would like to mention my long-term office mate Alexander Große, as well as Maren Maus, Stephan Zieger and Bastian Kogel, who have always been my first points of contact for academic discussions.

My thanks also go to my second work family at the graduate school UnRAVeL for the joint workshops and your openness to unfamiliar topics. I have greatly benefited from the interdisciplinary exchange with all of you. I would especially like to thank Maria Anapolska, Alexander Bork, Paul Brunzema and Katharina Eickhoff for the many fruitful research discussions.

I also gratefully acknowledge funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 2236/2. Computational experiments were performed with computing resources granted by RWTH Aachen University under projects rwth1413 and rwth1635.

Finally, I would like to thank my friends and family for their support and understanding throughout this challenging period. Especially to you, Michi: Thank you for always being there to support me and lift me up after difficult days.

Table of Contents

Abstract . . . . . **III**

Kurzfassung . . . . . **V**

Acknowledgements . . . . . **VII**

Table of Contents . . . . . **IX**

List of Papers . . . . . **XIII**

List of Figures . . . . . **XIV**

List of Tables . . . . . **XVI**

1. Introduction **1**

2. Background **5**

2.1. Fundamentals 5

2.1.1. Planning Levels 5

2.1.2. Elements of a Railway System 8

2.1.3. Capacity Concepts 10

2.1.4. Capacity Consumption 11

2.1.5. Key Figures for Performance Analysis 16

2.2. Related Work 18

2.2.1. Construction 18

2.2.2. Delay Propagation Analysis 19

2.2.3. Analysis of Operational Data 20

2.2.4. Simulation 20

2.2.5. Optimization 21

2.2.6. Queueing-based Models 22

2.2.7. Summary and Discussion 24

2.3. Motivation and Research Goal 25

3. Evaluating Railway Junction Infrastructure: A Queueing-Based, Timetable-Independent Analysis . . . 27
Abstract . . . 27
3.1. Introduction . . . 27
3.2. Related Work . . . 29
3.2.1. Terminology of railway performance analysis . . . 29
3.2.2. Literature Review . . . 30
3.2.3. Motivation and Contextualisation . . . 35
3.3. A Continuous-Time Markov Chain Model For The Queueing System In A Railway Junction . . . 36
3.3.1. Junction Layout . . . 36
3.3.2. Problem Formulation . . . 37
3.3.3. Queueing System . . . 39
3.3.4. Modelling railway junctions . . . 42
3.3.5. Scalability . . . 48
3.3.6. Determination of the estimated queue length . . . 49
3.3.7. Threshold values . . . 53
3.3.8. Methodical Procedure . . . 53
3.3.9. Contributing Factors . . . 56
3.4. Model Performance . . . 59
3.4.1. Simulation Architecture . . . 59
3.4.2. Validation . . . 60
3.5. Computational Study . . . 63
3.5.1. Timetable Capacity Analysis . . . 63
3.5.2. Infrastructure Layout Planning . . . 69
3.6. Discussion and Outlook . . . 79

4. Utilizing Phase-Type Distributions for Queueing-Based Railway Junction Performance Determination . . . 83
Abstract . . . 83
4.1. Introduction . . . 83
4.2. Related Work . . . 85
4.3. Problem Description . . . 88
4.4. Methods . . . 93
4.4.1. A Queueing Based Railway Junction CTMC . . . 93

4.4.2. Threshold Values . . . 95
4.4.3. Approximation Formulas . . . 96
4.4.4. Phase-Type Distributions . . . 97
4.4.5. Queue-Length Analysis . . . 102
4.4.6. Capacity Determination Algorithm . . . 103
4.5. Validation . . . 104
4.5.1. Computation Time . . . 106
4.5.2. Queue-Length Approximation Quality . . . 107
4.5.3. Performance Determination Quality . . . 110
4.6. Case Study . . . 113
4.6.1. Setup . . . 114
4.6.2. Timetable Capacity Results . . . 116
4.7. Discussion and Conclusion . . . 118

5. A Queueing-Based Approach for Timetable-Independent Railway Station Performance Analysis . . . 121
Abstract . . . 121
5.1. Introduction . . . 121
5.2. Queueing System Model of a Railway Station . . . 124
5.2.1. General Discussion on Queueing Systems . . . 125
5.2.2. Approximation Formula . . . 126
5.2.3. Approach to Railway Station Partitioning . . . 126
5.2.4. Continuous-Time Markov Chain Formulation . . . 129
5.3. Determining the Timetable Capacity of Railway Stations . . . 134
5.3.1. The Timetable Capacity Performance Metric . . . 135
5.3.2. Required Data . . . 135
5.3.3. Acquiring Queue-Length Estimations . . . 138
5.3.4. Defining Thresholds . . . 138
5.3.5. Procedure . . . 139
5.3.6. Scalability . . . 141
5.4. Case Study . . . 142
5.4.1. Operational Scenario Analysis . . . 142
5.4.2. Influence of Service Processes . . . 145
5.4.3. Dependence on the Number of Waiting Positions . . . 146
5.5. Conclusion . . . 148

6. Summary and Discussion . . . 151
6.1. Summary of Key Findings . . . 151
6.2. Limitations and Implications for Further Research . . . 152
6.2.1. Scalability . . . 152
6.2.2. Thresholds and Quality Indicators . . . 153
6.2.3. Validity of Queue-Length Estimations . . . 154
6.2.4. Modeling Arrival and Service Processes . . . 155
6.2.5. Priority Service . . . 155
6.2.6. Further Adaptations and Enhancements . . . 156
6.3. Conclusion . . . 157

Bibliography . . . 159

Appendix A . . . 172
Appendix A.1: Literature Summary . . . 172
Appendix A.2: Supplementary Boxplots . . . 173
Appendix A.3: Table With Exemplary Results . . . 175
Appendix A.4: Full Computational Study . . . 176
Appendix A.5: Full Comparison with Capacity Utilization . . . 180

Appendix B . . . 183
Appendix B.1: Capacity Determination Algorithm . . . 183
Appendix B.2: Performance Determination for Simulations . . . 185
Appendix B.3: Computation Times Case Study . . . 187

Appendix C . . . 188
Appendix C.1: Combining Coefficients of Variation . . . 188

List of Papers

This cumulative dissertation is based on three peer-reviewed journal articles that the doctoral candidate Tamme Emunds co-authored with the doctoral supervisor Nils Nießen.

For all three articles included in this dissertation, the contributions were as follows:

Tamme Emunds led the conceptualization of the study, developed the methodology, implemented the software, conducted the formal analysis, and prepared, revised, and finalized the manuscript. Nils Nießen contributed through conceptual guidance, supervision of the research process, critical review and editorial feedback.

The following three articles form part of this dissertation and have already been published:

1. Tamme Emunds and Nils Nießen. *Evaluating railway junction infrastructure: A queuing-based, timetable-independent analysis*. Transportation Research Part C, 2024. [see also Chapter 3 and (Emunds and Nießen, 2024a)]
2. Tamme Emunds and Nils Nießen. *Utilizing phase-type distributions for queueing-based railway junction performance determination*. Journal of Rail Transport Planning & Management, 2025. [see also Chapter 4 and (Emunds and Nießen, 2025c)]
3. Tamme Emunds and Nils Nießen. *A queueing-based approach for timetable-independent railway station performance analysis*. Journal of Rail Transport Planning & Management, 2025. [see also Chapter 5 and (Emunds and Nießen, 2025b)]

List of Figures

2.1. Phases of railway transport system planning over time until the provision of transport services. Own illustration based on (UIC, 2013; Weik, 2020; Cats et al., 2025). 6
2.2. Selected railway infrastructure elements in the schematic illustration of a fictional example. 8
2.3. Components of the blocking time in a time-distance diagram. 12
2.4. Blocking and dwell times during track occupancy for a stopping train . . 14
2.5. The relevant block section for determining the minimum headway using two blocking time stairways. 15

3.1. Track-layout for a double-track junction 37
3.2. Markov Chain for a $M/M/1/\infty$ Queueing System 42
3.3. Conflicting routes in railway infrastructure 45
3.4. Markov Chain modelling the Queueing System of the single track segment in Figure 3.3a 46
3.5. Markov Chain modelling the Queueing System of the crossover segment in Figure 3.3b 47
3.6. Application of the introduced method to choose an optimal infrastructure layout 55
3.7. Comparison of the required computation times 61
3.8. Accuracy of the analytical approach when compared to the confidence interval of the performed simulations 62
3.9. A track-layout plan for the triangle of Gagny. Own illustration, adapted from Pellegrini et al. (2014). 64
3.10. Obtained estimated queue lengths for every scenario (a-c) and comparison of the maximal queue length of any route per scenario (d). . . . 68
3.11. Track-layout for a double-track junction with an overpass 70
3.12. Overview of the considered operating program combinations 72
3.13. Estimated queue length on route r_3 for the combination of a local train main line and a long-distance train branching line 73
3.14. Calculated maximum mean service times in minutes for the scenario without an overpass 75

3.15. Calculated maximum mean service times in minutes for the scenario with an overpass . . . 75
3.16. Comparison between the obtained maximum mean service times of the two infrastructure scenarios . . . 76
3.17. Occupancy rates (following equation (3.38)) for the scenario without an overpass, assuming mean service times from Figure 3.14 . . . 78

4.1. Examplary infrastructure of a railway junction . . . 89
4.2. Examples of the service process with exponential (M) or phase-distributed (Ph) service times . . . 97
4.3. Computation times of all considered models . . . 106
4.4. short . . . 108
4.5. Comparison of the performance determination by the share of main line traffic p_{main} for the different methods . . . 111
4.6. Traffic scenario of the case study . . . 113
4.7. Infrastructure of the considered railway junction . . . 113
4.8. The number of trains n_{total} in the junction with a traffic split of $p_{\text{main}} = 0.5$ for all iterations . . . 116
4.9. Timetable capacity n_{max} of the two models PH/PH and M/M and quality factor per route for varying p_{main} . . . 117

5.1. An example for a railway station infrastructure . . . 127
5.2. Dependencies in the queueing systems for two routes in opposite directions. . . . 128
5.3. The two possibilities of the arrival transition to an empty route ($q_r = 0$) and a combined representation. . . . 131
5.4. An example for the choice transitions between the entry processes of two conflicting routes r_1 and r_4, assuming non-empty queues ($q_{r_1}, q_{r_4} > 0$) for both routes. . . . 133
5.5. A CTMC for the arrival, entry, stopping and departure process on one route with $m = 2$. . . . 134
5.6. The railway station infrastructure with swapped stopping platform tracks for routes r_2 and r_4 . . . 142
5.7. Comparison of the queue-length estimations for all routes in both scenarios . . . 144
5.8. Sensitivity of the timetable capacity to service time scaling . . . 146

5.9. Comparison of Capacity and Computation Times for Different Number of Waiting Positions m 147

A.1. Comparison between the analytical solution and the simulations with an unlimited m 173
A.2. Comparison between the analytical solution and the simulations with a limit of $m = 5$ 173
A.3. Comparison between the analytical solution and the simulations with an unlimited m, assuming different numbers of trains per route and a fixed service rate $\mu_i = 0.3$ for all $r_i \in R$. 174
A.4. Comparison between the analytical solution and the simulations with a limit of $m = 5$, assuming different numbers of trains per route and a fixed service rate $\mu_i = 0.3$ for all $r_i \in R$. 174
A.5. Overview of all operating program combinations 177
A.6. Calculated maximum mean service times in minutes for the scenario without an overpass 178
A.7. Calculated maximum mean service times in minutes for the scenario with an overpass 179
A.8. Occupancy rates (following equation (3.38)) for the scenario with an overpass, assuming mean service times from Figure 3.15 180
A.9. Occupancy rates (following equation (3.38)) for the scenario without an overpass, assuming mean service times from Figure A.6 181
A.10. Occupancy rates (following equation (3.38)) for the scenario with an overpass, assuming mean service times from Figure A.7 182

List of Tables

2.1. Influencing factors of the considered capacity concepts 11
2.2. Selected key figures for describing the quality of railway systems 17
2.3. Queueing-based methods for timetable capacity analysis on railway nodes 23
2.4. Applicability of Procedure Classes within Different Capacity Concepts . 24

3.1. Literature considering railway performance estimations 31
3.2. Analytical capacity determination methodology 33
3.3. Routes through the junction 37
3.4. List of Notations 40
3.5. Number of states by the number of waiting positions in Queuing Systems for the examples in Figure 3.3. 48
3.6. Requirements for the described method 56
3.7. Model assumptions 58
3.8. Traffic share $p_{i,\theta}$ of routes r_i in every scenario $\theta \in \Theta$. 65
3.9. Assumed minimum headway times $h_{i,j}$ in seconds for the junction of Gagny 66
3.10. Service time $b_{i,\theta}$ of route r_i in every scenario $\theta \in \Theta$ in minutes. 66
3.11. Timetable capacity for all considered scenarios and selected parameters of the significant operating program A^*_θ. 69
3.12. Number of trains by traffic type for the considered operating programs . 71

4.1. Analytical junction capacity determination methodology 87
4.2. List of notations. 90
4.3. Parameters of the studied models 105
4.4. Parameters of the approximation settings 108
4.5. Selected indicators for the different methods 110
4.6. Traffic in the case study 114
4.7. Assumed minimum headway times $h_{i,j}$ in minutes for the considered junction 114
4.8. Service Rates, Variance Coefficients and size of the CTMC Model (PH/PH) at different p_{main} Values 115

5.1. Methods for queueing-based railway station performance analysis . . . 123

5.2. List of Variables and Notations 136
5.3. Route Node and Platform Usage per Scenario 143
5.4. Blocking and Stopping Times in Minutes 143

A.1. Literature considering railway performance estimations 172
A.2. Estimated queue length on route r_3 for the combination of a local train main line and a long-distance train branching line 175
A.3. Selected operating programs 176
A.4. Selected traffic code parameters from (European Comission, 2019) . . 176
B.1. Performance Metrics for Different Model Types at Various p_{main} Values . 187

1. Introduction

Railway transportation continues to develop into one of the key aspects of increasing the sustainability of transportation for passengers and goods. In their Mobility Strategy, the European Commission (2020) set the goal of doubling traffic of passenger rail by 2030, and of freight rail by 2050. More recently, the European Commission committed to the ambition of facilitating an European high-speed rail network (European Commission, 2025) and the Council of the European Union (2024) recognized the requirement of better management and increased use of Europe's rail infrastructure. In Germany, specifically, the largest infrastructure manager states 35 *overloaded railway line segments* (dt. *überlasteter Schienenweg*) (DB InfraGO, 2025), highlighting the need for substantial expansions of the infrastructure in order to facilitate today's traffic alone, even without considering the expected growth in demand due to political objectives.

In order to design effective and demand-driven railway networks, infrastructure managers need to be able to identify the performance capability of all infrastructure elements. In practice, many infrastructure managers choose to apply timetable-based performance indicators, i.e. maximal occupation ratios (UIC, 2013), in an effort to ensure sufficient operational quality, necessitating the assumption of train sequences or timetables in the planning process. However, both the expected additional traffic demand and unanticipated changes in European and national regulations raise questions regarding the reliability of long-term timetable planning. Given that railway infrastructure is designed to last for several decades, effective planning processes may need to consider indicators that describe infrastructure performance without the knowledge of a fixed timetable.

Realizing this requirement, previous research introduced *queueing-based* methods to cope with the uncertainty structure of train demand for long-term planning problems. Corresponding methods for analyzing line (Schwanhäußer, 1974), route node (Schwanhäußer, 1978) or station (Potthoff, 1962; Hertel, 1984) capacity have been implemented in practical applications for over a decade (DB InfraGO, 2009, 2022). Several advancements have been made to introduce multi-channel models for route node analysis (Nießen, 2008, 2013; Schmitz et al., 2017), however they require advanced system knowledge of the user to further partition the infrastructure into multiple single channel systems.

While the infrastructure along railway lines primarily determines travel times, node areas are typically more congested. At passenger platforms, dwell times facilitate passenger exchange and may synchronize connections with other services. Furthermore, switches and at-grade intersections of railway tracks require exclusive occupation, restricting the simultaneous movement of multiple trains. Consequently, railway junctions and stations are particularly prone to the emergence of waiting times, which are necessary to ensure safe operations by preventing conflicting train movements. Additionally, although waiting times may originate from line occupation outside the station, it may be beneficial for trains to spend this additional waiting time within stations. This allows for uninterrupted travel to the next stopping point and enables further passenger exchange during the extended platform dwell. As a result, node areas represent critical infrastructure elements within the rail network, and their thorough performance analysis is crucial for identifying capacity constraints and understanding the performance capability of the entire system.

This dissertation introduces three queueing-based methods to analyze multi channel performance of railway nodes by modeling dependencies between different routes within a queueing system of the node infrastructure. Major contributions contain the description of a novel framework to formulate the capacity allocation process on railway route nodes as Continuous-Time Markov Chains to obtain performance parameters by applying probabilistic model-checking, and two extensions of it. In the first extension, phase-type distributions are utilized to enable the explicit consideration of non-exponential arrival and service processes in railway nodes. A second extension corresponds to the first (to the best of our knowledge) queueing-based method to analyze the performance of an entire railway station in one combined multi-channel queuing system. All three methods decompose the railway node by infrastructure-routes and are therefore easily applicable to a broad range of scenarios without requiring a wide range of input data. In particular, they do not require the input of a timetable, making them especially useful to be applied in long-term planning problems.

The remainder of this thesis is structured into five chapters. Chapter 2 outlines the basics of railway performance analysis and characterizes the state of the art. It is followed by three chapters describing the three main methods, each of them has been published in a peer-reviewed journal paper. In Chapter 3 (Emunds and Nießen, 2024a) the novel framework to formulate railway junction capacity with Continuous-Time Markov Chains is introduced. The extension to utilize phase-type distributions

for a more accurate modeling of arrival and service processes is presented in Chapter 4 (Emunds and Nießen, 2025c). Subsequently, Chapter 5 (Emunds and Nießen, 2025b) illustrates the new approach for timetable-independent performance analysis of railway stations. Finally, Chapter 6 closes the thesis by summarizing and discussing the three introduced approaches.

2. Background

Prior to presenting the methodological advancements in the following chapters, this chapter provides some relevant background on the performance analysis of railway systems. Section 2.1 introduces fundamental concepts in planning and capacity analysis of railway networks. Section 2.2 then surveys related methods for performance assessment. Finally, Section 2.3 articulates the research gap and motivates the contributions of this dissertation.

2.1. Fundamentals

In the following, the various levels of railway system planning (Section 2.1.1), different elements of railway networks (Section 2.1.2), general concepts of capacity (Section 2.1.3), the capacity consumption of train operations (Section 2.1.4), and key figures for performance analysis (Section 2.1.5) are introduced.

2.1.1. Planning Levels

The planning of general transportation systems and in particular of railway systems can be divided into different levels of planning (Bussieck et al., 1997; UIC, 2013; Cats et al., 2025). Commonly, a distinction is made between the *strategic*, *tactical*, and *operational* levels of planning.

Figure 2.1 provides an overview of the individual phases of the planning process. The planning elements are arranged chronologically, ranging from strategic decisions made several decades prior to actual transport operations to short-term ad hoc decisions in operational management.

Strategic planning encompasses the long-term development of entire transport networks (*network design*) or specific infrastructure elements (*infrastructure planning*), such as stations or railway lines. Given the long service life of railway infrastructures, often spanning several years or decades, and the extensive duration of authorization and construction processes, detailed planning must be undertaken well in advance.

Once an infrastructure scenario is established, *line planning* can be carried out to

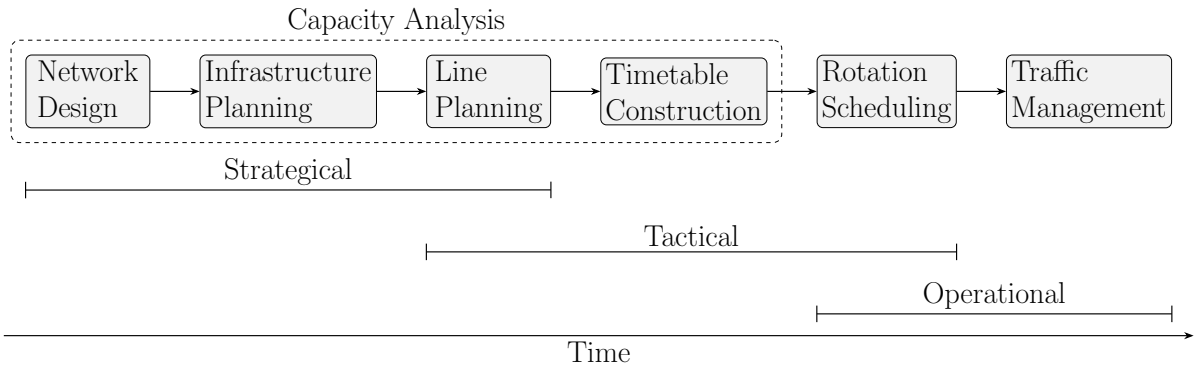


Figure 2.1.: Phases of railway transport system planning over time until the provision of transport services. Own illustration based on (UIC, 2013; Weik, 2020; Cats et al., 2025).

develop concepts for line routes and service frequency. Taking place multiple years in advance, this process must take into account both the long-term transport concept and demand forecasts, as well as the rolling stock available. Line planning thus serves as a transition between strategic and tactical levels of planning.

During the tactical phase, a *timetable* is constructed, implementing the line concept into individual train services with assigned departure and arrival times. Approximately one year before the timetable period, infrastructure managers collect train path requests from railway operators and develop a conflict-free annual timetable for the upcoming year. Within the timetable period, this timetable may be revised several months or weeks prior to operation to accommodate construction works. In order to satisfy on-demand freight traffic, some infrastructure managers even allow the booking of additional freight services only days or hours before their departure time. Through *crew* and *vehicle rotation scheduling*, rolling stock and personnel are allocated to planned train services. Short-term disruptions such as staff leave or vehicle maintenance may necessitate adjustments to duties or rotations; therefore, these elements are also associated with operational planning. Operational planning additionally includes all tasks related to disposition and traffic management.

Dividing the overall process into distinct phases is not only justified by varying time horizons but also by differences in data availability: As the actual provision of transport services approaches, more detailed information becomes accessible. Early stages thus require dealing with uncertainties regarding input parameters, such as infrastructure data or planned traffic services. Likewise, data granularity increases with decreasing duration until the actual operation of the transport services.

Therefore, varying representations of infrastructure data may be employed at different stages of the planning process. For strategic network planning, *macroscopic* models are commonly used. Here, entire stations are represented as vertices, which are connected by edges that correspond to railway line sections. In contrast, timetable construction should employ *microscopic* models to ensure conflict-free scheduling of individual train movements. In such a representation, necessary infrastructure components, like main signals, distant signals, switches (points), or even changes in gradient, are modeled explicitly as vertices in a graph model of the infrastructure. Depending on the requirements of the considered planning stage, different elements may be relevant for the analysis.

Some applications benefit from *mesoscopic* models that aggregate certain nodes from microscopic representations. Therefore, a mesoscopic model might be able to accurately model route relationships within route nodes without the need to include details about signals or track vacancy detection devices.

The conceptualization of a transport system typically proceeds through each phase separately. However, feedback from later stages may necessitate a revision of earlier steps. In reality, transport networks are subject to constant change: for instance, shifts in demand may necessitate modifications of line concepts, and technological innovations might require infrastructure redesigns. Including all these dependencies into a single planning problem would substantially increase the complexity and hence exceed current planning capabilities. Therefore, most research treats individual planning problems independently, even though some recent works combine selected planning phases.

A comprehensive overview of operations research (OR) applications supporting public transportation systems is provided by Cats et al. (2025). For railway systems specifically, Correia Duarte et al. (2025) review related work in timetabling and vehicle rotation planning. Recent research also addresses integrated approaches for multiple phases. For instance, Schöbel (2017) introduces the 'Eigenmodel', which iteratively solves line planning, timetabling and vehicle rotation problems. Other studies focus on integrating line planning with timetabling (Liebchen and Möhring, 2007; Burggraeve et al., 2017; Fuchs et al., 2022), or timetabling with vehicle rotation scheduling (Li et al., 2023).

The in early phases designed infrastructure network should support the requirements arising from line and timetable concepts. Therefore, strategic and tactical planning

levels are the most relevant phases for the application of railway capacity methodology. During the operational phase, constructed timetables should not result in major challenges when facing minor disturbances. In order to ensure the stability of the railway system, it is therefore essential to integrate consideration about the performance of individual elements and the overall system into the entire planning process.

2.1.2. Elements of a Railway System

A railway system consists of various components that collectively determine its overall performance capability. The infrastructure of a railway system is composed of nodes and lines; an example infrastructure is shown in Figure 2.2. Traditionally (DB InfraGO, 2022), a network is partitioned into *operating points*, which include all infrastructural elements within a node or a section of the line.

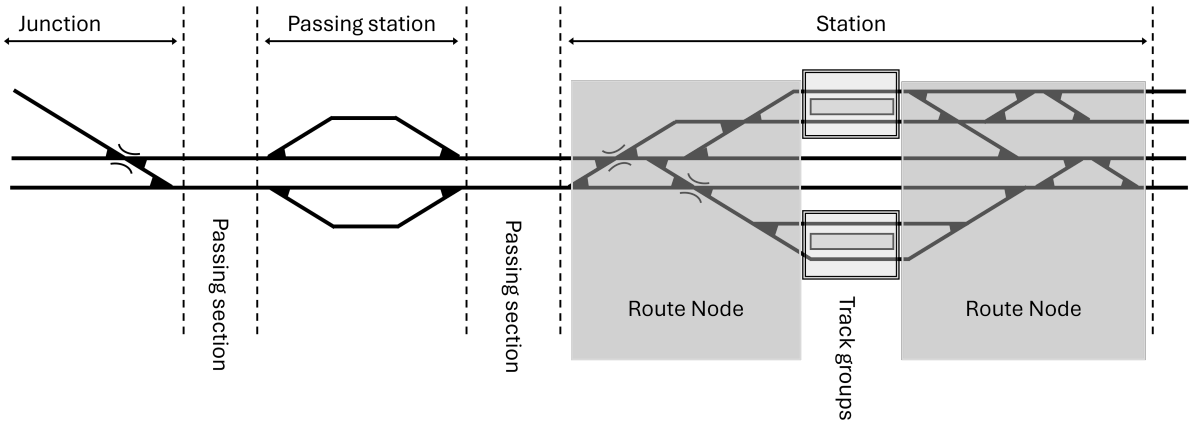


Figure 2.2.: Selected railway infrastructure elements in the schematic illustration of a fictional example.

Nodes comprise both *stations* and *junctions*. A *line* represents the connection between two major stations. For capacity analysis purposes, lines are further subdivided into *passing sections* between two *passing stations*. A passing station is defined as a station equipped with passing tracks that allow trains to overtake and therefore change their order for following infrastructure.

The location where a train comes to a stop is referred to as a *stop position*. In stations with multiple tracks, the various stop positions are partitioned into *track groups*. All tracks within a track group are intended to serve as substitutes for each other. In principle, any train should be able to use any of the tracks within a given track group.

In addition to the track group(s), a station also comprises one or more *route nodes*. Within a route node, various *train routes* can be established via the interlocking system. These routes ensure that a specific track section is exclusively allocated to a single train movement. A train route refers to the secured segment for a particular train operation, for example, the section between the entry signal and the stop position during an arrival. Further information on *train control systems* can be found in Pachl (2022).

Depending on the infrastructure within a route node, multiple train routes may be occupied in parallel by different train movements. Utilizing information about the infrastructure and train control system within a route node, it is possible to specify for each combination of two train routes whether both can be occupied simultaneously or if a *route exclusion* applies. This information can be represented in an *exclusion matrix* (Potthoff, 1962).

When assessing the capacity of a route node, a detailed analysis of all mutual route exclusions can lead to considerable complexity. Therefore, some methods for capacity determination partition the considered route node into several *serial route nodes* or *sub route nodes* (Schwanhäußer, 1978). Unlike the overall route node, all train routes within a serial route node (SRN) are mutually exclusive: Only one train movement can occupy the SRN at any given time. In an approach to analyze the capacity of a route node, each SRN could be modeled individually as a single-channel service system. However, dependencies between SRNs are ignored in such an approach. Therefore, some research studies (Nießen, 2008, 2013; Schmitz et al., 2017) model SRNs as resources that train movements exclusively reserve, but consider performance parameters on the whole route node.

It is not only the infrastructure that influences capacity. Capacity consumption (see also Section 2.1.4) often depends on train dynamics and the specific *train occupation sequence*. Individual *train occupation sequence cases* are distinguished, each describing a particular combination of a leading and a following train movement.

Due to the dependence of capacity on occupation sequences, some methods incorporate not only infrastructure but also timetables or schedules as input data. Other approaches rely solely on information about the frequency of different train types, which can be specified via an *operating program* or *procedure*. Such methods are referred to as *timetable-independent*.

The following section introduces the various concepts used to describe the capacity of a railway system.

2.1.3. Capacity Concepts

Various terms and metrics are used to analyze the performance of a railway system. As a key indicator, *capacity* describes the maximum number of trains that can be operated within a specified time horizon under certain conditions. In the following, three different concepts of railway capacity are distinguished.

- *Theoretical capacity* refers to the maximum number of trains that can be scheduled on a given infrastructure without any conflict, considering the traction characteristics of the rolling stock and the signaling and control systems. No quality requirements are taken into account.
- *Timetable capacity* describes the maximum number of trains that can be scheduled on a given infrastructure in acceptable quality for a specific train mix, taking traction dynamics, as well as signaling and control systems into account.
- *Operational capacity* (also referred to as *practical capacity*) denotes the maximum number of trains that can be operated on a given infrastructure while maintaining a defined level of service quality for a specific train mix. In addition to traction dynamics and train control, this concept also accounts for the effects of secondary delays with respect to a predefined delay threshold.

The considered influencing factors for the capacity concepts differentiated in this section are listed in Table 2.1.

In the literature, these capacity concepts are not always distinguished in exactly this way. For example, Abril et al. (2008) differentiate only between theoretical and practical capacity, where practical capacity serves as a generalization of both timetable and operational capacity by supplementing theoretical capacity with quality requirements. In Jensen et al. (2020), an additional concept of *maximum capacity* is introduced alongside theoretical capacity; maximum capacity does not consider quality thresholds but assumes a fixed train mix and heterogeneous operations.

The concept of timetable capacity used here incorporates quality considerations by accounting for waiting processes that occur during timetable construction (see also Wendler (2007)). When constructing a timetable, a railway infrastructure manager

Table 2.1.: Influencing factors of the considered capacity concepts

	Theoretical Capacity	Timetable Capacity	Operational Capacity
Infrastructure	✓	✓	✓
Train Control	✓	✓	✓
Traction Dynamics	✓	✓	✓
Train Mix	(✓)	✓	✓
Quality Standard	-	✓	✓
Timetable	-	(✓)	(✓)
Delays	-	-	✓
Dispatching	-	-	(✓)

must coordinate requests from various railway undertakings to use its infrastructure. These requests typically specify both the desired time and location for track occupancy, referred to as the *requested train path*. In pursuit of maximizing the number of requested train paths that can be scheduled on the infrastructure, the infrastructure manager may adjust requested paths within certain limits. This results in *scheduled waiting times*, defined as the temporal difference between the requested and actually assigned train path. An appropriately dimensioned infrastructure can thus be characterized by ensuring that waiting times arising in the timetabling process remain below a threshold value, allowing train paths to be allocated close to the initially requested time and therefore sustaining their attractiveness for transport operators.

2.1.4. Capacity Consumption

The capacity of railway transport systems depends particularly on how much capacity individual train movements consume on various infrastructure elements. The following sections introduce several key metrics related to this topic.

2.1.4.1. Blocking Times

To ensure safe train operations, conventional signalling systems divide the infrastructure into so-called *block sections*. A block section can only be traversed if it is not occupied by another train. During timetable construction, the infrastructure manager

must prevent multiple trains from occupying the same block section simultaneously. The duration for which a block section is occupied by a train is referred to as the *blocking time*, see also Happel (1959); Hansen and Pachl (2014).

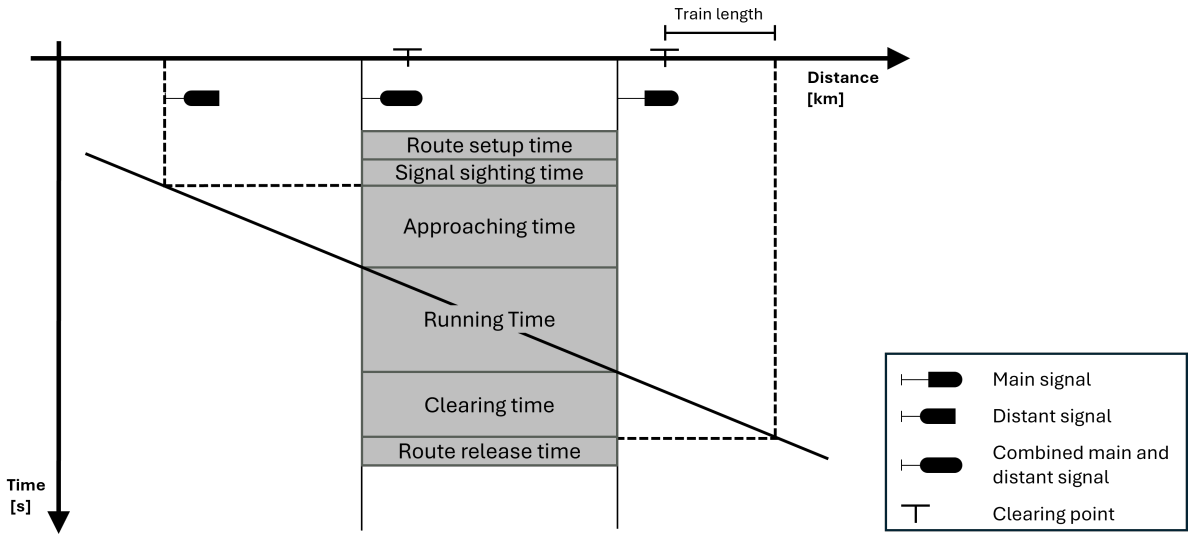


Figure 2.3.: Components of the blocking time in a time-distance diagram.

The blocking time consists of the following components:

- The *route setup time* denotes the required duration until the train route is established and secured against conflicting train movements. This duration depends primarily on the type of interlocking.
- The *signal sighting time* refers to the time needed for the train driver to perceive the signal. For train control systems without physical signals, *system reaction time* is used instead.
- The *approaching time* describes the time taken by the front of the train to travel from the distant signal to the main signal at the entry into the block section. In systems without fixed signals (e.g., LZB, ETCS), the braking initiation point is considered instead of a distant signal.
- The *running time (within the block)* represents the duration for which the front of the train moves between entering a block section and reaching its endpoint, such as at the next main signal.
- The *clearing time* denotes how long it takes for the entire train to clear a block section. This is determined by when the rear of the train passes the correspond-

ing clearing point.

- The *route release time* refers to the period from complete clearance of a block section until restoration of default settings for infrastructure elements used in that route.

Figure 2.3 illustrates the blocking time of a block section for a passing train movement in a time-distance diagram. Further explanations and the calculation of individual components can be found in Nießen et al. (2023) and Pachl (2022).

2.1.4.2. Track Occupancy in Stations

For scheduled stops in stations, infrastructure occupancy can be divided into three main phases:

- During the arrival phase, the train enters the designated track. It occupies the route from the entry signal to the stop position. The time consumed during this process is quantified as the *arrival blocking time*.
- During the actual stop, the train remains at the stop position for passenger exchange or freight operations. The *dwell time* does not necessarily represent only the operationally required duration (e.g. passenger exchange time), but may also include additional synchronization times to facilitate connections or waiting times for clearance of subsequent line sections.
- During departure, the train leaves the station track onto a line towards the next operating point. It occupies both the stop position and the route from the departure signal to its defined clearing point. The duration of this occupancy is referred to as *departure blocking time*.

Figure 2.4 illustrates the different time intervals associated with track occupancy for a stopping train.

The capacity assessments of route nodes requires information about the corresponding arrival or departure blocking times, which are generally composed of the same elements as described in Section 2.1.4.1. However, depending on the specific train occupation sequence case, different clearing points may be relevant when determining the clearing time. Additionally, in certain stopping scenarios, the signal sighting time within the departure blocking time must be replaced by (system) reaction times (Nießen et al., 2023; DB InfraGO, 2009).

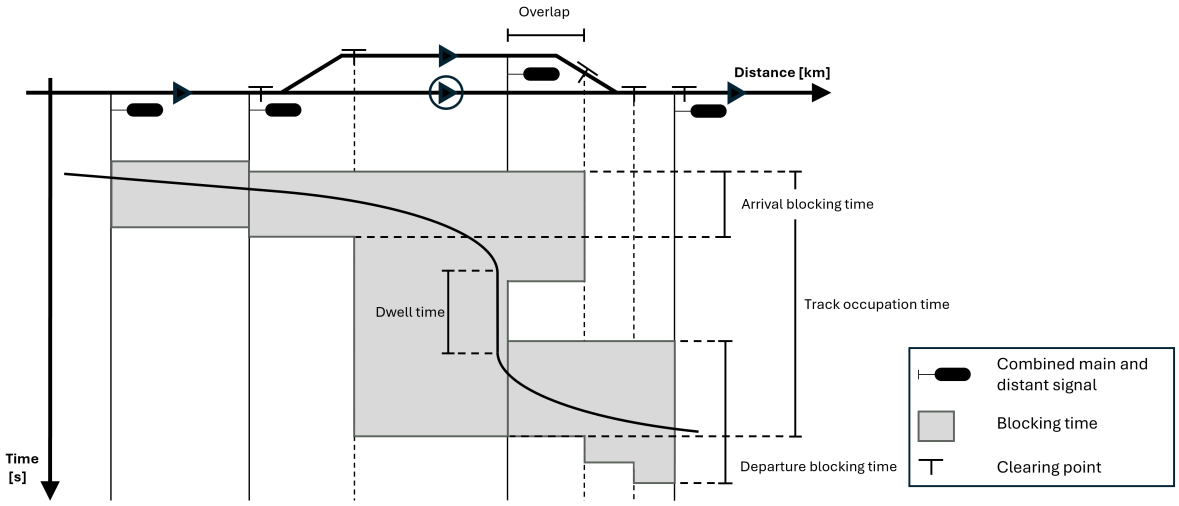


Figure 2.4.: Blocking and dwell times during track occupancy for a stopping train

For the dimensioning of track groups, the *track occupancy time* is required. This period begins with the start of the arrival blocking time and ends with the departure blocking time. In certain cases, as illustrated in Figure 2.4, the stop position may be released early at an additional clearing point, allowing the track occupancy time to end slightly before the departure blocking time. This is only possible if the subsequent train is traveling in the same direction as the considered train.

For some applications, track occupancy time is described as the sum of arrival blocking time, dwell time, and departure blocking time. In this context, arrival and departure blocking times are defined up to train standstill (or from train departure), differing from the description in Section 2.1.4.1. The dwell time then represents only the duration during which the train remains stationary at the stop position, ensuring that overlapping occupancy periods are not counted multiple times within the track occupancy time.

2.1.4.3. Minimum Headway Times

The required blocking times for individual block sections vary according to the resistance parameters and operational characteristics of train movements. In timetable construction, it is desirable that a train remains in continuous motion on the route between two passing stations and does not need to stop on open line segments between operating points to wait for clearance of a subsequent block section. Therefore, the

entire shared running path between two passing stations is considered to determine the separation of two consecutive train movements. For a given train, the blocking times of all relevant block sections form the *blocking time stairway* (Pachl, 2022). When scheduling consecutive trains, it must be ensured that their respective blocking time stairways do not overlap.

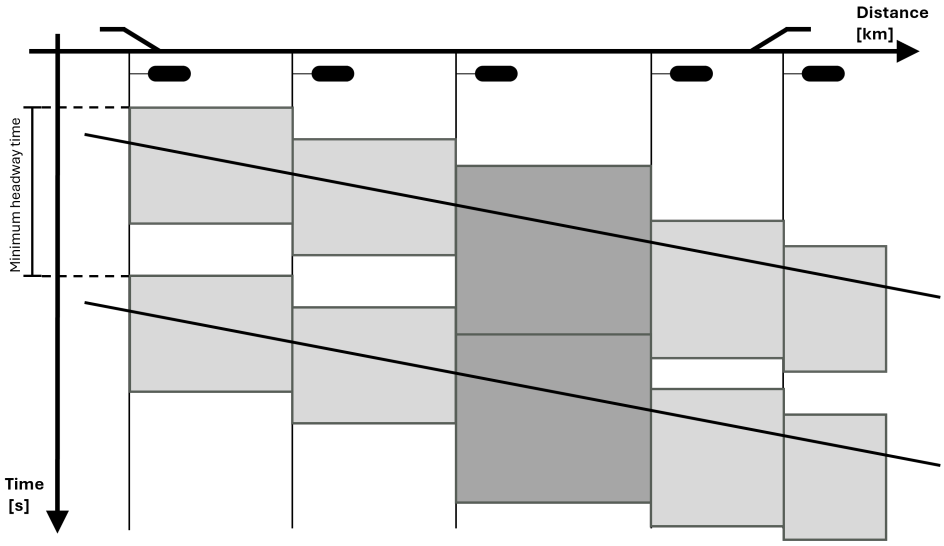


Figure 2.5.: The relevant block section for determining the minimum headway using two blocking time stairways.

The time difference between the start of the first block in the blocking time stairway of the preceding train and that of the following train is referred to as the *headway time*, and the shortest possible headway is called the *minimum headway time*. For the latter, there must be at least one block section where occupancy occurs consecutively in the diagram of both blocking time staircases, as shown in Figure 2.5. This block section is referred to as the *critical block section*, and its position is dependent on the considered train occupation sequence. For two identical trains, the section with the longest blocking time will usually be identified as the critical block section, as it is the case in Figure 2.5. However, an occupation sequence of more heterogeneous trains and stopping patterns might influence the position of the critical block section.

2.1.4.4. Supplements and Buffers

In the operation of complex systems, disruptions cannot always be avoided. To ensure smooth operations, it is therefore essential to incorporate additional times into

the timetable of a railway system. A distinction is made between *supplements* and *buffer times*. Supplements are added to the technically shortest running or dwell time to compensate for minor uncertainties in infrastructure conditions, train dynamics, or passenger exchange processes. Buffer times are scheduled to prevent the propagation of delays. They correspond to the minimum temporal separation between the blocking times of two consecutive blocking time staircases and thus represent the difference between headway and minimum headway times.

Some planning guidelines dictate the same minimum buffer times regardless of the specific line segment (Nießen et al., 2023). However, recent research shows that optimizing the distribution of buffer times can increase timetable robustness (Sels et al., 2016). Further references include review articles on robustness (Lusby et al., 2018) and on timetable construction in general (Correia Duarte et al., 2025).

2.1.5. Key Figures for Performance Analysis

The capacity of a railway system is defined by the relationship between quality and transport performance (see Section 2.1.3). To quantify quality, various quality indicators are used, which may differ depending on the specific capacity concept applied, see also (DB InfraGO, 2022). Table 2.2 provides an overview of selected key figures, which are briefly introduced below.

The *occupation ratio* describes the relationship between utilized and available capacity. Various methods and threshold values exist for its calculation and classification, one example being the UIC Code 406 (UIC, 2013).

The average *waiting time* refers to the time incurred due to the postponement of a requested occupancy, either during planning or in operation. Waiting times in planning are referred to as *scheduled*, while waiting times during operation are termed *unscheduled*. Unscheduled waiting times may also be referred to as *knock-on delays*, explicitly describing their occurrence when a delayed train causes following train(s) to wait for their occupation slot longer than anticipated during timetabling.

The *queue length* indicates the number of requests waiting to be served. In timetable construction, for any given moment within the analyzed period, the queue length corresponds to the number of train path requests whose desired start time was before the considered time but had to be postponed to a later time.

Table 2.2.: Selected key figures for describing the quality of railway systems

Name	Type	Planning	Operation	Application
Occupation ratio	Utilization	✓	✓	Line, Route Node
Waiting time	Temporal	✓	✓	Line, Route Node
Queue length	Number of trains	✓	✓	Line, Route Node, Station
Waiting risk	Probability	✓	(✓)	Track Group, Station
Loss risk	Probability	✓	-	Line, Route Node, Track Group, Station
Transport time ratio	Temporal	✓	✓	Line, Network
Concatenation value	Probability	✓	-	Route Node, Station
Punctuality	Rate	-	✓	Line, Station, Network
Achieved connections	Rate	-	✓	Station, Network
Change in delay	Temporal	(✓)	✓	Line, Station

The *waiting risk* describes the probability that a requested train path cannot be scheduled without postponement. The *loss risk* represents the likelihood that a requested train path cannot be fulfilled. For models in which a rejection of train paths is generally not intended, the loss risk can serve as an indicator for the accuracy of the model.

The *transport time ratio* describes the relationship between the shortest possible transport time and the transport time actually achieved for a request, either in planning or operation. A transport time ratio of 1.0 indicates that the achieved transport time matches the minimum possible transport time (planning) or the planned transport time (operation). In operation, a transport time ratio less than 1.0 may occur when the running time is faster than anticipated during planning, indicating that running time supplements have not been fully utilized.

The *concatenation value* (sometimes also referred to as *connectivity factor*) is an indicator of the complexity of a route node. The higher the concatenation value, the more likely it is that two train movements are mutually exclusive. A concatenation value of 1.0 means that all train movements are mutually exclusive, i.e., no parallel

service is possible.

In addition to these quality indicators, which are particularly suitable for strategical and tactical planning stages, further metrics can be formulated for the analysis of operations. These include indicators such as the average *punctuality*, the *number of achieved connections*, or the *change in delay* at isolated operating points. Both punctuality and the number of achieved connections may refer to the entire area under study, or they can be analyzed for more temporal or spatial specified components. The change in delay describes the increase or decrease in delay after passing an operating point compared to the value before that point. This can be analyzed either for individual train runs, or as an average for all trains within a given period or on a specific line.

2.2. Related Work

Various methods have been developed to analyze the performance of railway systems. This section gives an overview regarding some approaches grouped into six categories. Additionally, Section 2.2.7 discusses the applicability of methods from those six categories for the previously introduced capacity concepts.

2.2.1. Construction

One of the most widely used methods for assessing the capacity of railway lines is the compression of blocking time stairways for a given timetable. This *concatenation method* was introduced by Adler (1968) and is described in the recommendations of UIC (2004, 2013) for determining track capacity. In this approach, the blocking time stairways of all trains in a timetable on the considered line within the analyzed period are shifted as far forward in time as possible without causing conflicts or changing their sequence. The *occupation ratio* between the compressed occupation time and the analysis period describes the *capacity utilization* of the analyzed combination of infrastructure and timetable. This value can be compared with threshold values that differ according to line category to assess the quality of the timetable.

The described procedure for lines is also applied in some research (UIC, 2013; Landex and Jensen, 2013; Weik et al., 2020b) to the capacity assessment of route nodes. While UIC (2013) proposes to split nodes and separately analyze track groups and

route nodes, Landex and Jensen (2013) develop an approach considering the whole station, recognizing limitations of the compression approach with respect to shunting operations or multiple routing options in large stations. Furthermore, extensions have been proposed (Landex and Jensen, 2013; Jensen et al., 2017; Weik et al., 2020b) that suggest a timetable-independent variant of this concatenation method. They assume probability distributions over different timetables and calculate corresponding expected values of utilization metrics across individual timetable cases.

For fixed timetables, capacity utilization can be efficiently determined using *Max-Plus automata* (Goverde, 2007; Bešinović and Goverde, 2018). In this model, blocking time stairways are concatenated via matrix-vector multiplications within the Max-Plus algebra, allowing the computation of a minimum cycle time for the entire timetable. This approach also enables an analysis of delay propagation effects (Goverde, 2007, 2010).

2.2.2. Delay Propagation Analysis

The analysis of delay propagation does not necessarily require the specification of a timetable. For railway lines, Schwanhäußer (1974) introduced the *StreLe* method (german: Strecken-Leistungsfähigkeit), which calculates average unscheduled waiting times. For single-track lines, Kraft (1988) describes additional models for timetable-independent capacity analysis. While these models are particularly applied by North American operators, DB InfraGO (2022) employs the StreLe method to calculate the nominal capacity of railway lines.

Other approaches use timetable data to describe delay propagation on lines (Goverde, 2010; Şahin, 2017; Corman and Kecman, 2018) or within networks (Büker, 2011). Delay transmission is modeled using Max-Plus automata (Goverde, 2010), Markov Chains (Şahin, 2017), Bayesian networks (Corman and Kecman, 2018), or the convolution of probability distributions (Büker, 2011). In some cases, operational data is used to determine the characteristics of modelled delay processes (Şahin, 2017; Corman and Kecman, 2018).

2.2.3. Analysis of Operational Data

Systematic *operational data analysis* provides another approach to assess system performance. An important application is the data-driven identification of bottlenecks in infrastructure or timetables, which enables infrastructure managers and railway undertakings to implement targeted improvements. One method involves comparing timetable and operational data, as applied in the Swiss railway network (Graffagnino, 2012) or at selected stations in Great Britain (Armstrong and Preston, 2017). Cerreto et al. (2018) employ cluster analysis techniques to identify patterns within delay data on a Danish railway line.

For the analysis of macroscopic relationships, Corman and Henken (2022) and Weik (2022) propose plotting key figures such as train density (number of trains per route), flow (number of trains per time unit), delay, or speed for a large set of data points. This can be achieved either by directly utilizing operational data (Corman and Henken, 2022), or by generating the required dataset through simulations (Weik, 2022) for different scenarios.

2.2.4. Simulation

The calculation of key figures through *simulations* is a practical approach for capacity analysis of individual timetables or infrastructure scenarios (Abril et al., 2008). Simulations can represent various levels of infrastructure detail. Microscopic simulations (Schmidt and Martin, 2010) may be used to analyze the capacity of junction areas or small sub networks. For instance, they can be used to quantify the impact of dispatching decisions on operational quality (Liang et al., 2017).

Other methods, in which occupancy conflicts are primarily modeled by enforcing minimum headways, are referred to as macroscopic simulations. Such approaches enable network-wide analyzes of delay propagation for specific timetables (Franke et al., 2013), or timetable-independent assessments of how buffer time distributions affect the operational quality of railway lines (Zieger et al., 2018).

A recent paper (Becker et al., 2025) describes a tool for conducting network-wide Monte Carlo simulations. In this approach, inaccuracies in resolving deadlocks are accepted in order to reduce simulation complexity and therefore enable an analysis of the entire German railway network. As a result, a single simulation run does not

necessarily correspond to a conflict-free schedule of train operation. Nevertheless, an evaluation using forecasts and actual punctuality data from 2022 demonstrates the tool's suitability for qualitative assessments at the network level.

2.2.5. Optimization

Conflict-freeness, however, is one of the central constraints in *optimization* approaches for capacity analysis.

In Harrod (2009) and Yaghini et al. (2014), the theoretical capacity of a line is determined by scheduling as many train movements as possible without conflicts. Both sources identify the train mix as an important factor influencing the capacity of the considered infrastructure.

The theoretical capacity of railway nodes is analyzed in Delorme et al. (2001) by assigning station routes to as many train movements as possible. This subproblem can be addressed using algorithms for *node-packing* (Zwaneveld et al., 1996) or *set-packing* (Lusby et al., 2011) problems.

Optimization problems are also formulated to determine the maximum throughput in networks. For example, Burdett and Kozan (2006) calculate the theoretical capacity of a railway network with a fixed train mix using an optimization approach. In an extension (Burdett, 2015), multi-objective optimization problems are formulated, including the calculation of a Pareto front for theoretical capacity in terms of the number of freight and passenger trains.

An extension to other capacity concepts (see Section 2.1.3) is presented in the work of De Kort et al. (2003), where a Max-Plus algebra is used to determine the starting time of the blocking time for the last train. Subsequently, they calculate the probability that the start of the blocking time for the last train falls within the analysis period for a given level of delay. In a case study on a railway line with several single-track sections, De Kort et al. (2003) further determine the maximum number of trains for different threshold values of this probability. Mussone and Wolfler Calvo (2013) propose an optimization problem for determining the capacity of railway infrastructure, in which the maximum number of trains is calculated as the sum of *regular* and *affected* trains. While theoretical capacity is set as an upper limit for lines, at nodes, affected trains receive additional time penalties, which, together with the regular occupation times, should not exceed the total investigation period.

Further research includes, for example, the integration of decisions regarding block division or network expansion (Burdett, 2016) while considering theoretical capacity, as well as the additional consideration of vehicle rotations (Liao et al., 2021) in calculating the maximum throughput for a section of a long-distance rail network in the region around Zhengzhou, China. To investigate the operational *residual capacity* of a railway network, Borndörfer et al. (2016); Meirich (2017); Weik et al. (2020a); Nikolayzik et al. (2024) determine the number of freight trains that can be accommodated in addition to an existing timetable while maintaining a specified level of operational quality.

Some of the described articles (Delorme et al., 2001; Burdett and Kozan, 2006; Harrod, 2009; Lusby et al., 2011; Mussone and Wolfler Calvo, 2013; Yaghini et al., 2014; Burdett, 2015, 2016) employ the technique of *timetable saturation*, which refers to the generation of maximally dense timetables into which no additional train movement can be inserted without conflict. There is thus a close relationship with optimization problems for timetable construction. An overview of corresponding research can be found in the review papers by Cacchiani and Toth (2012) and Correia Duarte et al. (2025).

2.2.6. Queueing-based Models

In contrast, the publications summarized in this work consider methods that are based on modeling capacity allocation using *queueing systems*. Queueing systems have been used for several decades to describe transport systems, early examples are given in Potthoff (1962) and Fischer and Hertel (1990). Fischer and Hertel (1990) characterize a queueing system by the property that both the arrival time of demands to be served and the required duration of the actual service process are randomly distributed. Due to these stochastic elements, queues may arise, and their expected length can be used as an indicator to describe the performance of the system (see also Section 2.1.5). To calculate the expected average queue length (and other indicators), corresponding queueing systems can (approximately) be formulated as *Continuous-Time Markov Chains*.

Table 2.3 combines some information on queuing-based methods to give further insights into the contribution of this thesis. Here, the symbol $(\checkmark)$ indicates a not fully comparable scope of the freight-station approach in Bychkov et al. (2021), trade-offs

in the route-based decomposition for single-channel approaches (Schwanhäußer, 1978; Weik, 2020) or another decomposition evaluated on routes (Nießen, 2008, 2013).

Table 2.3.: Queueing-based methods for timetable capacity analysis on railway nodes

Method	Scope		Model details			
	Route node analysis	Track group analysis	Route-based decomposition	Explicit non exp.-processes	Multi-channel model	Solution method
Schwanhäußer (1978)	✓		(✓)			Closed-form formula
Nießen (2008, 2013)	✓		(✓)		✓	Iterative formula
Schmitz et al. (2017)	✓			✓	✓	Matrix-vector equations
Weik (2020)	✓		(✓)	✓		Matrix-vector equation
Potthoff (1962)		✓			✓	Iterative formula
Hertel (1984)		✓			✓	Closed-form formula
Bychkov et al. (2021)	(✓)	(✓)			(✓)	Simulation
Chapter 3 (of this thesis)	✓		✓		✓	Probabilistic Model-Checking
Chapter 4	✓		✓	✓	✓	Probabilistic Model-Checking
Chapter 5	✓	✓	✓		✓	Probabilistic Model-Checking

The next paragraphs give further information regarding existing work tackling track group analysis with queueing-based methods.

Potthoff (1962) and Hertel (1984) introduce approaches to determine the required number of tracks within the analyzed track group. For this, they calculate the waiting probability (see Section 2.1.5) for occupation requests of a track modelled as a service channel within a queueing system. Potthoff (1962) considers a randomized arrival process and constant track occupation times (see Section 2.1.4.2) to model filling and service functions of the station. He assumes that all requests without an immediate occupation slot are lost and calculates the loss probability as an approximation of the waiting probability. In a more general analysis of queueing systems with multiple channels and unlimited waiting positions, Hertel (1984) acquires a formula to estimate waiting probabilities and queue lengths for systems with general randomized

arrival and service processes on the basis of results for exponential distributed inter-arrival and service times. Using the track occupation times as service times, this can be applied to determine the performance capability of track groups within a railway station.

To analyze the performance of entire freight railway stations, Bychkov et al. (2021) formulate a network of queuing systems, splitting it into multiple yards. They consider a simultaneous arrival of freight cars in batches and calculate performance indicators such as the loss probability for two different freight stations. However, they refrain from explicitly formulating the underlying states due to the complexity of the system and obtain results with a simulation approach instead.

2.2.7. Summary and Discussion

In summary, railway infrastructure managers can already select from a broad spectrum of available methods when assessing the performance capability of their network. Table 2.4 evaluates the applicability of the discussed methods for the capacity concepts from Section 2.1.3.

Table 2.4.: Applicability of Procedure Classes within Different Capacity Concepts

	Theoretical Capacity	Timetable Capacity	Operational Capacity
Construction	✓	(✓)	-
Delay Propagation Analysis	-	-	✓
Operational Data Analysis	-	-	✓
Simulation	-	(✓)	✓
Optimization	✓	(✓)	(✓)
Queueing-based Model	-	✓	(✓)

Construction or Optimization approaches might be the first choice when determining whether the inclusion of another train into an already existing timetable is feasible without occupation conflicts. While it is certainly possible to include quality metrics into such considerations, other methods may be more suitable. When considering the

operational quality, infrastructure managers may theoretically analyze delay propagation in their network and evaluate actual historic data or simulate operations on new infrastructure. The main realization of traffic capacity however takes place during the timetabling stage. The efficiency of infrastructure is determined during the process where occupation requests from railway operators actually get assigned to time-slots. Therefore, infrastructure managers might be mainly concerned about the timetable capacity of their infrastructure. Here, queueing-based methods have the advantage of not requiring previous knowledge about a timetable while delivering deterministic results. However, in practice their application has been limited to railway lines and approximations of railway route nodes. So far, infrastructure managers analyzing the capacity of railway stations have typically either relied on simulation models or decomposed the station into track groups and route nodes for independent analysis.

2.3. Motivation and Research Goal

The evaluation of related work in the last section included some existing approaches to analyze the timetable capacity of railway nodes. However, it also concluded that existing methods from both suitable classes, simulation and queuing-based models, have certain drawbacks that limit their applicability.

Simulation approaches, for example, will simulate only one interpretation of the possible outcomes of the random processes in every run. Therefore, users need to perform a high number of individual iterations, in order to come up with meaningfull results. In comparison, queueing-based models return a deterministic result for a given infrastructure instance. But established techniques for railway station analysis require separate consideration of route nodes and track groups. Furthermore, existing methods either consider a single-channel simplification of complex multi-channel systems at railway route nodes (Schwanhäußer, 1978; Weik, 2020) or require additional data and expert knowledge to further decompose route node infrastructure into single-channel ressources, such as serial route nodes (Nießen, 2008, 2013; Schmitz et al., 2017). This creates substantial effort for infrastructure managers if they are to apply these methods across their entire network. Additionally, applied solution methods ususally require the formulation of complex (matrix-vector) equation systems for every scenario, which can be quite cumbersome and may be prone to errors.

In this work, novel queueing-based methods to estimate timetable-independent per-

formance metrics for railway junctions and stations are introduced. Timetable capacity is calculated by abstracting the capacity allocation process with multi-dimensional Continuous-Time Markov Chains, specifically designed to model the parallel servicability of different routes within the considered infrastructure. With this route-based decompositon, performance analysis requires only the knowledge of route-exclusions in addition to information regarding service or arrival processes.

The here proposed methods obtain quality indicators by directly applying probabilistic model-checking to explicitly formulated Continuous-Time Markov Chains. With established software for probabilistic model-checking, multiple performance parameters can be checked with the same Markov Chain representation. This not only enables efficient iteration over different input parameters, but also allows the inclusion of additional parameters, for instance for model validation. One major contribution (see also Section 3.2.3) is therefore the development of formal representations of Continuous-Time Markov Chain models for railway node capacity determination, enabling the use of state-of-the-art probabilistic model checking software, and their embedding into frameworks for railway performance analysis.

The publication (Emunds and Nießen, 2024a) in Chapter 3 of this thesis introduces Continuous-Time Markov Chain models, designed for the timetable-independent performance analysis of railway junctions. The model is validated against simulations and implemented within two main application scenarios: The determination of the timetable capacity of an existing junction infrastructure and the planning of an optimal infrastructure layout for different operating scenarios.

This approach is extended with explicit representations of non-exponential arrival and service processes in the publication (Emunds and Nießen, 2025c) in Chapter 4. An improvement of the approximation quality can be quantified with a comparison to simulation results. Furthermore, the applicability is demonstrated in a route-based capacity analysis for an exemplary railway junction.

Another extension is the consideration of entire railway stations (Emunds and Nießen, 2025b) in Chapter 5. To the best of our knowledge, the presented method is the first to enable the queueing-based analysis of an entire railway station in one combined multi-channel model. It can be utilized to identify bottlenecks in routes or service processes on considered railway stations. A case study highlights its capabilities to assist in the decision-making of infrastructure managers.

3. Evaluating Railway Junction Infrastructure: A Queueing-Based, Timetable-Independent Analysis

Abstract

Many infrastructure managers have the goal to increase the capacity of their railway infrastructure due to an increasing demand. While methods for performance calculations of railway line infrastructure are already well established, the determination of railway junction capacity remains a challenge. This work utilizes the concept of queueing theory to develop a method for the capacity calculation of railway junctions, solely depending on their infrastructure layout along with arrival and service rates. The implementation of the introduced approach is based on probabilistic model-checking. It can be used to decide which infrastructure layout to build, i.e. whether an overpass for the analysed railway junction is needed. The developed method addresses the need for fast and reliable timetable-independent railway junction capacity evaluation, catering specifically to the long-term strategic planning of junction infrastructure.

3.1. Introduction

The amount of goods and passengers transported by rail is predicted to increase significantly, hence infrastructure managers not only need to build new railway infrastructure but also explore enhancements for traffic volume on existing lines and nodes. Both of these tasks require sufficient methods to determine the capacity of all sub-units, i.e. lines, junctions and stations, in a railway network.While Capacity Determination techniques for railway lines, junctions and stations have already been developed and continue to contribute to more efficient rail transportation today, only some methods describe timetable-independent approaches, allowing for sophisticated infrastructure assessment in the early stages of the planning process, which is traditionally done in multiple steps (UIC, 2013). Strategic design decisions on a

network level may take place decades before operation and are an ongoing process for most infrastructure managers. The technical infrastructure planning does take some time and is done several years in advance, while the definitive timetable construction may be done several months or years prior to operation. Therefore, timetable-independent methods determining the capacity of railway infrastructure are of particular importance for the early process stages including 'strategical network design' and 'technical infrastructure planning'.

The main contribution of this work is the introduction of a novel Continuous-Time Markov Chain (*CTMC*) formulation, that

- extends queueing-based capacity research to parallel utilizable infrastructure,
- enables timetable-independent capacity determination by solely depending on route conflicts, traffic shares and route service times, and
- is capable of highlighting the influence of different routes on the performance of analysed infrastructure.

It estimates the number of trains waiting for the assignment of their requested infrastructure in the timetabling process, by calculating state probabilities in a CTMC modelling the queuing system of a multi-route railway junction. This work utilizes state-of-the-art probabilistic model-checking software to perform the computations. The presented approach does not require timetable data to determine the capacity of a railway junction, enabling early-stage infrastructure planning. Consequently, technical planning decisions, i.e. whether a railway overpass is to be built, can be assisted by capacity calculations with the described approach. In comparison with other approaches to determine timetable-independent capacity, the presented method does not rely on simulations or sampling of possible sequences for timetable-compression, but rather on queueing systems, formulated for multiple service channels. With its versatile range of parameters, multiple applications of capacity research, i.e. infrastructure planning for a fixed traffic demand or capacity determination of a fixed junction infrastructure, may be realised.

While Section 3.2 includes a literature review regarding other approaches to the determination of railway capacity, Section 3.3 introduces formal problem formulations and proposes the new capacity calculation method. In Section 3.4, the proposed method is compared to a simulation approach concerning computation times and accuracy of the obtained solutions. Two Case Studies, highlighting the applicability of

the introduced approach, are performed in Section 3.5. This work is concluded with a summary and discussion in Section 3.6.

3.2. Related Work

This section provides an overview of the state-of-the-art for railway capacity analysis. After the definition of some key types of performance analysis and frequently used terms, selected examples of literature are matched regarding their associated methodology and capacity definition.

3.2.1. Terminology of railway performance analysis

With their various requirements to the grade of detail, different stages of the planning process can require an analysis of diverse definitions of railway capacity. While a first definition of railway capacity as the *maximal number of trains traversing a given infrastructure in a given time period under some fixed assumptions* summarizes the concept in a straightforward manner, the interpretation of the included assumptions determines different levels of capacity. In detail, three types of railway capacity may be distinguished (see also Jensen et al., 2020):

- *Theoretical capacity*: The maximum number of requests (i.e. trains, train-route enquiries), that can be scheduled without conflicts on the given infrastructure under consideration of driving dynamics and installed railway control systems.
- *Timetable capacity* (sometimes referred to as *maximal capacity*): The maximum number of requests, that can traverse the given infrastructure in acceptable quality when compared to a specified threshold, not only considering driving dynamics, railway control systems, but also operating program specific settings, such as train-mix and arrival processes.
- *Operational capacity* (sometimes referred to as *practical capacity*): The maximum number of trains, that can traverse the given infrastructure in acceptable operational quality when compared to a specified threshold, considering driving dynamics, railway control systems, operating programs, and additionally respecting disturbances and (knock-on) delays.

Additionally, research has not only been focused on determining the capacity (of any

kind), but also on the calculation of the *capacity utilization*, i.e, the amount of available capacity, consumed in a given timetable.

While methods determining capacity utilization are mostly dependent on a timetable (*timetable dependent*) or at least a fixed order of a given train set, some methods are capable of determining performance indicators without the need for a previously set timetable (*timetable independent*), making them valuable for infrastructure assessment in early stages of infrastructure planning processes. Some approaches build on a given timetable and find maximal subsets of a set of trains that can additionally be scheduled (*timetable saturation*), building a category in between both dependency expressions. In contrast to a timetable, an *operating program* specifies the demand of train types on given lines or routes, hence being indispensable for most performance analyses.

Depending on the stage of railway infrastructure planning, a timetable may already be available, such that methodologies like *timetable compression*, i.e. routing trains through the infrastructure with a separation of only the minimum headway possible, are useable.

The various railway infrastructure planning stages make use of varying capacity definitions and hence distinct methodologies for the analysis of railway infrastructure. They additionally differ in terms of timetable dependency or granularity of the described infrastructure. During the following subsection, the methodologies *max-plus-algebra*, *optimisation*, *operational data analysis*, *simulation* and *analytical* are differentiated.

Some additional distinctions of relevant railway performance analysis are made regarding their analysed infrastructure – lines, junctions, stations or networks –, their infrastructure decomposition and their utilized solution methods, f.e. *mixed integer programming* (MIP) or matrix calculations.

3.2.2. Literature Review

The described terminology is used in Table 3.1 to partition some relevant related research regarding their associated category. Additionally, selected utilization, optimisation and delay-propagation methods are briefly introduced in the following, while our contribution is classified along the strongly related analytical approaches. An additional table, categorizing the considered literature can be found in the Appendix in

Table A.1.

Table 3.1.: Literature considering railway performance estimations

methodology	capacity type	timetable independent	junction evaluations	stochastic arrival process	stochastic service process	application specific infrastructure decomposition	solution technique
timetable compression	util	no	(yes)	no	no	no	compression
randomized timetable compression	util	yes	(yes)	no	no	no	combinatorial optimisation
max-plus-algebra	util	(yes)	yes	no	no	no	matrix calculations
timetable saturation	theo	(yes)	(yes)	no	no	no	mostly MIP
capacity optimisation	theo, util	(yes)	(yes)	no	no	no	mostly MIP
operational data analysis	op	no	no	(yes)	(yes)	no	stat. analysis and machine learning
delay propagation analysis	op	(yes)	no	(yes)	yes	no	matrix calculations and iterative formula
simulation	op, tt	(yes)	yes	(yes)	yes	no	simulation computations
analytical line capacity	tt, op	yes	no	yes	yes	no	closed formula
(prior) analytical node capacity	tt, op	yes	yes	yes	yes	yes	matrix calculations and iterative formula
introduced here	tt	yes	yes	yes	yes	no	probabilistic model-checking

Remarks: Please note that a statement in bracelets means that this feature is partially supported, i.e. for some methods within the group or in an incomplete interpretation only. Furthermore, capacity types utilization (util), theoretical (theo), operational (op) and timetable (tt) are abbreviated.

Describing the determination of **capacity utilization** and introducing threshold values for sufficient operational quality, the UIC Code 406 (UIC, 2004, 2013) is widely used internationally for capacity assessments on railway lines (Abril et al., 2008; Landex, 2009; Goverde et al., 2013) and stations (Landex and Jensen, 2013; Weik et al., 2020b). Utilizing the theory of capacity occupation, Max-Plus Automata (Goverde, 2007; Bešinović and Goverde, 2018) give measures for the assessment of railway infrastructure and timetable structure. While traditional compression methods (UIC, 2004; Goverde, 2007; Abril et al., 2008; Landex, 2009; Goverde et al., 2013) require a timetable to calculate capacity consumption, other approaches have been proposed utilizing randomly generated sequences of train types to overcome timetable dependencies (Jensen et al., 2017, 2020; Weik et al., 2020b), focusing on the determination of timetable capacity. An overview and comparison with other capacity consumption methods can be found in Zhong et al. (2023).

Furthermore, **optimisation methods** for the estimation of theoretical capacity have been developed. They mostly formulate (linear) mixed integer programming problems, including approaches for railway lines (Harrod, 2009; Yaghini et al., 2014), stations (Zwaneveld et al., 1996, 2001; Delorme et al., 2001; Lusby et al., 2011) and networks (Burdett and Kozan, 2006; Burdett, 2015, 2016). Some approaches rely

on solutions to the railway timetabling problem (Cacchiani and Toth, 2012; Leutwiler and Corman, 2022), building timetables utilizing a given infrastructure in a 'best', as defined by objective functions, manner. Methods may also estimate the capacity occupation of its solution while creating a timetable (Zhang and Nie, 2016). Heuristic solution techniques for job-shop formulations of the railway timetabling problem have been used for railway junctions and crossovers (Burdett and Kozan, 2009).

Other approaches are based on the saturation of given timetables, which may also include empty schedules, and optimise the amount of additional traffic (Burdett and Kozan, 2006; Harrod, 2009; Lusby et al., 2011; Liao et al., 2021). Expanding their work (Burdett and Kozan, 2006), Burdett (2016) introduces MILP approaches to calculate optimal track sub-divisions on a railway line or determine which connections within a railway network can benefit from track duplications, considering the theoretical capacity of railway infrastructure.

Other optimisation methods may incorporate rolling stock information (Liao et al., 2021) to the construction of a saturated timetable or estimate the effects of emerging first-order delays to following trains (Mussone and Wolfler Calvo, 2013), handling the propagation of so called *knock on* delays. For this, Mussone and Wolfler Calvo (2013) extend an approach by De Kort et al. (2003), utilizing max-plus algebras to calculate capacity on lines with some single-track sections, i.e. tunnels.

More detailed insights into operational parameters of specific timetables and detailed infrastructure, rolling stock and delay distribution data can be obtained by utilizing **simulations**. D'Acierno et al. (2019) provide a comprehensive literature review. While being subject to large computational times, simulations are versatile in their use case. They can simulate train operations based on a given timetable or even include arrival and service processes as random variables. The influence on model performance of different parameters can be analysed, i.e. different buffer time distributions to operational capacity estimations (Zieger et al., 2018), analysing the propagation of delays on railway infrastructure.

An investigation of **delay propagation** properties has also been part of further analytical (Goverde, 2010; Büker, 2011; Burdett and Kozan, 2014b,a) research and subject of machine-learning models (Şahin, 2017; Corman and Kecman, 2018), trained on historical operational data. Şahin (2017) calculates probabilities for different delay states with the help of a Markov Chain, while Corman and Kecman (2018) utilize Bayesian networks. Interested readers are referred to Spanninger et al. (2023) for a

detailed review.

Recent approaches make additional use of **operational data** to identify capacity bottlenecks. While some research introduces train traffic data mining approaches to analyse actual operation on railway lines (Graffagnino, 2012) or stations (Armstrong and Preston, 2017), others (Weik, 2022; Corman and Henken, 2022) discuss the applicability of macroscopic fundamental diagrams. For this, Weik (2022) introduces a simulation approach to highlight their benefit for further macroscopic research and Corman and Henken (2022) provide an overview of open research applications.

Taking randomly distributed inter-arrival and service times into consideration, **analytical methods**, based on queueing theory, have been developed for an efficient timetable or operational capacity analysis in the early planning stages. Table 3.2 contains a comparison and summary of their features related to this work for a selection of analytical methods.

Table 3.2.: Analytical capacity determination methodology

Literature	junction evaluations	timetable capacity	route-based decomposition	route-based quality assessment	solution method
(Potthoff, 1962)		✓	✓		iterative formula
(Schwanhäußer, 1974)			(✓)*	(✓)*	closed-form formula
(Schwanhäußer, 1978)		✓	✓		closed-form formula
(Wakob, 1984)		✓	(✓)*	(✓)*	closed-form formula
(Wendler, 2007)		✓	(✓)*	(✓)*	matrix-vector equations
(Nießen, 2008, 2013)	✓	✓		(✓)	iterative formula
(Schmitz et al., 2017)	✓	✓			matrix-vector equations
(Weik, 2020)	(✓)	✓	✓		matrix-vector equations
introduced here	✓	✓	✓	✓	probabilistic model-checking

Remarks: (✓)* indicates that the method is used for line capacity and does hence support route-based decomposition and quality assessment natively.

In (Potthoff, 1962) railway lines and stations are dimensioned, analysing loss probabilities depending on filling and service functions utilizing probability distributions for the arrival process of trains to the analysed station. While Schwanhäußer (1974); Wakob (1984); Wendler (2007) introduce measures for an analytical determination of the capacity of lines, Schwanhäußer (1978); Nießen (2008, 2013); Schmitz et al. (2017) analyse junctions, incorporating randomly distributed inter-arrival and service

durations.

Schwanhäußer (1974) formulates the STRELE method to calculate expected values for the waiting times of trains in a line infrastructure, which is extended to junction infrastructure in Nießen (2008). They hence give measures to calculate the **operational capacity** of lines and junctions by comparing the estimated waiting times with threshold values (Schwanhäußer and Schultze, 1982).

Potthoff (1962); Schwanhäußer (1978); Wendler (2007); Wakob (1984); Nießen (2008, 2013); Schmitz et al. (2017); Weik (2020), however, implement results for the **timetable capacity** of the analysed infrastructure by matching corresponding limits with estimated waiting times for trains without taking delays into account.

Calculating the expected waiting times via probabilities for the loss of an arriving train, Potthoff (1962); Nießen (2008, 2013) formulate methods for station (Potthoff, 1962) and junction infrastructure. For junction capacity estimations, Nießen (2008, 2013) tackles queueing systems with multiple channels, while Schwanhäußer (1978) approximates the waiting times in a junction via a single-channel system. This single-channel system approximation is based on route-exclusion probabilities and is adapted in Weik (2020), joining it with line-capacity advancements in Wendler (2007), therefore utilising multi-state service processes and hence more flexible probability distributions when modelling the service process.

Parameter estimations have been done (Wakob, 1984) to obtain a closed formula for approximating the actual waiting times on a railway line with general independent service and inter-arrival time distributions.

Also directly modelling general independent service and arrival processes, Schmitz et al. (2017) formulate multi-dimensional Markov Chains and include phase-type distributions in their approximation. However, their approach is limited to already partitioned junction infrastructure, hence dependent on additional input and deep system knowledge of the algorithm operator. Additionally, they distinguish between train and request types in a queue, resulting in issues with computational memory and scaling, when analysing a more complex infrastructure and/or operating program.

3.2.3. Motivation and Contextualisation

Overall, the methodology landscape is still lacking research combining major advantages of analytical and other timetable-independent junction capacity methods: Calculating timetable capacity for multi-channel railway junctions while being easily applicable to a broad range of problem formulations.

This work introduces a novel methodology that calculates timetable capacity with the help of problem statements that solely depend on the routes through the considered infrastructures, the parallel serviceability of those routes and the possible operating programs. A more formal introduction to the considered problem statements can be found in Section 3.3.2. The introduced approach can therefore be used at early stages in the planning process and does not require any additional system-specific knowledge to further partition the junction infrastructure.

Following other analytical models, timetable capacity is analysed with the help of Continuous-Time Markov Chains. However, the introduced multi-dimensional Markov Chain formulation is obtained by utilizing a novel route-based decomposition. It is therefore easily applicable to more complex junctions by further extending the multi-dimensional state-space with additional dimensions for the modelling of supplementary routes.

While other methods implementing multi-dimensional Markov Chains (Schmitz et al., 2017; Weik, 2020) require a set-up of Matrix-Vector equations at the modelling step and use more general frameworks to solve those equation systems on the equation level, the introduced method applies probabilistic model-checking methodology directly to the obtained Continuous-Time Markov Chain. Hence, the full broadband of graph-theoretical and numerical optimized algorithms for probabilistic model-checking may be utilized to acquire the solution. Furthermore, by explicitly formulating the Continuous-Time Markov Chain, multiple properties, describing different characteristics of the analysed infrastructure, may be checked with the same Markov Chain model. This can be utilized for more efficient parameter iteration, i.e. for different arrival- and service rates in the presented approach, or in further research even for reachability or deadlock analysis of the formulated model, which might be suitable as additional model validation steps.

In this work, infrastructure performance is measured by a timetable-independent approach analysing the timetable capacity, modelled with a multi-dimensional Continuous-

Time Markov Chain and calculated with a probabilistic model-checking framework. The later discussed model features a route-based infrastructure decomposition with fixed route allocations, hence being easily applicable to more complex junctions, while respecting arising resource conflicts without the need for additional infrastructure partitioning.

3.3. A Continuous-Time Markov Chain Model For The Queueing System In A Railway Junction

In this section, a CTMC formulation, modelling the timetable process of a railway junction, will be introduced. The first section (Section 3.3.1) will illustrate an exemplary junction layout and the second section (Section 3.3.2) will introduce a detailed problem definition. Afterwards, queuing systems in railway line performance modelling will be described (Section 3.3.3) and the CTMC model for a railway junction will be introduced in Section 3.3.4. Following Section 3.3.5, discussing the scalability of the introduced approach, a method to perform the estimations with the modelled CTMC will be proposed (Section 3.3.6). Furthermore, threshold values for the determination of junction capacity will be discussed in Section 3.3.7. This Chapter will conclude with sections discussing the methodical procedure (Section 3.3.8) and contributing factors (Section 3.3.9) of the introduced approach.

3.3.1. Junction Layout

To evaluate the capacity of a railway system, all sub-units of the network need to be assessed. A common differentiation is made between line and node capacity, while methods describing node capacity can additionally be partitioned into methods describing junction and methods describing station capacity.

A *railway line* is a connection between two origins, usually equipped with single- or double-track infrastructure. If some traffic shares only part of the route and connects to a third origin, a *railway junction* will be installed to divide the traffic regarding its destination. Unlike at junctions, trains may start and end their journey at *railway stations*. They can be used for passenger- and good-exchange or fulfil primarily operational needs, such as ensuring a possibility for over-takings.

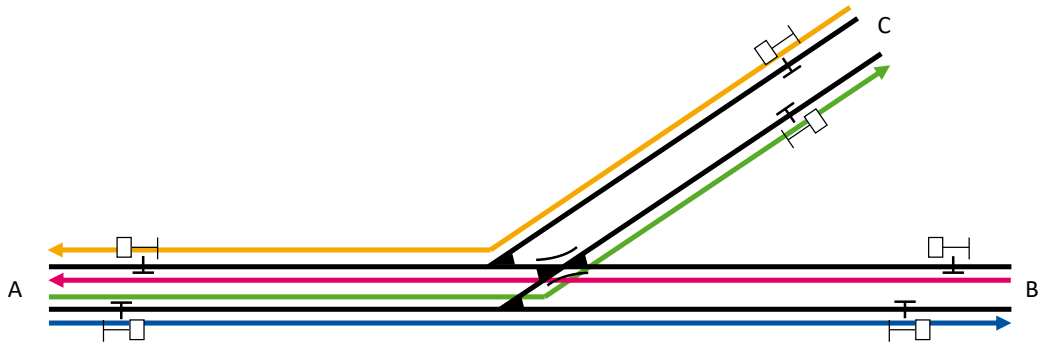


Figure 3.1.: Track-layout for a double-track junction

In Figure 3.1 an infrastructure design for a railway junction of two double-track lines is given. The *main line* spans between the two origins A and B, while the *branching line* connects C with A.

As in most European countries, we consider operation to be defaulting to right-hand traffic, which suggests the usage of route r_1 for the direction A to B and routing traffic from B to A via route r_3 for the main line. Consequently, routes r_2 and r_4 are used accordingly for traffic between A and C.

Table 3.3.: Routes through the junction

Name	Origin	Destination	Conflicts
r_1	A	B	r_2
r_2	A	C	r_1, r_3
r_3	B	A	r_2, r_4
r_4	C	A	r_3

When dividing traffic on the junction into routes, those four different trails may be considered (Table 3.3). The infrastructure can be abstracted along those routes to model its operation.

3.3.2. Problem Formulation

A railway junction infrastructure $J = (R, C)$ consists of a set R of k routes $r \in R$ and a *conflict-matrix* $C \in \{0,1\}^{k \times k}$. Two routes $r_i, r_j \in R$ are described as *conflicting*, i.e. they cannot be used at the same time, by denoting $C_{i,j} = 1$, and as not conflicting by $C_{i,j} = 0$. Per default, the same route cannot be used twice at a time, hence $C_{i,i} = 1$ for all $i \in \{1, \ldots, k\}$.

An *operating program* on the railway junction J is a set A of *demands* $a = (r, n)$ corresponding to a route $r \in R$ and a total number of trains $n \in \mathbb{R}$ of that demand in the time horizon U. To service a train of demand $a = (r, n) \in A$, the service time $t_{\text{service}}(a)$ may be calculated using microscopic railway software. The service time of demand a is dependent on the infrastructure and safety technology in J (and adjacent railway lines and junctions), as well as some train-specific characteristics, such as braking percentages, acceleration capabilities, or total mass.

The problem of planning of railway junctions can be formulated by two inverse statements:

(I) Given a railway junction J and a set of feasible operating programs $\mathcal{A} = \{A_1, \ldots, A_l\}$. Determine $A^* \in \mathcal{A}$, so that the equation

$$\sum_{(r,n)\in A^*} (n) = \max_{A\in\mathcal{A}} \sum_{(r,n)\in A} (n) \tag{3.1}$$

holds.

(II) Given a set of feasible infrastructure layouts $\mathcal{J} = \{J_1, \ldots J_l\}$ with the same set of route R, an operating program A on them and a cost function $\phi : \mathcal{J} \to \mathbb{R}$. Determine $J^* \in \mathcal{J}$, so that the equation

$$\phi(J^*) = \min_{J\in\mathcal{J}} \phi(J) \tag{3.2}$$

holds.

Note that an objective function can be formulated for both problem statements. Specifically, statement (I) corresponds to finding the operating program with the highest total number of trains, while statement (II) might model the determination of the junction infrastructure layout with the least associated costs (f.e. for infrastructure construction).

While problem statement (I) can be used to assess the theoretical maximal capacity of a railway junction, railway infrastructure operators could be mostly interested in solutions to the statement (II): The operational program can often be assessed beforehand, i.e. when fixed by external factors, such as governments, promoting a modal shift to railways (Europe's Rail, 2015), or increasing demands when establishing new industry locations.

For both problem statements, the selection from the set of feasible solutions regarding their respective objective function is straightforward. However, assessing whether the combination of an operating program A and a junction layout J is feasible requires the determination of performance indicators, which can be compared to thresholds, to verify feasibility.

In this contribution, the timetable capacity of a railway junction is modelled (see Section 3.2.1 for a definition). It can be interpreted as the capacity of railway infrastructure in the timetabling process, i.e. the process in which the infrastructure operator finds conflict-free allocations for the service of different requests, submitted by transport operators. Since the requests are randomly distributed, all possible timetables that could emerge from the so-distributed request collection are covered. Conflicts may arise with this procedure, as the requests are made randomly, so it may be necessary to set waiting times for some requests. Additionally, some requests might have to be declined, i.e. if an allocation without a conflict is impossible. The estimated number of requests that *wait* for their allocated occupancy of a route, can serve as a performance indicator, i.e. the *estimated queue length* E_{LW}, for the infrastructure. While the proposed method technically represents the allocation of occupation requests in a railway schedule, we sometimes refer to those requests as *trains* to facilitate readability.

The method introduced below models infrastructure and operating program in a queuing system and analyzes its characteristics to evaluate the feasibility of such a combination. For further reference, a list of notations has been given in Table 3.4

3.3.3. Queueing System

Queueing theory (Bolch et al., 2006; Zukerman, 2013) has been extensively used to describe the processes in a transportation system. An analysed entity is usually divided into a *service* and a *waiting* part (*queue*). Incoming duties may either be assigned to a processing *channel* or, if none is available, to the next slot in the queue. The Kendall notation (Kendall, 1953) has been developed to abbreviate definitions of Queueing Systems with

$$A/B/n/m.$$

Arrival (A) and service processes (B) can be modelled with arbitrary probability dis-

Table 3.4.: List of Notations

R	Set of routes
r_i	Route i
$C \in \{0,1\}^{k \times k}$	Conflict matrix
$J = (R, C)$	Junction infrastructure
$\mathcal{J}$	Set of junction infrastructures
$A = \{a \mid a \text{ is demand}\}$	Operating program
$\mathcal{A}$	Set of operating programs
$a = (r, n)$	Demand
U	Time horizon
$t_{\text{service}}(a)$	Service time of a demand a
E_{LW}	Expected queue length
m	number of waiting slots
λ	Arrival rate
μ	Service rate
ET_A	Expected inter-arrival time
ET_S	Expected service time
ρ	Occupancy rate
v_A	Variation coefficient arrival process
v_S	Variation coefficient service process
σ	Standard deviation
p_{loss}	Probability of loss
S	Set of states
$s = (q_1, b_1, \ldots, q_k, b_k)$	State
q_i	Number of requests in queue of route r_i
b_i	Status of Service of route r_i
M	Maximum rate
$MC = (S, T)$	Continuous-Time Markov Chain
T	Set of transitions
$t = (u, v)$	Transition
$L_{W,limit}$	Threshold value for sufficient quality
$h_{i,j}$	headway time of the sequence train j after train i
$\bar{b}$	Mean service time
b_{max}	Maximum mean service time
n_r	number of trains on route r
ψ	Concatenation value
ρ_{conn}	Connected occupation rate

tributions. The described Queuing System can contain one or more service channels (n) and any natural number of waiting slots (m) - including none or infinitely many. In this work, Exponential (M) and General independent (GI) distributions are considered to describe arrival and service processes, but generally, other probability distributions can be utilized as well.

Modelled Queueing Systems are analysed by a set of relevant parameters. For *Markovian* (or *Exponential distributed*) models (*Markov Chains*), the state at any point in time is the only dependency to the future evolution of the process (see Zukerman (2013, Ch. 2.4)). Hence, transition rates between the states in a Markov Chain can be given. We denote *arrival rates* by λ and *service rates* by μ. They can be calculated with the use of expected values for the inter-arrival time ET_A

$$\lambda = \frac{1}{ET_A}, \tag{3.3}$$

and service times ET_S

$$\mu = \frac{1}{ET_S}. \tag{3.4}$$

Further, we describe the *occupancy rate* of a queuing system by

$$\rho = \frac{\lambda}{\mu}. \tag{3.5}$$

Furthermore, variation coefficients of the arrival process v_A and of the service process v_B may be used for estimating supplementary characteristics. Additional parameters include the *estimated queue length* E_{LW} and the *probability of loss* p_{loss}, describing the probability that incoming requests can not be considered in the system, as all service and waiting slots may be occupied.

An example of the use of a Queueing System is the determination of the capacity of a railway line (Wakob, 1984; Wendler, 2007) , which is usually described with a $GI/GI/1/\infty$ system. As an analytical solution to those systems has not yet been found, one can model it as an $M/M/1/\infty$ system and use an approximation formula (Gudehus, 1976) for the calculation of the expected queue length

$$E_{LW}(GI/GI/1/\infty) \approx E_{LW}(M/M/1/\infty) \cdot c = \frac{\rho^2}{1-\rho} \cdot \frac{v_A^2 + v_B^2}{2} \tag{3.6}$$

with a factor $c = \frac{v_A^2+v_B^2}{2}$.

Further approximations for the lengths of the queue in general independent arrival and/or service processes have been developed in Wakob (1984); Fischer and Hertel (1990); Wendler (2007); Weik (2020).

Figure 3.2 introduces a graphical representation of a Continuous-Time Markov Chain (see Ross (2014) for a definition) modelling the $M/M/1/\infty$ system. It consists of a root state (without a label) as well as states with one currently serviced unit (denoted by 's') and some units in the queue (each denoted by 'w'). Ordering the states by the number of units in the system, as of being serviced and currently waiting for units, transitions between subsequent states correspond to the arrival of a unit into the Queueing System (denoted by the arrival rate 'λ') or the termination of the service of a unit (denoted by the service rate 'μ').

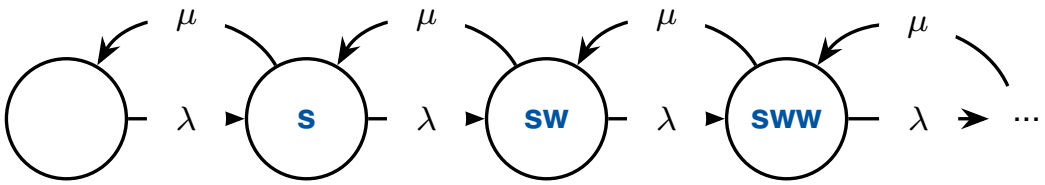


Figure 3.2.: Markov Chain for a $M/M/1/\infty$ Queueing System

The following Section generalizes this concept to railway junctions, which allow more complex resource allocations than on a railway line without multiple possible routes.

3.3.4. Modelling railway junctions

In this work, railway junctions are considered, which differ from railway lines in one central aspect: It may be feasible for multiple trains to use the infrastructure at the same time - depending on the routes, they are scheduled to take.

3.3.4.1. States

Modelling those infrastructures as a Continuous-Time Markov Chain $MC = (S, T)$, used states $s \in S$ need to be distinguishable regarding the currently serviced and waited-for route(s).

The general set of possible states

$$\hat{S} = \{(q_1, b_1, \ldots, q_k, b_k) \,|q_i \in \{0, \ldots, m\}, b_i \in \{0, 1\}\} \tag{3.7}$$

can be obtained by utilizing the information in the set of routes R of a junction $J = (R, C)$. A State $s = (q_1, b_1, \ldots, q_k, b_k) \in \hat{S}$ contains information regarding the number of trains $q_i \in \{0, \ldots, m\}$ waiting in the queue for route r_i and whether a route r_i is currently in service $b_i \in \{0, 1\}$.

The state-space $\hat{S}$ can be further restricted to

$$S = \left\{ (q_1, b_1, \ldots, q_k, b_k) \in \hat{S} | \sum_{i=1}^{k} \sum_{j=1}^{k} (C_{i,j} b_i b_j) = 0 \right\} \tag{3.8}$$

by applying the conflicts described in the conflict matrix C. Furtheron, entries q_i or b_i in a State $s = (q_1, b_1, \ldots, q_k, b_k) \in S$ are also referenced by $q_{s,i}$ or $b_{s,i}$.

3.3.4.2. Transitions

Transitions $(u, v) = t \in T$ between two states $u, v \in S$ can correspond to either the *arrival* of a train for route r_i, the *completion of service* on route r_i or the *choice* which holding train to service next.

The used transition rates of arrival-transitions are given by the arrival rate λ_i, of service-transitions by the service rate μ_i, and of choice-transitions by the maximum rate M. Those transition rates can be obtained by the described operational program A and the time horizon U.

An arrival transition between

$$u = (q_1, b_1, \ldots, q_i, b_i, \ldots, q_k, b_k) \tag{3.9}$$

and

$$v = (q_1, b_1, \ldots, q_i + 1, b_i, \ldots, q_k, b_k) \tag{3.10}$$

utilizes the arrival rate

$$\lambda_i = \frac{n_{r_i}}{U}, \tag{3.11}$$

where the total number of trains using route r_i, $n_{r_i} = \sum_{(r_i,n) \in A}(n)$, is divided by the time horizon U .

Transitions for a service process can be modelled between

$$u = (q_1, b_1, \ldots, q_i, 1, \ldots, q_k, b_k) \tag{3.12}$$

and

$$v = (q_1, b_1, \ldots, q_i, 0, \ldots, q_k, b_k)\,, \tag{3.13}$$

utilizing the service rate

$$\mu_i = \frac{1}{\frac{\sum_{(r_i,n)\in A}(t_{\text{service}}((r_i,n))\cdot n)}{n_r}}, \tag{3.14}$$

which corresponds to the reciprocal of the average service times of requests on the route r_i, weighted by the number of trains in this demand.

Choice-transitions $t = (u, v)$ exclusively start at states $u \in S$ with

$$\sum_{i=1}^{k} b_{u,i} = 0 \tag{3.15}$$

and at least two conflicting sets of routes $R_i, R_j \subset R$, with

$$\begin{aligned} q_{u,r_o} > 0,\ \forall r_o \in R_i \\ q_{u,r_p} > 0,\ \forall r_p \in R_j, \end{aligned} \tag{3.16}$$

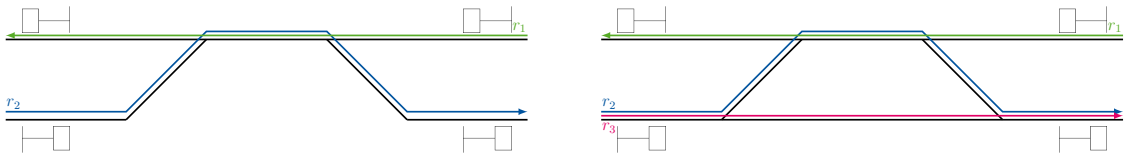
which are not operable simultaneously, and end at states $v \in S$, with at least one serviced route.

Hence, if the system is in a state s_0 where multiple routes are serviceable next, there is one choice-transition $(s_0, s_{\text{service},r_i})$ for every possible choice of the next serviced route r_i starting at this state s_0 to a corresponding state s_{service,r_i}. They correspond to the choice of which route to operate next when multiple options are possible. While they are modelled by ordinary transitions in the Continuous-Time Markov Chain, they connect to all possible route options utilizing identical transition rates, giving all options the same probability.

Since the choice transitions should not induce additional time in the system, the maximum rate M should be large enough such that the additional expected time in the system per choice transition $1/M$ is sufficiently small. In this work, choice transitions with identical transition rates are included to model the different decisions between the route(s) to be serviced next. Hence, the obtained results can be assumed to be independent of disposition strategies. The maximum rate is further chosen as $M = 600$, corresponding to a rate of 10 per second, which is deemed a competitive approximation.

3.3.4.3. Example

The introduced Continuous-Time Markov Chain $MC = (S, T)$ can be used to model Queueing Systems with more complex service regulations. In Figure 3.3, two different examples of railway track layouts are presented. While layout 3.3a corresponds to a short single-track segment with two possible routes, layout 3.3b illustrates the infrastructure of a crossover segment with three routes, two of which (r_1 and r_3) being operable simultaneously.



(a) Two routes on a short single track segment (b) Three routes on a crossover segment

Figure 3.3.: Conflicting routes in railway infrastructure

Figures 3.4 and 3.5 give graphical representations of Continuous-Time Markov Chains modelling the single track segment (Figure 3.3a) and the crossover segment (Figure 3.3b). Starting at the initial state, arrivals to their routes r_i are possible with their respective arrival rate λ_i. Since the Queueing System is empty at the initial state, the first arriving train can start its service time immediately after arriving. While every other train, arriving in the single track system (Figure 3.3a) during the service time of the first train, will be allocated to a waiting slot (indicated by '**w**'), the assignment of a second train in the Queueing System for the crossover segment does depend on the used routes. If the first train uses r_1 or r_3 and the second train uses the remaining route, the second train can traverse the block-section immediately (states indicated with '**ss**'). If the two trains share the same route or one of them is using route r_3, the second train has to be assigned to a waiting slot.

Those implications of route exclusions continue in both graphs and all combinations of arriving and serviced trains are modeled with one limitation: While in theory, unlimited state spaces and therefore the modelling of $m = \infty$ waiting positions are possible, an analysis of those models is very complex and in practical operation, the number of waiting trains will always be limited due to space restrictions. Hence, the examples in Figure 3.4 and 3.5 contain only one waiting slot per route. Trains arriving to a route while all waiting slots are occupied are therefore neglected.

Table 3.5 lists the number of states for the two models, for a selection of values of m.

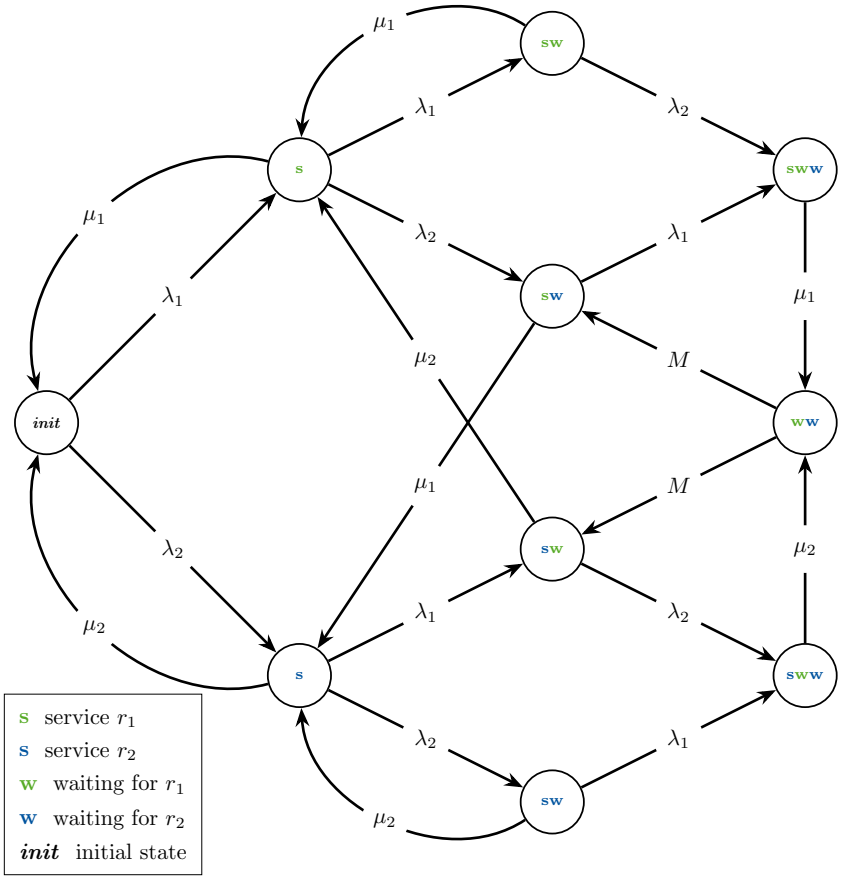


Figure 3.4.: Markov Chain modelling the Queueing System of the single track segment in Figure 3.3a

Since the probability of loss $p_{loss} = p_{loss}(m)$ is decreasing with m, a sufficient number of waiting positions may be calculated by identifying a limit value p^*_{loss} and determining the lowest $m \in \mathbb{N}$ satisfying

$$p_{loss}(m) \leq p^*_{loss} \tag{3.17}$$

iteratively or by testing relevant expressions of m.

A Model for the junction in Figure 3.1 does incorporate more complex route conflicts and additional parameters, leading to substantially higher state numbers, i.e. $|S| = 128$ for an instance with only one waiting slot $m = 1$ per route. Hence, we refrain from including a graphical representation of the induced Continuous-Time Markov Chain. In the online repository (Emunds and Nießen, 2023) however, a full description of the used model, obtained with the definition in Section 3.3.4, is given.

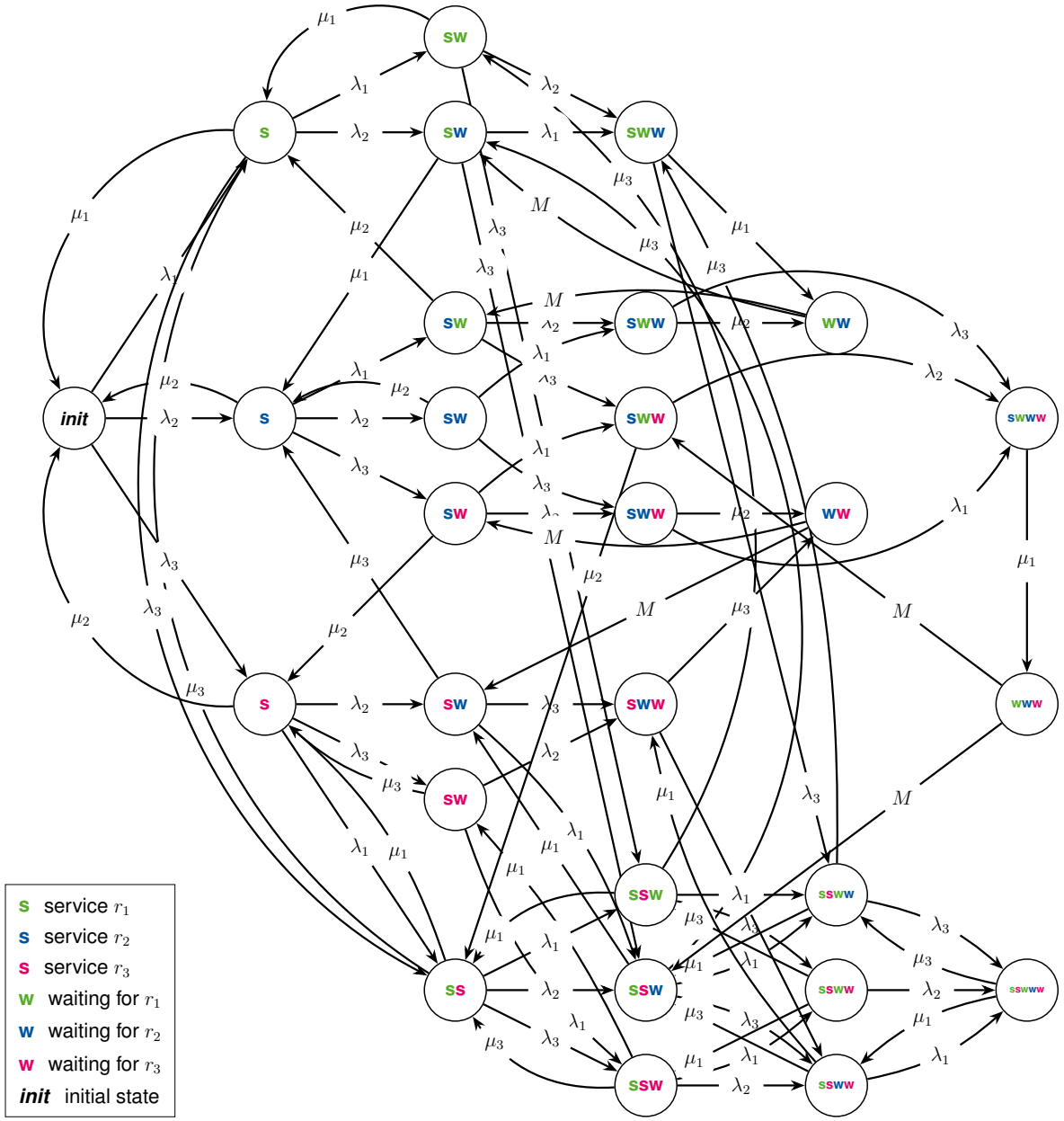


Figure 3.5.: Markov Chain modelling the Queueing System of the crossover segment in Figure 3.3b

Table 3.5.: Number of states by the number of waiting positions in Queuing Systems for the examples in Figure 3.3.

number of waiting positions m	1	2	4	8	16
single track segment	10	23	67	227	835
crossover segment	28	89	397	2261	15013

3.3.5. Scalability

To further analyse the scalability of the introduced model, an upper bound for the number of states can be given.

According to the definition in Section 3.3.4.1, a state consists of $k = |R|$ variables $q_1, \ldots, q_k$, with $q_i \in \{0, \ldots, m\}$ modelling the amount of requests waiting for route r_i, and $k = |R|$ variables $b_1, \ldots, b_k$, with $b_i \in \{0, 1\}$ modelling whether route r_i is currently utilized.

Hence, the amount of states $|S|$ is limited by

$$|S| \leq (m+1)^k \cdot 2^k \,. \tag{3.18}$$

However, not every combination of values is possible, due to the limitations induced by the conflict matrix (see Section 3.3.4.1). Let L indicate the maximum number of parallel usable routes, then the number of potential combinations of parallel used routes can be bounded by k^L. Therefore, the total number of states is bounded by

$$|S| \leq (m+1)^k \cdot k^L \,. \tag{3.19}$$

Since the required time and memory to solve a Continuous-Time Markov Chain scales with the number of states in a Markov-Chain, this number can be interpreted to estimate the scalability of the introduced model. According to the Benchmark in Budde et al. (2021), the largest solvable CTMCs possess state spaces in the order of magnitude of 10^7 up to 10^8 states. Under the assumption, that a maximum of $L = 3$ routes are usable in parallel, a model with $k = 8$ routes and a maximum of $m = 5$ waiting positions per route would require up to $8.6 \cdot 10^8$ states according to the bound in equation (3.19) and it therefore might be barely solvable.

Section 3.5 introduces two different case studies, the first of which analyses a complex railway junction, including 8 routes, with $L = 4$ being parallel usable. To efficiently

analyse this junction, the number of waiting positions per route is decreased to $m = 4$ in the model, making it more likely for a request to be declined. To ensure the soundness of the approximation, the probability of loss p_{loss} is checked on the calculated solutions.

For some even more complex junctions or when analysing entire railway stations, other measures might be needed to restrict the size of the corresponding model. For instance, it might be feasible to divide junctions with more than 8 routes into multiple smaller junctions that can be analysed independently, depending on the conflict matrix. This and the development of other approximation techniques are subject to further research.

However, the introduced method is capable of modelling railway junctions with up to 8 routes without any adjustments. The applicability can be further described by studying the number of routes in railway junctions. In the case of two double-track railway lines merging in a railway junction, the number of routes will probably be between 4 and 6, one for every combination of directions. Similarly, an intersection between two double-track railway lines might consist of 4 to 12 routes. Depending on the number of parallel usable routes, those infrastructures can be analysable, for example by reducing the amount of modelled waiting positions per route.

In any case, the modelled CTMC is utilized to determine the estimated queue length of every route.

3.3.6. Determination of the estimated queue length

While closed-form analytical solutions exist for the estimated queue length E_{LW} for some Queueing Systems (i.e. equation (3.6), more examples in Fischer and Hertel (1990)), the analysis of more complex systems remains challenging.

3.3.6.1. Utilizing Steady-State Distribution

For multidimensional Markov Chains, recent research (Schmitz et al., 2017; Weik, 2020) utilizes Matrix-Vector equations to determine the steady-state distribution, i.e. probabilities $p : S \to [0, 1]$ that the system is in state $s \in S$ in the long run.

Given those probabilities , the expected queue length E_{LW,r_i} of a route r_i can be

calculated by summing the probabilities of all states containing elements in the queue

$$E_{LW,r_i} = \sum_{\Psi_i(s)>0} p(s) \cdot \Psi_i(s), \tag{3.20}$$

utilizing the function $\Psi_i : S \to \mathbb{N}^0$, giving the number of elements in the queue of the route r_i for a state $s \in S$.

This approach has been used extensively for the determination of the timetable capacity of railway lines (Weik and Nießen, 2017) and junctions (Schmitz et al., 2017) and even been applied to model infrastructure unavailability with an additional stochastic process (Weik, 2020).

However, in this work, the techniques of probabilistic model-checking are explored regarding their potential for the determination of the estimated queue length.

3.3.6.2. Probabilistic Model-Checking

Much scientific research is based on a representation of natural or technical processes in a model. To analyse the properties of such systems, theoretical computer scientists developed formal methods to analyse their behaviour. By specifying the desired requirements in a formal temporal logic, the satisfaction of those requirements to a formalized model can be verified with the help of model-checking techniques.

Since in reality, many processes contain some kind of probabilistic behaviour, an extension, *probabilistic model-checking* has been introduced. This technique can analyse different types of Markov models, which are a way to include probabilistic behaviour into the representation of analysed processes. Since a detailed introduction and explanation of the techniques of probabilistic model-checking would extend the scope of this work, interested readers are referred to the articles of Kwiatkowska et al. (2010) and Katoen (2016).

This work utilizes probabilistic model-checking because of its capability and extensive development for the analysis of Continuous-Time Markov Chains, a modelling approach that has been thoroughly utilized for the analytical determination of railway capacity. The model formulation as a Continuous-Time Markov Chains has been described in Section 3.3.4 and can be represented in a model file, f.e. in the PRISM modelling language (Parker et al., 2000; Kwiatkowska et al., 2011). An example of a model file of the junction considered in this work has been provided in the online

repository Emunds and Nießen (2023).

Additionally to the model file, a formal description of the property has to be established. The probabilistic model-checking tools utilized for this work are capable of parsing properties in the PRISM modelling language. This language can be used to declare *path formulae* and *state formulae*, which allow for a formalization of the required logical expressions. Let X_{MC} be the random variable describing the state of the Continuous-Time Markov-Chain MC. The probability that the Markov-Chain corresponds to a state $s \in S$ in the long run is then given by

$$Pr(X_{MC} = s) = p(s) \tag{3.21}$$

By formulating a state formula Φ, i.e. a condition that can be satisfied on a state, the probability that this condition is satisfied in the model in the long run can be given by

$$Pr(X_{MC} = s, \Phi(s)) = \sum_{s \in S, \Phi(s)} p(s). \tag{3.22}$$

As an example, the condition whether a state representing a system case, in which at least one train is waiting for the service at route r_1, can be given as $\Phi_{waiting,1}(s) = q_{s,r_1} > 0$ for the formulated model.

The probability $Pr(X_{MC} = s, \Phi_{waiting,1}(s))$ that the system is in a state that satisfies this condition can then be obtained by checking the corresponding property with a probabilistic model-checking tool.

An important extension for properties is information about *rewards*. A reward can be formulated as a function $r : S \to \mathbb{R}, s \mapsto r(s)$ mapping the reward $r(s)$ to every state. Probabilistic model-checking tools can be utilized to determine the expected value of a reward under a certain condition. Regarding the application discussed here, the number of trains on a waiting slot for route r_i in a state $s \in S$ of the Continuous-Time Markov Chain (S, T) from Section 3.3.4 can be given as $r_{qL,r_i} : S \to \mathbb{R}, s \mapsto r_{qL,r_i}(s)$ with

$$r_{qL,r_i}(s) = \begin{cases} q_{s,r_i} & q_{s,r_i} > 0 \\ 0 & otherwise. \end{cases} \tag{3.23}$$

and checked with a corresponding property on the Continuous-Time Markov Chain

from Section 3.3.4, to determine the estimated queue length E_{LW,r_i} for route $r_i \in R$.

More information regarding property definition can be found in Parker et al. (2000) and the properties utilized in this paper are accessible via the online repository Emunds and Nießen (2023).

3.3.6.3. Approximating non-exponential processes

Notice that the calculation of the expected queue length E_{LW,r_i} on an arbitrary route r_i in Sections 3.3.6.1 and 3.3.6.2 relies on the use of Markov Chains, solely capable of modelling queueing system of type $M/M/s/\infty$.

To analyse systems with general independent (GI) arrival or service processes, factors, utilizing the variation coefficient of the described process, can be introduced. According to Fischer and Hertel (1990), the expected length of a queue in a Queueing system of type $GI/GI/s/\infty$ can be analysed by using a $M/M/s/\infty$ system and modifying the results

$$E_{LW,r_i}(M/M/s/\infty) \cdot \frac{1}{\gamma} \approx E_{LW,r_i}(GI/GI/s/\infty) \tag{3.24}$$

accordingly, using

$$\gamma = \frac{2}{c \cdot v_B^2 + v_A^2} \tag{3.25}$$

and

$$c = \left(\frac{\rho}{s}\right)^{1-v_A^2} \cdot (1 + v_A^2) - v_A^2. \tag{3.26}$$

Here, v_A and v_B correspond to the coefficients of variation for the arrival and the service process respectively.

Following the equations (3.24 - 3.26), a greater coefficient of variation in either the arrival or the service process yields a larger expected queue length. In addition to the introduced coefficients of variation, factor γ depends on the occupancy rate ρ (see identity (3.5)) and the number of parallel service channels s. Here, the queue length E_{LW,r_i} corresponds to arrivals on one route r_i only, $s = 1$ can hence be fixed.

This modification can be implemented in the design of threshold values that have been developed for the capacity analysis of railway infrastructure.

Regarding the modelled railway infrastructure (see Section 3.3.4), E_{LW,r_i} has to be

calculated for every route r_i. Using the threshold values introduced in the next section, every E_{LW,r_i} can be verified regarding its sufficiency for acceptable operating quality on the infrastructure.

3.3.7. Threshold values

Different threshold values for theoretical and operational capacity have been discussed in the literature. In the UIC Code 406 (UIC, 2013) maximum occupancy rates have been introduced to limit capacity utilization. Potthoff (1962) introduces limits for the loss probabilities in railway stations, which are still in practical use for the determination of track numbers in a railway station, i.e. at the German infrastructure manager DB InfraGO (2009).

In this work, the threshold value $L^*_{W,limit}$, introduced in Schwanhäußer and Schultze (1982) and likewise still in practical use (DB InfraGO, 2009), is utilized. It corresponds to the maximum number of trains that are to wait in the analysed line section at any given time for a sufficient performance of the infrastructure. With the ratio of passenger trains in all considered trains p_{pt}, the threshold value

$$L^*_{W,limit} = 0.479 \cdot \exp(-1.3 \cdot p_{pt}) \tag{3.27}$$

can be specified. A threshold value of the approximating queuing system

$$L_{W,limit} \approx \gamma \cdot L^*_{W,limit} = \gamma \cdot 0.479 \cdot \exp(-1.3 \cdot p_{pt}) \tag{3.28}$$

can be obtained by utilizing (3.24) and (3.27).

Hence, the performance of railway infrastructure is judged based on the arrival and service processes, including their rates and variation, as well as on the operating program during the analysed time horizon.

3.3.8. Methodical Procedure

Two different problem formulations that can be solved with the introduced method have been defined in Section 3.3.2. The Problem of planning an infrastructure out of a set of possible layouts (II) can be solved by iterating through all possible service times while calculating the corresponding estimated queue lengths at a given route

$E_{LW,r}$ by utilizing the introduced Continuous-Time Markov Chain model. A detailed flow chart is given in Figure 3.6.

The calculation utilises an input consisting of a set of infrastructure layouts I and an operating program A, corresponding to a set of demand tuples (r, n). For every infrastructure layout $I_i \in I$, a Continuous-Time Markov Chain model can be formulated according to the definitions in Section 3.3.4. By iterating over all routes $r \in R$ in this infrastructure layout and all mean service rates μ_k for $k \in K$ in an index-set K, the estimated queue length $E_{LW,i,r}(\mu_k)$ can be obtained from probabilistic model-checking, yielding the expected value of the in Section 3.3.6.2 formulated reward $r_{qL,r}$.

Utilising the operating program A, a threshold to the queue length $L_{W,limit}$ (see Section 3.3.7) can be obtained. If the calculated $E_{LW,i,r}(\mu_k)$ is below the computed threshold of $L_{W,limit}$, the assumed service rate μ_k can be saved in a List L, from which, after the value iteration over all μ_k has been completed, the minimum service rate $\mu_{min,i,r}$ for the route r can be obtained.

Since this minimum service rate $\mu_{min,i,r}$ corresponds to the minimum rate with which service of the requests on this route r has to be completed, it gives a lower bound to the total minimum service rate $\mu_{min,i}$ in the infrastructure layout I_i. By building the maximum $\mu_{min,i} = \max_{r \in R} \mu_{min,i,r}$ of all route-based lower bounds, the total minimum service rate can be obtained.

Therefore, by calculating $b_{max,i} = 1/\mu_{min,i}$, an upper bound to the mean service time over all routes in the infrastructure, that needs to be fulfilled for a sufficient quality, can be obtained. Please note that $b_{max,i}$ does not necessarily correspond to practical realisable minimum headway or service times, but rather to a limit to the average service time value. Hence, after verifying that the mean service time in the junction $\bar{b}$ is smaller than the obtained maximum mean service time $b_{max,i}$ for every infrastructure layout $I_i \in \tilde{I} \subset I$, the better layout $\hat{I}$ can be selected.

This architecture translates directly to the determination of the timetable capacity of a railway junction (I), by iterating over all possible arrival rates λ_k and utilizing a given share of passenger trains directly to specify the threshold $L_{W,limit}$. In that case, the estimated service times are needed to determine the fixed service rates μ_r for every route $r \in R$ in the analysed junction infrastructure.

In both cases, the quality of a junction is determined by the calculation of an estimated queue length $E_{LW,r}$ for a given route $r \in R$ and comparing it to a threshold value. An

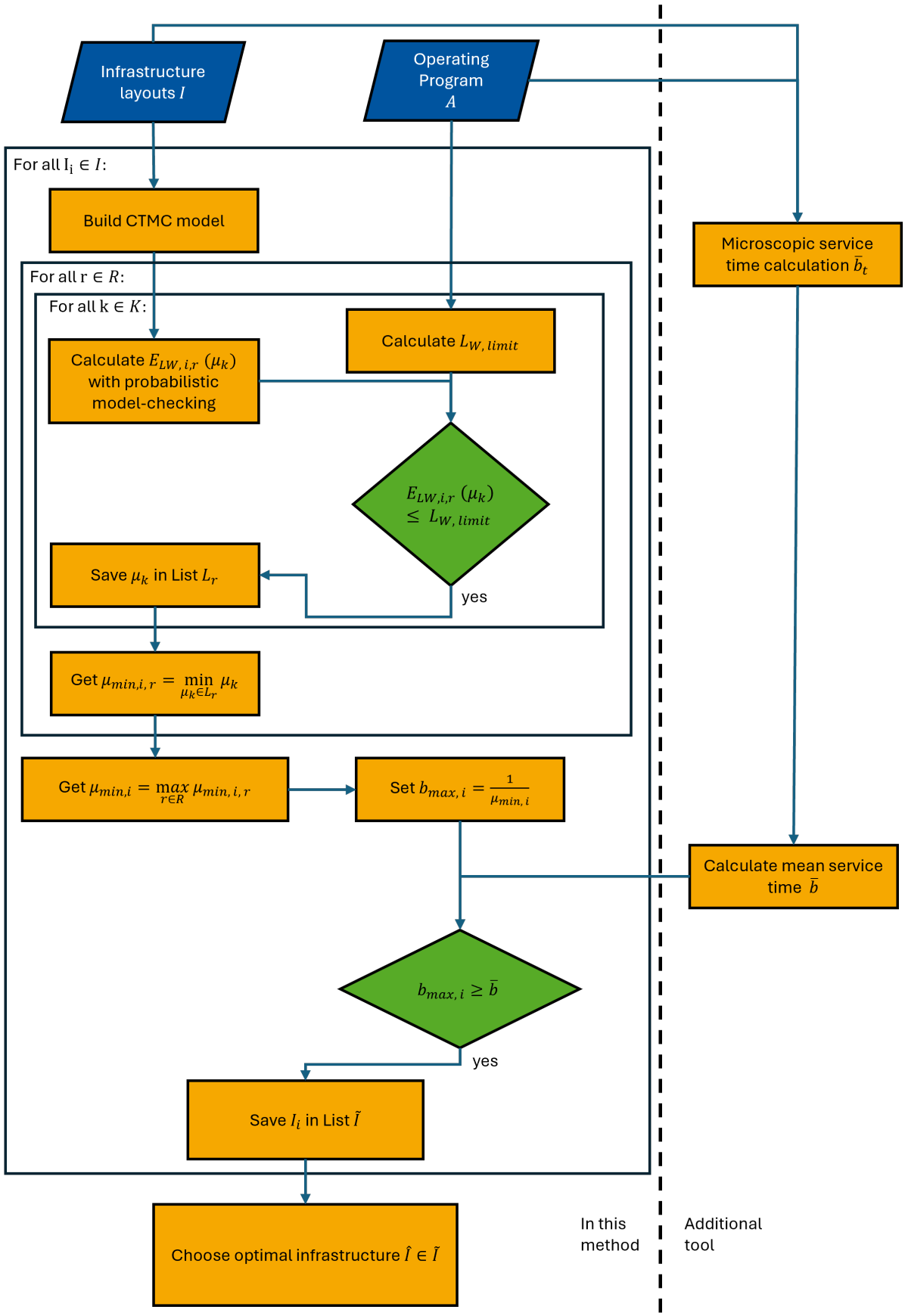


Figure 3.6.: Application of the introduced method to choose an optimal infrastructure layout

example can be found in the online repository Emunds and Nießen (2023).

While some parts are essential for the construction of the introduced model, others, such as additional microscopic service time calculations, that can be used to calculate the precise mean service time of the different infrastructure layouts, can assist in the determination of the optimal infrastructure layout.

Different parameters may contribute to the formulation and calculation of those processes and therefore influence the timetable capacity of a railway station.

3.3.9. Contributing Factors

The introduced method needs to take different parameters into account when assisting in the planning process of a railway junction. For example, not only the number of trains on every route in the analysed time horizon but also train-specific characteristics may influence the achieved quality, and not all parameters might be known at the time of early planning stages.

Table 3.6.: Requirements for the described method

Parameter	Required for capacity determination (I)	Required for infrastructure planning (II)
routes	✓	✓
route exclusions	✓	✓
Share of passenger trains	✓	✓
operating program(s)	(+)	✓
estimated service times	✓	(+)
microscopic infrastructure data	(+)	(+)
train-specific parameters	(+)	(+)
train control system	(+)	(+)

Remarks: Please note that a strong requirement is indicated by ✓, while an enhancement to the qualitative decision making, f.e. in later planning stages, is indicated by (+).

In Table 3.6, different parameters and their requirement for the two different planning problems – capacity determination (I) and infrastructure planning (II) – are described.

While the knowledge of routes, their exclusions and the share of passenger trains are needed for both problem formulations, estimated service times are strongly required

for the determination of the timetable capacity (I) for a considered infrastructure only. If an infrastructure layout has to be selected (II), an estimation of service times on every route/train combination is not necessary but can be used to evaluate other planning decisions. The knowledge of the detailed operating program is required for the infrastructure planning (II) formulation only. For the determination of the timetable capacity (I) only the share of passenger trains is required to be given, as it is needed in the definition of a threshold for sufficient quality (see Section 3.3.7). However, it might be advisable to define a set of possible operation programs, to allow small deviations in the share of passenger trains, which would otherwise be fixed to the given value and hence substantially restrict the solution space.

Additional parameters, such as microscopic infrastructure data, train-specific parameters and the considered train control system may be utilized in combination with a microscopic running and blocking time calculation. Blocking times can be utilized to determine the minimum headway time between the service of two consecutive route/train combinations. Further information regarding their definition and calculation may be found in related literature (Happel, 1959; Hansen and Pachl, 2014). By building the weighted average $\bar{b}_a$ of all minimum headway times h_{t_1,t_2} between a route/train combination t_1 and any successor t_2, the service time $t_{\text{service}}(t_1) = \bar{b}_{t_1}$ can replace the service time estimations for every route/train combination in the railway junction or yield key points for the evaluation of the method's results.

This enables decision-makers to compare the impact of different parameters, such as infrastructure layout, speed restrictions or train control systems and choose the combination that meets all their requirements. As such, the capacity of a railway junction can be increased in manifold ways, f.e. by decreasing the utilised service time, either through relaxed speed restrictions or by implementing alternative train control systems, and by choosing a different infrastructure layout that reduces the number of route exclusions.

With the approach described in this work, all considered parameters may be analysed regarding their influence on the formulated problems by using the described input data in the modelling of a queueing system and analysing its characteristics.

Additionally to the parameters in Table 3.6, which may be extracted from input data, some choices regarding the model itself have to be made. In particular, Table 3.7 gives an overview of the modelling constants and their assumed value in the remainder of this work.

Table 3.7.: Model assumptions

Parameter	Assumed value	Reasoning	Further information
Amount of maximal parallel occupied waiting positions per route m	5 (or 4 in Section 3.5.1)	Limiting the size of the Markov Chain	Section 3.3.4.3
Maximum Rate M	600	Corresponding to 10 ticks per second	Section 3.3.4.2
Variation coefficient arrival process v_A	0.8	Standard value in literature	(Wendler, 1999)
Variation coefficient service process v_B	0.3	Standard value in literature	(Schwanhäußer, 1974)
Threshold for sufficient quality $L^*_{W,limit}$	$0.479 \cdot \exp(-1.3 \cdot p_{pt})$	Standard value in literature and practice	Section 3.3.7

Here, the maximal amount of parallel occupied waiting positions per route is set to $m = 5$ to restrict the number of states in the underlying Continuous-Time Markov Chain and hence speed up the computation. To model near instantaneous transitions over the choice transitions (see Section 3.3.4.2), the choice rate per minute has been set to $M = 600$.

Other parameters have been fixed according to standard values in literature, such as $v_A = 0.8, v_B = 0.3$ for the coefficients of variation of the arrival and service process (see (Wendler, 1999; Schwanhäußer, 1974)). This assumption is due to the reasoning that the arrival process is not as random as the use of an exponential distribution presumes since traffic demands usually underlie some kind of periodicity. The service times are even more stable as they are majorly determined by a fixed set of physical properties, such as the train control system and train-specific parameters, which are usually similar due to the use of homologous rolling stock for one railway line. The threshold for sufficient quality is assumed to be given by $L^*_{W,limit} = 0.479 \cdot \exp(-1.3 \cdot p_{pt})$ following DB InfraGO (2009) (see also Section 3.3.7).

In addition to those fixed value presumptions, further modelling assumptions have been made in the remainder of this work. Thus, the arrival and service processes have been supposed to be independent and exponentially distributed for the initial calculation with the Continuous-Time Markov Chain. A more general distribution has been assumed to be approximable by utilizing the factors in Section 3.3.6. Furthermore, the threshold values (see Section 3.3.7) have only been formulated for railway line capacity; they have been assumed to also give a reasonable quality limit for the estimated queuing length of a route in a railway junction.

To validate the introduced Continuous-Time Markov Chain model for the queueing system of a railway junction, Section 3.4 compares simulation results with analytical results obtained by applying the described method.

3.4. Model Performance

Using the formulation of Section 3.3.4, a Continuous-Time Markov Chain for a railway junction can be obtained. In this work, a maximum of $m = 5$ waiting slots per route has been utilized, as a trade-off between tractability (dependent on the model size) and accuracy of the model (see Section 3.3.4.3). By calculating the estimated length of a queue and comparing it to the obtained threshold values (see Section 3.3.7), the capacity of modelled railway junctions may be assessed. To ensure the quality of the obtained model, the validity of the generated queue length estimations has to be verified.

Aiming to survey solely the performance of the solution process using the introduced Continuous-Time Markov Chain, $M/M/s/\infty$ systems have been considered. For this, simulations of a railway junction with multiple incoming railway lines have been built and run on sample data.

3.4.1. Simulation Architecture

The Simulations have been implemented in Python 3.10.9 (Van Rossum and Drake Jr, 1995; Python Software Foundation, 2022), utilizing SimPy (Team SimPy Revision, 2023) and SciPy (Virtanen et al., 2020). A model of the junction in Figure 3.1 has been built, including four different routes with route-specific arrival and service processes.

To estimate inter-arrival times for every route, a pseudo-random number generator yields the next inter-arrival time within a specified exponential distribution with an expected value $ET_{A,r}$ equal to the reciprocal value of the mean arrival rate $\lambda_r = \frac{1}{ET_{A,r}}$ of the modelled route. Trains are stored in a first-in-first-out queue for the route and serviced according to the service time acquired by a second pseudo-random number generator, utilizing another exponential distribution with an expected value ET_S equal to the mean service rate $\mu = \frac{1}{ET_S}$ of the system. During the service of a route r, it is ensured, that no conflicting route $r' \in R$ can start service by utilizing shared resources for every pair of routes $(r, r') \in R \times R$.

An implementation can be found in the online repository (Emunds and Nießen, 2023). By performing Kolmogorov-Smirnov tests on the sampled inter-arrival and service times, the correspondence of the generated data with the implied distributions has been verified. The obtained p-values, describing the goodness of the fit for every single simulation run, are summarized in the online repository. Furthermore, the validity of the simulated schedule has been proven for every simulation iteration by computing the number of conflicting trains with a suitable algorithm.

To assess the performance of the simulated process, a snapshot of every route's queue length is taken in every simulated minute. Hence, the mean length of a queue can be obtained easily for every simulation run.

3.4.2. Validation

Both, implementations of the simulation and the queueing-length estimations with the formulated Continuous-Time Markov Chain, have been run on a single core of an Intel Core i7-10700 Processor (2.90 GHz), utilizing a maximum of 32 GB working memory.

Two different simulation setups have been considered:

1. A simulation with no limit to the number of trains being able to wait in the queues of their requested routes
2. A simulation with a maximum of 5 trains per route being able to wait at the same time for the service on their requested routes

In the second simulation and in the analytical setting, an arriving train is rejected, i.e. not inserted into the respective queue, if its arrival time lies within a time frame where 5 trains are already waiting for the release of the requested route. Here, all routes are capable of having 5 trains waiting for service, hence a total maximum of 20 trains may wait at the same time.

All three solution approaches, both simulation setups and the analytical method, have been set to compute the estimated queue length $E_{LW,r}$ for the route r_3 (see Table 3.3), conflicting with routes in direction of A and C (see Figure 3.1). Route r_3 has been selected as it is one of the routes with the most conflicts (with r_2 and r_4) and it would directly benefit from an overpass construction.

Since those computations have been done to compare the results and running times of the approaches, only a small set of 10 different service rates, between 0.1 and 1.0

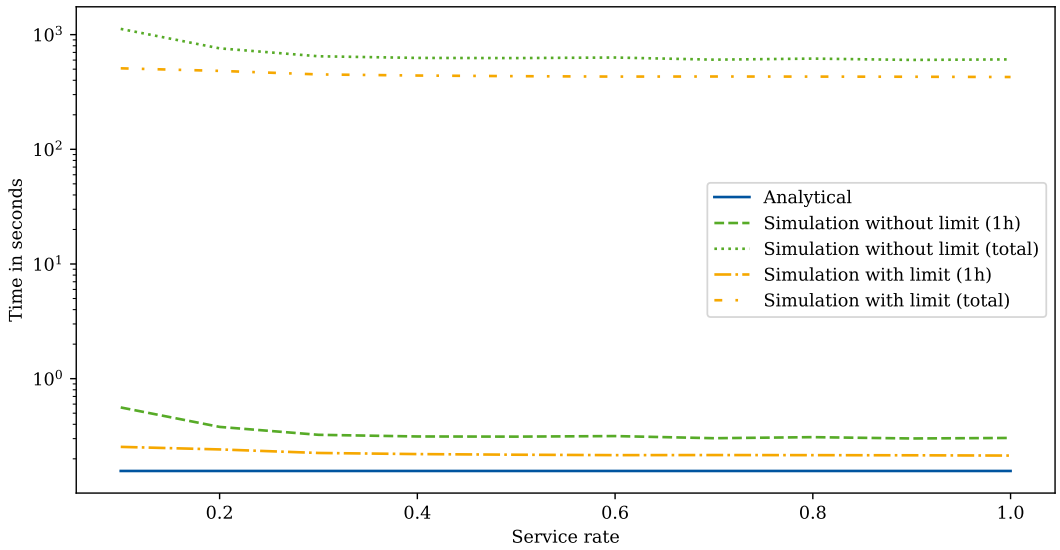


Figure 3.7.: Comparison of the required computation times

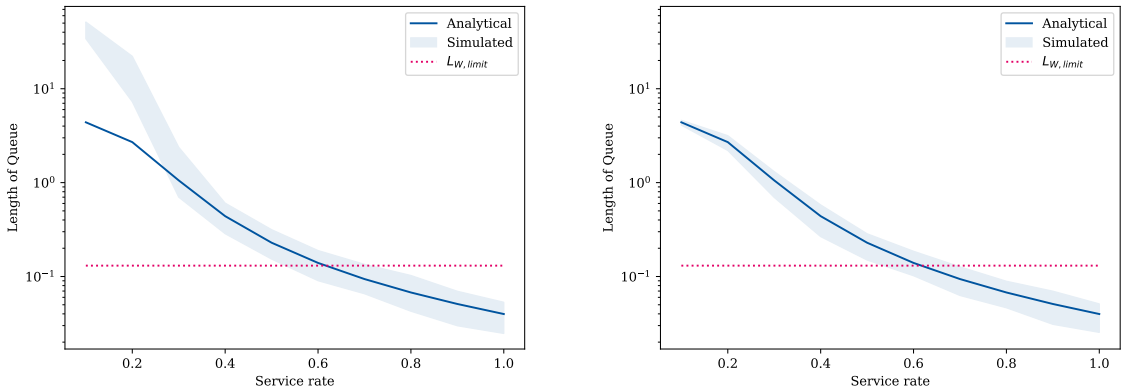
trains per minute, has been considered. An arrival rate of 0.1 trains per minute has been set for every route in all computed instances.

Each simulation has been run 100 times for every considered service time and for every run 22 hours have been simulated. From those 22 hours, the first and last have not been considered, resulting in an evaluated time of 20 hours per simulation and 2000 hours in total for every service time investigated.

For both simulation setups, the mean computing time per hour and the mean total run time per 22-hour simulation have been evaluated. Those can be found in Figure 3.7, which additionally includes the running time of the analytical solution for reference.

The computing time results in Figure 3.7 indicate that the analytical approach is faster than the tested simulation implementation, even if compared to the run of a single simulation hour. Noting the logarithmic scale, for small service rates the analytical approach is faster by a factor of 5 to 10, depending on the considered service rate. Since simulations have to be conducted multiple times to receive sufficient results, a comparison with the total simulation times might be more realistic, yielding factors of 10^3 up to 10^4.

In general, the computing times of the simulation approach without a limit are significantly increasing with a decreasing service rate. This is probably due to a higher amount of trains in the network at any given time, as the mean service time increases with a decreasing service rate, leading to more trains waiting for the release of their



(a) Simulation with unlimited m

(b) Simulation with a limit of $m = 5$

Figure 3.8.: Accuracy of the analytical approach when compared to the confidence interval of the performed simulations

requested route. Similarly, the simulation runs with an unlimited m required up to 50% more computing time compared to the simulation scenarios with a limit of $m = 5$.

Additional observations can be made by evaluating the accuracy of the introduced analytical approach. This can be done by comparing the results obtained from the simulation methods with the results of the queueing-based analytical approach. In Figure 3.8 the results obtained from the computations E_{LW,r_3} are depicted by including the exact analytical results and the standard deviation area of the simulation results, i.e. the area $\left[\overline{E}_{LW} - \sigma, \overline{E}_{LW,r_3} + \sigma\right]$, surrounding the mean $\overline{E}_{LW,r_3}$ by the standard deviation σ. Supplementary boxplots, including a study varying the number of trains per route for a fixed service rate, are available in Appendix A.2.

Noting the logarithmic scale in Figure 3.8, the results show significant differences between the simulation setups. For the setup with a limit of $m = 5$ on the number of trains in a queue (Figure 3.8b), matching the setting for the queueing-based analytical approach, the results of the simulation and the analytical solution are similar. Contrarily, when comparing the analytical solution to the simulation setup without a limit on the number of trains in a queue (Figure 3.8a), serious discrepancies between the simulation and the analytical results can be found for service rates of less than 0.3.

Hence, an assumption, that a high probability of the queue length exceeding the limit of 5 requests exists, can be made for the considered railway junction, utilizing the fixed predefinitions and an averaged service rate of less than 0.3. Consequently, arriving

trains are likely to have to wait, resulting in very poor service quality.

Taking the limit for $M/M/s/\infty$ queueing systems (see Section 3.3.7), in Figure 3.8 with '$L_{W,limit}$' annotated, into account, the accuracy of the introduced queueing-based analytical approach is sufficient for the use in practical applications.

While the introduced analytical method is sufficiently accurate with only one computation, the considered simulation methodology would require multiple computations to describe the queuing process satisfactorily, since it can only compute samples for process iterations and does not consider the stochastic process in its entirety. Furthermore, the simulations would need to be rerun for every change in parameters, and since both described problem statements need to evaluate the queue lengths at multiple parameter configurations, such as various service or arrival rates, this would substantially increase the total computation time. Therefore, the case studies in the following section have been conducted with the introduced analytical approach.

3.5. Computational Study

The introduced method for the calculation of queue lengths can be used to guide infrastructure managers when planning railway junctions. In this section, two different railway junctions have been considered to highlight the application of the proposed method for both introduced problem statements (see Section 3.3.2).

While Subsection 3.5.1 determines the timetable capacity of a given railway junction, Subsection 3.5.2 focuses on the selection of an infrastructure layout for fixed operating program demands.

3.5.1. Timetable Capacity Analysis

To describe the performance capabilities of a complex railway junction in the timetabling process, its timetable capacity can be calculated with the introduced method. In this section, a junction located in the east of Paris, referred to as the *triangle of Gagny*, is analysed.

Since the junction has already been subject to railway-specific research, i.e. the design of real-time traffic management algorithms in Pellegrini et al. (2014), a track diagram (see Figure 3.9) could be obtained from the existing literature. The junction

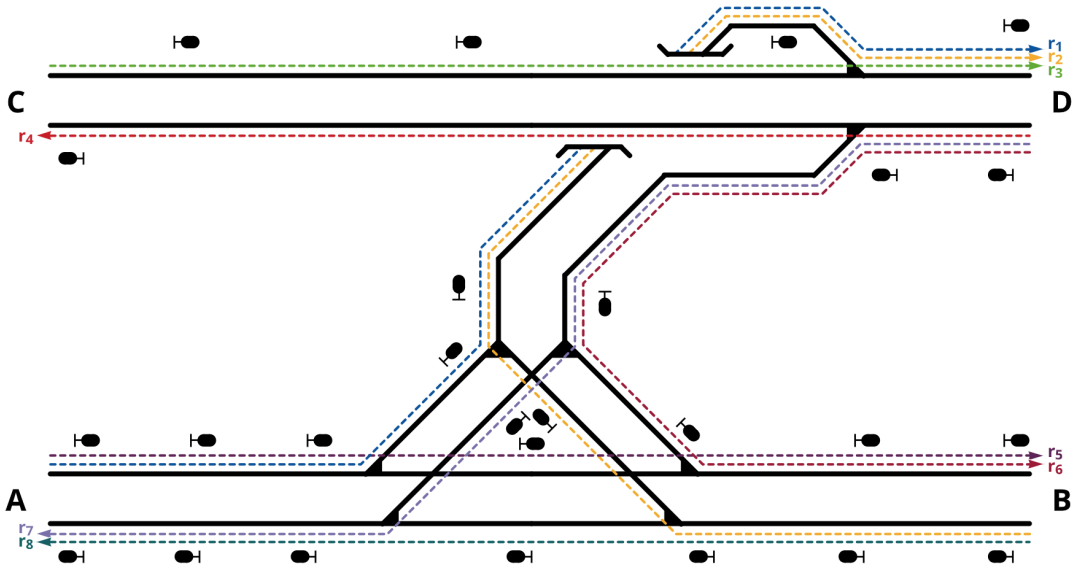


Figure 3.9.: A track-layout plan for the triangle of Gagny. Own illustration, adapted from Pellegrini et al. (2014).

infrastructure

$$J_{\text{Gagny}} = (R = \{r_1, \ldots, r_8\}, C) \tag{3.29}$$

can be described with the conflict matrix

$$C_{i,j} = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{pmatrix} . \tag{3.30}$$

We assume that an infrastructure manager plans to utilise all 8 routes with passenger traffic only and discuss three different distributions of traffic shares on requested routes.

The set of possible scenarios $\Theta = \{a, b, c\}$ includes the following scenarios:

(a) a homogeneous distribution of requests to all 8 routes,

(b) a heterogenous distribution, assuming two times more traffic on the direct routes

(r_3, r_4, r_5, r_8),

(c) a heterogenous distribution, assuming two times more traffic on the branching routes (r_1, r_2, r_6, r_7).

The three distributions are listed in Table 3.8 by their traffic share $p_{i,\theta}$ in the scenario θ.

Table 3.8.: Traffic share $p_{i,\theta}$ of routes r_i in every scenario $\theta \in \Theta$.

	1	2	3	4	5	6	7	8
a	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$	$1/8$
b	$1/12$	$1/12$	$1/6$	$1/6$	$1/6$	$1/12$	$1/12$	$1/6$
c	$1/6$	$1/6$	$1/12$	$1/12$	$1/12$	$1/6$	$1/6$	$1/12$

These scenarios are further investigated by estimating the queue length of every route for a set $\mathcal{A}_\theta$ of possible operating programs for every scenario $\theta \in \Theta$. A possible operating program $A \in \mathcal{A}_\theta$ consists of demands $(r_i, n_i) \in A$, so that the numbers of trains per routes n_i correspond to the traffic share distribution of the scenario θ.

For this example, we consider the set $\mathcal{A}_\theta = \{A_\theta(c) | c \in \{1, \ldots, 100\}\}$ of operating programs

$$A_\theta(c) = \{(r_i, c \cdot p_{i,\theta}) \, | r_i \in R\} \tag{3.31}$$

for every scenario θ, corresponding to a total number of 1 up to 100 trains per hour in the system.

While the arrival rates λ_r can be obtained from the operating program A by applying equation (3.11), the service rates are calculated from the service times of every route in the infrastructure. Usually, this would be done through a microscopic minimum headway time calculation. Unfortunately, no microscopic data has been made available for this example, therefore the minimum headway times have been estimated in Table 3.9, assuming only one type of rolling stock per route.

The service time $b_{i,\theta}$ of a route r_i can be obtained by building the weighted average

$$b_{i,\theta} = \sum_{j=1}^{8} h_{i,j} \cdot w_{i,j,\theta}, \tag{3.32}$$

Table 3.9.: Assumed minimum headway times $h_{i,j}$ in seconds for the junction of Gagny

i \ j	1	2	3	4	5	6	7	8
1	180	180	140		130			
2	180	180	140		90		90	130
3	90	90	90					
4				90		90	90	
5	110	90			110	110	90	
6				160	160	160	160	
7		90		180	90	160	180	180
8		110					110	110

with the weights

$$w_{i,j,\theta} = \begin{cases} p_{i,j,\theta} & C_{i,j} = 1 \\ 0 & otherwise, \end{cases} \tag{3.33}$$

corresponding to the probability

$$p_{i,j,\theta} = \frac{p_{j,\theta}}{\sum_{j,\ C_{i,j}=1}(p_{j,\theta})} \tag{3.34}$$

that route r_j follows route r_i in scenario θ if r_j has a conflict with r_i. The calculated service times $b_{i,\theta}$ are listed in Table 3.10.

Table 3.10.: Service time $b_{i,\theta}$ of route r_i in every scenario $\theta \in \Theta$ in minutes.

	1	2	3	4	5	6	7	8
a	2.63	2.25	1.5	1.5	1.7	2.67	2.44	1.83
b	2.5	2.17	1.5	1.5	1.72	2.67	2.46	1.83
c	2.75	2.33	1.5	1.5	1.69	2.67	2.43	1.83

Utilizing the service time of a selected scenario θ, the service rate μ_i can be calculated with $\mu_i = 1/b_{i,\theta}$.

Together with the arrival rate λ_i, a CTMC, modelling the timetabling process on the triangle of Gagny, can be described according to the model in Section 3.3.4. Since the evaluated junction infrastructure utilizes a total of $|R| = 8$ routes, only $m = 4$ waiting positions have been assumed for every route in the queueing system. Here, the maximum rate is modelled with $M = 600$, similar to the models utilised for validation in Section 3.4. In total, the analysed $MC_{\text{Gagny}} = (S, T)$ consists of $|S| = 1.25 \cdot 10^7$

states and $|T| \approx 1.35 \cdot 10^8$ transitions.

The solving process has been automated using Python 3.10.9 (Van Rossum and Drake Jr, 1995; Python Software Foundation, 2022) and probabilistic model-checking Software. For this work, the model-checker *Storm* (Hensel et al., 2022) and its Python-Interface *Stormpy* (Junges and Volk, 2023) have been utilised. Used Models have been formulated according to Section 3.3.4, expressed in the PRISM modelling language (Parker et al., 2000; Kwiatkowska et al., 2011). Detailed information regarding the solving process and the used model files can be found in the online repository (Emunds and Nießen, 2023). The solution time for a model is between 15 and 20 min on six cores of an Intel Xeon Platinum 8160 Processor (2.1 GHz), utilizing $23\ 400$ MB working memory.

The results of the probabilistic model-checking algorithms are the expected queue lengths $E_{LW,r_i}(M/M/s/m) = E^*_{LW,i}$ on the routes r_i when describing exponentially distributed inter-arrival and service times. By assuming v_A and v_S according to Table 3.7, the expected queue lengths $E_{LW,r_i}(GI/GI/s/m) = E_{LW,i}$ can be estimated with equation (3.24).

In Figure 3.10, the estimated queue lengths for all 8 routes are depicted for the three scenarios. Furthermore, the threshold

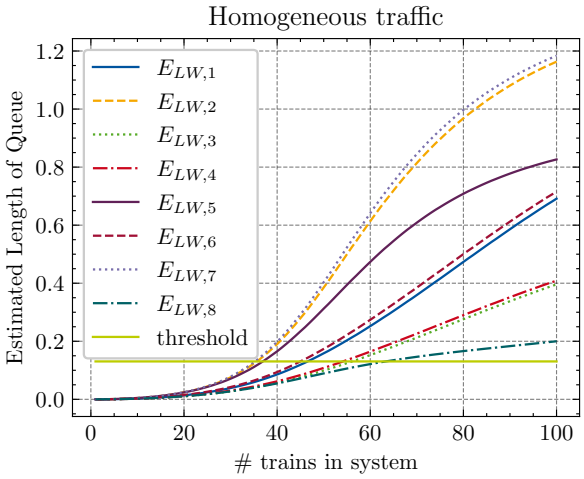
$$L^*_{W,limit} = 0.479 \cdot \mathsf{exp}(-1.3 \cdot 1) \approx 0.131 \tag{3.35}$$

for the queue length in a system with only passenger trains (see Section 3.3.7) is shown.

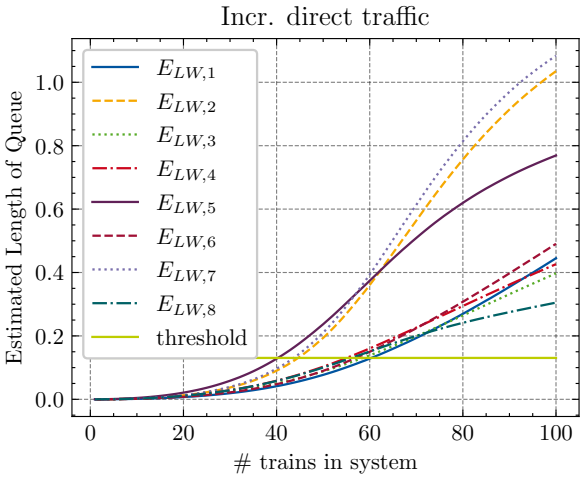
Since the traffic distribution varies between the calculated scenarios, curve characteristics differ for the plotted queue length estimations.

In the homogeneous traffic scenario, the longest queue is estimated for route r_7 (crossing multiple other routes on its way from D to A). Another crossing route, r_2, has a similar progression. As the straight route r_5 has multiple conflict points with crossing routes, its estimated queue length is higher than the one of the other branching (r_1 and r_6) and direct (r_3, r_4 and r_8) routes.

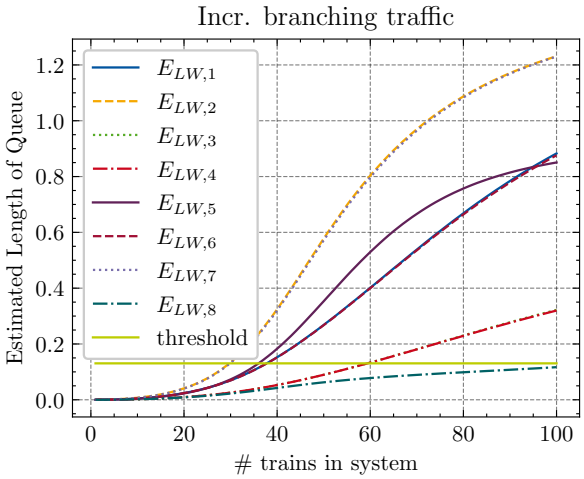
In the scenario with increased direct traffic, route r_5 is the route with the highest estimated queue length for a total number of trains of less than 60. If more than 60 requests are to be scheduled in an hour, the routes r_7 and r_2 are estimated with a



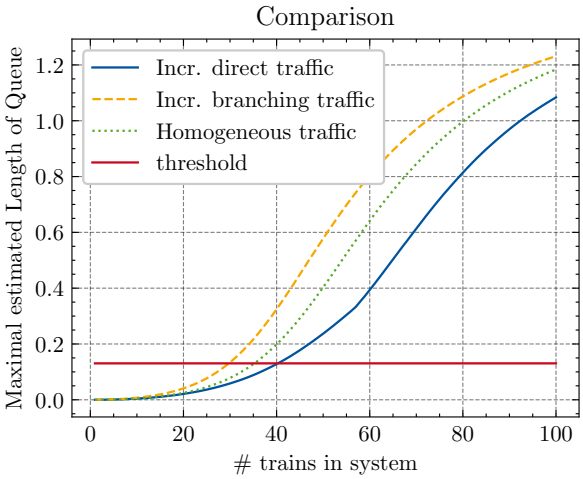
(a) Expected queue lengths in the homogeneous traffic distribution scenario



(b) Expected queue lengths in the scenario with increased direct traffic



(c) Expected queue lengths in the scenario with increased branching traffic



(d) Maximal queue lengths for every scenario in comparison with the threshold $L^*_{W,limit}$

Figure 3.10.: Obtained estimated queue lengths for every scenario (a-c) and comparison of the maximal queue length of any route per scenario (d).

higher queue length than r_5.

For the scenario with more traffic on the branching routes, $E_{LW,2}$ outgrows the queue length of every other route, while $E_{LW,7}$ is only marginally smaller.

When comparing the maximal queue length $\max_i E_{LW,i}$ of all routes for a given scenario with the threshold $L^*_{W,limit}$ for sufficient quality (Figure 3.10d), the maximal number of requests per hour can be assessed by selecting the corresponding operating program A^*_θ. This timetable capacity differs for the different scenarios, significant values can be read in Table 3.11.

Table 3.11.: Timetable capacity for all considered scenarios and selected parameters of the significant operating program A^*_θ.

	n_{total}	n_1	n_2	n_3	n_4	n_5	n_6	n_7	n_8	significant route	$p_{\text{loss,max}}$
a	35	4.38	4.38	4.38	4.38	4.38	4.38	4.38	4.38	r_7	0.0103
b	40	3.333	3.33	6.67	6.67	6.67	3.333	3.333	6.67	r_5	0.0118
c	29	4.83	4.83	2.42	2.42	2.42	4.83	4.83	2.42	r_2	0.0098

Summarizing the results, timetable capacity differs depending on the selected traffic share. While only $m = 4$ waiting positions per route have been included in the created CTMC due to memory restrictions, the maximum probability of a loss $p_{\text{loss,max}}$ on all routes has been calculated as ≈ 0.01 for every considered scenario. This can serve as evidence for the soundness of the presented model, even though the model has been restricted to this amount of waiting positions.

In conclusion, the introduced method enables infrastructure managers to estimate the performance of junction infrastructure even before planning a detailed timetable, by restricting the required input to route-exclusions and traffic share.

3.5.2. Infrastructure Layout Planning

The introduced timetable capacity analysis allows the infrastructure operator to quantify the number of plannable trains in the timetable process considering a fixed infrastructure layout. It therefore utilizes pre-specified service times for each different route through the infrastructure.

However, infrastructure operators might be interested in not only evaluating a single infrastructure layout regarding its capacity but also in choosing one layout out of a

collection. With the following application of the introduced method, different infrastructure layouts can be compared with the performance indicator *maximum mean service time*. That indicator describes an upper bound to the average service time, that a request in a given operating program may utilize.

This section therefore particularly focuses on choosing the right infrastructure layout for a given operating program, i.e. problem statement (II) in Section 3.3.2.

3.5.2.1. Setup



Figure 3.11.: Track-layout for a double-track junction with an overpass

In detail, a case study, deciding whether or not an overpass should be built for a junction with a given operating program has been conducted. An overpass is a way to reduce the number of route conflicts in a railway junction, an example for the track layout of the junction in Figure 3.1 is given in Figure 3.11.

To show the applicability of the described analytical approach, 23 different operating programs have been considered. Note the definition of an operating program in Section 3.3.2 as the train count on a route per U. In total, the operating program of the railway junction is assumed to have the same number of trains on the routes of the main line and the same number of trains on the routes of the branching line. Thus, all combinations of 23 different operating programs for both to the railway junction adjacent railway lines have been built and analysed for their peak traffic hour, resulting in a total number of $23^2 = 529$ different examples. A detailed description of all considered operating program combinations can be found in Appendix A.4, we restrict ourselves to six exemplary railway lines here.

The operating programs have been selected according to different types of adjacent

railway lines, we distinguish between *mixed traffic lines*, *local train lines*, *long-distance train lines*, *freight train lines*, and *urban railway lines*. Additionally, two different loads have been considered for the mixed traffic line. Table 3.12 lists the selected operating programs and their corresponding share of passenger traffic p_{pt}. Since we refrain from distinguishing different service rates per train type, the influence of the train mix in the operating program is mainly the definition of route-specific arrival rates and the determination of the share of passenger trains for the calculation of the threshold $L_{W,limit}$ (see Section 3.3.7).

Table 3.12.: Number of trains by traffic type for the considered operating programs

operating program	# regional trains	# high speed trains	# freight trains	share of passenger trains p_{pt}
low intensity mixed	2	0	1	0.67
long distance	0	4	0	1
local train	4	1	0	1
freight train	0	0	5	0
high intensity mixed	4	2	2	0.75
urban railway	10	0	0	1

In Figure 3.12 the considered combinations of those operating programs have been listed. In every entry, the top bar corresponds to the operating program on the *main line*, i.e. the routes from A to B (r_1) and from B to A (r_3), while the bottom bar corresponds to the operating program on the *branching line*, i.e. the routes from (r_4) and to (r_2) C.

The total number of trains on a route n_r is utilized to determine the arrival rate $\lambda_r = \frac{1}{n_r}$ for this route. Next, Continuous-Time Markov Chains (see Section 3.3.4) can be formulated for any possible service rate $\mu \in (0, 1]$, using the defined arrival rates λ_r and the maximum number of waiting positions per queue $m = 5$, while also considering conflicting routes. To ensure sufficient precision while also maintaining efficient computability, all service rates $\mu \in [0.01, 1]$ with a step size of 0.01 have been taken into consideration in this work.

Hence, 100 Models have been analysed for every route in both junction infrastructure layouts with/without an overpass for every considered combination of operating programs. The solving process has been automated utilizing the same software as in the case study in Section 3.5.1, similarly, the model files are available in the online

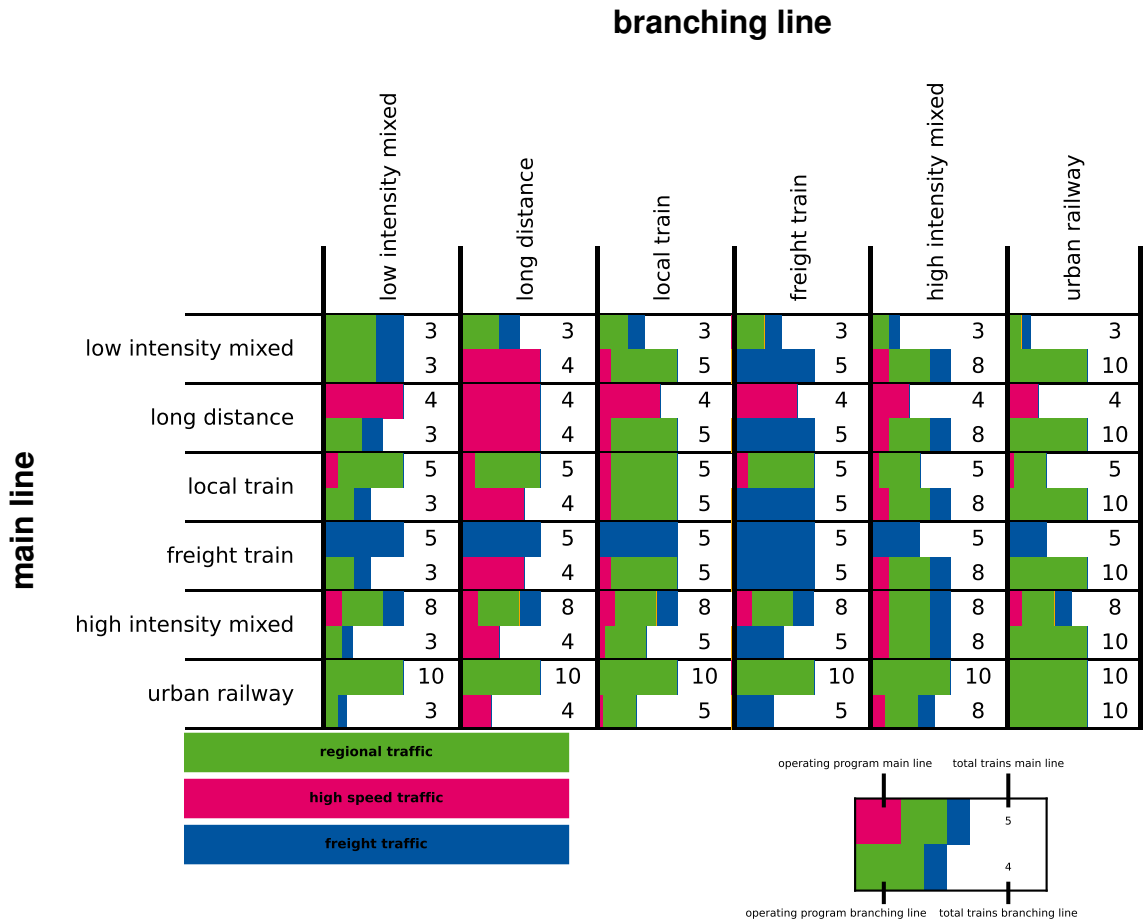


Figure 3.12.: Overview of the considered operating program combinations

repository (Emunds and Nießen, 2023).

Aiming to gain insights into the performance benefit of the infrastructure layout including the overpass (see Figure 3.11), the expected queue lengths at every route have been analysed. To facilitate a sufficient infrastructure quality, the most limiting service rate, hence the highest service rate of all routes in the junction, has been chosen.

Resulting from the solving process is a grid of expected lengths of the queue for both infrastructure layouts for every considered service rate $\mu \in [0.01, 1]$. In Figure 3.13 the expected queue length at route r_3 from B to A for both railway junction layouts of a local train main line and a long-distance train branching line is recorded. Selected values are listed in Appendix A.3.

Both operating programs consider passenger transport only, hence a ratio of passen-

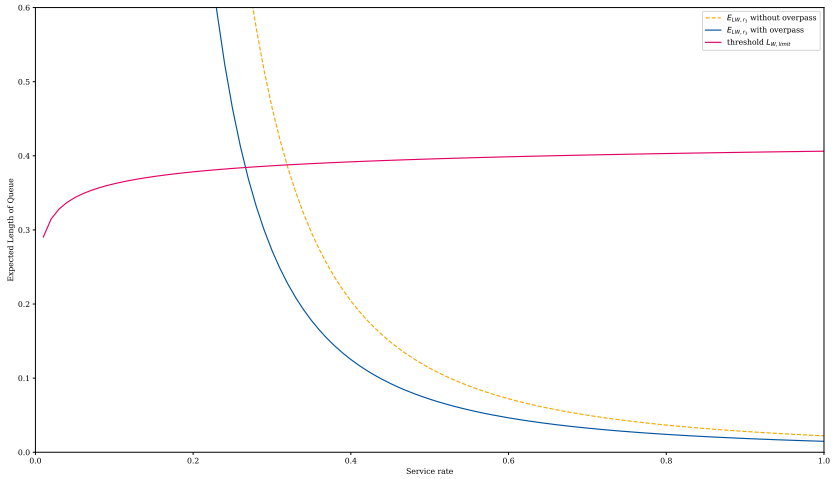


Figure 3.13.: Estimated queue length on route r_3 for the combination of a local train main line and a long-distance train branching line

ger trains in all considered trains of $p_{pt} = 1$ can be used for the calculation of the threshold value $L_{W,limit}$.

The combination is assumed to have to deal with an hourly load of 5 trains on the main line, from A to B and back, as well as 4 additional trains on the branching line, from A to C and back. This resumes in an arrival rate of $\lambda_r = 0.083$ for the routes $r \in \{r_1, r_3\}$ and of $\lambda_{r'} = 0.067$ for the routes $r' \in \{r_2, r_4\}$.

When observing the calculated estimations for the queue length E_{LW,r_3} at route r_3, the difference between the results for the infrastructure layouts with or without an overpass can be analysed. With a rising mean service rate μ in the infrastructure, the difference between the estimated queue length between the two infrastructure layouts drops significantly. This can be explained by the fact, that resource conflicts become less likely when less service time is consumed by the requests, indicated by a higher service rate. Hence the omission of the additional route conflict in the scenario with an overpass is much less relevant for high service rates.

Additionally, the threshold $L_{W,limit}$ is depicted, modified according to Section 3.3.7, depending on the occupancy rate of $\rho = \frac{\lambda_{r_3}}{\mu}$ for route r_3. The assumed parameters are listed in Table 3.7.

Utilizing the grid of resulting queue length estimations for every route and the calculated threshold, a minimum mean service rate μ_{min}, needed for sufficient infrastructure quality, may be obtained. Hence, the maximum mean service time $b_{\text{max}} = \frac{1}{\mu_{\text{min}}}$ for

the given operation program on the main and branching lines can be derived. This maximum mean service time can be used to investigate the needed infrastructure by comparing it to the actual achieved service times on the analysed junction, which are subject to train and control system specific parameters.

3.5.2.2. Results

The introduced derivation of a maximum mean service time has been applied to all 529 considered operating program combinations for the infrastructure settings with and without an overpass.

In Figures 3.14 and 3.15 the results for the selected 36 combinations are shown in heat-maps for both considered railway junction layouts. While both infrastructure settings share the same global distributions of service time requirements, results show that under the same load, as fixed by the operational program of the main and branching line, a railway junction including an overpass structure allows for a higher mean service time, while achieving the same operational quality as a railway junction without the overpass structure.

Considering the example of a railway junction (see Figure 3.13) with a local train main line (5 trains per direction) and a long-distance train branching line (4 trains per direction), the following results are obtained. To implement an infrastructure layout without an overpass a mean service time of at most 3.12 min needs to be fulfilled. Furthermore, the infrastructure layout with an overpass would allow for a mean service time of at most 3.7 min. By constructing an overpass, trains would be allowed to utilize 0.58 min (18.59%) more service time.

With the calculated maximum mean service times, the infrastructure operator can specify precise planning guidelines for infrastructure planners when to plan for an overpass structure, or when to decrease average mean service times by implementing more efficient railway control systems. After developing the microscopic railway junction design scenario, infrastructure planners can decide whether the specified upper bound is adhered to with a microscopic minimum headway time calculation.

Further investigating the difference between the two infrastructure settings, Figure 3.16 includes the maximum mean service times (Figure 3.16a) for all 529 computed combinations of main and branching lines (see also Appendix A.4) and a histogram, representing the distribution of relative differences in the calculated maximum mean

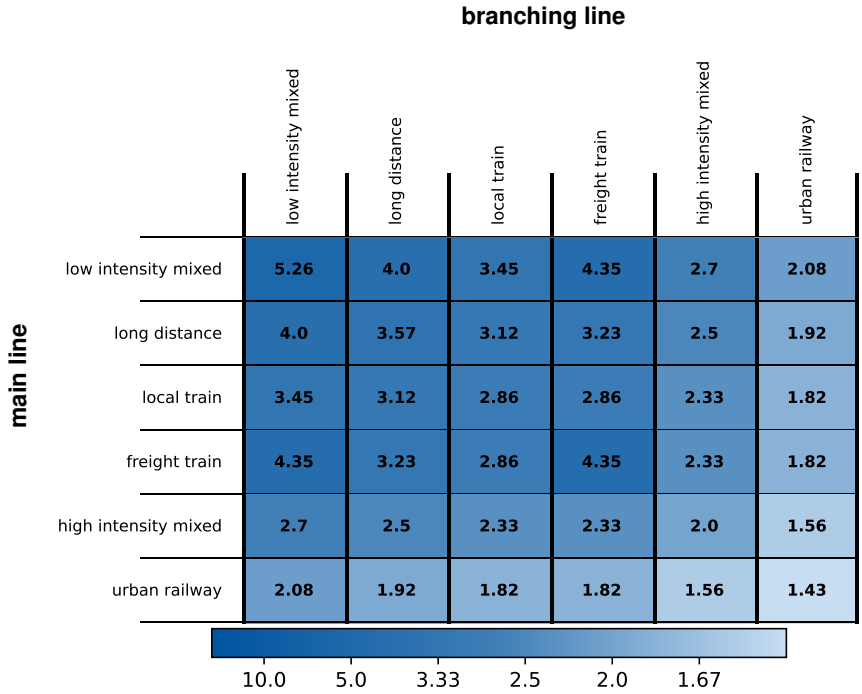


Figure 3.14.: Calculated maximum mean service times in minutes for the scenario without an overpass

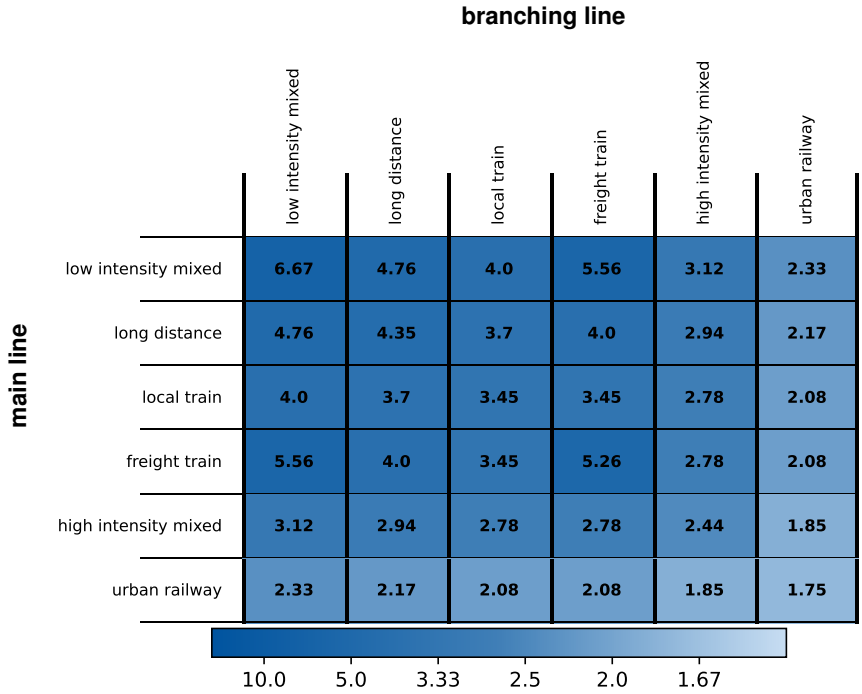
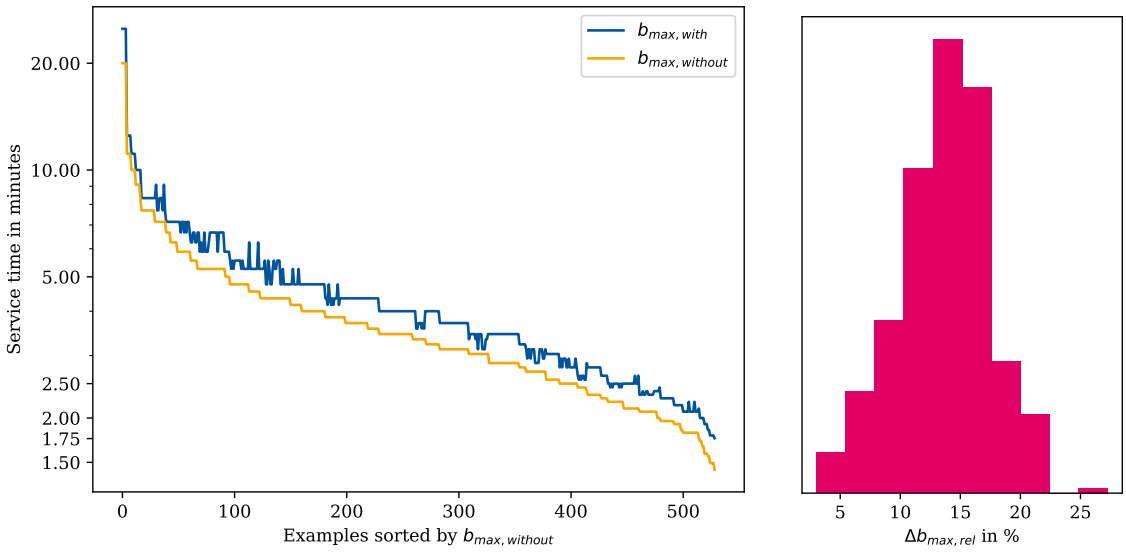


Figure 3.15.: Calculated maximum mean service times in minutes for the scenario with an overpass



(a) Distribution of maximum mean service times on all examples

(b) Relative difference in maximum mean service times

Figure 3.16.: Comparison between the obtained maximum mean service times of the two infrastructure scenarios

service times (Figure 3.16b).

Taking the logarithmic scale in Figure 3.16a into account, a wide range of calculated maximum mean service times required for sufficient operational quality can be recognized. While some operating programs, i.e. examples with a low number of total trains in the considered peak hours, are granting sufficient operational quality even for mean service times as high as 10 to 20 minutes, the majority of considered examples require a mean service time of under 5 minutes, with very densely operated main and branching lines demanding maximum mean service times of even under 2 minutes.

By including an overpass in the layout of a planned railway junction, the required maximum mean service times can be decreased by a significant margin. For the majority of considered examples, the relative difference in the maximum mean service time lies between 10% and 20 %, but the achieved reduction of maximum mean service time is diverse. Some examples show only a marginal difference between the considered infrastructure layouts, others report maximum mean service time decreases by up to 25 %.

Crucially, service times are dependent on various factors, some of which are not

under the influence of infrastructure managers. Hence, substantial limitations to the lower bound of service times are relevant, affected by physical properties such as length or acceleration and braking performance of the rolling stock.

In conclusion, differences between the considered infrastructure layouts have been analysed with the introduced method. For this, the calculation of many different operating programs and service rate configurations has been conducted, resulting in substantiated estimations of infrastructure quality requirements. The results clearly show the applicability of the introduced approach to quantify the effect of overpass structures in railway junctions.

3.5.2.3. Comparison with capacity utilization

The introduced method calculates the estimated queue length on a route through the junction infrastructure for different service times. It compares the results to the threshold of the maximum queue length (see Section 3.3.7) to obtain the maximum possible mean service time on all routes b_{max} for sufficient quality. While this threshold is in practical use for the timetable capacity calculation in Germany (DB InfraGO, 2009), internationally the UIC 406 Method (UIC, 2013) is applied more often, planning railway infrastructure according to the capacity utilization of a fixed timetable.

For this, the occupation rate ρ_{UIC} for the infrastructure is calculated by compressing the service times of all trains $b^* = \sum_i b_i$ and dividing it by the time period t_U

$$\rho_{UIC} = \frac{\sum_i b_i}{t_U}. \tag{3.36}$$

For this method, an occupation rate between 60% and 80%, $\rho_{UIC} \in [0.6, 0.8]$, is defined as a sufficient quality in railway node switch areas.

While the introduced methodology is timetable-independent, the UIC
Method is based on a fixed timetable and can therefore utilize the order of train runs and routes to calculate the minimum headway times between subsequent train runs on a railway junction. However, in previous research (Schwanhäußer, 1978; Hansen and Pachl, 2014) a timetable-independent method to calculate the *connected occupation rate* ρ_{conn} of a railway junction by utilizing the *concatenation value* ψ has been proposed. The concatenation value of a railway junction can be obtained with the operating program $A = \{(r_i, n_i) | i \in R\}$ by calculating the frequency of a sequence i-j,

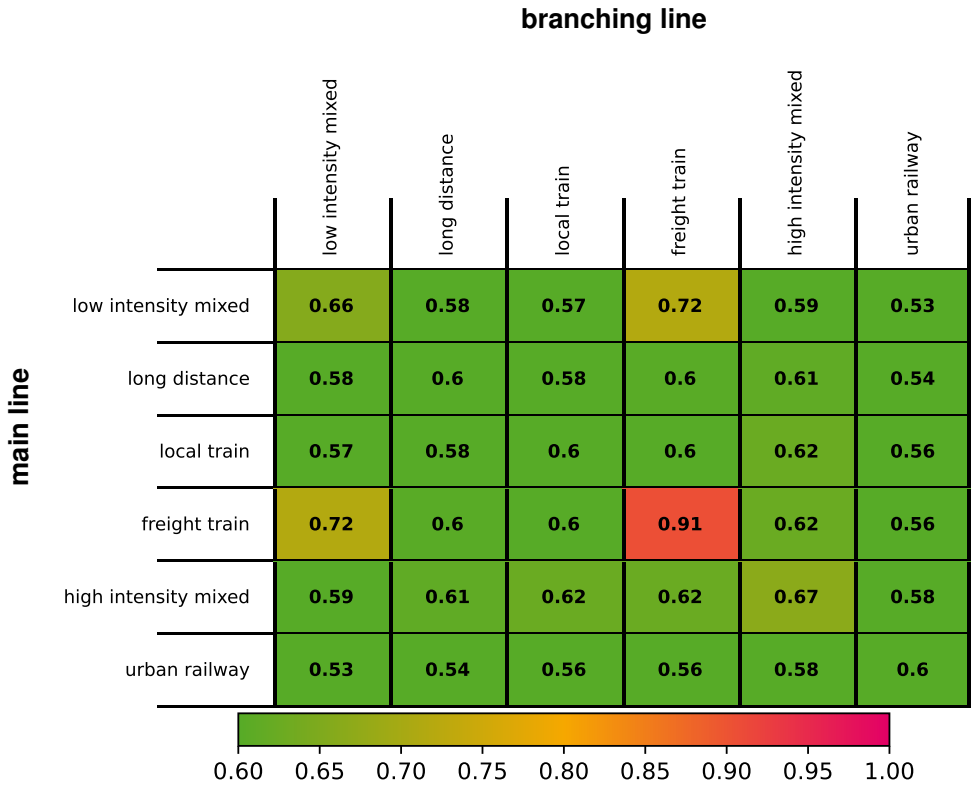


Figure 3.17.: Occupancy rates (following equation (3.38)) for the scenario without an overpass, assuming mean service times from Figure 3.14

$p_{i,j} = (n_i * n_j)/(n_{total}^2)$, with the total number of trains in the operating program n_{total} and

$$\psi = \sum_{i,j} p_{i,j} \cdot C_{i,j}, \tag{3.37}$$

applying the conflicts described in the conflict-matrix C (Potthoff, 1962). The connected occupancy rate can further be obtained with the knowledge of a mean service time b with

$$\rho_{conn} = \frac{b \cdot n_{total} \cdot \psi}{t_U}. \tag{3.38}$$

Figure 3.17 depicts the connected occupancy rates of a railway junction, considering the maximum mean service rates computed in Section 3.5.2.2 for the infrastructure setup without an overpass. It indicates the similarity between the utilized threshold, describing the limit on the queue length on the route r_3, and the capacity utilization threshold formulated in UIC (2013).

For the considered examples, most calculated maximum mean service times corre-

spond to capacity occupations in the desired range of 60% to 80%. However, for operating programs with a lot of freight trains, the occupation factors are substantially higher, reaching up to 91%. This can be explained by the dependence of the utilized threshold $L^*_{W,limit}$ from the ratio of passenger trains p_{pt} in the operating program.

Some maximum mean service time results yield capacity occupation factors of less than 0.6, with the lowest value at 0.53. This can be observed for some operating programs with a very small calculated maximum mean service time, such as the instances for urban railways. In those cases, the amount of traffic on the main line is greater than the amount of traffic on the branching line. Hence, the measured quality on the main line is worse than in the combined junction. Therefore, a small maximum mean service time is needed to adhere to the quality threshold on the main line.

Similar observations can be made for the results with an overpass and when considering all operating program combinations (see Appendix A.5).

Noting the difference in the objectives, timetable-independent vs. timetable-dependent analysis, and the additional train mix sensitivity of the utilized threshold, the introduced method yields plausible results for the considered operating program combinations. Since the introduced method provides route exact results, it is sensitive to differences in the occupation of different routes through the infrastructure. Therefore, the calculated values of all routes should be considered for infrastructure planning, in particular for instances with diverging main and branching line operating programs.

3.6. Discussion and Outlook

This work introduced a novel method for analysing the timetable capacity of railway junctions based on queueing theory. It is applicable for solving both formulated problem statements (Section 3.3.2), timetable capacity determination of a given railway junction infrastructure (I) and planning of junction infrastructure for a fixed operating program (II). By modelling railway junction routes as parts of a queueing system, while respecting their parallel service possibilities and considering resource conflicts, timetable independent analyses of examined infrastructure are enabled.

Extending classical queueing theory concepts, well established for railway line performance analysis (see Section 3.3.3), timetable capacity is determined by comparing queue length estimations through a novel Continuous-Time Markov Chain formulation

representing the considered railway junction (see Section 3.3.4) with threshold values (see Section 3.3.7) depending on parameters of considered service and arrival process distributions along with operating program specifics. In this work, estimations of queue lengths have been carried out by probabilistic model-checking software, enabling a fast and reliable computation for complex infrastructure dependencies.

With the introduced CTMC, railway junctions of up to 8 routes can be described without further amendments, however, railway line intersections of more than 2 lines and railway stations may consist of a higher number of routes. While the capabilities of the presented modelling approach will advance with increasing memory availability in future computer architecture generations, it is unlikely that those performance developments can fully account for the route-dependent exponential scaling of the introduced state set. It is therefore crucial to enforce the development of approximation techniques and decomposition approaches in further research.

The performance of the introduced approach has been studied by exemplary comparing the computation times and results of the analytical solution with simulations (see Section 3.4). When compared to simulations with an unlimited number of waiting positions, differences between the calculated queue lengths can be identified. Considering that this difference decreases with a reduction of the occupation ratio, the accuracy of the introduced method is accurate enough for a sufficient analysis within the neighbourhood of and below the threshold values, which correspond to the quality, where practical operation of railway traffic usually takes place. The conducted simulations indicate the high accuracy of the introduced method when limiting the number of waiting positions. Furthermore, the described calculation process determines the analytically correct solution to the presented model, while also being significantly faster than the applied simulations, which can only combine individual representations of the randomized timetabling process.

In an implementation of the novel analytical model for two separate case studies (see Section 3.5), the two problem formulations, determining the timetable capacity of a railway junction and selecting an infrastructure layout for a given operating program, have been tested. As an initial example, the timetable capacity of a complex railway junction in the west of Paris has been estimated for different traffic share scenarios. Additionally, a case study describing the selection of sufficient junction infrastructure, assuming 529 different operating program combinations, has been conducted, leading to substantive estimations regarding the effect of overpass structure on the

timetable capacity of railway junctions. In a comparison with an approach calculating the capacity occupation of the junction infrastructure, the route-sensitive accuracy of the obtained results has been highlighted. Hence, the introduced method has been proven to be able to create accurate solutions for manifold application scenarios, ranging from an abstract analysis of the general effect of overpass infrastructures to a practical implementation in the planning process of railway infrastructure operators when evaluating the performance of infrastructure layouts.

Even though the concept is already applicable, additional research could still yield substantial benefits. As such, the modelling of general independent stochastic processes with Markov processes and approximation factors might be improvable, i.e. by including phase-type distributions. Similarly, utilized thresholds for the expected length of a queue have been introduced for use in the context of railway lines, updating these to innovative measures could improve the real-world applicability. Additionally, the introduced concept could be extended to implementations for railway stations and eventually railway networks. By including delay distributions and propagation in the model, similar measures for operational capacity could be enabled.

However, by utilizing the developed timetable independent method, infrastructure managers can identify bottlenecks in the early stages of the planning process. With the achieved computing time benefits when compared to simulation approaches, they are set up to give junction designers substantiated indicators for the required infrastructure. Hence, the introduced analytical timetable independent approach has been proven to form a valuable addition to the capacity method landscape.

4. Utilizing Phase-Type Distributions for Queueing-Based Railway Junction Performance Determination

Abstract

To ensure the effective and objective development of transportation networks, it is crucial to identify performance limitations across various subsystems. A timetable-independent assessment of infrastructure capacity at railway junctions is a fundamental aspect of long-term rail network planning. While recent research introduced queueing-based methods to quantify route-based railway junction performance, modelling arrival and service processes has been limited to exponential distributions. This work utilizes Phase-Type Distributions to propose an extension to a previously described Continuous-Time Markov Chain model. In a comparison between assumed distribution combinations, the effect of a more detailed stochastic process modelling is described. Furthermore, an analysis of the differences to a simulation method is conducted for an exemplary railway junction. The introduced method enables infrastructure managers to accurately model stochastic processes for performance determination in the early stages of the strategic planning phase.

4.1. Introduction

Due to the increasing demand for environmentally friendly modes of transportation, railway infrastructure managers worldwide need to develop their infrastructure further. Performance analysis of individual facilities serves as a valuable method to evaluate existing infrastructure and compare expansion and new construction scenarios. Some well established methods are particularly suited for analysing current infrastructure and timetables, for example those designed to determine capacity utilisation (UIC, 2013). On the other hand, efficient resource management requires a

focus on long-term planning. For this purpose, timetable-independent methods can be particularly useful.

While such methods for railway lines are already well established in their practical implementation, methods for railway junctions or nodes are still the subject of ongoing research. The currently utilized systems in practice employ single-channel approximations for multichannel service systems, such as railway route nodes and junctions. However, those systems with parallel utilisable routes are particularly critical points that need to accommodate the traffic from multiple railway lines.

In a preceding work (Emunds and Nießen, 2024a), a multichannel method was introduced to analyze the performance of railway junctions and estimate the necessity for an overpass structure to mitigate route dependencies. In that study, the arrival and service processes were modeled using only exponential distributions, necessitating the use of approximation formulas to accommodate other probability distributions.

In contrast, this work introduces a novel approach, modelling general independent arrival and service processes with phase-type distributions in a Continuous-Time Markov Chain for multi-channel railway systems. This model can be used to obtain more accurate approximations of the queue-lengths of the different routes through the infrastructure, thereby enabling timetable-independent quantifications of the occupancy of each route, which allows for a detailed bottleneck analysis of the infrastructure.

The contribution of this work is manifold, including:

- A novel Continuous-Time Markov Chain model that utilizes phase-type distributions for modeling arrival and service processes in a railway junction.
- An algorithm designed to efficiently compute timetable capacity by comparing queue-length estimations to their respective thresholds.
- A comprehensive study analyzing the influence and approximation quality of different queuing models for varying distributions of traffic types across routes.
- A case study demonstrating the usability of route-based analysis to evaluate the performance of a railway junction infrastructure that accommodates both freight and passenger traffic.

The remainder of this work is structured as follows. Section 4.2 presents the current state of research. In Section 4.3, a formal description of the junction capacity de-

termination problem is provided. Subsequently, the novel model and the employed approximation formulas are introduced in Sections 4.4.3 to 4.4.5, along with an explanation of the algorithm for efficient capacity determination in Section 4.4.6. Section 4.5 offers a validation of the proposed method and a comparison with other models under different traffic distribution scenarios. Finally, an example railway junction combining a freight and a passenger traffic line is analyzed in Section 4.6.

4.2. Related Work

This section provides a summarized overview of the state-of-the-art for railway capacity analysis, focusing on key methodologies and their applications.

Railway capacity is defined in various ways depending on the planning stage and requirements (Jensen et al., 2020; Emunds and Nießen, 2024a). The *theoretical capacity* refers to the maximum number of trains that can be scheduled without conflicts, considering driving dynamics and control systems. The *timetable capacity* or *maximal capacity* includes additional factors such as train-mix and schedule quality. Finally, the *operational capacity* or *practical capacity* accounts for disturbances and delays, ensuring acceptable operational quality. Research also focusses on calculating the *capacity utilization*, which measures how much of the available capacity is used in a given timetable.

Various methodologies have been employed for railway performance analysis and vary in their dependency on timetables; some are timetable-dependent while others are not, making them useful for early infrastructure planning stages.

These methodologies differ based on the analyzed infrastructure (lines, junctions, stations, networks), infrastructure decomposition, and solution methods like mixed integer programming (MIP) or matrix calculations.

The *UIC Code 406* (UIC (2004, 2013)) is widely used internationally for assessing railway line and station capacities by determining capacity utilization through compression methods requiring a timetable or randomly generated sequences of train types to overcome timetable dependencies (Goverde (2007); Goverde et al. (2013); Abril et al. (2008); Bešinović and Goverde (2018); Jensen et al. (2020)).

Optimisation methods estimate theoretical capacity through linear mixed-integer programming problems for lines, stations, and networks. They solve railway timetabling

problems to build optimal timetables based on objective functions (Zwaneveld et al. (1996); Burdett and Kozan (2006); Burdett (2016); Harrod (2009); Lusby et al. (2011); Cacchiani and Toth (2012); Cacchiani et al. (2016); Yaghini et al. (2014)). Some optimisation approaches incorporate rolling stock information or model delay propagation effects using max-plus algebras (De Kort et al. (2003); Mussone and Wolfler Calvo (2013); Liao et al. (2021)).

Simulations provide detailed insights into operational parameters by simulating train operations based on given timetables or random variables representing arrival and service processes (Zieger et al. (2018); D'Acierno et al. (2019)).

Analytical methods based on queueing theory efficiently analyze timetable or operational capacities during early planning stages by considering inter-arrival and service time distributions, thus typically remaining independent of a predefined schedule. Methods to determine the performance of railway lines (Schwanhäußer (1974); Schwanhäußer and Schultze (1982); Wendler (2007); Weik and Nießen (2017)), railway junctions (Schwanhäußer (1978); Wakob (1981); Nießen (2008, 2013); Schmitz et al. (2017); Weik (2020); Emunds and Nießen (2024a)) and track groups (Potthoff (1962); Fischer and Hertel (1990)) have been introduced.

These timetable-independent models inherently address uncertainties in arrival and service processes, such as train precedence, buffer times or minimum headway times. To incorporate these uncertainties into UIC Code 406 or simulation based methodology, multiple timetable samples must be considered (Jensen et al., 2020; Weik et al., 2020b). Alternatively, optimization methods require additional complexity to ensure *robust* solutions (Liebchen et al., 2009).

The timetable-independent method proposed here is based on queuing theory. Table 4.1 compares other queueing-based methods for evaluating railway junction performance with the model introduced in this paper.

For example, Schwanhäußer (1978) introduces a single-channel approximation to obtain performance indicators of railway junctions by scaling route-based waiting times with the occurrence probability of route conflicts. While Schwanhäußer considers exponentially distributed inter-arrival times, Wakob (1981) extends this model for Erlang distributed arrival processes.

Nießen (2008, 2013) analyses interlocking nodes for their timetable and operational capacity using loss probabilities in service systems. To this end, exponentially dis-

Table 4.1.: Analytical junction capacity determination methodology

Literature	multi-channel model	timetable capacity	route-based decomposition	route-based quality assessment	explicit non exponential process modelling	solution method
(Schwanhäußer, 1978)		✓	(✓)			closed-form formula
(Wakob, 1981)		✓	(✓)		(✓)	iterative formula
(Nießen, 2008, 2013)	✓	✓		(✓)		iterative formula
(Schmitz et al., 2017)	✓	✓			✓	matrix-vector equations
(Weik, 2020)		✓	(✓)		✓	matrix-vector equations
(Emunds and Nießen, 2024a)	✓	✓	✓	✓		probabilistic model-checking
introduced here	✓	✓	✓	✓	✓	probabilistic model-checking

Remarks: Please note that a statement in bracelets means that this feature is partially supported.

tributed inter-arrival and service times are employed and results are scaled to approximate non-exponential service times, utilizing the variation coefficient.

In Schmitz et al. (2017), a railway junction is divided into two service stations and is modelled as a Markov Chain. Notably, the exponential scaling of the state space due to the storage of the request type, as well as the assumption of an existing subdivision into independent service stations, are significant properties of this approach. This method allows a railway junction to be considered as a multi-channel service system and analysed using phase-type distributed arrival and service processes.

To analyse the significance of switch connections in station yards, Weik (2020) examines interlocking nodes by approximating them as a series of interconnected service stations in a single-channel system, following the methodology of Schwanhäußer (1978); Wakob (1981). In this model, the arrival, service, and an additional repair processes are represented using phase-type distributions.

In contrast, this paper is based on the model proposed by Emunds and Nießen (2024a). This model enables the performance evaluation of various junction infrastructures by discretizing the utilized infrastructure based on operational route paths and modelling the arrival and service of requests on this infrastructure using a Markov chain. Initially, only exponentially distributed inter-arrival and service processes were considered using the Markov chain model, which had to be converted into processes with variation coefficients not equal to 1 using approximation formulas for multi-channel service systems according to Fischer and Hertel (1990).

In this work, models are derived that enable the computation of phase-type distributions. In particular, several influencing factors on the accuracy of the approximation are presented and analyzed.

4.3. Problem Description

This work introduces a method to evaluating the performance of a double-track railway junction. The performance is assessed using the metric known as *timetable capacity*, which gauges the infrastructure's ability to accommodate train schedules, see also Wendler (2007); Emunds and Nießen (2024a). Specifically, it pertains to the timetabling process, wherein the infrastructure manager allocates service slots to various requests submitted by train operators. With this approach, random distributed requests can be modelled, encompassing all potential timetables that could result in such a procedure. Conflicts may arise due to the inherent randomness of the requests, which might necessitate imposing waiting times for some of these requests. Furthermore, it may be necessary to decline certain requests if no conflict-free allocation is possible. The performance of the infrastructure can be assessed by the estimated number of requests that are pending, referred to as the *estimated queue length* L. Although the method we propose is designed to model the allocation of scheduling requests, for clarity, we often refer to these requests as *trains*.

A *railway junction* describes the infrastructure that connects railway lines to more than two different destinations. It usually consists of entry- and exit-signals from and to every direction, as well as one or multiple switches to set the path for the different routes between origins and destinations. In comparison with a *railway station*, trains are not allowed to have scheduled stops in a railway junction, they only stop if the signalling system indicates occupied infrastructure in their planned route.

However, in the remainder of this work, a simple example of a double-track railway junction with four different routes will be considered. These routes can be distinguished regarding their direction, A-B (r_1) and A-C (r_2) start in A towards B or C, while B-A (r_3) and C-A (r_4) end in A, coming from B or C. This exemplary infrastructure is depicted in Figure 4.1.

Since no use of turnout tracks is required for the two routes between A and B, we refer to the line between A and B as the *main line* and the line between A and C as

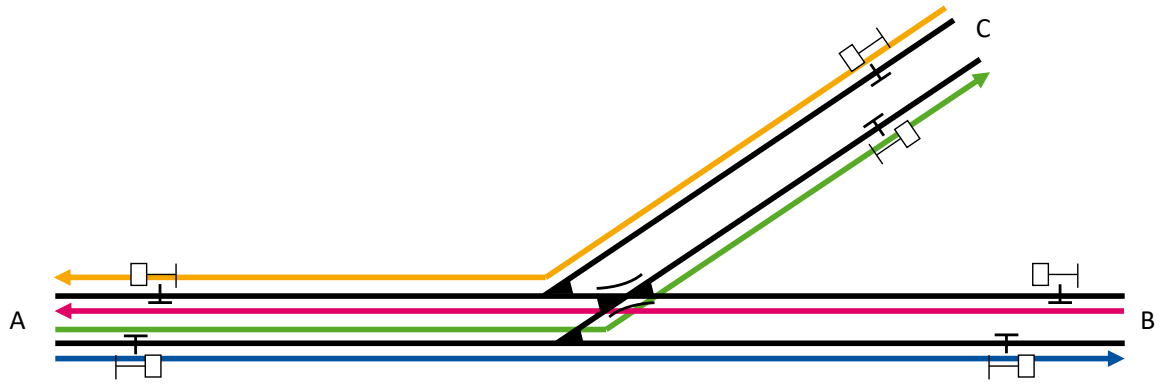


Figure 4.1.: Examplary infrastructure of a railway junction

the *branch line*. The proportion of the main line in the total traffic volume is then given by

$$p_{main} = \frac{n_{main}}{n_{total}}, \tag{4.1}$$

where it depends on the number of trains on the main line (n_{main}) and the total number of trains $n_{total} = n_{main} + n_{branch}$ on main and branch line (n_{branch}).

We express the railway junction as a tuple $J = (R, C)$ of a set R of k routes and a *conflict matrix* $C \in \{0,1\}^{k \times k}$, describing whether two routes $r, r' \in R$ can be used at the same time ($C_{r,r'} = 0$), or are *conflicting* ($C_{r,r'} = 1$).

In addition to the infrastructure, performance can depend on the used train types T. We therefore formulate an occupation request type $o = (r, t) \in O \subset R \times T$ as a combination of a route $r \in R$ and a train $t \in T$.

For some examples, it might be suitable to give the distribution of a given total number n_{total} of requests to the number of requests of same type $o = (r, t)$. This can be formalized by a function

$$\theta : \mathbb{R} \to \mathbb{R}^{|O|} : n_{\text{total}} \mapsto \theta\left(n_{\text{total}}\right), \tag{4.2}$$

yielding the number of requests $n_{r,t} = \theta_{r,t}\left(n_{\text{total}}\right)$ for every occupation request type $o = (r, t)$.

Determining the timetable capacity of a railway junction involves the definition and calculation of multiple parameters, some of which can already serve as an indicator about the quality of the planned junction. Table 4.2 gives a list of the used notations.

Table 4.2.: List of notations.

R	Set of routes
r_i	Route i
$C \in \{0,1\}^{k \times k}$	Conflict matrix
$J = (R, C)$	Junction infrastructure
t_U	Time horizon
T	Set of trains
O	Set of occupation request types
θ	Request distribution
λ	Arrival rate
μ	Service rate
ρ	Occupancy rate
v_A	Variation coefficient arrival process
v_S	Variation coefficient service process
S	Set of states
q_i	Number of requests in queue of route r_i
s_i	Status of Service of route r_i
$p_{A,i}$	Phase of arrival process on route r_i
$p_{S,i}$	Phase of service process on route r_i
M	Maximum rate
$MC = (S, Q_{\text{rates}}, t_{\text{init}}, AP, \mathcal{L})$	Continuous-Time Markov Chain
Q_{rates}	Transition rate function
(u, v)	Transition from state u to state v
L_r	Expected queue length of route r
m	number of waiting slots
L_{limit}	Threshold value for sufficient quality
p_{pt}	Share of passenger trains
p_{main}	Share of main line traffic
n_{total}	Total number of trains
n_{max}	Timetable capacity
p_{suburban}	Share of suburban trains
$p_{\text{regional freight}}$	Share of regional freight trains
$n_{(r,t)}$	Train numbers on route r and for traffic type t
$h_{i,j}$	Minimum headway time of the sequence train j after train i
b	Service time
n_r	Number of trains on route r

Performance analysis of railway infrastructure is typically conducted with respect to a fixed *time horizon* t_U, which defines the duration of the investigation period in minutes. The *arrival rate* $\lambda_r = \frac{n_r}{t_U}$ represents the average number of trains per route that request service on route r per minute.

The performance of railway infrastructure is significantly influenced by the service times planned for train usage. For this purpose, the *minimum headway times* $h_{(r_i,t_i),(r_j,t_j)}$ describe the minimum time interval required for the route-train combination (r_j, t_j) to initiate service after the service of the preceding route-train combination (r_i, t_i) has commenced. These headway times are dependent not only on the infrastructure and the specific routes that the trains t_i and t_j are scheduled to traverse, but also on certain rolling stock parameters, such as acceleration and braking behavior, as well as train length.

For comprehensive details on obtaining blocking and minimum headway times, readers are referred to Hansen and Pachl (2014). In this paper, it is assumed that the minimum headway times are provided; typically, these would have been calculated using a microscopic tool prior to conducting a detailed infrastructure analysis. However, microscopic data might not be available for long planning horizons, and assumptions or other measures to estimate blocking times might be necessary.

Let $n_{r,t}$ denote the number of trains on route r of train type t. For a given pair r, r' of conflicting routes ($C_{r,r'} = 1$), the minimum headway time $h_{(r,t),(r',t')}$ of all possible sequences $(r,t),(r',t')$ of train-route combinations for this route pair r, r', can be weighted with the total number of possible sequences for r and r', $n_{r,r'} = \sum_t \sum_{t'} n_{r,t} \cdot n_{r',t'}$, to obtain the average minimum headway time

$$h_{r,r'} = \sum_t \sum_{t'} n_{r,t} \cdot n_{r',t'} \cdot h_{(r,t),(r',t')} \cdot \frac{1}{n_{r,r'}} \tag{4.3}$$

between all train pairs t, t' on route r and r' respectively. By further weighting the average minimum headway times $h_{r,r'}$ for each conflicting route pair with the total number of trains $n_{r'}$ per conflicting route r', the average *service time* for a route r can be calculated as:

$$b_r = \sum_{\substack{r' \\ C_{r,r'}=1}} \frac{n_{r'}}{n_{r,\text{conflict}}} \cdot h_{r,r'}, \tag{4.4}$$

where $\frac{n_{r'}}{n_{r,\text{conflict}}}$ describes the probability of the pair (r, r'), i.e. a sequence describing

any train on the route $r' \in R$ following any train on the conflicting route r. Here, the total number of conflicting trains for route r is given by

$$n_{r,\text{conflict}} = \sum_{\substack{r' \\ C_{r,r'}=1}} n_{r'}. \tag{4.5}$$

Using this service time b_r for each route, the *service rate* $\mu_r = \frac{1}{b_r}$ describes the average number of trains that can be serviced on route r per minute.

The *occupation ratio* ρ_r of a route can be determined by comparing the arrival- and service rates

$$\rho_r = \frac{\lambda_r}{\mu_r}. \tag{4.6}$$

Furthermore, to determine the timetable capacity of a railway junction, the expected queue-length L_r can be calculated for every route $r \in R$ utilizing the arrival rate λ_r and service rate μ_r.

To enforce a sufficient timetable quality, a limit to the expected queue length $L_{\text{limit},r}$ can be set by the infrastructure manager, see Section 4.4.2 for a detailed description. This limit depends on the share of passenger trains and can therefore vary between every route, depending on its operating program. However, the expected queue-length of every route can be compared to the route-specific threshold value to determine the quality factor

$$qf_r = \frac{L_r}{L_{\text{limit},r}} \tag{4.7}$$

of this specific route. Therefore, the bottleneck in the analysed railway junction infrastructure layout can be identified by comparing the quality factors of the different routes.

The timetable capacity of a given railway junction infrastructure J can be formulated as the maximal number of train requests n_{max} that can be scheduled, under consideration of a fixed distribution θ to routes, while respecting given minimum headway times $h_{(r_i,t_i),(r_j,t_j)}$ and not exceeding a set limit $L_{\text{limit},r}$ of the expected queue-length per route.

A method to calculate the route-based queue length has been introduced in Emunds and Nießen (2024a). It corresponds to modelling the railway junction as a Continuous-Time Markov Chain (CTMC) and calculating the state probabilities utilizing probabilis-

tic Model-Checking.

Since this model utilizes Markov Chains to model the arrival and service processes, it assumes exponentially distributed inter-arrival and service times. Therefore, the arrival and service processes have been supposed to be completely random, expressed by their coefficient of variation of $v_A = v_B = 1$. In correspondence with the Kendall notation (see Kendall (1953)), we call this fully exponential setting (M/M).

However, the coefficient of variation can vary between different settings and general planning rules specify the coefficient of variation for the arrival process to $v_A = 0.8$ and for the service process to $v_B = 0.3$. Hence, more general distributions need to be analysed in a queueing system for railway infrastructure. We call this general independent setting (GI/GI), also following the Kendall notation.

The next section extends the Continuos-Time Markov Chain model using phase-type distributions and further explains quality thresholds.

4.4. Methods

In this section, the methods for calculating the timetable capacity of a railway junction are presented. First, the Continuos-Time Markov Chain model for exponential arrival and service processes is described. Afterwards, the threshold values that ensure sufficient infrastructure quality are defined in Section 4.4.2. Next, the approximation methods used to adapt exponential distributions to general independent settings are discussed in Section 4.4.3. The novel model, which incorporates phase-type distributions for the arrival and service processes, is introduced in Section 4.4.4. Finally, the algorithm designed to efficiently compute timetable capacity using queue-length estimations (Section 4.4.5) is described in Section 4.4.6.

4.4.1. A Queueing Based Railway Junction CTMC

Following the approach in Emunds and Nießen (2024a), a CTMC model can be obtained by describing the railway junction as a queueing system.

For this, we model the queuing system as a stochastic process $\{X(t)|t \in \tau\}$, where $t \in \tau$ describes a point in continuous Time τ and $X(t)$ corresponds to a random variable with values, called *states*, in a state set S.

This state set S can be used to formalize the stochastic process as a Continuous-Time Markov Chain $MC = (S, Q_{\text{rates}}, t_{\text{init}}, AP, \mathcal{L})$ with a transition rate function $Q_{\text{rates}} : S \times S \to \mathbb{R}_0^+$, an initial distribution $t_{\text{init}} : S \to [0,1]$ with $\sum_{u \in S} t_{\text{init}}(u) = 1$, a set of atomic propositions AP and a labelling function $\mathcal{L} : S \to 2^{AP}, u \mapsto \mathcal{L}(u) \subset AP$. To obtain the estimated queue length of every route, this information needs to be accessible via the labelling function. Therefore, we define the set of atomic propositions as the set of $|R|$-Tuples $AP = \left\{(q_r)_{r \in R} \,|q_r \in \mathbb{N}^0\right\}$, which represents all possible combinations of queue-lengths $(q_r)_{r \in R}$ for the different routes. The labelling function L therefore maps a state $u \in S$ to the tuple $\mathcal{L}(u) = (q_r\,(s))_{r \in R}$, with $q_r(u)$ as the queue length of route r at state u. Finally, we utilize a simple initial distribution

$$t_{\text{init}}(u) = \begin{cases} 1 & \text{if } u = u_0 \\ 0 & \text{else,} \end{cases} \tag{4.8}$$

with one starting state $u_0 \in S$, describing an empty system.

In the former contribution (Emunds and Nießen, 2024a), the state set has been described as

$$\hat{S}_{M,M} = \{(q_1, s_1, \ldots, q_k, s_k)\,|q_r \in \{0, \ldots, m\}, s_r \in \{0,1\}\}\,. \tag{4.9}$$

Hence, the number of waiting requests $q_r \in \{0, \ldots, m\}$, and the state of use of the track $s_r \in \{0,1\}$ are described for each route r. Regarding the transitions $(u,v) \in S \times S$, three different types have been distinguished: *Arrival* transitions $(\ldots, q_i, s_i, \ldots) \mapsto (\ldots, q_i + 1, s_i, \ldots)$ correspond to an arrival to route i; *Service* transitions $(\ldots, q_i, 1, q_{i+1} \ldots) \mapsto (\ldots, q_i, 0, q_{i+1} \ldots)$ correspond to the service on route i; and *Choice* transitions $u \mapsto v$ correspond to the choice which route to serve next. The transition rate function

$$Q_{M,M}(u,v) = \begin{cases} \lambda_r & \text{if arrival to } r \\ \mu_r & \text{if service at } r \\ M & \text{if choice transition,} \end{cases} \tag{4.10}$$

maps the transitions (u,v) to their rates. In this model, the choice rate M corresponds to a very high rate, in order to introduce only negligibly small delays to the system. Note that a state might have a single outgoing choice transition, corresponding to the start of the service on a route if the associated queue is not empty and no other

conflicting route is being served. We still call this a choice transition, even though it technically does not correspond to a choice between multiple possibilities. It is therefore possible that multiple non conflicting routes might be served in parallel. The number of channels of the modelled queuing system depends on the currently served requests. However, we calculate the estimated queue length L_r for every route and therefore assume a single-channel queue in the approximation formulas introduced below.

To model non-exponential inter-arrival and service time distributions, we extend this formulation with phase-type distributions in Section 4.4.4. Before, thresholds for the performance determination are introduced in the next section.

4.4.2. Threshold Values

The calculated queue-length L_r for a route can be compared with limit values for the queue-length $L_{limit,r}$ to assess the quality of the provided transport service. For long-term infrastructure planning, the largest german infrastructure manager, DB InfraGO (2009), utilizes the threshold

$$L_{\text{limit, r}} = 0.479 \cdot \exp(-1.3 \cdot p_{\text{pt},r}) \tag{4.11}$$

(Schwanhäußer and Schultze, 1982).

This formula is dependent on the share of passenger trains

$$p_{\text{pt},r} = \frac{\text{number of passenger trains on route } r}{\text{total number of trains on route } r} \tag{4.12}$$

on route r. The more passenger trains are in the operating program of this route, the higher the threshold for sufficient timetable quality. Some intuition into this can be obtained by looking at allowances for time deviations in the timetabling process in Germany (DB InfraGO, 2023). While the timetable constructor might only move passenger trains up to 3 min, freight trains may be shifted by up to 30 min - depending on the ordered type.

4.4.3. Approximation Formulas

Different techniques have been developed to approximate the general independent queueing system (GI/GI). In this section, we focus on two well-known approximation formulas in Hertel (1984) (Section 4.4.3.1) and Kingman (1961) (Section 4.4.3.2), while the utilizing of phase-type distributions in the Markov Chain to model general independent stochastic processes is presented in Section 4.4.4.

4.4.3.1. Hertel Approximation

To achieve the coefficients of variation of $v_A = 0.8$ and $v_B = 0.3$, the calculated queue-lengths can be scaled with the approximation formula

$$L_r(M/M) \cdot \frac{1}{\gamma} \approx L_r(GI/GI) \tag{4.13}$$

from Hertel (1984) (see also (Fischer and Hertel, 1990)).

For this, the parameters

$$\gamma = \frac{2}{c \cdot v_B^2 + v_A^2} \tag{4.14}$$

and

$$c = \left(\frac{\rho_r}{s}\right)^{1-v_A^2} \cdot (1 + v_A^2) - v_A^2. \tag{4.15}$$

are determined in dependence on the occupation ration $\rho_r = \lambda_r/\mu_r$, the amount of channels $s = 1$ and the coefficients of variation v_A, v_B. This has been used in Emunds and Nießen (2024a) to model non-exponential arrival and service time distributions.

4.4.3.2. Kingman Approximation

Another Approximation, that has been widely used for single-channel systems (see f.e. Gudehus (1976)), is the Kingman approximation formula (Kingman, 1961)

$$L_r(M/M) \cdot \left(\frac{v_A^2 + v_B^2}{2}\right) \approx L_r(GI/GI). \tag{4.16}$$

It gives a good approximation for single-channel scenarios with a high occupation ratio $\rho \to 1$.

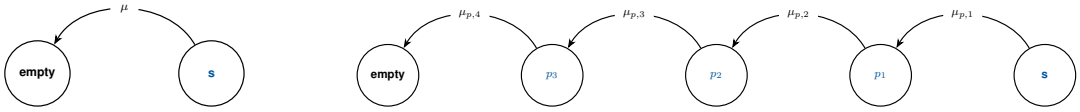
4.4.4. Phase-Type Distributions

To use the estimated queue lengths L_r with the Markovian model, they are scaled with one of the approximation formulas in Section 4.4.3 to facilitate arbitrary coefficients of variation for the interarrival and service distributions.

In this section, we introduce a novel Continuous-Time Markov Chain formulation, enabling a direct modelling of general independent distributions.

For this, phase-type distributions (Cox, 1955), a type of distribution function, whose parameters can be fitted to approximate any other arbitrary probability distribution (Asmussen, 2003), can be used. Additionally, they can be represented by a Continuous-Time Markov chain, which facilitates their integration into the described model.

Specifically, the service and arrival processes on railways can be modelled with hypoexponential distributions because their assumed coefficients of variation $v_X < 1,\ X \in \{A, B\}$ are less than 1 (cf. Weik (2020)).



(a) M (b) Ph

Figure 4.2.: Examples of the service process with exponential (M) or phase-distributed (Ph) service times

Figure 4.2 illustrates two different types of service processes: one with exponentially distributed service times (Figure 4.2a) and one with hypoexponentially distributed service times (Figure 4.2b). In the Markov chain for the latter distribution, three intermediate states were introduced, which can be reached through transitions with different rates from the preceding states.

The number k of the minimum required phases to approximate a distribution X with coefficient of variation v_X can be determined using

$$k \geq \lceil \frac{1}{{v_X}^2} \rceil \ , \tag{4.17}$$

(see David and Larry (1987)).

4.4.4.1. Hypoexponential Parameter Fitting

To adapt the hypoexponential distribution $\hat{X}$ to a distribution X, both the number of phases k and the transition rates μ_{p_i} of the corresponding phases can be used. Weik (2020) adapts an approach by Sommereder (2011), which divides the phases into two successive Erlang distributions X_1, X_2. The k phases are determined using the coefficient of variation according to (4.17) and divided into $k_{X_1} = \lceil \frac{k}{2} \rceil$ and $k_{X_2} = k - k_{X_1}$ phases.

In a first step, a combined distribution (X_1^*, X_2^*) of two artificial Erlang distributions with k_{X_1}/k_{X_2} phases is fitted to the coefficient of variation v_X. For this, the expected value of X_1^* is fixed to $E(X_1^*) = 1$.

By determining the expected value

$$E({X_2}^*) = \frac{k_{X_1} k_{X_2} {v_X}^2 + \sqrt{k_{X_1} k_{X_2} ({v_X}^2 (k_{X_1} + k_{X_2}) - 1)}}{k_{X_1} (1 - {v_X}^2 k_{X_2})} \tag{4.18}$$

for the artificial Erlang distribution ${X_2}^*$, the target coefficient of variation

$$v_X = \frac{\sqrt{Var({X_1}^*) + Var({X_2}^*)}}{E({X_1}^*) + E({X_2}^*)} = \frac{\sqrt{1/k_{X_1} + E({X_2}^*)^2/k_{X_2}}}{1 + E({X_2}^*)} \tag{4.19}$$

can be approximated together with the artificial Erlang distribution ${X_1}^*$.

To correctly estimate the expected value $E(X)$ of the distribution X, the expected values of the artificial Erlang distributions are then adjusted via

$$E(X_1) = \frac{E(X)}{1 + E({X_2}^*)} \tag{4.20}$$

and

$$E(X_2) = \frac{E(X) E({X_2}^*)}{1 + E({X_2}^*)}. \tag{4.21}$$

Finally, the transition rate μ_{p_i} between phases $i-1$ and i can then be determined by

$$\mu_{p_i} = \begin{cases} \frac{k_{X_1}}{E(X_1)} & i \leq k_{X1} \\ \frac{k_{X_2}}{E(X_2)} & i > k_{X1} \ . \end{cases} \tag{4.22}$$

4.4.4.2. Example

In the example from Figure 4.2b, the following parameters would be used to define a corresponding hypoexponential distribution for a service process with coefficient of variation $v_B = 0.5$ and a mean service time of $\bar{b} = 3$: In total,

$$k = 4 = \frac{1}{0.5^2} \tag{4.23}$$

phases need to be considered, with $k_{X_1} = k_{X_2} = 2$ states for the two artificial Erlang distributions. Using (4.18) - (4.22), it follows that

$$\mu_{p_i} = \begin{cases} 1.33 & 1 \leq i \leq 2 \\ 1.33 & 2 < i \leq 4 \end{cases} \tag{4.24}$$

for the transition rates between the states of the 4 phases, where the first transition represents the transition from state **s** to state p_1, and the last transition represents the transition from state p_3 to state **empty**. Note that the states may have additional specifications, such as the number of waiting requests, which are omitted here.

4.4.4.3. States

To model hypoexponential distributions for the service and/or arrival processes, the current phase of the service and/or arrival processes must be added to the state description for each route.

The state set $\hat{S}_{M,M}$ with exponentially distributed inter-arrival and service times from equation (4.9) can be extended to the state set

$$\hat{S}_{Ph,Ph} = \left\{ (q_1, s_1, p_{A,1}, p_{S,1} \ldots, q_k, s_k, p_{A,k}, p_{S,k}) \,|\, (q_r, s_r, p_{A,r}, p_{S,r}) \in \tilde{S}_r \right\} \tag{4.25}$$

where with

$$\tilde{S}_r = \{0, \ldots, m\} \times \{0, 1\} \times \{0, \ldots, k_{A,r} - 1\} \times \{0, \ldots, k_{S,r} - 1\} \tag{4.26}$$

the set of tuples can be specified that describe the state of this track for a route r.

Thus, the number of waiting requests $q_r \in \{0, \ldots, m\}$, the state of use of the track $s_r \in \{0, 1\}$, as well as the current phase of arrival $p_{A,r} \in \{0, \ldots, k_{A,r} - 1\}$ and service

process $p_{S,r} \in \{0, \ldots, k_{S,r} - 1\}$ are described for each route r. Here, $k_{A,r}$ and $k_{S,r}$ denote the number of phases in the arrival or service process.

For readability, the notations $q_r(u)$ for the queue-length q_r, $s_r(u)$ for the service status s_r, and $p_{A,r}(u), p_{S,r}(u)$ for the phase of the arrival/service process $p_{A,r}, p_{S,r}$ of route r at state u are additionally used in the following.

Furthermore, the states-space $\hat{S}_{Ph,Ph}$ can be restricted to

$$S_{Ph,Ph} = \left\{ u \in \hat{S} | \sum_{i=1}^{k} \sum_{\substack{j=1 \\ j \neq i}}^{k} (C_{i,j} \cdot s_i(u) \cdot s_j(j)) = 0 \right\} \tag{4.27}$$

by using the conflict matrix C to exclude states $u^* \in \hat{S}_{Ph,Ph} \setminus S_{Ph,Ph}$, that can not be visited because of conflicting ($C_{i,j} = 1$) routes, being serviced at the same time $s_i(u^*) = s_i(u^*) = 1$.

By formulating the transition rate function Q_{rates}, the railway junction can be modelled as a Continuous-Time Markov chain $MC = \big(S_{Ph,Ph}, Q_{\text{rates}}, t_{\text{init}}, AP, \mathcal{L}\big)$. To map the hypoexponential distribution in the arrival or service process, the arrival or service transitions in (Emunds and Nießen, 2024a) are replaced by transitions between individual phases in the respective process.

4.4.4.4. Transitions

The three different transition types from the model in Section 4.4.1 can also be formulated on the set of states $S = S_{Ph,Ph}$.

The **arrival** process to a route r has been described by a transition $t = (u, v)$ with $q_r(v) = q_r(u) + 1$ in the M/M model. However, for the use of a phasetype distribution with a total of $k_{A,r}$ phases, a change in rates after $k^*_{A,r}$ phases and the two rates $\lambda_{r,a}$ and $\lambda_{r,b}$, multiple transitions (u, v) need to be formulated for the arrival of a train. They can be distinguished according to the arrival phase $p_{A,r}(u)$ from the start state of a transition. For $p_{A,r}(u) \leq k^*_{A,r}$, transitions from u to v with

$$p_{A,r}(v) = p_{A,r}(u) + 1, \tag{4.28}$$

are applied with a rate of $Q_{Ph,Ph}(u, v) = \lambda_{r,a}$. Furthermore, for $k^*_{A,r} \leq p_{A,r}(u) \leq k_{A,r}$,

transitions of

$$p_{A,r}(v) = p_{A,r}(u) + 1, \tag{4.29}$$

with rate $Q_{Ph,Ph}(u, v) = \lambda_{r,b}$ are added. Lastly, transitions starting the service or adding the train to the queue are formulated for $p_{A,r}(u) = k_{A,r}$. When no conflicting routes are serviced in the start state u, the service of route r can start in state v

$$s_r(v) = 1, \tag{4.30}$$

otherwise a train is added to the queue

$$q_r(v) = q_r(u) + 1 \tag{4.31}$$

of the route. Both possible transitions reset the phase of the arrival process to

$$p_{A,r}(v) = 1 \tag{4.32}$$

in the end state v and use a rate of $Q_{Ph,Ph}(u, v) = \lambda_{r,b}$.

Service transitions between states u, v, with $s_r = 1$, are similarly modelled. Let $k_{S,r}$ be the number of phases in the service process with a change from rate $\mu_{r,a}$ to $\mu_{r,b}$ after $k^*_{S,r}$ phases. Then, for $p_{S,r}(u) \leq k^*_{S,r}$ transitions

$$p_{S,r}(v) = p_{S,r}(u) + 1, \tag{4.33}$$

with a rate of $Q_{Ph,Ph}(u, v) = \mu_{r,a}$ are used. For $k^*_{S,r} \leq p_{S,r}(u) < k_{S,r}$, the rate changes to $Q_{Ph,Ph}(u, v) = \mu_{r,b}$ and the same property (4.33) for u and v. Finally, if $p_{S,r}(u) = k_{S,r}$, the service process is terminated by resetting

$$p_{S,r}(v) = 1. \tag{4.34}$$

The last type, **choice** transitions, are used to model the initiation of service for a route when multiple routes $r \in R$ can be serviced due to the absence of conflicting routes being active in the initial state u. For each such possible route r^*, a transition (u, v) with an artificial rate $Q_{Ph,Ph}(u, v) = M$ is introduced, ensuring that

$$s_{r^*}(v) = 1 \tag{4.35}$$

and

$$q_{r^*}(v) = q_{r^*}(u) - 1 \tag{4.36}$$

are satisfied.

Ideally, these choice transitions would not introduce any additional time into the system. Therefore, the rate M should be selected to be sufficiently high, such that the expected induced time $1/M$ remains negligibly small. For all models presented in this paper, a rate of $M = 600$ has been selected, corresponding to a delay of $1/600$ minutes, which is considered to provide an adequately precise approximation (see also (Emunds and Nießen, 2024a)). Because all choice transitions occur at the same rate from any given state, there are no priorities assigned to different routes or requests.

Additionally, transitions from other processes must also be incorporated into the introduced intermediate states. Consequently, the arrival of a request on a route i' can be modeled during phase j of the service process for another route i. A full description of a model in the PRISM modelling language can be found in the online repository (Emunds and Nießen, 2024b).

The modeling approach introduced here, which employs hypoexponential distributions for the arrival and/or service processes, is evaluated for accuracy in Section 4.5.

4.4.5. Queue-Length Analysis

The introduced Continuous-Time Markov Chain $MC = (S_{Ph,Ph}, T) = (S, T)$ models the process of scheduling train requests on the considered railway junction. In order to assess the performance of this junction, the length L_r of the queue can be determined for all routes r by analysing the probability $p(u)$ of the states $u \in S$ with certain queue-lengths in the stationary distribution, i.e. the probability that the system is in state u in the long run. Hence, the expected length of the queue can be calculated with

$$L_r = \sum_{\substack{u \in S \\ q_r(u) > 0}} p(u) \cdot q_r(u). \tag{4.37}$$

The same basic concept has already been used by other analytical approaches in the performance determination of railway infrastructure (Schmitz et al., 2017; Weik and

Nießen, 2017; Weik, 2020; Emunds and Nießen, 2024a). Note that this estimation with stationary distributions assumes static rates in the CTMC, and therefore corresponds to a theoretical analysis of static traffic distributions for long time horizons. In such a capacity analysis, the influence of partial timespans with less or more traffic is therefore neglected. While various solution methods to determine the state probabilities $p(u)$ have been used, Emunds and Nießen (2024a) introduced an approach that builds the CTMC in the formal PRISM language (Parker et al., 2000) and applies probabilistic model-checking (Hensel et al., 2022) to obtain the expected queue-lengths L_r. The same mechanism is used for this work.

4.4.6. Capacity Determination Algorithm

To determine the timetable capacity of a railway junction, multiple n_{total} values must be tested to find the maximum train count n_{max}, where the expected queue lengths L_r do not exceed the limit $L_{\text{limit},r}$ (Section 4.4.2). With the introduced approach of modeling phase-type distributions directly within the CTMC, the number of states increases significantly (see also Section 4.5), leading to considerably longer computation times for the expected queue lengths L_r. It is therefore crucial to minimize the number of iterations in which the queue lengths must be recalculated, i.e., the number of tested n_{total} values.

To achieve this, a root finding problem can be formulated. Let the function

$$L_r : N_{\text{total}} \to \mathbb{R}, n_{\text{total}} \mapsto L_r(n_{\text{total}}) \tag{4.38}$$

describe the estimation of the expected queue-length $L_r(n_{\text{total}}) = L_r$ for a given train count $n_{\text{total}} \in N_{\text{total}}$ by utilizing probabilistic model-checking on a formulated CTMC.

By comparing it to the limit, the quality factor

$$qf_r(n_{\text{total}}) = \frac{L_r(n_{\text{total}})}{L_{\text{limit}}} \tag{4.39}$$

of a route r can be obtained. Since the limit has to be adhered to for every route, the optimal $n^*_{\text{total}} = n_{\text{max}}$ fulfills the property

$$qf_{\text{max}}\left(n^*_{\text{total}}\right) = \max_{r \in R} qf_r\left(n^*_{\text{total}}\right) = 1, \tag{4.40}$$

such that the maximum quality factor qf_{max} is exactly 1, and therefore the queue length at one route is exactly at the limit for this route.

The function

$$\phi : N_{\text{total}} \to \mathbb{R} : n_{\text{total}} \mapsto \phi\left(n_{\text{total}}\right) = qf_{\text{max}}\left(n_{\text{total}}\right) - 1 \tag{4.41}$$

can now be formulated, whose root $\phi\left(n^*_{\text{total}}\right) = 0$ describes the timetable capacity $n^*_{\text{total}} = n_{\text{max}}$ of the railway junction.

To determine the root of ϕ, the algorithm from Brent (1973), as implemented in `scipy` (Virtanen et al., 2020), was utilized. This algorithm efficiently computes the root of a function $f : \mathbb{R} \to \mathbb{R}$ within a specified interval $[a, b]$, where f changes its sign, achieving a solution up to a chosen level of accuracy. In the validation implementation (Section 4.5.3) and the case study (Section 4.6), a termination criterion was set to ensure both a total error of less than 10^{-3} and a relative error below 10^{-3}. An Algorithm, describing how to obtain the function evaluation $\phi(n_{\text{total}})$ for a given total train count n_{total} is given in Appendix B.1.

This algorithm is then called consecutively from within Bent's method for varying n_{total}, until the set tolerances are met and n_{max} has been found.

4.5. Validation

To analyze the different distribution functions, the following distinguishes between four different applications.

- Exponentially distributed inter-arrival and service times (M/M)
- Phase-type distributed inter-arrival and exponentially distributed service times (PH/M)
- Exponentially distributed inter-arrival and phase-type distributed service times (M/PH)
- Phase-type distributed inter-arrival and service times (PH/PH)

For each distribution of arrival and service processes, a model of the junction from Figure 4.1 was created.

In Table 4.3, the parameters of the different models are presented. Particularly noteworthy is the influence of the hypoexponential distribution on the size of the model.

Table 4.3.: Parameters of the studied models

Arrival process	Service process	Number of phases Arrival	Number of phases Service	v_A	v_S	Number of waiting slots m	Number of states	Number of transitions
M	M	1	1	1	1	5	10 368	63 688
Ph	M	2	1	0.8	1	5	141 108	829 825
M	Ph	1	12	1	0.3	5	623 376	3 664 703
Ph	Ph	2	12	0.8	0.3	5	8 192 448	46 447 056

While the number of states for the model with exponentially distributed inter-arrival and service times is still in the order of 10^4, more than 8 million states must be considered for the model with hypoexponentially distributed inter-arrival and service times.

The solving process has been automated using Python 3.10.9 (Van Rossum and Drake Jr, 1995; Python Software Foundation, 2022) and the probabilistic model-checker `Storm` (Hensel et al., 2022) with its Python-Interface `stormpy` (Junges and Volk, 2023). Furthermore, `scipy` (Virtanen et al., 2020) has been used for an implementation of the Brent's method (see Section 4.4.6). Used CTMC models (see Section 4.4.4) have been expressed in the PRISM modelling language (Parker et al., 2000; Kwiatkowska et al., 2011). The used model files can be found in the online repository (Emunds and Nießen, 2024b).

Results are compared to simulations, that have been implemented in Python 3.10.9 (Van Rossum and Drake Jr, 1995; Python Software Foundation, 2022), utilizing `SimPy` (Team SimPy Revision, 2023) and `Ciw` (The Ciw library developers, 2024; Palmer et al., 2019). An implementation can be found in the online repository (Emunds and Nießen, 2024b). In the simulation, inter-arrival times and service times are drawn randomly from phase-type distributions fitted according to Section 4.4.4.1 for every route. This process does not utilize any timetable information. Trains are managed using first-in-first-out (FIFO) queues for every route, and conflicting routes are prevented from starting service simultaneously by ensuring that shared resources are not utilized concurrently. The validity of this approach is verified through conflict analysis and the queue-length per route is monitored at each simulated minute and averaged for comparison purposes.

In the following sections, computation times and the approximation quality of the different experiments are analysed.

4.5.1. Computation Time

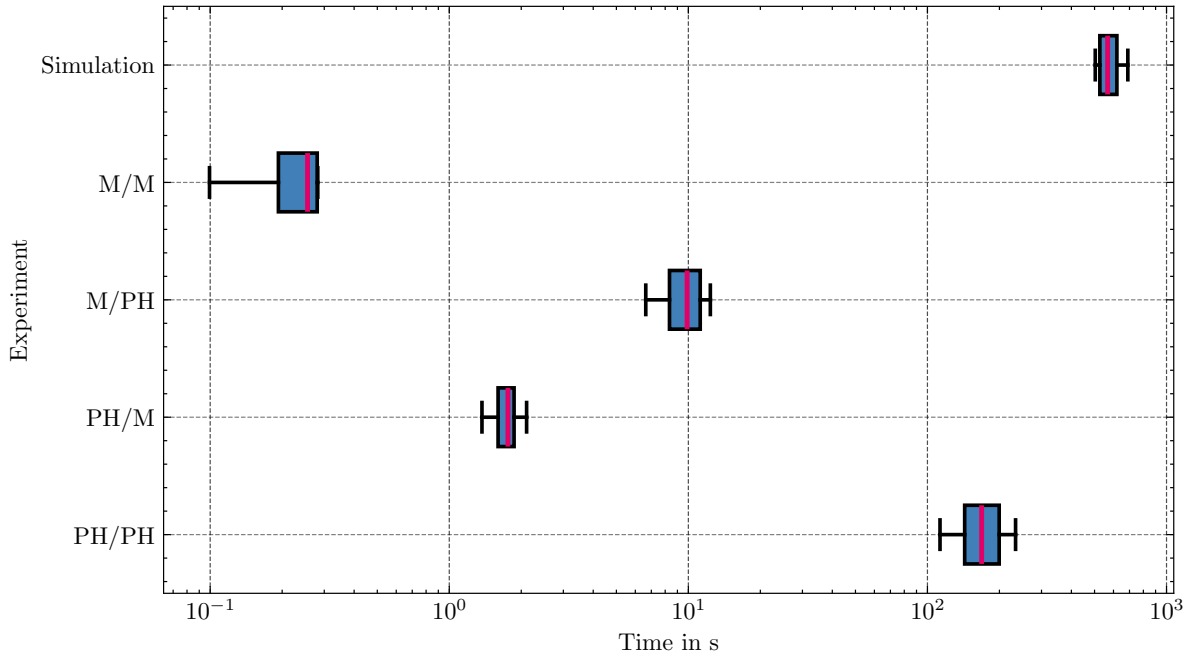


Figure 4.3.: Computation times of all considered models

All experiments have been conducted on an Intel Xeon Platinum 8160 Processor (2.1 GHz) and with a working memory of 32 GB. Additionally, 100 Simulations with 22 simulated hours each have been performed on the same processor with 4 GB of RAM for every considered train count n_{total}. To ensure the simulated system reaches a steady-state condition, characterized by minimal fluctuations after initial transient behaviours, the first and last hours of the 22-hour period were excluded from consideration.

When examining the computational times necessary to determine the queue lengths in the individual experiments, the size of the state-space becomes a significant factor. Figure 4.3 shows the distribution of the respective calculations for each model. Furthermore, Figure 4.3 includes the computational time required to serially simulate 2000 hours of traffic, which does not differ significantly between the assumed distributions of inter-arrival and service times. It is evident that the computation time increases substantially with the number of states.

Using the model for exponentially distributed inter-arrival and service times, approximately 0.22 seconds are required to compute the expected queue length on average. In contrast, computing the expected queue length using the model with arrival and service processes following a hypoexponential distribution requires over 140 sec-

onds. However, all analytical methods can calculate the queue length faster than a simulation of 2000 hours of traffic, which takes a mean time of over 550 seconds. The computation time for the simulation could be significantly reduced by employing parallel simulations.

4.5.2. Queue-Length Approximation Quality

In addition to the computation times of the different models, the approximated queue-lengths can be compared. For this, an exemplary railway junction with 4 routes (see Figure 4.1) has been chosen. For this example, we assume passenger traffic only, hence resulting in a quality threshold (see Section 4.4.2) of $L_{limit} = 0.479 \cdot \exp(-1.3) = 0.13$. The service rates have been assumed to have artificial values of $\mu_r = 0.3$ for every route $r \in R$, corresponding to a mean minimum headway time of $\bar{b} = 1/0.3 \approx 3.33$ min. To highlight the dependencies of the results on the occupancy rate ρ, the total number of trains in the infrastructure per hour $n_{\text{total}} = \sum_{r \in R} n_r$ has been varied from $n_{\text{total}} \in \{4, 8, \ldots, 40\}$ by setting the arrival rate $\lambda_r = n_r/60$ for every route $r \in R$. For a fixed traffic distribution of $p_{\text{main}} = 0.5$, the number of trains per route n_r is given by $n_r = n_{\text{total}}/4$ for every route. The arrival-rate per route λ_r has hence been set to values between $1/60$ and $10/60$.

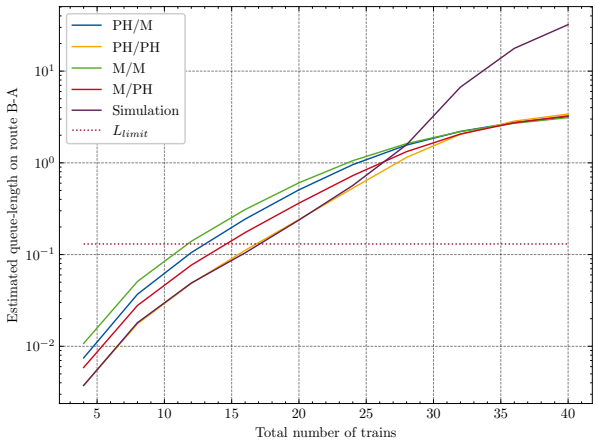
The expected queue length L_r for each route r has been calculated for every experimental distribution setting. Furthermore, for the models with exponential arrival and/or service processes, the approximation formulas by Hertel (Section 4.4.3.1) and Kingman (Section 4.4.3.2) have been applied to approximate the expected queue-length of the general independent system $L_r(GI/GI)$. In their formulas, the coefficients of variation v_A and v_B have been set according to the used distribution setting: If a coefficient of variation $v_X \neq 1$ has been used in the model, the corresponding coefficient of variation has been set to $v_X = 1$ in the approximation formula. Otherwise a coefficient of variation of $v_A = 0.8$ has been used for the arrival process and of $v_S = 0.3$ for the service process. Table 4.4 lists the values in the approximation formulas.

To analyse the approximation quality of a single route, a fixed traffic distribution of $p_{\text{main}} = 0.5$ has been set for varying n_{total}. Selecting route r_3 from B to A, Figure 4.4 shows the estimated values for the queue-length at this route and compares them to the mean of the simulation. The values have been calculated with the specified model and either not scaled at all (Figure 4.4a) or approximated with the Hertel (Figure 4.4b)

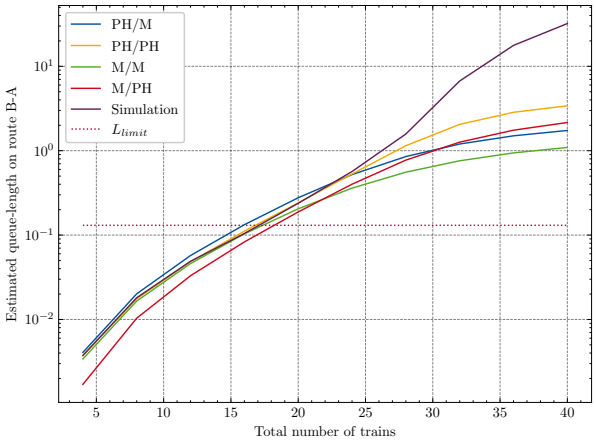
Table 4.4.: Parameters of the approximation settings

Arrival process	Service process	approx. formula v_A	approx. formula v_S	service rate per route μ_r	total train count n_{total}	quality threshold L_{limit}	share of main traffic p_{main}
M	M	0.8	0.3	0.3	$4, \ldots, 40$	0.13	0.5
Ph	M	1	0.3	0.3	$4, \ldots, 40$	0.13	0.5
M	Ph	0.8	1	0.3	$4, \ldots, 40$	0.13	0.5
Ph	Ph	1	1	0.3	$4, \ldots, 40$	0.13	0.5

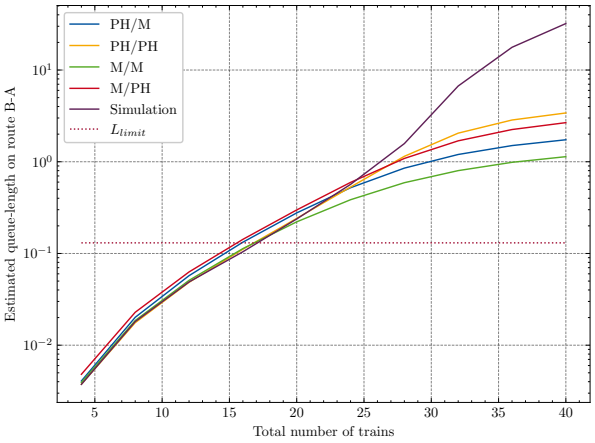
or Kingman (Figure 4.4c) method. In this example, all four routes $r \in R$ are assumed to have the same number of trains $n_r = n_{total}/4$ in an hour.



(a) Without Scaling Factor



(b) Hertel Approximation



(c) Kingman Approximation

Figure 4.4.: Comparison of the queue-length estimations at route r_3 for different arrival- and service-process distributions with and without scaling factors

In all three Figures 4.4a - 4.4c, the graphs of the PH/PH model and the simulation are almost congruent for a train count n_{total} below 23. At higher train numbers, the difference between the two calculates queue-lengths increases. This can be explained with the limitations in the maximum queue-length, or the number of waiting slots m in the queueing system, which has been set to $m = 5$ in the analytical experiments, but remained unlimited for the simulation setup. The probability of more than 5 trains having to wait for their service increases with an increasing number of trains. Consequently, the analytical models, which omit the state probabilities with more than 5 simultaneous waiting trains, underestimate the total queue-length. However, since the quality threshold is defined for a limit of $L_{limit} = 0.13$, the PH/PH model gives an adequate estimation in the neighborhood of the limit and can therefore be used without any scaling functions to evaluate the performance of the example railway junction.

For the graph (Figure 4.4a) without any applied scaling, it is clearly evident that there are significant differences between the simulation method and the models with exponential arrival and/or service processes. In these three experiments, only $12 - 14.5$ trains per hour would be permitted to meet the given quality threshold specified quality threshold, whereas both the simulation method and the PH/PH model would set this limit at approximately 17 trains per hour. Additionally, the M/PH model demonstrates greater alignment with the simulation and PH/PH model compared to the PH/M model, which still exhibits more accuracy than the basic M/M model. These discrepancies can be attributed to variations in the coefficients of variation v_A and v_S from 1 — the M/PH model accurately represents the service process (characterized by a smaller coefficient of variation), while the PH/M model provides a more precise description of the arrival process (with v_A closer to 1).

However, when an approximation method is applied, the M/M model closely aligns with the simulation and the PH/PH method. In the case of the Hertel approximation (Figure 4.4b), significant discrepancies become evident for a total train count of $n_{total} \geq 17$. Similarly, for the Kingman approximation (Figure 4.4c), differences become apparent for $n_{total} \geq 19$.

The models for the PH/M and M/PH settings consistently overestimate or underestimate the simulation results within the relevant region where the total train count is $n_{total} \leq 22$. In the case of the Hertel approximation, the PH/M model yields an overestimation while the M/PH model yields an underestimation. Alternatively, for the Kingman approximation, both models produce overestimations, albeit more closely

aligned with the simulation results.

4.5.3. Performance Determination Quality

While the results for the raw queue-length estimations have been analyzed in the previous section, this section focuses on the application of the introduced methods for performance assessment. In certain regions, such as those far from the quality threshold, a small deviation from the correct result in the analyzed parameter may not significantly impact the overall approximation quality of performance indicators.

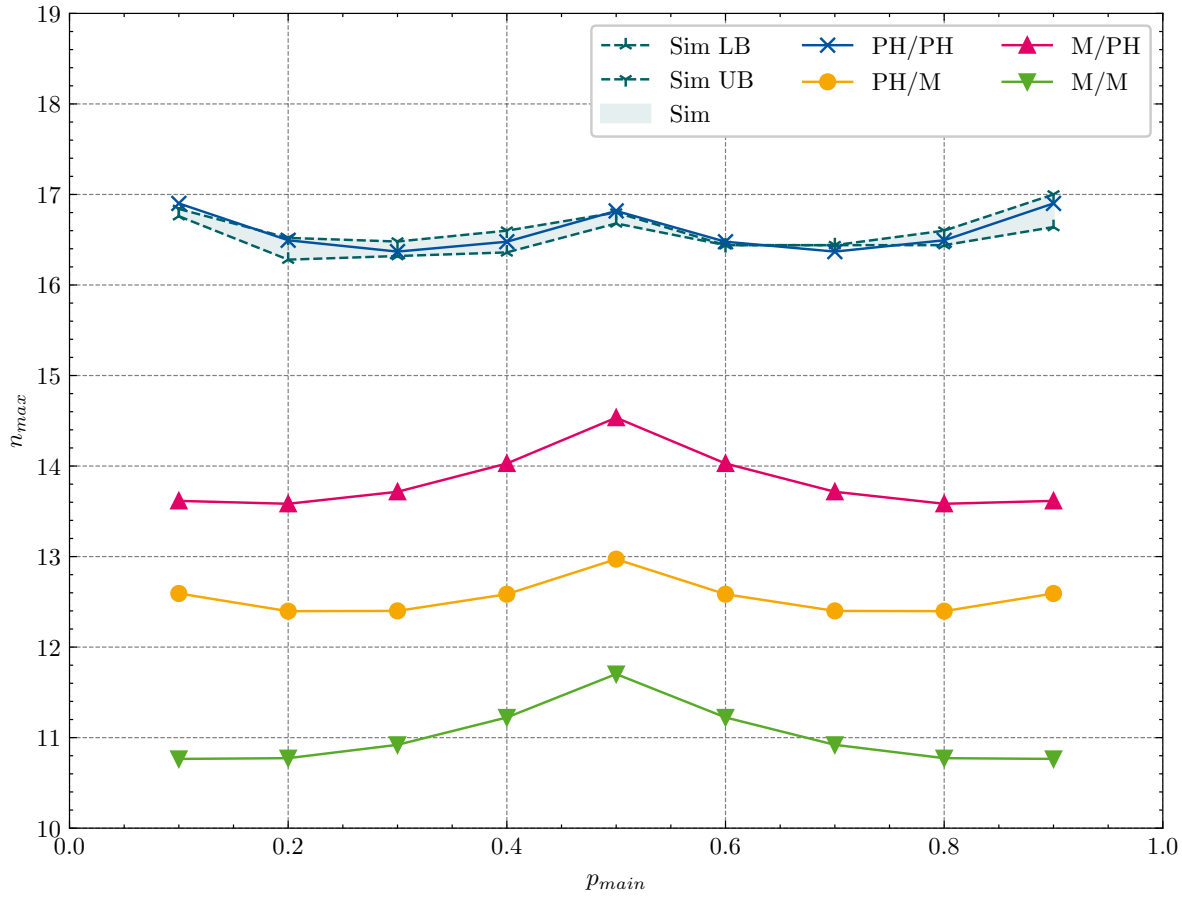
Therefore, we apply the capacity determination algorithm from Section 4.4.6 for all four analytical settings and both introduced approximation methodologies (see Section 4.4.3), the parameter settings can be found in Tables 4.3 and 4.4. Table 4.5 summarizes the results for the different calculation scenarios.

Table 4.5.: Selected indicators for the different methods

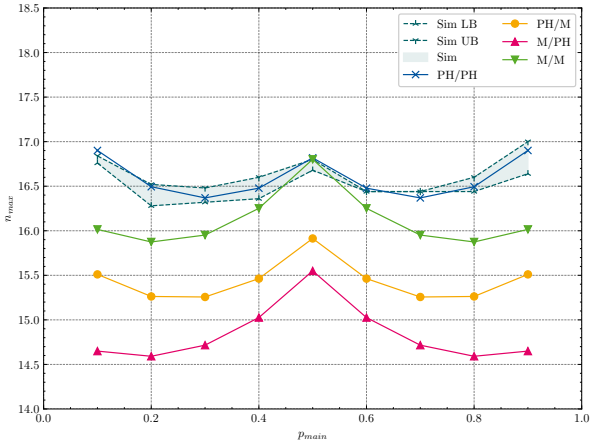
Approximation formula	Arrival process	Service process	Mean number of iterations	Mean total computation time	Maximum capacity $\hat{n}_{max}$	Traffic share at max. capacity
None	M	M	9	2.22 s	11.70	0.5
None	PH	M	9	17.04 s	12.97	0.5
None	M	PH	8.44	78.86 s	14.53	0.5
None	PH	PH	9	1432.20 s	16.90	0.1
Kingman	M	M	8.89	2.35 s	16.80	0.5
Kingman	PH	M	8	15.76 s	15.91	0.5
Kingman	M	PH	7.78	74.65 s	15.55	0.5
Hertel	M	M	9.33	2.31 s	17.29	0.5
Hertel	PH	M	8	15.70 s	15.91	0.5
Hertel	M	PH	9.22	87.33 s	18.17	0.5

The algorithm from Section 4.4.6 assumes monotonicity for the queue-length function $L_r : P_{\text{main}} \times N_{\text{total}} \to \mathbb{R}$, which cannot be guaranteed for the non"=deterministic simulations. Therefore, 100 simulations, each lasting 20 hours, are conducted for every combination of $p_{\text{main}} \in P_{\text{main}} = \{0.1, \ldots, 0.9\}$ and $n_{\text{total}} \in N_{\text{total}} = \{12.00, 12.04, \ldots, 19.92, 19.96\}$. From this, the bounds $n^*_{\text{max, LB}}$ and $n^*_{\text{max, UB}}$ can be identified to limit the optimal train-count n_{max} in the simulation that meets the quality thresholds (see Appendix B.2).

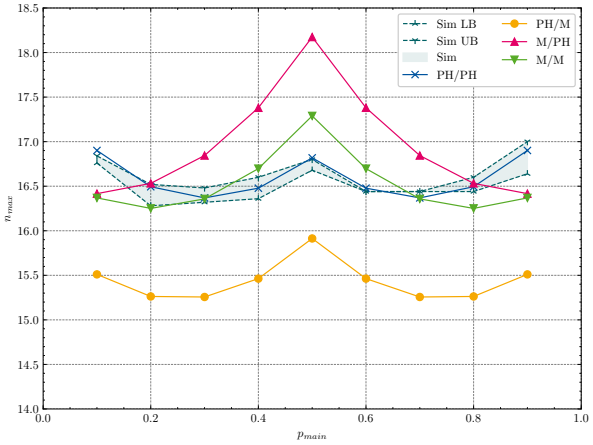
In Figure 4.5, the results of the performance determinations are depicted for the different methods.



(a) Without Scaling



(b) Kingman Approximation



(c) Hertel Approximation

Figure 4.5.: Comparison of the performance determination by the share of main line traffic p_{main} for the different methods

When comparing the simulation results to the PH/PH method (Figure 4.5a), some minor deviations are apparent. For example, the analytical solution exhibits a clear symmetry around the axis $p_{main} = 0.5$, which is only qualitatively reflected in the simulation results. This demonstrates that even the averaged simulation results represent only a subset of all possible outcomes, whereas the analytical method performs the capacity determination deterministically.

Overall, the simulation and PH/PH results are similar in terms of capacity n_{max} and exhibit the same general pattern: one local maximum for homogeneous traffic and two local maxima at the extremes, where nearly all traffic is concentrated on either the main or branch line.

For the raw analytical methods without any applied scaling factor (Figure 4.5a), the M/PH, PH/M, and M/M methods significantly underestimate the available capacity and fail to exhibit the same general pattern of two maxima at the extremes.

Applying the Kingman approximation (Figure 4.5b) significantly reduces the gap between the results of the M/M model and the PH/PH model, particularly for the homogeneous traffic scenario at $p_{main} = 0.5$. However, substantial differences in the calculated capacities persist in other traffic distributions. Notably, the order of approximation accuracy changes with the Kingman approximation: the M/M model offers the best approximation, followed by the PH/M model, and lastly the M/PH model.

Using the Hertel formula (Figure 4.5c), the M/M model also provides the best fit to the PH/PH and simulation results. Although the M/M model still fails to accurately capture the capacity increase for highly heterogeneous traffic distributions, its overall estimates are quite close, with differences of less than 0.5 trains per hour at the local extrema of the PH/PH results. The M/PH model overestimates capacity by up to 1.5 trains per hour in the homogeneous scenario, intersects the PH/PH model near the boundaries, and underestimates performance at both the lowest and highest traffic shares. In contrast, the PH/M model shows the exact same poor approximation quality as with the Kingman approximation, with a clear underestimation of capacity by at least 1 to 1.5 trains per hour.

In summary, the PH/PH model delivers the most accurate results, although it requires the highest computational time. On the other hand, the Hertel approximation enables the M/M model to achieve a good fit for homogeneous traffic scenarios in a fraction of the time. The use of PH/M or M/PH models is generally not advisable, as these

models are more inaccurate after applying an approximation formula, while also demanding more computational time. Depending on the application scenario, engineers may opt for the M/M model combined with the Hertel approximation for preliminary and rapid planning stages, while reserving the PH/PH model for more detailed analysis of junction infrastructure in selected long-term projects.

4.6. Case Study

In this Section, we consider a double track railway junction, where a mixed-traffic railway line is divided into a freight line from and to a freight yard and a line for passenger traffic only (see Figure 4.6).

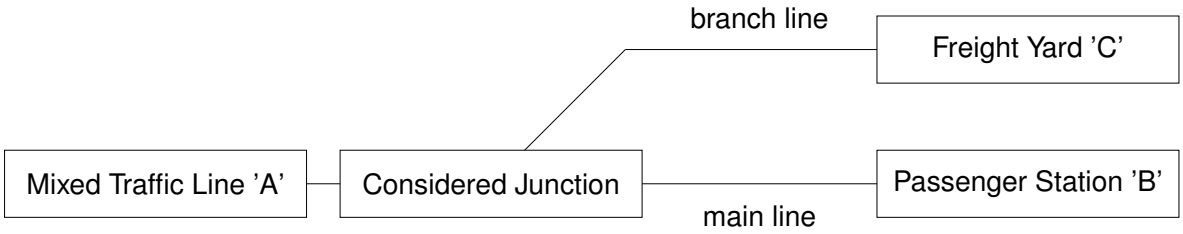


Figure 4.6.: Traffic scenario of the case study

Consistent with the introduced notation, the line to the passenger station 'B' is called *main line* and the line to the freight yard 'C' *branch line*. For the railway junction infrastructure itself, we consider the same infrastructure as in Figure 4.1, Figure 4.7 redraws it for convenience.

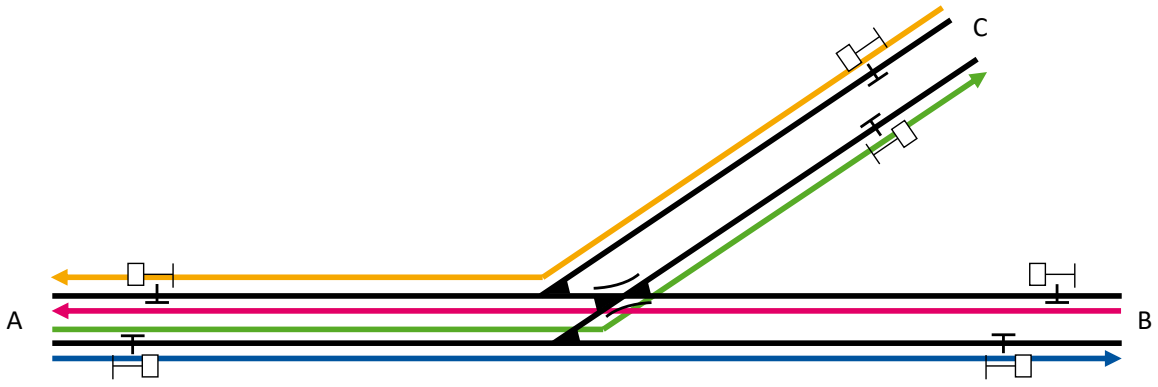


Figure 4.7.: Infrastructure of the considered railway junction

4.6.1. Setup

The traffic on this infrastructure can be described with four different train types: *Suburban* (s) and *regional* (r) trains cover passenger traffic, while cargo trains can be distinguished between *long-distance freight* (lf) and *regional freight* (rf) trains. Depending on their origin and destination, only certain routes are feasible for these types of trains (see Table 4.6).

Table 4.6.: Traffic in the case study

Train types	Route	Origin	Destination	Conflicts
suburban train, regional train	r_1	A	B	r_2
long-dist. freight train, regional freight train	r_2	A	C	r_1, r_3
suburban train, regional train	r_3	B	A	r_2, r_4
long-dist. freight train, regional freight train	r_4	C	A	r_3

To assess the junction's performance, minimum headway times of the different routes have to be considered. Usually, an infrastructure manager would utilise microscopic infrastructure data to calculate minimum headway times with their corresponding toolset. However, for this made-up example, no microscopic data exists, therefore minimum headway times have been estimated in Table 4.7.

Table 4.7.: Assumed minimum headway times $h_{i,j}$ in minutes for the considered junction

$(r_i, t_i) \setminus (r_j, t_j)$	(r_1, s)	(r_1, r)	(r_3, s)	(r_3, r)	(r_2, lf)	(r_2, rf)	(r_4, lf)	(r_4, rf)
(r_1, s)	2.5	5.5			5	5		
(r_1, r)	3	2			3	3		
(r_3, s)			2.5	5.5	1.5	1.5	5	5
(r_3, r)			3	2	1.5	1.5	3	3
(r_2, lf)	3.5	5.5	3	2	3	6		
(r_2, rf)	8	8.5	2.5	2.5	7	4		
(r_4, lf)			3.5	5.5			3	6
(r_4, rf)			8	8.5			7	4

Furthermore, the number of trains per route-train combination has to be specified to calculate the arrival rates in the Markov Chain model from Section 4.4.4. Together

with the service rates, calculated from the weighted minimum headway times, queue-lengths can be estimated for every route $r \in R$. Note that no priorities have been assigned to the different routes and train types, in correspondence with the definition of choice transition rates in Section 5.2.4.2.

In addition to the total number of trains n_{total}, the volume of traffic per combination of route-train $n_{(r,t)} \in \mathbb{R}$ depends on the share of main line traffic $p_{\text{main}} \in [0, 1]$, as well as on the share of suburban trains $p_{\text{suburban}} \in [0, 1]$ and regional freight trains $p_{\text{regional freight}} \in [0, 1]$. For the remainder of this case study, we assume a fixed rolling stock distribution of $p_{\text{suburban}} = 0.5$ for passenger and of $p_{\text{regional freight}} = 0.5$ for freight trains.

Therefore, we can formulate the functions $\theta_{(r,t)} : \mathbb{R} \to \mathbb{R}, n_{\text{total}} \mapsto n_{r,t}$ describing the number of trains per combination for every total number of trains.

To analyse the performance capabilities of this railway junction, the introduced CTMC model for phase-type inter-arrival and service time distributions can be formulated (Section 4.4.4). Since the model's size depends on the number of phases in the service process, the variance coefficient for the service process per route $v_{S,r}$ must be calculated.

Table 4.8.: Service Rates, Variance Coefficients and size of the CTMC Model (PH/PH) at different p_{main} Values

p_{main}	μ_1	μ_2	μ_3	μ_4	$v_{S,1}$	$v_{S,2}$	$v_{S,3}$	$v_{S,4}$	Number of states	Number of transitions
0.1	0.25	0.20	0.36	0.19	0.27	0.36	0.52	0.33	5.36×10^6	3.05×10^7
0.2	0.26	0.21	0.36	0.19	0.29	0.40	0.51	0.33	4.32×10^6	2.46×10^7
0.3	0.26	0.21	0.35	0.18	0.31	0.43	0.51	0.34	3.91×10^6	2.23×10^7
0.4	0.27	0.21	0.35	0.18	0.33	0.45	0.50	0.34	3.70×10^6	2.11×10^7
0.5	0.28	0.22	0.34	0.18	0.34	0.47	0.49	0.34	3.44×10^6	1.96×10^7
0.6	0.28	0.22	0.34	0.17	0.36	0.49	0.48	0.34	3.17×10^6	1.81×10^7
0.7	0.29	0.22	0.33	0.17	0.37	0.51	0.47	0.33	3.01×10^6	1.72×10^7
0.8	0.29	0.22	0.32	0.16	0.39	0.52	0.45	0.33	2.95×10^6	1.68×10^7
0.9	0.30	0.22	0.32	0.16	0.40	0.53	0.44	0.32	3.09×10^6	1.76×10^7

Table 4.8 summarizes the calculated service rates μ_r, variance coefficients $v_{S,r}$ and model size for the analysed traffic distributions of $p_{\text{main}} \in \{0.1, \ldots, 0.9\}$. Additionally, the variance coefficient for the arrival process has been set to $v_A = 0.8$.

The results of those calculations are described and analysed in the following Section 4.6.2.

4.6.2. Timetable Capacity Results

Using the algorithm described in Section 4.4.6, the performance capabilities of the railway junction with fixed $p_{\text{main}}, p_{\text{suburban}}, p_{\text{regional freight}}$ traffic shares can be estimated. For this, a lower bound of $n_{\text{total, LB}} = 1$ and an upper bound of $n_{\text{total, UB}} = 40$ has been chosen.

Figure 4.8 highlights the convergence of the method for a the described example with a main line traffic share of $p_{\text{main}} = 0.5$.

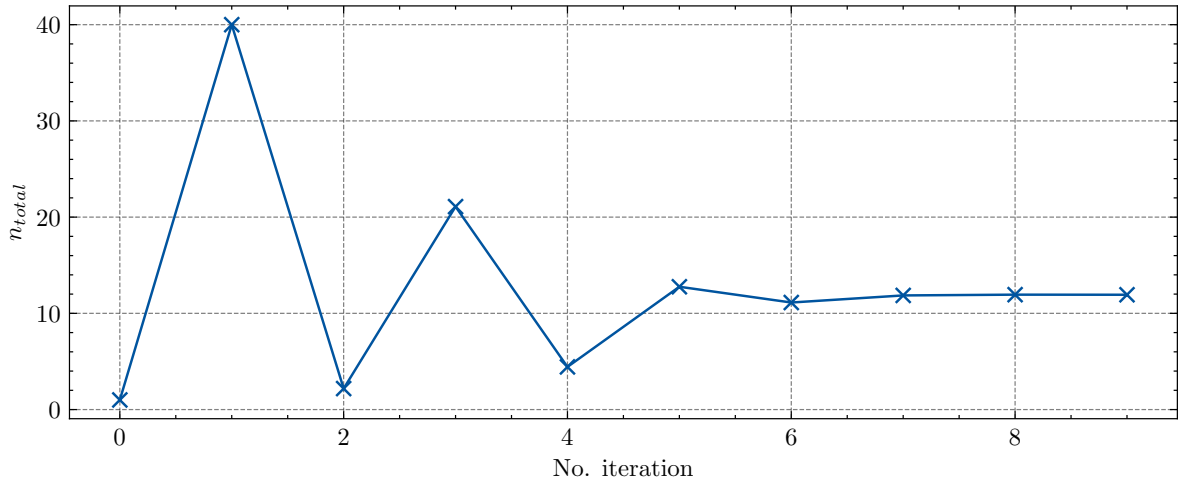
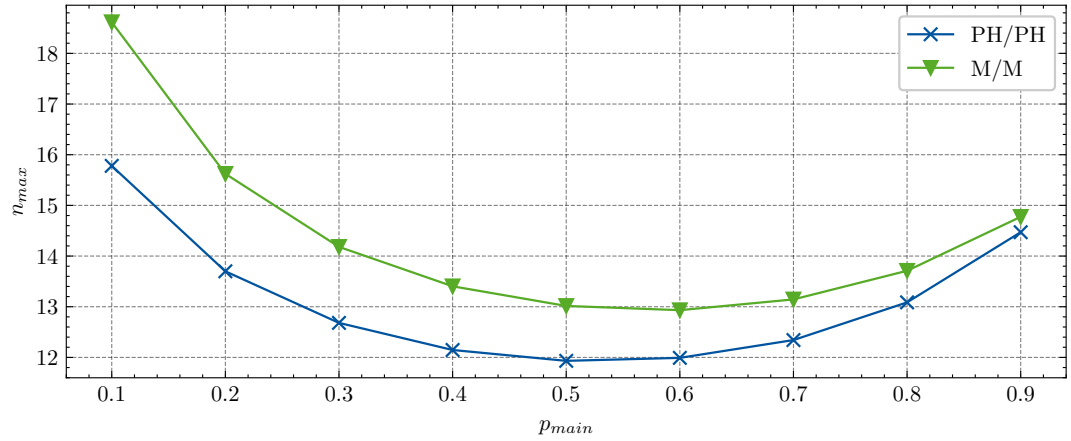


Figure 4.8.: The number of trains n_{total} in the junction with a traffic split of $p_{\text{main}} = 0.5$ for all iterations

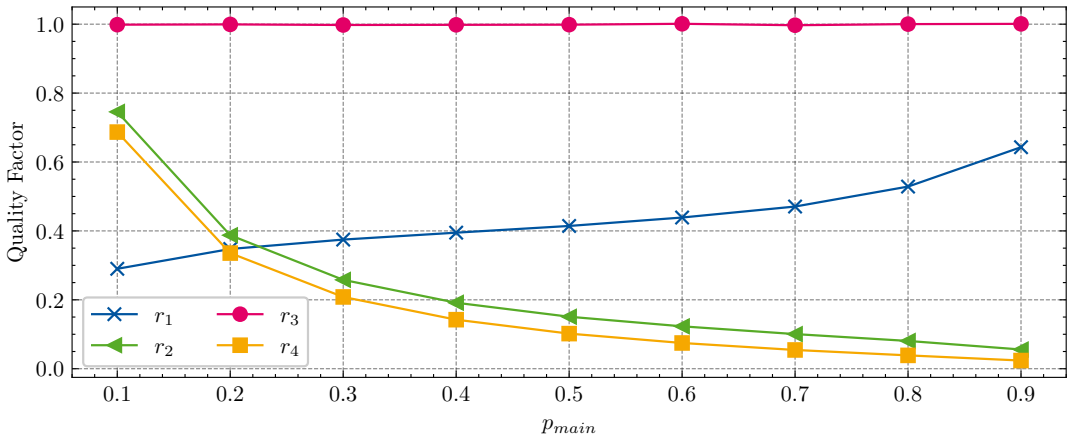
The figure shows the continuous convergence of the capacity determination algorithm for a growing number of iterations. While the given limits of 1 and 40 are explored in iteration 0 and 1, the algorithm terminates after 10 calls to the model-checking, respectively after 9 iterations at a value of $n_{\text{total}} = n_{\text{max}} = 11.93$, corresponding to the timetable capacity for the specified case.

The timetable capacity n_{max} for other main traffic shares $p_{\text{main}} \in \{0.1, \ldots, 0.9\}$ has been calculated similarly. Figure 4.9a presents the results for both the PH/PH model and the M/M model, with the latter further scaled using the Hertel approximation (see Section 4.4.3.1). Information about the computation times can be obtained from Appendix B.3.

In the PH/PH model, the highest timetable capacities are determined for scenarios involving asymmetric traffic, where movement predominantly occurs on either the branching or the main line. The maximum number of trains per hour is calculated



(a) Timetable Capacity



(b) Quality Factor

Figure 4.9.: Timetable capacity n_{max} of the two models PH/PH and M/M and quality factor per route for varying p_{main}

as $n_{\text{max}} = 15.78$ for a main track probability of $p_{\text{main}} = 0.1$, and $n_{\text{max}} = 14.47$ trains per hour for $p_{\text{main}} = 0.9$.

Compared to the PH/PH model, the M/M model significantly overestimates timetable capacity. While the difference is marginal for scenarios with a high main traffic (and therefore passenger traffic) share, it increases to a full train at $p_{\text{main}} = 0.5$ and exceeds 2.5 trains at $p_{\text{main}} = 0.1$.

Furthermore, Figure 4.9b depicts the quality factor for each route, which can be used to assess the infrastructure's bottleneck. In the given example, route r_3 consistently has a quality factor of 1 for every p_{main}, indicating that this route is the bottleneck for every described traffic scenarios.

This can be explained by the lower capacity limit for passenger traffic compared to freight traffic, and by the fact that route r_3 conflicts with both branching routes r_2 and r_4, which have high service times due to their exclusive utilization for freight traffic.

The influence of the traffic share on the other routes' quality is as expected: the freight traffic routes r_2 and r_4 improve their quality factors as traffic on these routes decreases, while route r_1 sees its quality factor increase inversely.

In summary, the analysis reveals that the introduced PH/PH model provides a more accurate estimation of timetable capacity. Its route-based application can be used to identify the bottleneck in the infrastructure for different traffic scenarios.

4.7. Discussion and Conclusion

In this work, a novel model for performance estimation of railway junctions has been introduced, utilizing Continuous-Time Markov chains with various distribution functions for inter-arrival and service times. The study examines the impact of using phase-type distributions compared to exponential distributions on the approximation accuracy of the models, with results compared against simulations.

Four different models, each describing distinct distributions for inter-arrival and service times, were analyzed for their approximation quality and computation time. Special attention was given to the differences in approximation quality for various traffic distributions between the two intersecting railway lines in the analyzed railway junction. Additionally, a case study was conducted to investigate the timetable capacity for a junction that splits a mixed traffic line into freight and passenger lines.

While the introduced model generally produces more accurate results, its computation times are significantly longer compared to models that utilize only exponential distributions. This is due to the use of phase-type distributions, which introduce additional states into the CTMC. As a result, the number of states becomes large even for relatively simple junctions. Consequently, the scalability of the introduced method is highly limited. To apply this approach to more complex railway junctions or even entire railway stations, further research is needed on decomposition techniques and alternative approximation methods. If such techniques allow for efficient sub-calculations of complex infrastructure without adding significant approximation errors, they would enable the use of models with phase-type distributions to benefit from their increased

accuracy. However, given the current computational demands, it may initially be more practical to explore more complex systems using purely exponential arrival and service process models and enhancing their results with approximation formulas.

A major parameter influencing both the model size and the timetable capacity of railway junctions is the variation coefficient of the service process. In some chapters of this work (see Section 4.5), a default value was used, while in the case study, it was calculated based on minimum headway times. For efficient use of the introduced method, it is essential to obtain an accurate estimation of this value to ensure the generation of reliable and usable results.

To assess the performance of the railway junction, queue lengths computed with CTMC models are compared to thresholds in the capacity determination algorithm (see Section 4.4.6). Although these thresholds are commonly used in practice, they were originally designed for line capacities. Therefore, new thresholds may need to be developed to appropriately assess the quality of multi-channel systems.

Future work may also explore techniques to incorporate priorities into the formulation. With the current route-based decomposition, priorities between routes could potentially be included by adjusting the rates of choice transitions. To incorporate priorities between different train types while maintaining the route-based decomposition, alternative techniques may be required, such as manipulating service times using priority coefficients, as demonstrated in the works of Schwanhäußer (1978) and Wakob (1981).

The proposed method, despite its longer computation times, allows for the substitution of approximation techniques (Section 4.4.3) in application scenarios that demand high accuracy. The introduced method demonstrates significantly shorter computation times compared to simulation approaches while producing deterministic results, making it more practical than the numerous simulation iterations needed to achieve sufficiently representative outcomes. However, a parallel implementation of simulation iterations might outperform the computation time of the introduced method.

The proposed model, which explicitly models phase-type distributions within the Continuous-Time Markov chain, can be utilized by infrastructure managers to obtain precise, timetable-independent estimations of the capacity of railway junctions. This enables a reduction in the reliance on computationally expensive simulations in long-term railway planning scenarios.

5. A Queueing-Based Approach for Timetable-Independent Railway Station Performance Analysis

Abstract

Railway stations serve as critical nodes within railway networks, facilitating connections across diverse travel directions. Traditionally, the analytical performance analysis of railway stations has been divided into two distinct components: the examination of stopping tracks and the evaluation of route nodes, the locations within a station where switches determine the direction of travel. This study introduces an innovative Continuous-Time Markov Chain model that represents a comprehensive queueing system encompassing the entire railway station. By deriving timetable-independent performance indicators, this model provides a robust framework for assessing station performance. Consequently, it equips infrastructure operators with a holistic tool for infrastructure planning and evaluation.

5.1. Introduction

With the rising demand for railway traffic, it is imperative that existing infrastructure undergo continuous evaluation and that large-scale infrastructure projects are planned with a long-term perspective. Railway infrastructure is designed to last for multiple decades, enabling infrastructure managers to make informed decisions based on comparative quality assessments of planned constructions is crucial.

Railway stations serve as critical nodes in the performance evaluation of railway networks. However, timetable-independent performance analyses have typically been conducted only on individual segments of station infrastructure, specifically dividing them into *route nodes* and *track areas*. The former (also known as *switch area*) refers to the section of the station where train routes from entry/exit signals to platform tracks may overlap. In contrast, the latter pertains to the group of platforms where trains stop to perform operational tasks such as (un)loading freight or facilitating passenger ex-

change.

To measure performance capability of railway infrastructure, three main types of railway capacity (see also Jensen et al., 2020) can be distinguished: *theoretical capacity*, *timetable capacity*, and *operational capacity*. *Theoretical capacity* is defined as the maximum number of trains or train-route inquiries that can be scheduled without conflicts on a given infrastructure, taking into account the driving dynamics and installed railway control systems. *Timetable capacity* (occ. referred to as *maximal capacity*), denotes the highest number of requests that can be scheduled on the infrastructure during the timetabling process while maintaining an acceptable quality level compared to a specified threshold. This includes considerations of driving dynamics, railway control systems, and operating program specifics such as train-mix and arrival processes. *Operational capacity* (occ. referred to as *practical capacity*, Abril et al., 2008) pertains to the maximum number of trains that can traverse the infrastructure with acceptable operational quality relative to a specified threshold. This definition encompasses all factors considered in timetable capacity while additionally accounting for the probability of disturbances and knock-on delays.

Various methodologies have been employed to analyse the capacity of railway infrastructure. The UIC (2013) introduces the *capacity occupation* metric, describing the share of actually utilized blocking times in a time horizon, by compressing the existing blocking-time staircases of a timetable (Bešinović and Goverde, 2018). Some other methods directly utilize operational data to describe quality during daily operations (Graffagnino, 2012; Armstrong and Preston, 2017; Weik, 2022; Corman and Henken, 2022). Others employ theoretical simulations (Abril et al., 2008; Schmidt and Martin, 2010; Liang et al., 2017; Zieger et al., 2018) or markov chains (Spanninger et al., 2023) to estimate operational capacity (Abril et al., 2008; Schmidt and Martin, 2010; Liang et al., 2017) or delay propagation effects (Zieger et al., 2018; Spanninger et al., 2023). Another approach involves using Mixed Integer Programming (MIP) to optimize train throughput within the infrastructure while considering different objective functions (Burdett and Kozan, 2006; Harrod, 2009; Mussone and Wolfler Calvo, 2013; Burdett, 2015; Liao et al., 2021). They aim to quantify infrastructure capacity by either utilizing an existing timetable or saturating it with additional train journeys.

Queueing-based methods provide timetable-independent metrics for assessing railway infrastructure performance. They typically describe the performance metric of timetable capacity, as introduced in Wendler (2007). These queueing-based meth-

ods can be further categorized based on their analysed infrastructure and solution techniques. Some approaches focus on the performance of railway lines (Schwanhäußer, 1974; Schwanhäußer and Schultze, 1982; Wendler, 2007; Weik and Nießen, 2017), while others examine route nodes (Schwanhäußer, 1978; Nießen, 2008, 2013; Schmitz et al., 2017; Weik, 2020; Emunds and Nießen, 2024a) or track groups (Potthoff, 1962; Hertel, 1986). Additionally, some studies analyse queueing networks to determine the capacity of freight railway stations (Bychkov et al., 2021) and railway networks (Huisman et al., 2002; Kazakov et al., 2023).

Method	Route node analysis	Track group analysis	Route-based decomposition	Multi-channel model	Solution method
(Potthoff, 1962)		✓		✓	Iterative formula
(Schwanhäußer, 1978)	✓		(✓)		Closed-form formula
(Hertel, 1986)		✓		✓	Closed-form formula
(Nießen, 2008, 2013)	✓		(✓)	✓	Iterative formula
(Schmitz et al., 2017)	✓			✓	Matrix-vector equations
(Weik, 2020)	✓		(✓)		Matrix-vector equations
(Bychkov et al., 2021)	(✓)	(✓)		(✓)	Simulation method
(Emunds and Nießen, 2024a)	✓		✓	✓	Probabilistic model-checking
Introduced here	✓	✓	✓	✓	Probabilistic model-checking

Table 5.1.: Methods for queueing-based railway station performance analysis

Table 5.1 provides a comparative analysis of various queueing-based methodologies for assessing railway station performance. It should be noted that any statement enclosed in parentheses indicates partial support for that feature. For route-based decomposition, this typically implies either the utilization of single-channel systems exclusively (Schwanhäußer, 1978; Weik, 2020) or a decomposition based on serial-route-nodes combined with an analysis of route-based quality metrics (Nießen, 2008, 2013).

The queueing network described by Bychkov et al. (2021) pertains to a freight railway station and, as such, has not been segmented into route nodes and track groups. Additionally, Bychkov et al. (2021) employ simulations to derive results from their formulated queueing network, whereas the other methods listed in Table 5.1 achieve exact

solutions through formulas, matrix-vector equations, or probabilistic model-checking.

The methodologies proposed by Potthoff (1962) and Hertel (1986) are specifically designed to determine the required number of platform tracks at a station. Other approaches focus on analysing the performance of route nodes; some use single-channel systems to approximate complex route nodes (Schwanhäußer, 1978; Weik, 2020), while others introduce multi-channel queueing systems that more directly account for parallel usability (Nießen, 2008, 2013; Schmitz et al., 2017; Emunds and Nießen, 2024a).

In this work, we introduce a novel Continuous-Time Markov Chain (CTMC) formulation that describes the system dynamics of a railway station. To the best of our knowledge, this is the first method, describing the timetable-independent performance of an entire railway station in one combined multi-channel queueing system. The main contributions of this study are:

- Presenting a novel exact queueing-based model for an entire railway station,
- Facilitating probabilistic model-checking and numerical root-finding methods to obtain timetable-independent performance metrics,
- Decomposing the queueing process by station routes and service processes to identify bottlenecks, and
- Applying the method to a case study from the perspective of an infrastructure manager, thereby aiding in decision-making between potential operating programs.

This work begins with the introduction of the formulated queueing system model in Section 5.2. Subsequently, the process for determining timetable capacity is described in Section 5.3. The case study presented in Section 5.4 demonstrates the practical applicability of the approach for infrastructure operators. A conclusion and discussion is provided in Section 5.5.

5.2. Queueing System Model of a Railway Station

In this section, we introduce the queueing system model of a railway station. First, we discuss queueing systems in general terms (Section 5.2.1). Next, we present an approximation formula relevant to our model (Section 5.2.2). Afterwards, we elucidate

our approach to partitioning railway stations within this framework (Section 5.2.3). Finally, we present a novel Continuous Time Markov Chain (CTMC) model for the queueing system of a railway station (Section 5.2.4).

5.2.1. General Discussion on Queueing Systems

A queueing system describes the arrival and service process of entities through a system. Since entities may arrive at arbitrary points in time, and their service may take an arbitrary amount of time, both the arrival and service processes are described using probability distributions. Furthermore, a queueing system also includes specifications regarding the number of waiting positions m, where arriving entities are 'parked' until they are serviced, and the number of channels capable of simultaneously servicing different entities (see also Kendall (1953)).

The arrival process of such a system can be described with the arrival rate λ and its coefficient of variation v_A, the service process with the service rate μ and its coefficient of variation v_S.

To evaluate the performance of such a system, different indicators can be distinguished. For example, the occupation ratio $\rho = \lambda/\mu$ provides an initial estimation regarding the stability of the system.

As a consequence of the arbitrary arrival and service processes, a queue may form in front of the first server. The average length of this queue, denoted as L, can be utilized to further analyse the system's performance. Furthermore, if the number of waiting positions m is finite, it might occur that the system is incapable of accommodating another arriving request because all waiting positions are already occupied. Consequently, incoming requests are lost, and this can be analysed through the probability of loss, denoted as p_{loss}.

While single-channel systems with exponentially distributed inter-arrival and service times (M/M) can be analysed using a closed-form solution, more advanced models incorporating multiple channels may require a system-specific analysis of the state probabilities within a Continuous-Time Markov Chain (CTMC) formulation. For certain railway infrastructures, such as railway lines (Schwanhäußer, 1974; Fischer and Hertel, 1990; Weik, 2020) and junctions (Schmitz et al., 2017; Emunds and Nießen, 2024a), corresponding formulations have already been analysed. However, in this work, we introduce the first CTMC formulation for railway stations (see Section 5.2.3).

To apply a formulation based on exponential arrival and service processes (M/M) to general independent processes (GI/GI), i.e., with non-unit coefficients of variation $v_A, v_S \neq 1$, an approximation formula can be used.

5.2.2. Approximation Formula

In this work, we use the approximation formula from Hertel (1986); Fischer and Hertel (1990),

$$L(M/M) \cdot \frac{1}{\gamma} \approx L(GI/GI), \tag{5.1}$$

with the parameters

$$\gamma = \frac{2}{c \cdot v_S^2 + v_A^2} \tag{5.2}$$

and

$$c = \left(\frac{\rho}{s}\right)^{1-v_A^2} \cdot (1 + v_A^2) - v_A^2, \tag{5.3}$$

where the scaling factor γ depends on the coefficients of variation, as well as the number of parallel service channels s (fixed at $s = 1$ for our formulation) and the occupation rate ρ.

The next section will introduce our approach for modelling a railway station.

5.2.3. Approach to Railway Station Partitioning

For the determination of the timetable capacity of a railway station, we decompose a station into its main parts, the switch and track area, and into the requested routes. An example of a railway station can be found in Figure 5.1.

It contains four different routes through a railway junction with four tracks. While route r_1 and r_4 enter via route node A and depart via route node B, route r_3 enters the station via route node B and leaves it via route node A. Notably, route r_2 only utilizes route node B since it corresponds to trains that turn on track 4.

To analyse the performance of such a railway station, we introduce multiple queues, one for every route. Therefore, the bottleneck of the station can be identified by comparing the estimated queue-lengths L_r per routes. Since every route corresponds to a queueing system, the parameters of those systems need to be formulated by the route. Assuming the operating programs, i.e. the number of trains per route n_r in

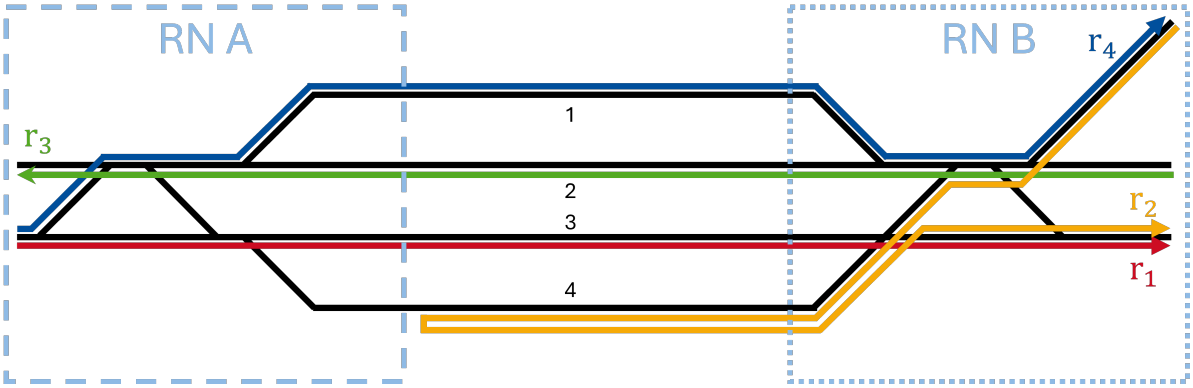


Figure 5.1.: An example for a railway station infrastructure

a specified time horizon U, are given, the arrival rate of the route is calculated as $\lambda_r = n_r/U$. Similarly, the coefficient of variation $v_{A,r}$ needs to be formulated for the arrival process on this route. For completely random arrival requests, an exponential distribution, with $v_{A,r} = 1.0$, could be assumed (see Wendler (1999, 2007)). However, since modern timetables continue to introduce more synchronised route occupation requests, we use a default value of $v_{A,r} = 0.8$ for every route in this work, in correspondence with best-practice recommendations (VIA Consulting & Development GmbH, 2021).

The model of the service process is significantly more complex, since the service of a route in a station infrastructure contains multiple parts. These consist of the *entry blocking*, *stopping* and *departure blocking* processes (this partitioning has already been used in Zwaneveld et al. (2001)). For every of those steps in a request's journey through the station infrastructure, different conflicts with other requests may arise. However, those conflicts are usually limited to one route node, or, in the case of entries, to one route node and one platform. Additionally, when modelling *passing*, i.e. not stopping, requests, a conflicts in both route nodes and one platform track might be also feasible. The separation between the two route nodes and the platform tracks is however still useful to get an understanding about the underlying blocking structure of the different service processes.

Figure 5.2 illustrates the dependencies between the various arrival and departure processes for the two routes, r_4 and r_3 , in the given example. It is important to note that only one queue per route is depicted, positioned just before the entry blocking process, representing the arrival until standstill at the platform track. Additionally, for each route, the different service processes — entry blocking, stopping, and departure

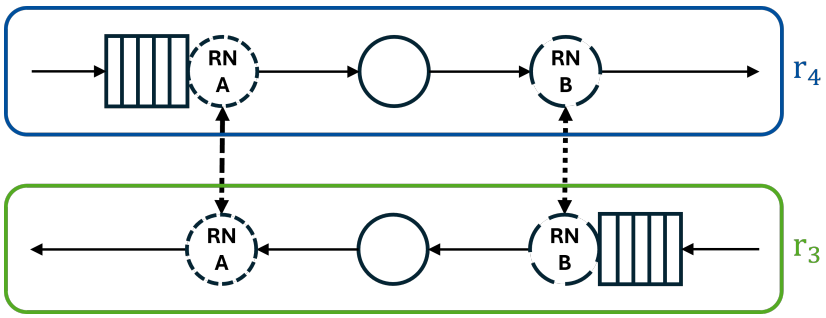


Figure 5.2.: Dependencies in the queueing systems for two routes in opposite directions.

blocking — are shown sequentially in this order.

Since r_4 starts at route node A and continues its journey to route node B, blocking both the entry of r_3 at route node B and its departure from route node A, a dependency between the corresponding service processes is depicted in the diagram.

All three service processes can be modeled with a corresponding service rate for each route: the service rate $\mu_{r,\text{entr}}$ for the entry blocking process, the service rate $\mu_{r,\text{stop}}$ for the stopping process, and the service rate $\mu_{r,\text{dep}}$ for the departure blocking process. Let $b_{r,\text{entr}}, b_{r,\text{dep}}, b_{r,\text{stop}}$ denote the average service times for the entry blocking, departure blocking, and stopping processes for route r, respectively. The corresponding service rates are then calculated as $\mu_{r,\text{entr}} = 1/b_{r,\text{entr}}$ for the entry blocking process and analogously for the departure blocking and stopping processes per route.

The required service times $b_{r,\text{entr}}, b_{r,\text{dep}}$ can be calculated by utilising any microscopic running time calculation tool, and enhancing it with the corresponding times for route-setting, reaction, approaching, clearing and route-release, forming the blocking-time (see f.e. Hansen and Pachl (2014)). To incorporate information about further exclusions in the line sections before or after the analysed stations, infrastructure operators can use the minimum headway times from or to the adjacent passing stations for all sequences of two route, rolling stock combinations sharing more than one signalling block, and averaging over all possible sequences with their relative frequency to obtain the average service time per route.

Since this work utilizes the Hertel formula (Section 5.2.2), all coefficients of variation for the three service processes, $v_{S,r,\text{entr}}, v_{S,r,\text{stop}}, v_{S,r,\text{dep}}$, need to be combined into a single coefficient of variation $v_{S,r}$ per route. This combined coefficient describes the total variation in all three service processes. We assume statistical independence of

the three service processes in order to calculate the combined coefficient of variation. Further details can be found in Appendix C.1. The variation coefficients of the arrival process v_A and of the combined service process v_S can then be used in the approximation formula to scale the estimated queue-length L_r per route.

Another challenge lies in addressing the mutual exclusions of the service processes for different routes. These can be described using the following formulation as a Continuous-Time Markov Chain.

5.2.4. Continuous-Time Markov Chain Formulation

A Continuous-Time Markov Chain (CTMC) $MC = (S, T)$ is described by a set of states S and a set of transitions T. By defining an initial state $s_0 \in S$ and transition rates for transitions $t = (s, s') \in T$, a *stationary distribution*, i.e., a probability distribution that remains unchanged as time progresses, can be calculated. This stationary distribution represents the long-term behavior of the system, where the probabilities of being in each state stabilize over time.

5.2.4.1. States

A state $s \in S$ needs to contain information regarding the current status $\pi_r \in \Pi_r$ in the service process of all routes and the current length of the queue q_r for all routes. Hence, the state set can be described as

$$S = \{(\pi_1, \mathsf{q}_1, \ldots, \pi_k, \mathsf{q}_k) \mid (\pi_r, \mathsf{q}_r) \in \Pi_r \times \{0, \ldots, m\}\}, \tag{5.4}$$

for a total number of k routes and the set of different service statuses Π_r.

The status $\pi_r \in \Pi_r = \{0, \ldots, 4\}$ of a route represents the progress through the station. A status of $\pi_r = 0$ corresponds to a route without any trains on it; after the arrival of a request, the status changes to $\pi_r = 1$. Furthermore, $\pi_r = 2$ indicates that the train is stopping, and $\pi_r = 4$ signifies the departure of the train. The remaining status, $\pi_r = 3$, is achieved only when a train has already completed its required stopping time but needs to wait until the departure route can be occupied.

Note that the status set of a passing route r' can be restricted to $\Pi_{r'} = \{0, 1, 4\}$ since the stopping status $\pi_{r'} = 2$ and consequently the waiting for departure status $\pi_{r'} = 3$

are omitted. This implies that both the arrival and departure routes of a passing train must be unoccupied for service to commence.

In the following, we extend the notation to $\pi_r(s)$, which describes the status of route r in state s, and $\mathsf{q}_r(s)$, which denotes the queue-length of route r in state s.

Furthermore, since the state set S corresponds to a composition of the status and queue-length settings for every route $r \in R$, certain combinations can be ruled out to further restrict the state set. For example, a state s cannot contain $\pi_{r_3} = 1$ and $\pi_{r_4} = 4$ if the entry blocking of route r_3 and the departure blocking of route r_4 are conflicting. In the implementation of this work, this process is automatically handled by the parsing algorithms of the model files describing the CTMC, by analysing the reachability of the different states as defined by the transitions T.

5.2.4.2. Transitions

A transition $t = (s, s') \in T$ between two states $s, s' \in S$ describes the transfer from state s to state s'. Its duration can be expressed with the transition rate, which depends on the type of the transition. The different types can include an *arrival* to a route, one of the various *service processes* for a route, or additionally, a *choice* of which route to serve next.

When representing the transitions in a figure, not all dimensions of the state space can be represented, in the following we therefore select two dimensions per figure. Furthermore we distinguish between invariant and conditional transitions to enhance the representations. For a fixed route r, *invariant* transitions (s, t) correspond to transitions that do not depend on parameters $\pi_{r'}(s), q_{r'}(s)$ of other routes $r' \neq r$. Conversely, *conditional* transitions do depend on other parameters.

An arrival transition $s \to s'$ utilizes the arrival rate λ_r of route r and defines

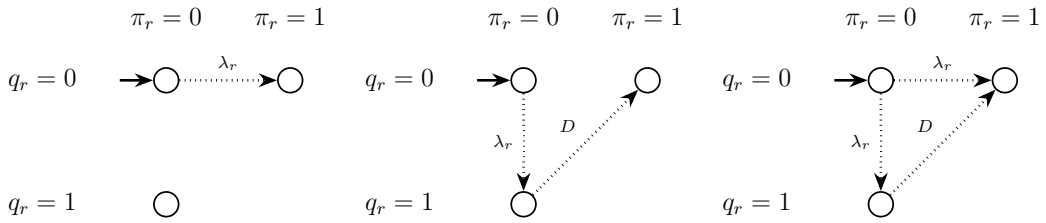
$$\pi_r(s') = 1, \tag{5.5}$$

if no conflicting route is currently in service in the corresponding route node. Otherwise, it defines

$$\mathsf{q}_r(s') = \mathsf{q}_r(s) + 1, \tag{5.6}$$

as long as $\mathsf{q}_r(s') \leq m_r$ holds and the arrival can therefore still be modeled.

Figure 5.3 gives a representation of both possible arrival transitions. Either the entry blocking service can directly start ($\pi_r(s') = 1$, Figure 5.3a) or it is blocked by service on another route, therefore staying in the same service phase ($\pi_r(s') = 0$) and adding the train to the queue ($q_r(s') = 1$), see Figure 5.3b. The arrival transition is therefore of the conditional transition type. In order to represent all possibilities of a conditional transition in a single figure, a combined representation, where every conditional transition might be taken from the origin state s, depending on the status $\pi_{r'}(s)$ of the service process in conflicting (and not necessarily visualised) routes r', can be used. An example of the discussed arrival transition is shown in Figure 5.3c.



(a) direct entry (b) blocked entry (c) combined

Figure 5.3.: The two possibilities of the arrival transition to an empty route ($q_r = 0$) and a combined representation.

For the service transitions, the entry blocking process is using the service rate $\mu_{r,\text{entr}}$ for the transition $s \to s'$ with $\pi_r(s) = 1$ and

$$\pi_r(s') = 2. \tag{5.7}$$

The stopping transition $s \to s'$ with rate $\mu_{r,\text{stop}}$ depends on conflicts between the departure blocking of route r and the arrival or departure blockings of other routes r' in state s, with $\pi_r(s) = 2$. If no conflicts are present, it maps the status of s' to

$$\pi_r(s') = 4. \tag{5.8}$$

Otherwise, it maps the status to

$$\pi_r(s') = 3. \tag{5.9}$$

The final service transition $s \to s'$ is of the kind departure blocking. It corresponds to the transfer between s with $\pi_r(s) = 4$, to s', fulfilling

$$\pi_r(s') = 0, \tag{5.10}$$

with a rate of $\mu_{r,\mathsf{dep}}$.

Notably, states s with $\pi_r(s) = 3$ do not have an outgoing service transition for the route r. This is due to their origin in a conflict with the departure blocking of route r. Hence, their departure service is initiated only after the initial conflicting service has been completed. Since other conflicting routes might also be waiting for their next service step, a choice transition is used to model a random decision on which route to serve next. Therefore, every state s with a route r, for which $\pi_r(s) = 3$ holds, has an outgoing choice transition to s' with

$$\pi_r(s') = 4, \tag{5.11}$$

using an artificial constant transition rate D. This rate is chosen sufficiently high so that the induced delay $1/D$ to the system is not significant.

Similarly, states s with $\pi_r(s) = 0$ and $\mathsf{q}_r(s) > 0$ also have an outgoing choice transition (with rate D) to a state s' with an activated entry blocking

$$\pi_r(s') = 1, \tag{5.12}$$

and hence one less item in the queue

$$\mathsf{q}_r(s') = \mathsf{q}_r(s) - 1 \tag{5.13}$$

for route r.

Figure 5.4 highlights π_{r_1} and π_{r_4} in an example describing the entry blocking service of two conflicting routes.

For the initial state s with $\pi_{r_1}(s) = 0$ and $\pi_{r_4}(s) = 0$, either the transition (s, t), with $\pi_{r_1}(t) = 1$ and $\pi_{r_4}(t) = 0$, or the transition (s, t'), with $\pi_{r_1}(t') = 0$ and $\pi_{r_4}(t') = 1$, is selected first.

Since all choice transitions use the same constant transition rate D, every possible choice in such a state is used with equal probability, and hence independent on the sequence of preceding events.

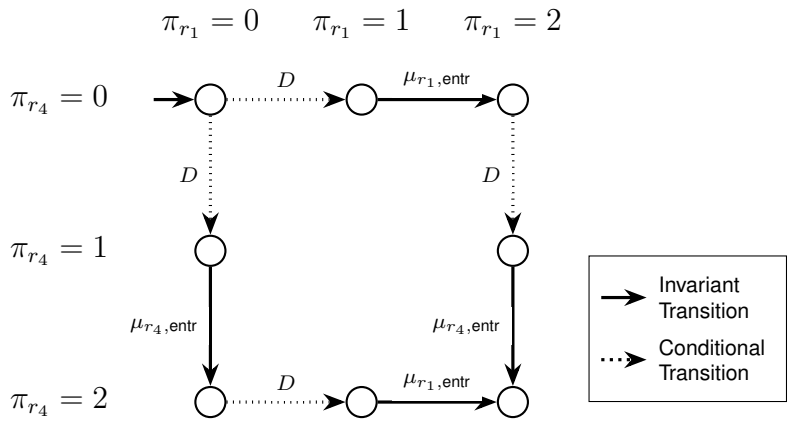


Figure 5.4.: An example for the choice transitions between the entry processes of two conflicting routes r_1 and r_4, assuming non-empty queues ($q_{r_1}, q_{r_4} > 0$) for both routes.

5.2.4.3. Example

All transitions share the characteristic that only one route is affected by each transition. For the arrival transitions, only one route gains another item in its system. Similarly, for the service transitions, the various service operations are carried out on a single specified route. For the choice transitions, either the commencement of the entry blocking or departure blocking is initiated for a single route.

The different states and transitions can therefore be easily visualized when limiting the visualization to a fixed route r. Figure 5.5 thus shows the dependencies between the states of a single route.

For an initial state s with $\pi_r(s) = 0$ the arrival transition (s, t) is an example of a conditional transition, as further explained in Figure 5.3. However, for states s' with any other current service state $\pi_r(s') > 0$, the arrival transition (s', t') to route r always adds to the corresponding queue $q_r(t') = q_r(s') + 1$.

Another conditional transition (s, t) is that of the stopping process. Depending on whether or not the departure blocking process is in conflict with any other routes, the transition's destination state t either fulfills $\pi_r(t) = 3$ (conflict) or $\pi_r(t) = 4$ (no conflict).

Furthermore, both the departure blocking process and correspondingly, the entry blocking process are invariant to other states. They simply connect states s and t

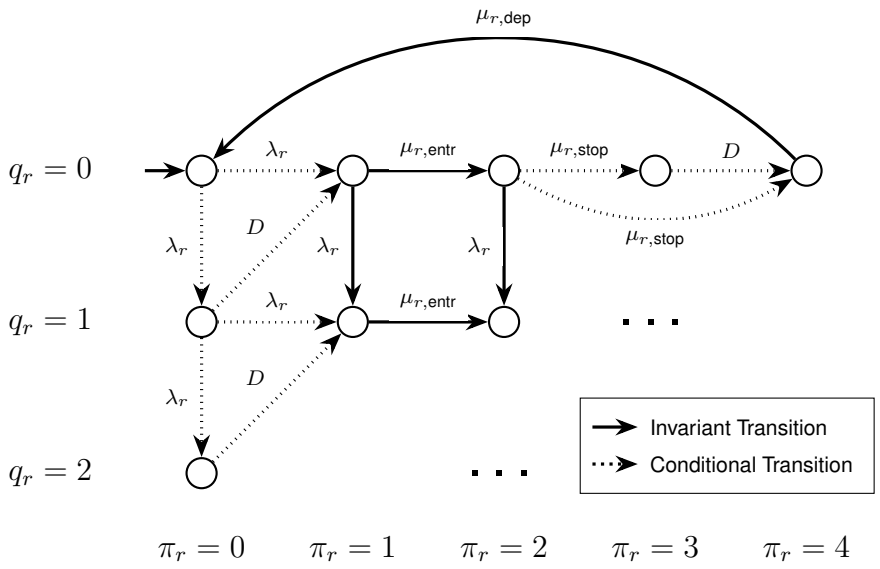


Figure 5.5.: A CTMC for the arrival, entry, stopping and departure process on one route with $m = 2$.

with

$$\pi_r(t) = \begin{cases} 2 & ;\text{if } \pi_r(s) = 1, \\ 0 & ;\text{if } \pi_r(s) = 4. \end{cases} \tag{5.14}$$

All choice transitions (s, t) are conditional since they are usually applied when conflicts between routes may exist. They correspond to the choice of the next service transition, an example can be found in Figure 5.4. Consequently, any type of service transition on a route r holds only if no other route r' is currently in a conflicting entry blocking, stopping or departure blocking process.

5.3. Determining the Timetable Capacity of Railway Stations

While the previous section introduced the Continuous-Time Markov Chain model, this section focuses on obtaining the timetable capacity of a railway station. After an introduction to timetable capacity (Section 5.3.1), the handling of the described CTMC model is discussed - from the required data (Section 5.3.2), to calculating the queue-length estimations (Section 5.3.3), defining the thresholds (Section 5.3.4) and outlining the procedural flow (Section 5.3.5). Furthermore, Section 5.3.6 gives insights into

the scalability of the introduced model.

5.3.1. The Timetable Capacity Performance Metric

Since the timetable of a railway system is more dynamic than its infrastructure, timetable-independent performance metrics are of high significance. In this work, we employ the notation of timetable capacity, as articulated in Wendler (2007). This capacity metric can be understood as the performance potential of railway infrastructure during the timetabling process, wherein train operators request resources for their planned train journeys from the infrastructure manager. It is important to note that multiple journeys may request identical or conflicting infrastructure routes for the same point in time, necessitating the rescheduling of certain journeys to achieve a conflict-free schedule. For each route, rescheduled journeys can be represented as a queue awaiting access to the requested infrastructure route. At any given point in time, the queue-length of a given route corresponds to the number of requests pending access to the infrastructure. Although the proposed method technically pertains to the allocation of occupation requests within a railway schedule, we sometimes refer to these requests as 'trains' for the sake of readability.

In this approach, the arrival times and required occupation durations of requests are subject to uncertainties and can thus be effectively modelled using queueing systems (see Section 5.2). Consequently, the estimated value of the average queue length per route during the timetabling process provides an estimation of infrastructure quality. To ensure adequate quality, thresholds for this queue length can be established (see Section 5.3.4). The maximum number of requests, for which the timetabling process meets this quality requirement for a given traffic distribution, is denoted as the timetable capacity of the railway infrastructure.

The following section outlines the data requirements necessary to determine the timetable capacity of a railway station.

5.3.2. Required Data

To determine the timetable capacity of a railway station using the introduced model, some data needs to be provided in order to obtain the results.

First, the time horizon U, which describes the analysed time span in minutes, needs

Table 5.2.: List of Variables and Notations

Variable/Notation	Description	Input Data	Default Value
Infrastructure and Time Horizon			
U	Time horizon	✓	-
I	Infrastructure	✓	-
R	Set of routes	✓	-
G	Set of route nodes	✓	-
P	Set of platform tracks	✓	-
C	Conflicts between routes	✓	-
n_{total}	Total number of requests im time horizon	-	-
Queueing and Service Process			
m	Number of waiting positions	-	6 (excl. Section 5.4.3)
λ	Arrival rate	-	-
μ	Service rate	-	-
ρ	Occupation ratio ($\rho = \lambda/\mu$)	-	-
L	Average length of the queue	-	-
p_{loss}	Probability of loss	-	-
$p(s)$	Steady-state probability distribution over states	-	-
Scaling Variables			
γ	Scaling factor in approximation formula	-	-
c	Intermediate variable for calculating γ	-	-
s	Number of parallel service channels in the approximation formula	-	fixed at 1
v_A	Coefficient of variation of the arrival process	-	-
v_S	Coefficient of variation of the service process	-	-
Route-Specific Parameters			
L_r	Estimated queue-length per route	-	-
ψ_r	Base count factor per route	✓	-
n_r	Number of trains per route in a specified time horizon ($n_r = \psi_r \cdot n_{\text{total}}$)	-	-
$b_{r,\text{entr}}$, $b_{r,\text{stop}}$, $b_{r,\text{dep}}$	Average service times for entry blocking, stopping, and departure blocking processes respectively	✓	-
$v_{A,r}$	Coefficient of variation for the arrival process on a specific route	✓	0.8
$v_{S,r,\text{entr}}$, $v_{S,r,\text{stop}}$, $v_{S,r,\text{dep}}$	Variation coefficient for entry blocking, stopping, and departure blocking processes respectively	✓	0.3
$v_{S,r}$	Combined coefficient of variation for all three service processes on a specific route	-	-
Additional CTMC Parameters			
π_r	Status in the service process for route r (e.g., 0: no service; 1: entry; 2: stopping; 3: waiting to depart; 4: departing)	-	-
$\pi_r(s)$	Status of route r in state s	-	-
$q_r(s)$	Queue length of route r in state s	-	-
m_r	Maximum queue length (number of waiting positions) for route r	-	-
λ_r	Arrival rate for a specific route ($\lambda_r = n_r/U$)	-	-
$\mu_{r,\text{entr}}$, $\mu_{r,\text{stop}}$, $\mu_{r,\text{dep}}$	Service rates for entry blocking, stopping, and departure blocking processes respectively	-	-
D	Artificial constant transition rate used to model random decisions between conflicting routes.	-	600

to be specified. Furthermore, the station infrastructure I is required. This consists of a set of routes R, a set of route nodes G, a set of platform tracks P, and a function describing the conflicts between routes and their service processes:

$$C : R \times \Pi \times R \times \Pi \to \{0,1\}, (r,\pi,r',\pi') \mapsto \begin{cases} 1 & \text{, conflict between } r \text{ in } \pi \text{ and } r' \text{ in } \pi' \\ 0 & \text{, otherwise} \end{cases}. \tag{5.15}$$

Notably, $\Pi = \{1,2,4\}$ describes the three service processes: entry blocking (1), stopping (2), and departure blocking (4) for a route.

Furthermore, additional characteristics for each route need to be provided. Specifically, the arrival rate λ_r and the service rates $\mu_{r,\text{entr}}$, $\mu_{r,\text{stop}}$, and $\mu_{r,\text{dep}}$ are required as input to the CTMC.

However, since we are optimizing for the maximum number of requests n_{max}, the total train count n_{total} changes during the optimization process (see Section 5.3.5). Hence, the number of trains per route n_r, used for calculating $\lambda_r = n_r/U$, is determined using a base count factor ψ_r with $n_r = \psi_r \cdot n_{\text{total}}$. This factor ψ_r thus describes the distribution of traffic across individual routes and is assumed to be fixed for a given instance. Hence, we denote the collection $\{\psi_r | r \in R\}$ of all base count factors as *traffic distribution*.

In addition, the service times $b_{r,\text{entr}}, b_{r,\text{stop}}, b_{r,\text{dep}}$ for each service step need to be provided in order to calculate the corresponding service rates. An infrastructure operator would typically determine these using blocking time theory on microscopic infrastructure data.

Since different rolling stock configurations may traverse a given route, a set of rolling stock configurations T needs to be specified. This work utilizes the different routes to decompose the queueing system of the considered junction; hence, the required arrival and service processes are formulated per route and thus always describe a weighted average over the rolling stock configurations. To enhance readability, we assume that the aggregation of the different rolling stock configurations per route has been completed prior to the procedure described in this work. Therefore, we differentiate only between routes.

To establish the queue-length threshold per route (see Section 5.3.4), it is necessary to specify the share of passenger trains on each route $p_{\text{pt},r}$. Moreover, to describe non-exponential arrival or service processes, the corresponding variation coefficients

must also be included.

Table 5.2 lists all the variables and notations, highlighting the required input data as well as the default values used in this work. Noteably, the default values correspond to the coefficients of variation for both, the arrival and the service processes. These default values are recommended in general planning rules (see f.e. VIA Consulting & Development GmbH (2021)) for the determination of timetable capacity.

5.3.3. Acquiring Queue-Length Estimations

The formulated Continuous-Time Markov Chain (see Section 5.2.4) $MC = (S, T)$ can be used to calculate the steady-state probabilities $p(s)$ of every state $s \in S$. In combination with the queue lengths $\mathsf{q}_r(s)$ per route r for each state, the expected queue length per route

$$L_r = \sum_{s \in S} p(s) \cdot \mathsf{q}_r(s) \tag{5.16}$$

can be obtained.

While some research utilizes the underlying Kolmogorov equations to directly obtain the state probabilities via matrix-vector equations (Weik, 2020) or iterative formulas (Nießen, 2008), recent publications use probabilistic model checking (Emunds and Nießen, 2024a) to directly calculate the estimated queue lengths per route. In this work, the latter approach has been adopted by implementing Python scripts that utilize the Python interface `stormpy` (Junges and Volk, 2023) for the probabilistic model checker `STORM` (Hensel et al., 2022).

The expected queue lengths per route L_r obtained in this manner needs to be scaled with the approximation formula (see Section 5.2.2) and compared to a threshold to ensure sufficient timetabling quality.

5.3.4. Defining Thresholds

In published research, various thresholds have been established to ensure adequate quality during the planning phase. While the UIC (2013) introduces maximum occupation rates, the largest European infrastructure manager, DB InfraGO (2009), employs thresholds based on the expected queue length.

The formula to determine the maximum admissible queue length, $L_{\text{limit, r}}$, was developed in Schwanhäußer and Schultze (1982) by conducting a survey under railway dispatchers to obtain thresholds for the quality of a railway line. As each route could be seen as a single service channel within the railway station, we utilize this threshold independently for each route in this work.

Depending on the proportion of passenger traffic $p_{\text{pt},r}$ on a route r, the limit is defined as

$$L_{\text{limit, r}} = 0.479 \cdot \exp(-1.3 \cdot p_{\text{pt},r}) \tag{5.17}$$

.

Computing this limit for railway stations with passenger traffic only ($p_{\text{pt},r} = 1$), results in a maximum queue-length of $L_{\text{limit, r}} = 0.1305$ per route.

In this study, this limit is compared to the estimated queue length per route L_r using the quality factor

$$\text{qf}_r = \frac{L_r}{L_{\text{limit, r}}}, \tag{5.18}$$

which quantifies the quality on route r. The quality threshold is thus defined by a quality factor of $\text{qf}_r = 1$.

In the subsequent section, the procedure for determining the timetable capacity of a railway station by comparing the estimated queue lengths per route (see Section 5.3.3) to the here introduced limits $L_{\text{limit},r}$ is described in detail.

5.3.5. Procedure

Given an instance, describing a railway station with the Infrastructure $I = (R, G, P, C)$ over the time horizon U, and the desired train mix, in detail the share of passenger traffic $p_{\text{pt},r}$ and the base count factor ψ_r per route r, the timetable capacity can be determined with the following general steps.

1. Determine the required minimum headway times h_{r_i,r_j} for every sequence of routes in their arrival or departure process to/from the corresponding platform track $p_{r_i,r_j} \in P$ in all route nodes G.
2. Determine the required stopping times per route.
3. Determine the maximum number of trains n_{max} in multiple steps for each con-

sidered n_{total}:

3.1. Calculate the number of trains per route $n_r = n_{\text{total}} \cdot \psi_r$.

3.2. Determine the required mean service times $b_{r,\text{entr}}$, $b_{r,\text{stop}}$, and $b_{r,\text{dep}}$ on the routes by averaging the minimum headway and stopping times of the routes.

3.3. Determine variation coefficients $v_{A,r}$ for the arrival processes on the routes.

3.4. Determine variation coefficients $v_{S,r,\text{entr}}, v_{S,r,\text{stop}}, v_{S,r,\text{dep}}$ for the service processes on the routes.

3.5. Formulate the CTMC in a `PRISM` model file.

3.6. Obtain L_r for every route from the CTMC and scale according to Section 5.2.2.

3.7. Calculate the quality factor qf_r for each route.

3.8. Determine $\text{qf}_{\text{max}} = \max_r(\text{qf}_r)$.

3.9. Select next considered value of n_{total} if precision has not been met.

To calculate the quality factors qf_r in step 3.6, the expected queue lengths L_r need to be obtained from the formulated CTMC. For this purpose, the CTMC is described in a `PRISM` model file (`.pm`) (Kwiatkowska et al., 2011) and evaluated with the probabilistic model-checking tool `stormpy` (Junges and Volk, 2023).

One evaluation step can take several seconds up to minutes, depending on the size of the CTMC. It is therefore crucial to restrict the number of iterations of step 3 by utilizing efficient methods for determining the next considered n_{total} (step 3.9).

The function

$$\phi : \mathbb{R} \to \mathbb{R},\ n_{\text{total}} \mapsto \phi(n_{\text{total}}) = \text{qf}_{\text{max}} - 1 \tag{5.19}$$

can be formulated to describe the difference between the maximum quality factor qf_{max} in step 3.8 and the desired value of 1.

With ϕ, determining timetable capacity can be formulated as a root-finding problem, i.e., finding $n^*_{\text{total}} = n_{\text{max}}$ such that $\phi(n^*_{\text{total}}) = 0$. This problem can be solved numerically, for example with Brent's method (Brent, 1973). In this work, we utilize the `scipy` (Virtanen et al., 2020) implementation of Brent's method with a relative tolerance of $\text{rtol} = 0.001$ and an absolute tolerance of $\text{xtol} = 0.001$ as the precision. Details about the implementation and the precision parameters are available at the `scipy` docu-

mentation (SciPy Developers, 2023).

The next section discusses the scalability of this procedure.

5.3.6. Scalability

As mentioned in the previous section, the computational requirements of the introduced procedure mainly depend on two factors: The number of necessary iterations of step 3 and the size of the CTMC model. The necessary iterations can be minimized with the discussed numerical solution to a root-finding problem, but the required computational power for a single iteration depends on the size of the CTMC model, in particular on the number of elements in the state set.

Recall the definition of the state set

$$S = \{(\pi_1, \mathsf{q}_1, \ldots, \pi_k, \mathsf{q}_k) \mid (\pi_r, \mathsf{q}_r) \in \Pi_r \times \{0, \ldots, m\}\}, \tag{5.20}$$

from Section 5.2.4. All states $s \in S$ therefore correspond to a selection of $\pi_r \in \Pi_r \subseteq \{0, \ldots, 4\}$ and $\mathsf{q}_r \in \{0, \ldots, m\}$ for each route r out of k possible routes, i.e. k choices between $|\Pi_r| \leq 5$ and $m + 1$ elements. Therefore, the upper bound

$$|S| \leq |\Pi_r|^k \cdot (m+1)^k \leq (5m+5)^k \tag{5.21}$$

to the number of states $|S|$ can be formulated.

Hence, the number of states grows exponentially with the number of routes k. Furthermore, for a fixed number of routes k^*, the number of states $|S|$ is bounded by a polynomial in $\mathcal{O}\left((5m)^{k^*}\right)$. Note that the factor 5^{k^*} describes the number of π_r-combinations for all k^* routes and it can be further restricted depending on the conflict function C in the route nodes of the analysed railway station. In comparison, the factor $(m+1)^{k^*}$ is more tight since every queue length in $\{0, \ldots, m\}$ is obtainable for every route, independent of the service status π_r and queue length of other routes.

Therefore, to enhance the scalability of the introduced model, it is preferable to decrease the number of modelled routes k, for example by enhancing the abstraction level or decomposing one station into multiple track groups. If the number of routes is fixed, another approach to reduce the state set is to minimize the modelled queue length per route, possibly individually for every route and in dependence to other in-

fluencing factors.

Given the substantial computational requirements, a small representative railway station has been chosen to illustrate the application of the proposed procedure in the next section.

5.4. Case Study

This case study analyses the performance capabilities of the station depicted in Figure 5.1. One aspect of the analysis examines two different operational scenarios, varying the stopping positions of two routes (Section 5.4.1). In a subsequent sensitivity-analysis (Section 5.4.2), the influence of the duration of the different service processes on timetable capacity is tested. Furthermore, Section 5.4.3 investigates models with different values for the maximum number of waiting positions m regarding their performance.

5.4.1. Operational Scenario Analysis

In this analysis, two different operational scenarios are compared for the passenger railway station described in Figure 5.1. The corresponding infrastructure example can be found in Figure 5.6.

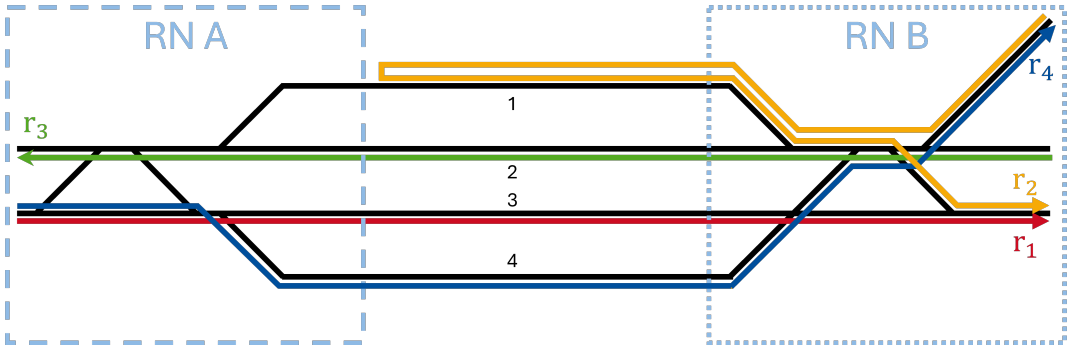


Figure 5.6.: The railway station infrastructure with swapped stopping platform tracks for routes r_2 and r_4

In the original Scenario 1, route r_2 requests a stop at platform track 4 and r_4 at platform track 1. In Scenario 2, these two tracks have been swapped in order to reduce the

number of conflicts during the entry process of all routes. Further details can be found in Table 5.3.

Table 5.3.: Route Node and Platform Usage per Scenario

Route	Scenario 1					Scenario 2				
	Entry Route Node	Platform Track	Departure Route Node	Entry Conflicts	Departure Conflicts	Entry Route Node	Platform Track	Departure Route Node	Entry Conflicts	Departure Conflicts
r_1	A	3	B	r_4	r_2	A	3	B	r_4	r_2 , r_4
r_2	B	4	B	r_1 , r_3 , r_4	r_1	B	1	B	r_3 , r_4	r_1 , r_3 , r_4
r_3	B	2	A	r_2 , r_4	r_4	B	2	A	r_2 , r_4	-
r_4	A	1	B	r_1 , r_3	r_2 , r_3	A	4	B	r_1	r_1 , r_2 , r_3

Notably, the track swap between the two routes leads to different entry and departure conflicts for all routes. In Scenario 2, for example, r_3 can depart without any conflicting route, whereas it is excluded by r_4 in Scenario 1. However, in Scenario 2, r_3 now conflicts not only with the entry of r_2 but also with its departure.

The service times for the entry, departure, and stopping processes are assumed to be equal for both scenarios. While the service times would typically be calculated using microscopic infrastructure and rolling stock data for a real-world application, we omit those details in this work and assume values instead. More information on the calculation of entry and departure blocking times can be found in Hansen and Pachl (2014). Table 5.4 lists the assumed service times in minutes.

Table 5.4.: Blocking and Stopping Times in Minutes

Route	Entry Blocking Time	Stopping Time	Departure Blocking Time
r_1	0.5	2.0	0.5
r_2	1.5	4.0	1.0
r_3	0.5	2.0	0.5
r_4	1.0	2.0	1.0

Notably, the arrival and departure blocking times for routes r_1 and r_3 are assumed to be 0.5 minutes each, as these routes do not need to traverse any switch branches. Furthermore, the stopping time of r_4 is assumed to be 4 minutes due to the necessity of turning on the platform track.

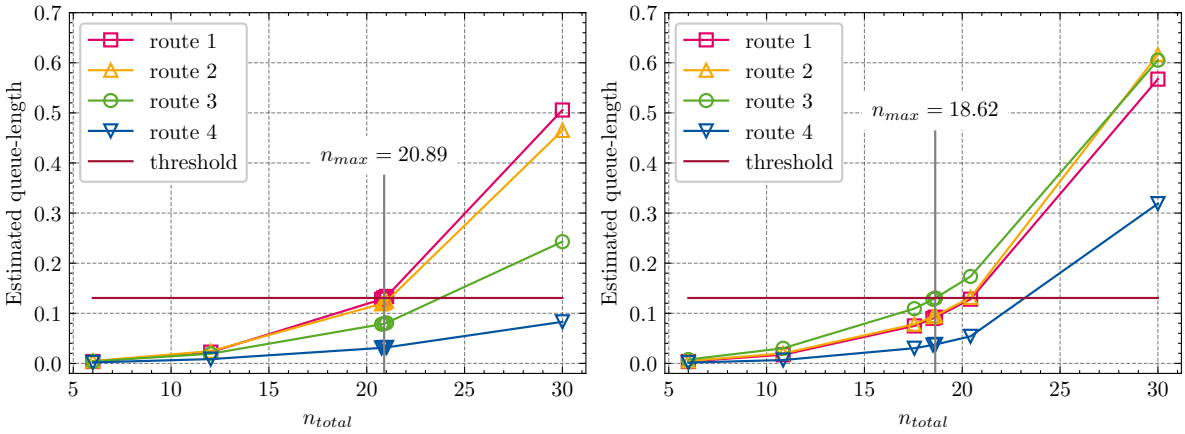
For all routes, only passenger traffic ($p_{\text{pt},r} = 1$) has been assumed. Furthermore, we

assume a traffic distribution of

$$\psi_r = \begin{cases} \frac{1}{3} & ;\text{if } r \in \{1,3\}, \\ \frac{1}{6} & ;\text{otherwise}, \end{cases} \tag{5.22}$$

thereby doubling the amount of traffic on the main line routes r_1 and r_3.

Using the default values for the variation coefficients of the arrival and service processes (see Section 5.3.2) and $m = 6$ waiting positions per route, the procedure from Section 5.3.5 has been performed for both scenarios. The resulting queue lengths (scaled according to Section 5.2.2) at each step of the method are depicted in Figure 5.7 and compared with the threshold $L_{\text{limit, r}} = 0.1305$ for passenger stations.



(a) Original Route Configuration

(b) Swapped Stopping Tracks

Figure 5.7.: Comparison of the queue-length estimations for all routes in both scenarios

While Figure 5.7a depicts the estimated queue lengths for the original Scenario 1, Figure 5.7b shows the estimated queue lengths for the scenario with swapped platform tracks.

For Scenario 1 (Figure 5.7a), the longest average queue is estimated for route r_1 , closely followed by the average queue length of r_2. Route r_3 falls in the mid-field, while route r_4 appears to be the least occupied. Therefore, the bottleneck can be assumed to be on route r_1 , resulting in a maximum capacity of $n_{\text{max}} = 20.89$ trains per hour.

In contrast, for Scenario 2 with swapped platform track stopping positions, r_3 is the most utilized route. While r_1 and r_2 follow closely, r_4 remains the least occupied route.

Utilizing the threshold for full-passenger-traffic routes (Section 5.3.4), the timetable capacity of $n_{\text{max}} = 18.62$ trains per hour can be concluded for this scenario.

The differences in the performance capabilities of the two scenarios can be explained by the high number of conflicts resulting from the re-routing of r_2. In the first scenario, the entry of route r_3 is only influenced during the entry of r_2 , and the departure of route r_1 is influenced by both the entry and departure of r_2.

In the second scenario, the entry process of r_3 is affected by both the entry and departure processes of r_2. While the influences on r_1 decrease, the increased conflicts of r_2 with the entry service process of r_3 significantly impact the queue length of r_3 , making it the bottleneck for the entire station.

An infrastructure operator should hence opt for the first scenario to fully utilize the stations performance capabilities.

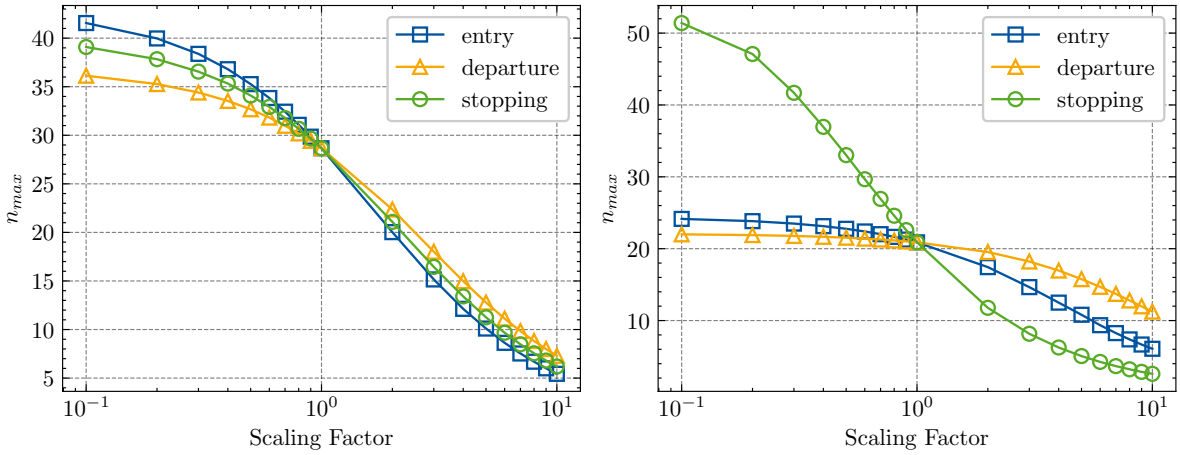
5.4.2. Influence of Service Processes

From the preceding results, the hypothesis that the duration of the entry blocking process is more influential to the timetable capacity of the railway junction, than the departure blocking process can be made. In this Section, a sensitivity analysis of the railway junction from scenario 1 to the duration of the service processes is performed.

For this, the service times $b_{r,\text{entr}}$, $b_{r,\text{stop}}$, $b_{r,\text{dep}}$ for the entry blocking, stopping and departure blocking process have each been scaled individually with a factor in the set $\{0.1, 0.2 \ldots, 0.9, 1, 2, \ldots, 9, 10\}$. The factor of both remaining service times has been set to 1. With this setting, the influence of the nominal value of the entry blocking time, stopping time or departure blocking time can be analysed. Note that the presented method utilizes an average service duration for each service step per route and does not distinguish between different train sequences. To obtain detailed results regarding headway times between specific conflicting train sequences, other methods, such as microscopic simulation tools, are to be preferred.

Figure 5.8 shows the resulting timetable capacity for the different scaling factors per service type.

To identify the sensitivity independently of the service time settings of the example railway station, Figure 5.8a assumes a base service time of 1 minute for every service process and scales it according to the scaling factor.



(a) Uniform Service Times

(b) Original Service Time Configuration

Figure 5.8.: Sensitivity of the timetable capacity to service time scaling

This sensitivity analysis reveals a significantly higher influence of the entry process duration, followed by the stopping duration, and finally the duration of the departure process.

In comparison, the scaling of the stopping time has the highest impact for the original service time configuration (Figure 5.8b). This can be explained by its significantly higher base setting of $2-4$ minutes in the original configuration, compared to $0.5-1.5$ minutes for the entry and departure blocking times. However, this assumption seems reasonable to allow a sufficient stopping duration for passenger exchange.

To increase the capacity of the example railway station, infrastructure operators should therefore focus on limiting the stopping time on the platform track and ensure sufficient railway control systems for a quick release of the entry route, in order to shorten the required blocking times in the entry process.

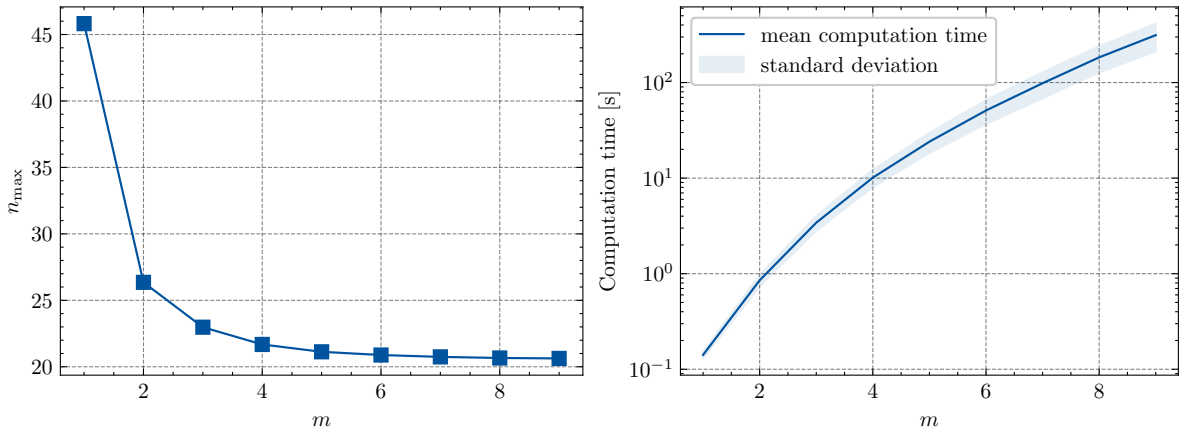
In the final analysis, the influence of the number of waiting positions m to the approximation quality of the introduced approach is investigated.

5.4.3. Dependence on the Number of Waiting Positions

The performance of the modelled queuing system is not only influenced by the arrival and service process, but additionally by the maximum number of requests that may be queued up for service (see also Section 5.2.1). While a constant number of $m = 6$ waiting positions per route has been assumed in the previous analysis, this section

examines the impact of different numbers of waiting positions.

For this, the timetable capacity of the original scenario 1 has been calculated for any $m \in \{1, \ldots, 9\}$. Figure 5.9 shows the resulting capacity and computation times for the different values.



(a) Timetable Capacity

(b) Computation Times

Figure 5.9.: Comparison of Capacity and Computation Times for Different Number of Waiting Positions m

The size of the state set S grows with the number of waiting positions per route, as every possible value combination must be considered. Consequently, the computation time increases substantially with increasing m (Figure 5.9b).

In contrast, for sufficiently small occupation ratios ρ, the probability of reaching a high number of parallel waiting requests per route might be insignificantly small, resulting in only incremental improvements for high values of m. For the analysed station example, the relative difference in timetable capacity between $m = 1$ and $m = 6$ exceeds 119%, whereas the difference between $m = 6$ and $m = 9$ is less than 2% of the value at $m = 6$.

Nonetheless, a higher number of available waiting slots increases the accuracy of the method. In order to fulfil the threshold to the expected average queue length (Section 5.3.4), the probability of experiencing such high queue lengths within the timetabling process needs to be negligible. Therefore, the determined capacity value decreases, in order to reduce the occupation rate per route and hence the probability of conflicting requests, directly influencing the amount of waiting trains.

In summary, a small number of waiting positions per route may result in faster solving

times but leads to a significant overestimation of timetable capacity with the introduced approach.

5.5. Conclusion

In this work, a queueing-based method to evaluate the performance of an entire railway station has been introduced. For this purpose, the station infrastructure has been partitioned along its routes, while the service process of each route has been further decomposed into entry, stopping, and departure processes.

This approach has been used to develop a novel Continuous-Time Markov Chain (CTMC) model that describes the timetabling process in railway stations (Section 5.2). With its multidimensional state-space, conflicts arising in all three service parts of different routes can be modeled. Additionally, the average number of trains per route waiting for their scheduled infrastructure occupation can be analysed.

By applying probabilistic model-checking, this number can be computed based on a fixed total train number, considering fixed traffic distributions and infrastructure properties (Section 5.3). A comparison with the formulated quality thresholds allows for the assessment of route-based quality factors to identify bottlenecks in the analysed infrastructure. Furthermore, numerical root-finding methods can be utilized to determine the timetable capacity of a given railway station's infrastructure.

The applicability of the introduced approach to practical questions posed by infrastructure operators has been discussed through a case study presented in Section 5.4. In this case study, the performance limits of two different operating procedures on an example infrastructure were analysed. The influence of different phases in the service life-cycle of a train within the railway station was also investigated. Finally, the introduced approach was applied for different numbers of waiting positions, highlighting the high computational requirements for an increasing number of waiting positions.

This observation supports the theoretical analysis of scalability discussed in Section 5.3.6, which associates computational complexity with the size of the state space in the CTMC. The state space not only depends on the number of waiting positions but also grows exponentially with an increase in the number of routes within the railway station. While the current implementation is effective for smaller stations, tackling

more complex nodes will require more sophisticated abstraction methods.

Other than enhancing abstraction levels or further decomposing the station, employing adaptive techniques to individually select the number of modelled waiting positions for each route might significantly enhance the scalability of the proposed model. This approach would involve estimating the error associated with formulations that use a fixed number of waiting positions per route. Such an estimation would facilitate the establishment of lower bounds on timetable capacity, thereby providing soundness guarantees. Consequently, this would enable the determination of the minimum number of waiting positions per route necessary to achieve a specified level of accuracy in the method.

Further exacerbating the need for better scalability is the issue of including an occupancy-based track assignment. In the current solution, the track-choice is fixed for every route and therefore for the traffic requests from train operators. Obtaining some flexibility here would certainly extend applicability, but it would correspond to introducing additional states in the CTMC formulation to monitor the occupation status of every track. Utilizing its occupation status, a track could be assigned when dequeueing a request to start its entry process. The choice between multiple available tracks could potentially be sorted by given track priorities per route. This procedure would not only necessitate to include information about track occupation into the state space and to reformulate transitions to reserve/repeal tracks when starting the entry process or terminating the departure process. In addition, conflicts between routes depend on the chosen tracks, hence further research would need to carefully model the dynamic parallel service capabilities of entry/arrival processes with flexible track assignments.

Regarding the accuracy, two main issues could be addressed. Firstly, the description of non exponential service processes is currently modelled by scaling results with the formula in Section 5.2.2. By utilising phase-type distributions for these processes, more accurate models could be built. However, this would also require additional states in the CTMC, which again negatively influences the scalability.

The second main accuracy extension is to further develop implemented quality thresholds. In its current formulation, the limit formula (5.17) has been obtained with railway line performance in mind. While it may be applicable to the individual routes of a railway station or junction, further research is needed to evaluate its precision for those application scenarios. In addition, current developments towards more synchronized timetables might necessitate the formulation of enhanced quality metrics, which could

address challenges posed by concentrated requests within short time intervals and ensure that the infrastructure allows for synchronized dwell times.

While the introduced method is designed to cope with uncertainties in the long-term strategic planning phase through its focus on abstract stochastic processes, it is not suitable to determine the performance of specific timetables. For such analyses, a simulation approach would be more appropriate, as it can provide detailed insights into timetable dependencies and evaluate performance metrics under various operational scenarios.

In summary, while future research can significantly enhance this initial method for comprehensive station performance analysis, the introduced model already serves as a valuable decision support tool for infrastructure managers. It extends existing methodologies by enabling a combined analysis of entire railway station infrastructure in the strategical planning phase.

6. Summary and Discussion

This chapter summarizes the findings of this work and discusses the proposed methods in order to determine further research needs.

6.1. Summary of Key Findings

In this thesis, multiple concepts to determine the performance capability of nodes in railway transportation networks have been introduced. For this, traffic volume has been quantified with the performance indicator of timetable capacity, ensuring sufficient quality within the timetabling process by limiting the average number of trains waiting for their occupation slot due to conflicting requests.

Chapter 3 introduced a novel queuing-based method to analyze the timetable capacity of railway junctions. Developed Continuous-Time Markov Chain formulations represent the allocation of occupation-slots in multi-channel junction infrastructures, enabling the application of probabilistic model-checking algorithms to obtain estimations of arising queueing lengths. Respecting the parallel serviceability of different routes through the railway junction, case studies demonstrated the ability to assist infrastructure managers with two critical tasks: Determining the timetable capacity and performance bottlenecks of a given infrastructure and adequately dimensioning railway junctions for a fixed operating program.

Furthermore, Chapter 4 extended the model with the possibility to utilize phase-type distributions to more accurately represent arrival and service processes. The introduced capacity determination algorithm allows an efficient computation of timetable capacity by formulating the optimization of the quality factor as a root-finding problem. This enabled a comprehensive comparison of different phase-type distribution implementations and simulation results, allowing to quantify the overall accuracy improvement by including phase-type models. A case study further highlighted that the gap to models without multiple arrival or service phases depends on the allocation of traffic volume to different routes and traffic types by applying the introduced method on a railway junction example where a mixed traffic line is split into freight and passenger lines.

Building on the approach for railway junctions, Chapter 5 presented a novel framework

to analyze the performance of entire railway stations. The main idea is to decompose the station occupation of one train request into entry, dwell, and departure processes. By introducing an auxiliary state before the departure, waiting processes after the dwelling phase have been incorporated into the framework, allowing to model the conflict resolving technique of increasing dwell times during the timetable construction. A Continuous-Time Markov Chain has been formulated as a combined representation of all route nodes and track groups, in order to estimate the average queue-length and hence quantify the timetable capacity with the previously described approach based on probabilistic model-checking. Applications at a small railway station showed the practicability for decision-making between operational scenarios and revealed differences in how the duration of different service processes may affect performance capabilities.

In summary, the here proposed methods introduced a novel framework to analyze the performance capability of nodes in a railway transportation network. They enable infrastructure managers to assess junction and station capacity by modeling train request relocation within timetable construction, the main capacity allocation process significantly determining transport volume. Nonetheless, several limitations and open questions remain, necessitating continued research efforts. These aspects are described in greater detail in the following section.

6.2. Limitations and Implications for Further Research

In order to obtain the presented results, certain assumptions were necessary and several limitations had to be acknowledged. This section examines these factors and their implications for further research.

6.2.1. Scalability

One major restriction that has been identified throughout all introduced methods is the limited scalability. Modeling parallel usage and queues for multiple routes leads to large state sets of the formulated Continuous-Time Markov Chains. However as of now, the here applied method of probabilistic model-checking internally requires building the entire Markov Chain, even though some states might be visited only rarely. A more efficient selection of actually analyzed states may enable the analysis

of more complex infrastructures. For this, further research could build on other works successfully abstracting state sets without modifying evaluation results, see Katoen (2016) for an overview on the abstraction landscape for probabilistic model-checking.

In this regard, simulations of queueing systems might be more effective as they only consider one particular state at every step and therefore visit rare states adequately infrequently. Thorough investigations require a high number of iterations in order to achieve a suitable accuracy. However, increasing computational capabilities make it feasible to run multiple simulations in parallel. This might make it more time-efficient to obtain performance capabilities via simulations instead of the here applied probabilistic-model checking. The simulation technique employed for the validation of the junction models (Sections 3.4.2 and 4.5) could be adapted, however this might require individual modifications for every infrastructure scenario.

Another approach to increase the applicability is to employ greater abstractions of the infrastructure. For route nodes, Schwanhäußer (1978) introduces a single-channel approximation. Recent work (Emunds and Nießen, 2025a) combines that idea with the presented method for railway stations (Chapter 5) to enable the assessment of larger railway stations.

6.2.2. Thresholds and Quality Indicators

In terms of the models applicability, employed thresholds are to be reconsidered. The utilized queue-length threshold has been derived from research on the operational capacity of railway lines (Schwanhäußer and Schultze, 1982). However, since that study was conducted, political and technical developments significantly reshaped European railway systems. A contemporary investigation could be able to establish dedicated thresholds for the conflict resolution during timetabling at stations and junctions.

More generally, the revised scope of multi-channel route nodes and entire stations motivates the examination of other quality indicators. One advantage of the here presented framework is that the Continuous-Time Markov Chain describes general relations between the different states in an independent model. The same model can hence be used to compute various quality indicators by checking correspondingly formulated properties (see Section 3.3.6.2). Therefore, the remaining challenge when implementing other quality indicators is to formulate the corresponding properties to the here introduced Continuous-Time Markov Chain models. With adapted proper-

ties, the presented methods may not only be used to compute average values in the long-run equilibrium, but can also exploit information about probability distributions across states or paths in the Continuous-Time Markov Chain formulations.

6.2.3. Validity of Queue-Length Estimations

With respect to the validity of the queue-length estimations, which are considered as the main performance indicators within this work, some remarks have to be made. First of, the scaling to non-exponential distributions with Hertel's Formula (Sections 3.3.6.3, 5.2.2) introduces inaccuracies into the calculations. Section 4.5.2 did assess the accuracy on a selected example, but a more comprehensive study might be necessary to fully understand the limitations of this approach. One solution, utilizing phase-type distributions, has already been discussed (Chapter 4), however this further expands the necessary state space of the underlying Continuous-Time Markov Chain model. Hence, further work may focus on developing approximation formulas for a multidimensional estimation of non-exponential distributions based on results for Markovian arrival and service processes. Following the example of Wakob (1984) for railway lines, such approximation formulas may be suitable to enhance the accuracy of queue-length estimations.

While the precision of estimations has been validated for introduced approaches considering railway junctions, the method for railway stations has not been examined in this regard. Therefore, subsequent research should follow Sections 3.4.2 and 4.5.3, utilizing simulation approaches in order to describe the accuracy of the obtained estimations. Furthermore, Section 5.4.3 investigated the influence of the number of modelled waiting positions to the accuracy and complexity of the described method for station performance analysis, highlighting its results as an upper bound to the actual timetable capacity. An examination of lower bounds in further research could enable the estimation of errors for different model sizes and hence build the basis of formulating soundness guarantees. This would be a crucial step towards the development of an adaptive algorithm that iteratively improves model detail where needed to efficiently compute the timetable capacity of railway node infrastructure.

6.2.4. Modeling Arrival and Service Processes

Considering the model of arrival and service processes, some additional limitations need to be respected. The arrival process assumes an individual and independent inquired start time of train requests, as well as a spread over the whole time horizon. However, political developments indicate a trend towards integrated symmetric timetables for passenger services. This would substantially condense traffic around selected symmetry minutes, especially in major railway stations. Further research may need to evaluate whether the introduced methods could be adapted to account for such modified arrival settings.

Regarding service processes, the introduced methods follow previous research on queueing-based railway performance estimations and model the service duration with blocking times or minimum headway times. Especially if multiple block sections are to be considered, practitioners may default to minimum headway times during application of the described methods. Here, it is crucial to consider the occurrence of special cases, i.e. where two fast trains with different train paths may follow a slower train. If one of the faster trains directly follows the route of the slower train, a large occupancy gap might arise. The other faster train could potentially utilize this gap, without influencing that first considered faster train. Related research (Niebel and Nießen, 2014; Weik et al., 2016; Nießen et al., 2023) describes this setting as the triangular gap problem.

Proposed solutions include the approximation of minimum headway time by accounting for the possibility of such an occurrence during minimum headway time calculation (Niebel and Nießen, 2014; Weik et al., 2016; Nießen et al., 2023) or the formulation of Semi-Markov processes to explicitly distinguish between different train sequences (Weik, 2020). Both approaches may be suitable to enhance the here described models, however an application of Semi-Markov processes would necessitate further enlargement of the state-space in the Continuous-Time Markov Chain model, resulting in even more challenging scaling behavior.

6.2.5. Priority Service

During timetable construction, infrastructure managers may need to prioritize certain traffic types and train routes in order to cope with regulatory frameworks or enhance quality in other network areas. This can substantially influence the estimated queue

lengths that are utilized to determine the timetable capacity in this work.

To model increased waiting times for lower priority traffic types, Schwanhäußer (1978) proposes *priority supplements* to the service times of train sequences with different priorities. While the same technique could be applied to the service times in this work, it assumes an occupation conflict for every occurrence of train sequences where the preceding train has a lower priority than the succeeding train, overestimating necessary additional waiting times for the low priority train.

Therefore, Weik (2020) includes transitions to preempt the service of a low priority train in case of the arrival of a higher priority train in a model with Semi-Markovian service processes. In order to accurately implement that strategy for different traffic type priorities, the here introduced models would need to distinguish between traffic types, further enlarging the state space of the underlying Continuous-Time Markov Chain. However, this method could be easily adapted to include route priorities into here discussed models, resetting low-priority routes from service back to the queue when a request with a higher priority route arrives. In addition, choice transitions (Sections 3.3.4.2 and 5.2.4.2) could be adapted to replicate priorities when determining which route or platform should be served next.

6.2.6. Further Adaptations and Enhancements

Recent European policy initiatives (Council of the European Union, 2024) enable infrastructure managers to designate preplanned infrastructure occupancy through *capacity planning objects* in a *capacity supply plan*. These capacity planning objects reserve infrastructure occupation slots for specific types of traffic services and may impose technical or rolling stock restrictions to optimize network use. Depending on the actual implementation of this policy by infrastructure managers, capacity allocation processes in member states of the European Union may be subject to major changes. Due to the formulation of capacity planning objects, the variability of blocking times and therefore the service process may decrease substantially, potentially reducing uncertainty during the timetabling process. This could have meaningful impact onto the arising waiting processes during timetable construction and therefore the timetable capacity of the infrastructure. Further research should aim to identify the impact of this regulation and carefully evaluate its implications to the here proposed and other methods that model the capacity allocation process in order to estimate

railway infrastructure performance capabilities.

A consideration of additional advancements may further enhance the applicability of the introduced methods. One example is to relax the static platform assignment in the railway station model, in order to more accurately represent conflict resolving possibilities during timetable construction. As previously discussed, a formulation of additional performance metrics or thresholds may introduce new perspective onto railway infrastructure performance. This may be particularly necessary when framework conditions change, i.e. with altering political guidelines. Further research with respect to decomposition techniques, more efficient adaptive algorithm designs or more aggressive abstractions could counteract the discussed issues regarding scalability and enable both: The application on more complex nodes or even small network areas and the enhancement of model accuracy by employing phase-type distributions or Semi-Markov models.

6.3. Conclusion

Even though certain limitations and assumptions have to be respected, the here proposed techniques may already be applied in real world scenarios. Due to their ability to include uncertainties about arrival and service processes into performance analysis they are particular suited to support infrastructure managers in the early stages of railway transportation planning.

The here introduced methods to analyze the performance capabilities of railway junctions and stations utilize the same general framework idea of formulating a queuing-based model of the capacity allocation process during timetabling to obtain a Continuous-Time Markov Chain representation and obtaining performance metrics by applying probabilistic model-checking. Resulting route-based queue length estimations represent a quality indicator during the timetable construction process on railway nodes, enabling the determination of timetable capacity for multi-channel systems such as railway junctions and stations. Utilizing standard probabilistic model-checking frameworks permits the separation of the model from the analyzed performance metric. This substantially improves the methods adaptability to future developments as it enables the formulation of other performance indicators without fundamentally changing the model or framework.

Furthermore, a decomposition based on routes and service phases allowed for the (to the best of our knowledge) first queuing-based model, capable of determining the timetable capacity of an entire railway station. By combining performance estimations of route nodes and track groups within a single model, infrastructure managers do not need to rely on distinguished methods and careful comparisons of the different results and performance indicators for the different elements within a single railway station in order to come up with consistent infrastructure capability assessments.

Bibliography

Abril, M., Barber, F., Ingolotti, L., Salido, M., Tormos, P., and Lova, A. (2008). An assessment of railway capacity. *Transportation Research Part E: Logistics and Transportation Review*, 44(5):774–806. doi:10.1016/j.tre.2007.04.001.

Adler, G. (1968). *Die Verkettung der Streckenbelegungen und der Belegungsgrad einer Gesamtstrecke*. PhD thesis, Hochschule für Verkehrswesen, Dresden.

Armstrong, J. and Preston, J. (2017). Capacity utilisation and performance at railway stations. *Journal of Rail Transport Planning & Management*, 7(3):187–205. doi:10.1016/j.jrtpm.2017.08.003.

Asmussen, S. (2003). *Applied probability and queues*, volume 2. Springer. doi:10.1007/b97236.

Becker, M., Betz, T., Geilke, M., Kabbe, G., Kaluza, A., Kössel, F., Terschlüsen, C., Zinser, M., and Höppner, S. (2025). Proton: A microscopic railway simulation for large scale networks. In *11th International Conference on Railway Operations Modelling and Analysis (RailDresden)*, Dresden, Germany.

Bešinović, N. and Goverde, R. M. P. (2018). Capacity Assessment in Railway Networks. In Borndörfer, R., Klug, T., Lamorgese, L., Mannino, C., Reuther, M., and Schlechte, T., editors, *Handbook of Optimization in the Railway Industry*, pages 25–45. Springer International Publishing. ISBN 978-3-319-72152-1. doi:10.1007/978-3-319-72153-8_2.

Bolch, G., Greiner, S., de Meer, H., and Trivedi, K. S. (2006). *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*. Wiley. ISBN 9780471791560.

Borndörfer, R., Klug, T., Schlechte, T., Fügenschuh, A., Schang, T., and Schülldorf, H. (2016). The Freight Train Routing Problem for Congested Railway Networks with Mixed Traffic. *Transportation Science*, 50(2):408–423. doi:10.1287/trsc.2015.0656.

Brent, R. (1973). Algorithms for minimization without derivatives. *Prentice-Hall, Englewood Cliffs NJ*.

Budde, C. E., Hartmanns, A., Klauck, M., Křetínský, J., Parker, D., Quatmann, T., Turrini, A., and Zhang, Z. (2021). On correctness, precision, and performance in quan-

titative verification. In Margaria, T. and Steffen, B., editors, *Leveraging Applications of Formal Methods, Verification and Validation: Tools and Trends*, volume 12479 of *Lecture Notes in Computer Science*, pages 216–241. Springer International Publishing, Cham. ISBN 978-3-030-83722-8. doi:10.1007/978-3-030-83723-5_15.

Büker, T. (2011). *Ausgewählte Aspekte der Verspätungsfortpflanzung in Netzen*. PhD thesis, Rheinisch-Westfälischen Technische Hochschule Aachen (german only).

Burdett, R. L. (2015). Multi-objective models and techniques for analysing the absolute capacity of railway networks. *European Journal of Operational Research*, 245(2):489–505. doi:10.1016/j.ejor.2015.03.020.

Burdett, R. L. (2016). Optimisation models for expanding a railway's theoretical capacity. *European Journal of Operational Research*, 251(3):783–797. doi:10.1016/j.ejor.2015.12.033.

Burdett, R. L. and Kozan, E. (2006). Techniques for absolute capacity determination in railways. *Transportation Research Part B: Methodological*, 40(8):616–632. doi:10.1016/j.trb.2005.09.004.

Burdett, R. L. and Kozan, E. (2009). Scheduling trains on parallel lines with crossover points. *Journal of Intelligent Transportation Systems*, 13(4):171–187. doi:10.1080/15472450903297608.

Burdett, R. L. and Kozan, E. (2014a). Determining operations affected by delay in predictive train timetables. *Computers & Operations Research*, 41:150–166. doi:10.1016/j.cor.2013.08.011.

Burdett, R. L. and Kozan, E. (2014b). Performance profiling for predictive train schedules. *Journal of Rail Transport Planning & Management*, 4(4):98–114. doi:10.1016/j.jrtpm.2014.11.002.

Burggraeve, S., Bull, S. H., Vansteenwegen, P., and Lusby, R. M. (2017). Integrating robust timetabling in line plan optimization for railway systems. *Transportation Research Part C: Emerging Technologies*, 77:134–160. doi:10.1016/j.trc.2017.01.015.

Bussieck, M. R., Winter, T., and Zimmermann, U. T. (1997). Discrete optimization in public rail transport. *Mathematical Programming*.

Bychkov, I., Kazakov, A., Lempert, A., and Zharkov, M. (2021). Modeling of Railway Stations Based on Queuing Networks. *Applied Sciences*, 11(5):2425. doi:10.3390/app11052425.

Cacchiani, V., Furini, F., and Kidd, M. P. (2016). Approaches to a real-world train timetabling problem in a railway node. *Omega*, 58:97–110. doi:10.1016/j.omega.2015.04.006.

Cacchiani, V. and Toth, P. (2012). Nominal and robust train timetabling problems. *European Journal of Operational Research*, 219(3):727–737. doi:10.1016/j.ejor.2011.11.003.

Cats, O., Gkiotsalitis, K., and Schöbel, A. (2025). 50 years of Operations Research in public transport. *EURO Journal on Transportation and Logistics*, 14:100160. doi:10.1016/j.ejtl.2025.100160.

Cerreto, F., Nielsen, B. F., Nielsen, O. A., and Harrod, S. S. (2018). Application of Data Clustering to Railway Delay Pattern Recognition. *Journal of Advanced Transportation*, 2018:1–18. doi:10.1155/2018/6164534.

Corman, F. and Henken, J. (2022). Estimating aggregate railway performance from realized empirical data: Literature review, a test case and a research roadmap. *Journal of Rail Transport Planning & Management*, 22:100316. doi:10.1016/j.jrtpm.2022.100316.

Corman, F. and Kecman, P. (2018). Stochastic prediction of train delays in real-time using bayesian networks. *Transportation Research Part C: Emerging Technologies*, 95:599–615. doi:10.1016/j.trc.2018.08.003.

Correia Duarte, P. J., Dauzère-Pérès, S., Huisman, D., Mannino, C., Sartor, G., Weik, N., and Widmann, P. (2025). 50 years of OR in railway timetabling and rolling stock planning. *EURO Journal on Transportation and Logistics*, 14:100155. doi:10.1016/j.ejtl.2025.100155.

Council of the European Union (2024). Proposal for a Regulation of the European Parliament and of the Council on the use of railway infrastructure capacity in the single European railway area. Technical Report 2023/0271(COD), Council of the European Union, Brussels. URL `https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CONSIL:ST_10938_2024_INIT`.

Cox, D. R. (1955). A use of complex probabilities in the theory of stochastic processes. In *mathematical proceedings of the Cambridge philosophical society*, volume 51, pages 313–319. Cambridge University Press. doi:10.1017/S0305004100030231.

D'Acierno, L., Botte, M., and Pignatiello, G. (2019). A simulation-based approach for estimating railway capacity. *International Journal of Transport Development and Integration*, 3(3):232–244. doi:10.2495/TDI-V3-N3-232-244.

David, A. and Larry, S. (1987). The least variable phase type distribution is erlang. *Communications in Statistics. Stochastic Models*, 3(3):467–473. doi:10.1080/15326348708807067.

DB InfraGO (2009). Richtlinie 405 Fahrwegkapazität. (german only).

DB InfraGO (2022). Richtlinie 405 Fahrwegkapazität. (german only).

DB InfraGO (2023). Nutzungsbedingungen Netz der DB Netz AG. (german only).

DB InfraGO (2025). Infrastrukturnutzungsbedingungen der DB InfraGO AG und der DB RegioNetz Infrastruktur GmbH (INB 2026). (german only).

De Kort, A. F., Heidergott, B., and Ayhan, H. (2003). A probabilistic (max, +) approach for determining railway infrastructure capacity. *European Journal of Operational Research*, 148(3):644–661. doi:10.1016/S0377-2217(02)00467-8.

Delorme, X., Rodriguez, J., and Gandibleux, X. (2001). Heuristics for railway infrastructure saturation. *Electronic Notes in Theoretical Computer Science*, 50(1):39–53. doi:10.1016/S1571-0661(04)00164-1.

Emunds, T. and Nießen, N. (2023). Research data and code: To build or not to build - a queueing-based approach to timetable independent railway junction infrastructure dimensioning. doi:10.5281/zenodo.8363461.

Emunds, T. and Nießen, N. (2024a). Evaluating railway junction infrastructure: A queueing-based, timetable-independent analysis. *Transportation Research Part C: Emerging Technologies*, 165:104704. doi:10.1016/j.trc.2024.104704.

Emunds, T. and Nießen, N. (2024b). Research Data and Code: Utilizing phase-type distributions for queueing-based railway junction performance determination. doi:10.5281/zenodo.14281662.

Emunds, T. and Nießen, N. (2025a). Approximation von Bediensystemen zur Leistungsfähigkeitsbestimmung von Eisenbahninfrastruktur in Bahnhöfen. In *Proceedings of the 5th International Railway Symposium Aachen (IRSA 2025)*, Aachen, Germany. RWTH Aachen University. (german only).

Emunds, T. and Nießen, N. (2025b). A queueing-based approach for timetable-

independent railway station performance analysis. *Journal of Rail Transport Planning & Management*, 36:100539. doi:10.1016/j.jrtpm.2025.100539.

Emunds, T. and Nießen, N. (2025c). Utilizing phase-type distributions for queueing-based railway junction performance determination. *Journal of Rail Transport Planning & Management*, 35:100523. doi:10.1016/j.jrtpm.2025.100523.

European Comission (2019). Technical specification for interoperability relating to the 'infrastructure' subsystem of the rail system within the european union: Comission regulation (EU) 2019/776.

European Commission (2020). Sustainable and Smart Mobility Strategy – putting European transport on track for the future. Communication COM(2020) 789 final, European Commission, Brussels. URL `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0789`.

European Commission (2025). Connecting Europe through High-Speed Rail. Communication COM(2025) 903 final, European Commission, Brussels. URL `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52025DC0903`.

Europe's Rail (2015). Shift 2 Rail. URL `https://projects.shift2rail.org/`.

Fischer, K. and Hertel, G. (1990). *Bedienungsprozesse im Transportwesen: Grundlagen und Anwendungen der Bedienungstheorie*. Transpress-Verlag. ISBN 3-344-00342-9. (german only).

Franke, B., Seybold, B., Büker, T., Graffagnino, T., and Labermeier, H. (2013). OnTime – Network-Wide Analysis of Timetable Stability. In *5th International Conference on Railway Operations Modelling and Analysis (RailCopenhagen)*.

Fuchs, F., Trivella, A., and Corman, F. (2022). Enhancing the interaction of railway timetabling and line planning with infrastructure awareness. *Transportation Research Part C: Emerging Technologies*, 142:103805. doi:10.1016/j.trc.2022.103805.

Goverde, R. M. P. (2007). Railway timetable stability analysis using max-plus system theory. *Transportation Research Part B: Methodological*, 41(2):179–201. doi:10.1016/j.trb.2006.02.003.

Goverde, R. M. P. (2010). A delay propagation algorithm for large-scale railway traffic networks. *Transportation Research Part C: Emerging Technologies*, 18(3):269–287. doi:10.1016/j.trc.2010.01.002.

Goverde, R. M. P., Corman, F., and D'Ariano, A. (2013). Railway line capacity consumption of different railway signalling systems under scheduled and disturbed conditions. *Journal of Rail Transport Planning & Management*, 3(3):78–94. doi:10.1016/j.jrtpm.2013.12.001.

Graffagnino, T. (2012). Ensuring timetable stability with train traffic data. In Brebbia, C. A., Tomii, N., Tzieropoulos, P., and Mera, J. M., editors, *Computers in Railways XIII*, WIT Transactions on The Built Environment, pages 427–438. WIT PressSouthampton, UK. doi:10.2495/CR120361.

Gudehus, T. (1976). Staueffekte vor Transportknoten. *Zeitschrift für Operations Research*, 20(6):B207–B252. doi:10.1007/BF01918395.

Hansen, I. A. and Pachl, J. (2014). *Railway Timetabling and Operations: Analysis, Modelling, Optimisation, Simulation, Performance, Evaluation*. Eurail press. ISBN 3-7771-0462-0.

Happel (1959). Sperrzeiten als Grundlage für die Fahrplankonstruktion. *Eisenbahntech. Rundsch. (ETR)*, 2:79.

Harrod, S. (2009). Capacity factors of a mixed speed railway network. *Transportation Research Part E: Logistics and Transportation Review*, 45(5):830–841. doi:10.1016/j.tre.2009.03.004.

Hensel, C., Junges, S., Katoen, J.-P., Quatmann, T., and Volk, M. (2022). The probabilistic model checker Storm. *International Journal on Software Tools for Technology Transfer*, 24(4):589–610. doi:10.1007/s10009-021-00633-z.

Hertel, G. (1984). Exakte Lösung zur Berechnung der Wartegleiszahl vor im Einrichtungsbetrieb befahrenen Streckengleisen bei Nicht-Poisson-Ankünften (G/M/1-Wartesystem). *Wissenschaftliche Zeitschrift der Hochschule für Verkehrswesen*, 31:195–205. (german only).

Hertel, G. (1986). *Analytische Modellierung und formelmässige Behandlung von Standard-Bedienungssystemen mit störanfälligen Kanälen*. PhD thesis, Hochschule für Verkehrswesen "Friedrich List", Dresden.

Huisman, T., Boucherie, R. J., and Van Dijk, N. M. (2002). A solvable queueing network model for railway networks and its validation and applications for the Netherlands. *European Journal of Operational Research*, 142(1):30–51. doi:10.1016/S0377-2217(01)00269-7.

Jensen, L. W., Landex, A., Nielsen, O. A., Kroon, L. G., and Schmidt, M. (2017). Strategic assessment of capacity consumption in railway networks: Framework and model. *Transportation Research Part C: Emerging Technologies*, 74:126–149. doi:10.1016/j.trc.2016.10.013.

Jensen, L. W., Schmidt, M., and Nielsen, O. A. (2020). Determination of infrastructure capacity in railway networks without the need for a fixed timetable. *Transportation Research Part C: Emerging Technologies*, 119:102751. doi:10.1016/j.trc.2020.102751.

Junges, S. and Volk, M. (2023). Stormpy - Python bindings for Storm: Version 1.7.0. URL `https://moves-rwth.github.io/stormpy/index.html`.

Katoen, J.-P. (2016). The probabilistic model checking landscape. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '16, page 31–45. Association for Computing Machinery. ISBN 9781450343916. doi:10.1145/2933575.2934574.

Kazakov, A., Lempert, A., and Zharkov, M. (2023). An approach to railway network sections modeling based on queuing networks. *Journal of Rail Transport Planning & Management*, 27:100404. doi:10.1016/j.jrtpm.2023.100404.

Kendall, D. G. (1953). Stochastic processes occurring in the theory of queues and their analysis by the method of the imbedded markov chain. *The Annals of Mathematical Statistics*, pages 338–354. URL `https://www.jstor.org/stable/2236285`.

Kingman, J. F. C. (1961). The single server queue in heavy traffic. *Mathematical Proceedings of the Cambridge Philosophical Society*, 57(4):902–904. doi:10.1017/S0305004100036094.

Kraft, E. R. (1988). Analytical models for rail line capacity analysis. *Journal of the transportation research forum*, 29(1):153–162.

Kwiatkowska, M., Norman, G., and Parker, D. (2010). Advances and challenges of probabilistic model checking. In *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1691–1698. doi:10.1109/allerton.2010.5707120.

Kwiatkowska, M., Norman, G., and Parker, D. (2011). PRISM 4.0: Verification of probabilistic real-time systems. In Gopalakrishnan, G. and Qadeer, S., editors, *Proc. 23rd International Conference on Computer Aided Verification (CAV'11)*, volume

6806 of *LNCS*, pages 585–591. Springer. doi:10.1007/978-3-642-22110-1_47.

Landex, A. (2009). Evaluation of railway networks with single track operation using the uic 406 capacity method. *Networks and Spatial Economics*, 9(1):7–23. doi:10.1007/s11067-008-9090-7.

Landex, A. and Jensen, L. W. (2013). Measures for track complexity and robustness of operation at stations. *Journal of Rail Transport Planning & Management*, 3(1-2):22–35. doi:10.1016/j.jrtpm.2013.10.003.

Leutwiler, F. and Corman, F. (2022). A logic-based benders decomposition for microscopic railway timetable planning. *European Journal of Operational Research*, 303(2):525–540. doi:10.1016/j.ejor.2022.02.043.

Li, C., Tang, J., Zhang, J., Zhao, Q., Wang, L., and Li, J. (2023). Integrated optimization of urban rail transit line planning, timetabling and rolling stock scheduling. *PLOS ONE*, 18(5):e0285932. doi:10.1371/journal.pone.0285932.

Liang, J., Martin, U., and Cui, Y. (2017). Increasing performance of railway systems by exploitation of the relationship between capacity and operation quality. *Journal of Rail Transport Planning & Management*, 7(3):127–140. doi:10.1016/j.jrtpm.2017.08.002.

Liao, Z., Li, H., Miao, J., and Corman, F. (2021). Railway capacity estimation considering vehicle circulation: Integrated timetable and vehicles scheduling on hybrid time-space networks. *Transportation Research Part C: Emerging Technologies*, 124:102961. doi:10.1016/j.trc.2020.102961.

Liebchen, C., Lübbecke, M., Möhring, R., and Stiller, S. (2009). The Concept of Recoverable Robustness, Linear Programming Recovery, and Railway Applications. In Hutchison, D., Kanade, T., Kittler, J., Kleinberg, J. M., Mattern, F., Mitchell, J. C., Naor, M., Nierstrasz, O., Pandu Rangan, C., Steffen, B., Sudan, M., Terzopoulos, D., Tygar, D., Vardi, M. Y., Weikum, G., Ahuja, R. K., Möhring, R. H., and Zaroliagis, C. D., editors, *Robust and Online Large-Scale Optimization*, volume 5868, pages 1–27. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-642-05465-5_1.

Liebchen, C. and Möhring, R. H. (2007). The Modeling Power of the Periodic Event Scheduling Problem: Railway Timetables — and Beyond. In Geraets, F., Kroon, L., Schoebel, A., Wagner, D., and Zaroliagis, C. D., editors, *Algorithmic Methods*

for Railway Optimization, volume 4359, pages 3–40. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-540-74245-6. doi:10.1007/978-3-540-74247-0_1.

Lusby, R., Larsen, J., Ryan, D., and Ehrgott, M. (2011). Routing trains through railway junctions: A new set-packing approach. *Transportation Science*, 45(2):228–245. doi:10.1287/trsc.1100.0362.

Lusby, R. M., Larsen, J., and Bull, S. (2018). A survey on robustness in railway planning. *European Journal of Operational Research*, 266(1):1–15. doi:10.1016/j.ejor.2017.07.044.

Meirich, C. (2017). *Berechnung Und Bewertung Der Gesamtleistungsfähigkeit von Eisenbahnnetzen*. PhD thesis, RWTH Aachen, Aachen.

Mussone, L. and Wolfler Calvo, R. (2013). An analytical approach to calculate the capacity of a railway system. *European Journal of Operational Research*, 228(1):11–23. doi:10.1016/j.ejor.2012.12.027.

Niebel, N. and Nießen, N. (2014). Berücksichtigung des Zacken-Lücken-Problems bei der analytischen Kapazitätsermittlung. *Eisenbahntech. Rundsch. (ETR)*, 63(12):34–36. (german only).

Nießen, N. (2008). *Leistungskenngrößen für Gesamtfahrstraßenknoten*. PhD thesis, Verkehrswissenschaftliches Institut der RWTH Aachen (german only). URL `https://publications.rwth-aachen.de/record/50214`.

Nießen, N. (2013). Waiting and loss probabilities for route nodes. In *5th International Conference on Railway Operations Modelling and Analysis (RailCopenhagen)*.

Nießen, N., Oetting, A., Ćosović, A., Etzold, R. D., Menne, D., Schüttler, J., Steinborn, U., Weinhold, W., Wolters, A., and Zwanziger, F. (2023). *Planungs- und Betriebsmanagement für das System Bahn*. DB-Fachbuch. Bahn Fachverlag, Berlin, 1st edition. ISBN 978-3-943214-29-1.

Nikolayzik, D., Maus, M., and Nießen, N. (2024). Investigation of Railway Network Capacity by Means of Dynamic Flows. *Applied Sciences*, 14(18):8233. doi:10.3390/app14188233.

Pachl, J. (2022). *Systemtechnik des Schienenverkehrs: Bahnbetrieb planen, steuern und sichern*. Springer Fachmedien Wiesbaden, Wiesbaden. ISBN 978-3-658-38265-0 978-3-658-38266-7. doi:10.1007/978-3-658-38266-7.

Palmer, G. I., Knight, V. A., Harper, P. R., and Hawa, A. L. (2019). Ciw: An open-source discrete event simulation library. *Journal of Simulation*, 13(1):68–82. doi:10.1080/17477778.2018.1473909.

Parker, D., Norman, G., and Kwiatkowska, M. (2000). Prism. URL `https://www.prismmodelchecker.org/`.

Pellegrini, P., Marlière, G., and Rodriguez, J. (2014). Optimal train routing and scheduling for managing traffic perturbations in complex junctions. *Transportation Research Part B: Methodological*, 59:58–80. doi:10.1016/j.trb.2013.10.013.

Potthoff, G. (1962). *Verkehrsströmungslehre - Die Zugfolge auf Strecken und in Bahnhöfen*. Transpress. (german only).

Python Software Foundation (2022). Python 3.10.9. URL `https://docs.python.org/release/3.10.9/index.html`.

Ross, S. M. A. (2014). *Introduction to probability models*. Elsevier. ISBN 978-0-12-407948-9.

Şahin, İ. (2017). Markov chain model for delay distribution in train schedules: Assessing the effectiveness of time allowances. *Journal of Rail Transport Planning & Management*, 7(3):101–113. doi:10.1016/j.jrtpm.2017.08.006.

Schmidt, C. and Martin, U. (2010). Erhöhung der Effektivität und Transparenz bei Leistungsuntersuchungen mit Simulationsverfahren. *Eisenbahntech. Rundsch. (ETR)*, 07-08/2010:463–468. (german only).

Schmitz, C., Weik, N., Zieger, S., Nießen, N., and Schmeink, A. (2017). Markov models for the performance analysis of railway networks. In *7th International Conference on Railway Operations Modelling and Analysis (RailLille)*, page 23, Lille, France.

Schöbel, A. (2017). An eigenmodel for iterative line planning, timetabling and vehicle scheduling in public transportation. *Transportation Research Part C: Emerging Technologies*, 74:348–365. doi:10.1016/j.trc.2016.11.018.

Schwanhäußer, W. (1974). *Die Bemessung der Pufferzeiten im Fahrplangefüge der Eisenbahn*. PhD thesis, Verkehrswissenschaftliches Institut der Rheinisch-Westfälischen Technischen Hochschule Aachen. URL `https://publications.rwth-aachen.de/record/77145`. (german only).

Schwanhäußer, W. (1978). Die ermittlung der leistungsfähigkeit von großen

fahrstraßenknoten und von teilen des eisenbahnetzes. *Archiv für Eisenbahntechnik*, 1978(38):7–18. (german only).

Schwanhäußer, W. and Schultze, K. (1982). Ermittlung von Qualitätsmaßstäben für die Berechnung der Leistungsfähigkeit eines Streckenabschnittes und Entwicklung eines Rechenverfahrens zur Ermittlung von Endverspätungen: Forschungsarbeit für die Deutsche Bundesbahn. (german only).

SciPy Developers (2023). scipy.optimize.brentq — SciPy v1.11.4 manual. Available at: `https://docs.scipy.org/doc/scipy-1.11.4/reference/generated/scipy.optimize.brentq.html`. Accessed 25 June 2025.

Sels, P., Dewilde, T., Cattrysse, D., and Vansteenwegen, P. (2016). Reducing the passenger travel time in practice by the automated construction of a robust railway timetable. *Transportation Research Part B: Methodological*, 84:124–156. doi:10.1016/j.trb.2015.12.007.

Sommereder, M. (2011). *Modelling of Queueing Systems with Markov Chains: An Introduction to Basic and Advanced Modelling Techniques*. BoD–Books on Demand. ISBN 978-3842348431.

Spanninger, T., Büchel, B., and Corman, F. (2023). Train Delay Predictions Using Markov Chains Based on Process Time Deviations and Elastic State Boundaries. *Mathematics*, 11(4):839. doi:10.3390/math11040839.

Team SimPy Revision (2023). Simpy 4.0.1. URL `https://simpy.readthedocs.io/en/4.0.1/`.

The Ciw library developers (2024). Ciw: v3.2.2. doi:10.5281/zenodo.12667778.

UIC (2004). *Code 406 - Capacity*. International Union of Railways, Paris, France.

UIC (2013). *Code 406 - Capacity*. International Union of Railways, Paris, France.

Van Rossum, G. and Drake Jr, F. L. (1995). *Python tutorial*. Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands.

VIA Consulting & Development GmbH (2021). *LUKS Handbuch*. (Software Handbook, german only).

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R.,

Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272. doi:10.1038/s41592-019-0686-2.

Wakob, H. (1981). Wahrscheinlichkeitsmodell zur Ermittlung von Wartezeiten bei Zugfolgefällen. *Archiv für Eisenbahntechnik*, 36:15–22. (german only).

Wakob, H. (1984). *Ableitung eines generellen Wartemodells zur Ermittlung der planmäßigen Wartezeiten im Eisenbahnbetrieb unter besonderer Berücksichtigung der Aspekte Leistungsfähigkeit und Anlagenbelastung*. PhD thesis, RWTH Aachen University, Aachen. URL `https://publications.rwth-aachen.de/record/74574`.

Weik, N. (2020). *Long-Term Capacity Planning of Railway Infrastructure – A Stochastic Approach Capturing Infrastructure Unavailability*. PhD thesis, RWTH Aachen University. doi:10.18154/RWTH-2020-06771.

Weik, N. (2022). Macroscopic traffic flow in railway systems — a discussion of the applicability of fundamental diagrams. *Journal of Rail Transport Planning & Management*, 23:100330. doi:10.1016/j.jrtpm.2022.100330.

Weik, N., Hemminki, E., and Nießen, N. (2020a). The Effective Residual Capacity in Railway Networks with Predefined Train Services. In Neufeld, J. S., Buscher, U., Lasch, R., Möst, D., and Schönberger, J., editors, *Operations Research Proceedings 2019*, pages 725–731. Springer International Publishing, Cham. ISBN 978-3-030-48438-5 978-3-030-48439-2. doi:10.1007/978-3-030-48439-2_88.

Weik, N., Niebel, N., and Nießen, N. (2016). Capacity analysis of railway lines in germany – a rigorous discussion of the queueing based approach. *Journal of Rail Transport Planning & Management*, 6(2):99–115. doi:10.1016/j.jrtpm.2016.06.001. URL `https://www.sciencedirect.com/science/article/pii/s2210970616300208`.

Weik, N. and Nießen, N. (2017). A quasi-birth-and-death process approach for integrated capacity and reliability modeling of railway systems. *Journal of Rail Transport Planning & Management*, 7(3):114–126. doi:10.1016/j.jrtpm.2017.06.001.

Weik, N., Warg, J., Johansson, I., Bohlin, M., and Nießen, N. (2020b). Extending UIC 406-based capacity analysis – New approaches for railway nodes

and network effects. *Journal of Rail Transport Planning & Management*, 15. doi:10.1016/j.jrtpm.2020.100199.

Wendler, E. (1999). *Analytische Berechnung Der Planmässigen Wartezeiten Bei Asynchroner Fahrplankonstruktion*. PhD thesis, RWTH Aachen University. URL `https://publications.rwth-aachen.de/record/60003`.

Wendler, E. (2007). The scheduled waiting time on railway lines. *Transportation Research Part B: Methodological*, 41(2):148–158. doi:10.1016/j.trb.2006.02.009.

Yaghini, M., Nikoo, N., and Ahadi, H. R. (2014). An integer programming model for analysing impacts of different train types on railway line capacity. *Transport*, 29(1):28–35. doi:10.3846/16484142.2014.894938.

Zhang, X. and Nie, L. (2016). Integrating capacity analysis with high-speed railway timetabling: A minimum cycle time calculation model with flexible overtaking constraints and intelligent enumeration. *Transportation Research Part C: Emerging Technologies*, 68:509–531. doi:10.1016/j.trc.2016.05.005.

Zhong, Q., Xu, C., Yang, R., and Zhong, Q. (2023). Equivalences between analytical railway capacity methods. *Journal of Rail Transport Planning & Management*, 25:100367. doi:10.1016/j.jrtpm.2022.100367.

Zieger, S., Weik, N., and Nießen, N. (2018). The influence of buffer time distributions in delay propagation modelling of railway networks. *Journal of Rail Transport Planning & Management*, 8(3-4):220–232. doi:10.1016/j.jrtpm.2018.09.001.

Zukerman, M. (2013). Introduction to queueing theory and stochastic teletraffic models. doi:10.48550/arXiv.1307.2968.

Zwaneveld, P. J., Kroon, L. G., Romeijn, H. E., Salomon, M., Dauzère-Pérès, S., Van Hoesel, S. P. M., and Ambergen, H. W. (1996). Routing Trains Through Railway Stations: Model Formulation and Algorithms. *Transportation Science*, 30(3):181–194. doi:10.1287/trsc.30.3.181.

Zwaneveld, P. J., Kroon, L. G., and Van Hoesel, S. P. (2001). Routing trains through a railway station based on a node packing model. *European Journal of Operational Research*, 128(1):14–33. doi:10.1016/s0377-2217(00)00087-4.

Appendix A.1: Literature Summary

Table A.1.: Literature considering railway performance estimations

reference	capacity type	timetable dependence	methodology	infrastructure type	infrastructure decomposition	solution technique
UIC (2004)	utilization	dependent	compression	line	block-section	closed-form formula
UIC (2013)	utilization	dependent	compression	line and junction	block-section and route-based	closed-form formula
Abril et al. (2008)	utilization	dependent	compression	line	block-section	closed-form formula
Landex (2009)	utilization	dependent	compression	line	block-section	closed-form formula
Landex and Jensen (2013)	utilization	dependent	compression	junction	block-section	closed-form formula
Goverde et al. (2013)	utilization	dependent	compression	line	train-path	closed-form formula
Jensen et al. (2017)	utilization	independent	compression	single-line network	train-path	combinatorial algorithms
Jensen et al. (2020)	utilization	independent	compression	single-line network	train-path	combinatorial algorithms
Weik et al. (2020b)	utilization	independent	compression	station	train-path	closed-form formula
Goverde (2007)	utilization	dependent	max-plus-algebra	network	timed-events	matrix-calculations
Bešinović and Goverde (2018)	utilization	dependent	max-plus-algebra	network	timed-events	matrix-calculations
De Kort et al. (2003)	operational	independent	max-plus-algebra	single-line network	timed-events	matrix-calculations
Harrod (2009)	theoretical	timetable-saturation	optimisation	line	train-path	MIP
Yaghini et al. (2014)	theoretical	independent	optimisation	line	timed-events	MIP
Zwaneveld et al. (1996)	theoretical	independent	optimisation	station	route-based	MIP
Zwaneveld et al. (2001)	theoretical	independent	optimisation	station	route-based	MIP
Delorme et al. (2001)	theoretical	independent	optimisation	junction	block-section	combinatorial algorithms
Lusby et al. (2011)	(theoretical)	timetable-saturation	optimisation	junction	block-section	MIP
Burdett and Kozan (2006)	theoretical	timetable-saturation	optimisation	network	train-path	MIP
Burdett (2015)	theoretical	independent	optimisation	network	train-path	combinatorial algorithms
(Cacchiani and Toth (2012))	(theoretical)	independent	optimisation	-	-	mostly MIP
Leutwiler and Corman (2022)	(theoretical)	independent	optimisation	network	timed-events	disjunctive programming
Zhang and Nie (2016)	utilization	independent	optimisation	line	timed-events	combinatorial algorithms
Liao et al. (2021)	theoretical	timetable-saturation	optimisation	line	timed-events	MIP
Mussone and Wolfler Calvo (2013)	utilization	independent	optimisation	network	train-path	MIP
Graffagnino (2012)	operational	dependent	operational-data-analysis	line	train-path	statistical analysis
Armstrong and Preston (2017)	utilization	dependent	operational-data-analysis	station	timed-events	statistical-analysis
Weik (2022)	operational	dependent	operational-data-analysis	line	block-section	statistical-analysis
(Corman and Henken (2022))	operational	dependent	operational-data-analysis	line	train-path	statistical-analysis
Goverde (2010)	(operational)	dependent	max-plus-algebra	network	timed-events	matrix-calculations
Büker (2011)	(operational)	independent	analytical	network	timed-events	iterative formula
Şahin (2017)	(operational)	dependent	analytical	line	timed-events	matrix-calculations
Zieger et al. (2018)	(operational)	independent	simulation	line	train-path	simulation
Corman and Kecman (2018)	(operational)	dependent	operational-data-analysis	line	timed-events	bayesian-network
(Spanninger et al. (2023))	(operational)	-	operational-data-analysis	-	-	-
Potthoff (1962)	timetable	independent	analytical	station	route-based	iterative formula
Schwanhäußer (1974)	operational	independent	analytical	line	train-path	closed-form formula
Schwanhäußer (1978)	timetable	independent	analytical	junction	route-based	closed-form formula
Wakob (1984)	timetable	independent	analytical	line	train-path	closed-form formula
Wendler (2007)	timetable	independent	analytical	line	train-path	matrix-calculations
Nießen (2008)	operational	independent	analytical	junction	individual	iterative formula
Nießen (2013)	timetable	independent	analytical	junction	individual	iterative formula
Schmitz et al. (2017)	timetable	independent	analytical	junction	individual	matrix-calculations
Weik (2020)	timetable	independent	analytical	line and junction	route-based	matrix-calculations
Introduced here	timetable	independent	analytical	junction	route-based	prob. model-checking

Remarks: Please note that literature referring to the *(operational)* capacity type is introducing methods for computing delay propagation, which is the basis for calculating operational capacity. Furthermore, literature referring to the *(theoretical)* capacity type is introducing methods for railway timetabling, which can be used to compute theoretical capacity. Literature written within bracelets *(reference)* does not introduce new methods but rather gives a comprehensible review of relevant methodology.

Appendix A.2: Supplementary Boxplots

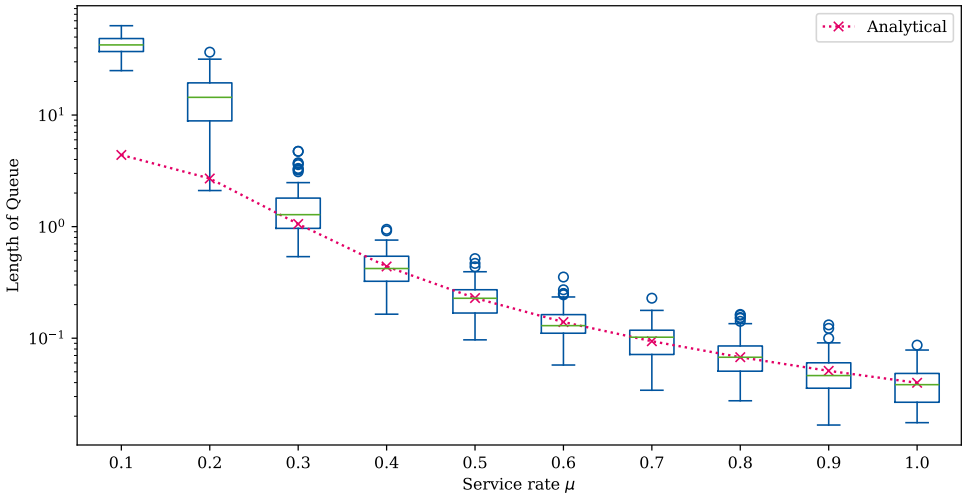


Figure A.1.: Comparison between the analytical solution and the simulations with an unlimited m

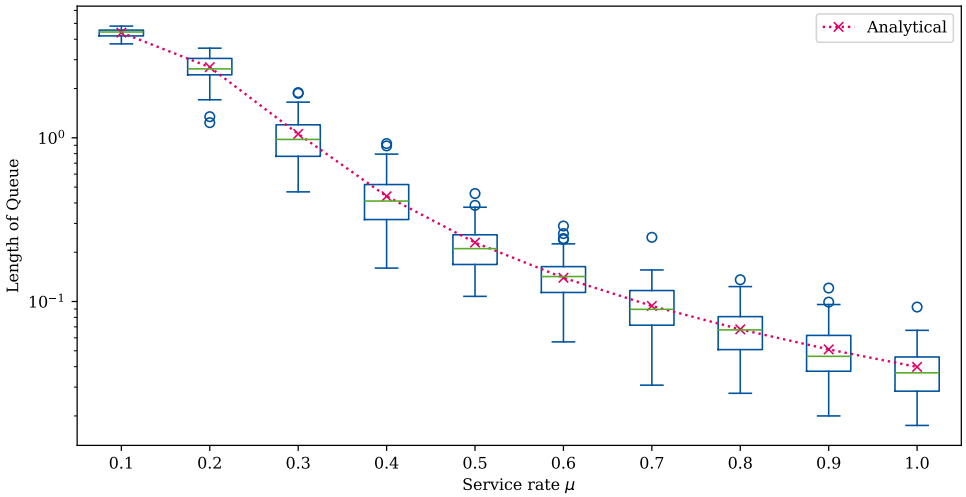


Figure A.2.: Comparison between the analytical solution and the simulations with a limit of $m = 5$

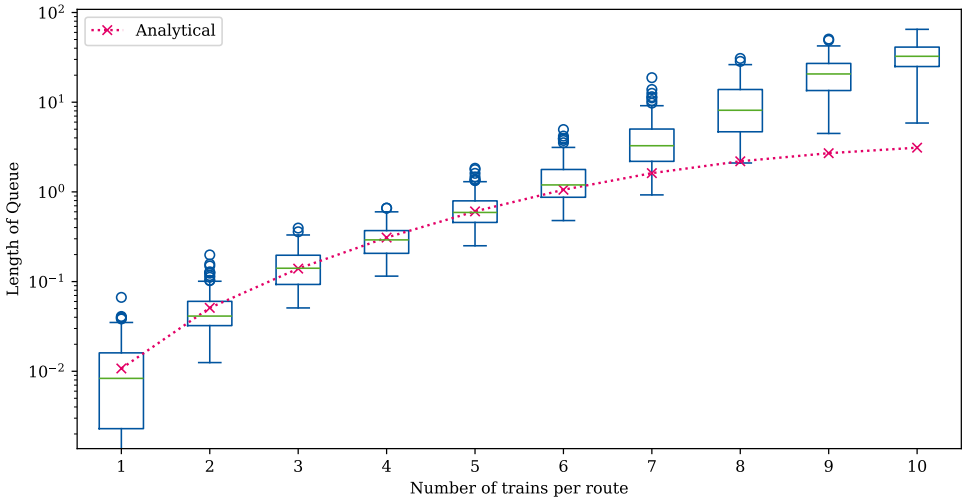


Figure A.3.: Comparison between the analytical solution and the simulations with an unlimited m, assuming different numbers of trains per route and a fixed service rate $\mu_i = 0.3$ for all $r_i \in R$.

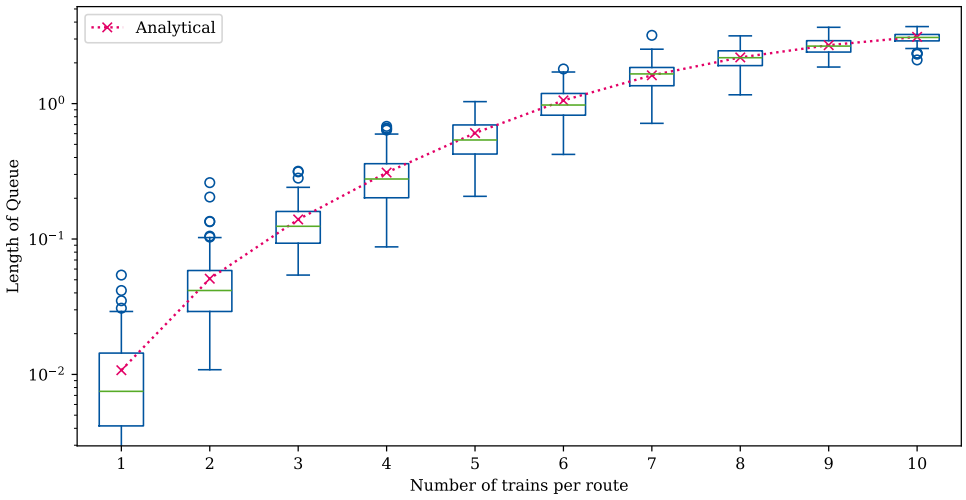


Figure A.4.: Comparison between the analytical solution and the simulations with a limit of $m = 5$, assuming different numbers of trains per route and a fixed service rate $\mu_i = 0.3$ for all $r_i \in R$.

Appendix A.3: Table With Exemplary Results

In Section 3.5.2, Figure 3.13 depicted the expected queue length at route r_3 (see Figure 3.1) from B to A for both railway junction layouts of a local train main line and a long-distance train branching line. Supplementary, selected results are listed in Table A.2.

Table A.2.: Estimated queue length on route r_3 for the combination of a local train main line and a long-distance train branching line

μ	$E_{LW,3}$ with overpass	$E_{LW,3}$ without overpass
0.01	4.9362	4.9575
0.1	3.6028	3.9855
0.2	0.9082	1.5143
0.3	0.2723	0.4654
0.4	0.1252	0.2042
0.5	0.0717	0.1135
0.6	0.0465	0.0722
0.7	0.0326	0.0500
0.8	0.0241	0.0367
0.9	0.0186	0.0281
1	0.0147	0.0222

Appendix A.4: Full Computational Study

The computational study explained in section 3.5.2 has been conducted for 529 operating program combinations in total. Since we refrained from showing their results in full detail in the text, some additional figures are given on the next pages. Please note that the results can additionally be found in the online repository (Emunds and Nießen, 2023).

Table A.3 lists all selected operating programs in section 3.5.2 and their respective names in the following figures.

Table A.3.: Selected operating programs

Name of railway line	Name in appendix
long-distance	P1, (0)
local train	P3, (0)
urban railway	P3-P5, (+)
freight train	F1, (0)
low intensity mixed	P3_F1, (0)
high intensity mixed	P3_F1, (+)

In Figure A.5 all used operating program combinations are depicted. Considered operating programs have been created as samples for demands to railway infrastructure and named in accordance with (European Comission, 2019), additionally distinguished between the three categories of exposure *base load* (-), *medium load* (0) and *high load* (+). Table A.4 lists a selection of traffic code parameters as specified in European Comission (2019).

Table A.4.: Selected traffic code parameters from (European Comission, 2019)

Traffic Code	Traffic type	Line Speed [km/h]	Usable length of platform [m]	Train length [m]
P1	passenger traffic	250-350	400	n.a.
P2	passenger traffic	200-250	200-400	n.a.
P3	passenger traffic	120-200	400-400	n.a.
P5	passenger traffic	80-120	50-200	n.a.
F1	freight traffic	100-120	n.a.	740-1050
F2	freight traffic	100-120	n.a.	600-1050
F3	freight traffic	60-100	n.a.	500-1050

Additionally, the corresponding results are depicted in Figures A.6 and A.7 for the calculations of maximum mean service times for the two infrastructure settings without and with an overpass.

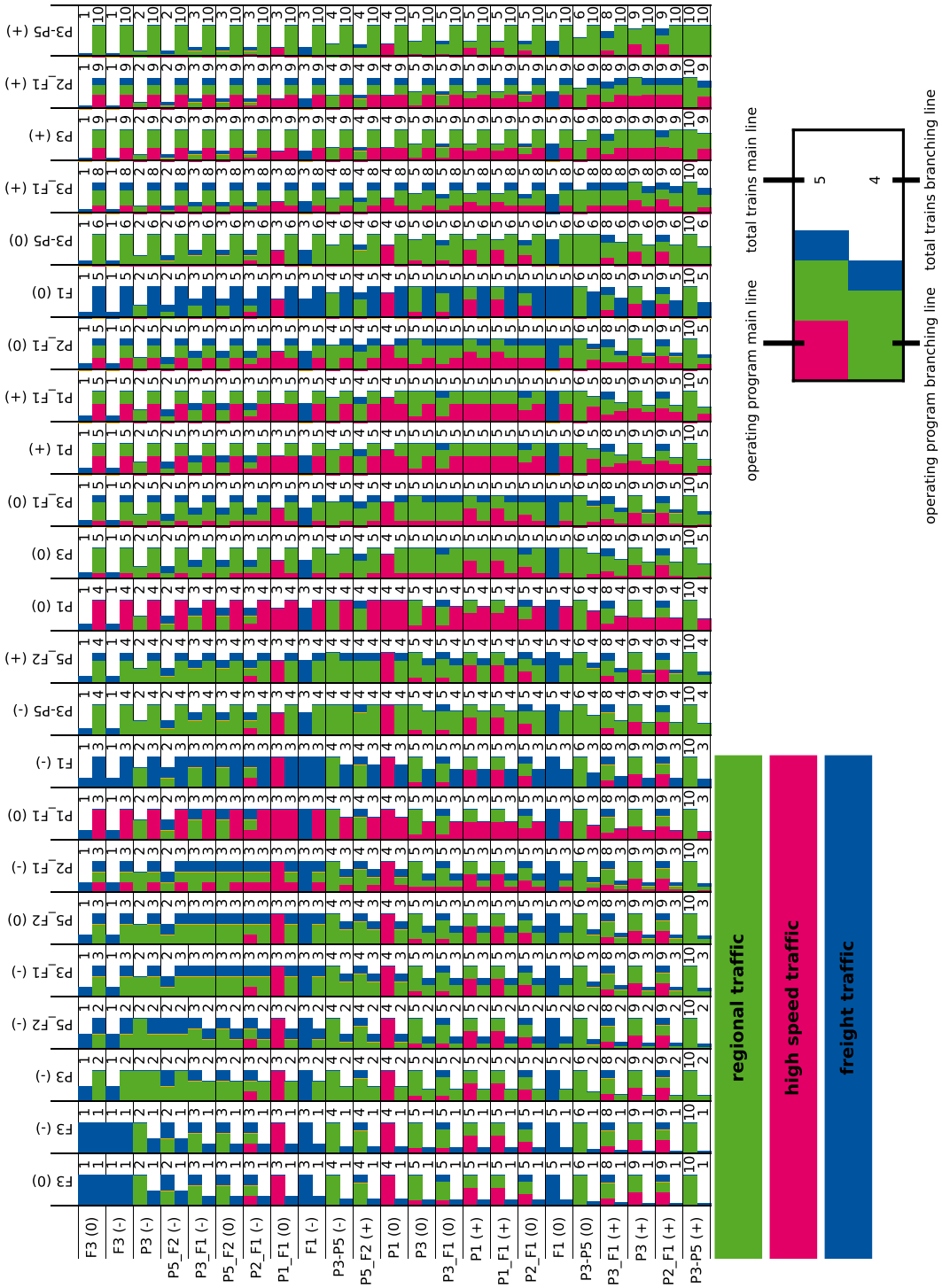


Figure A.5.: Overview of all operating program combinations

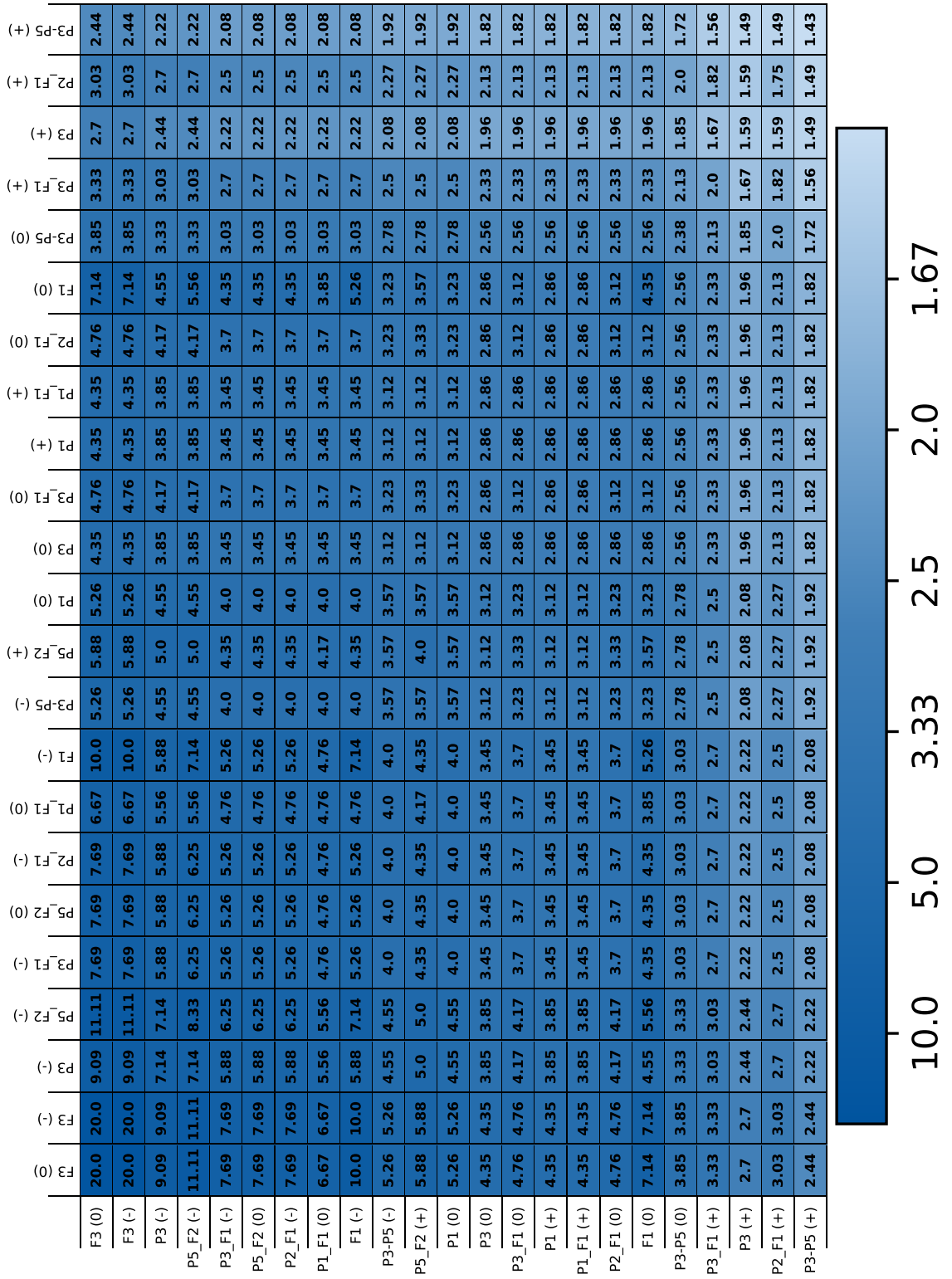


Figure A.6.: Calculated maximum mean service times in minutes for the scenario without an overpass

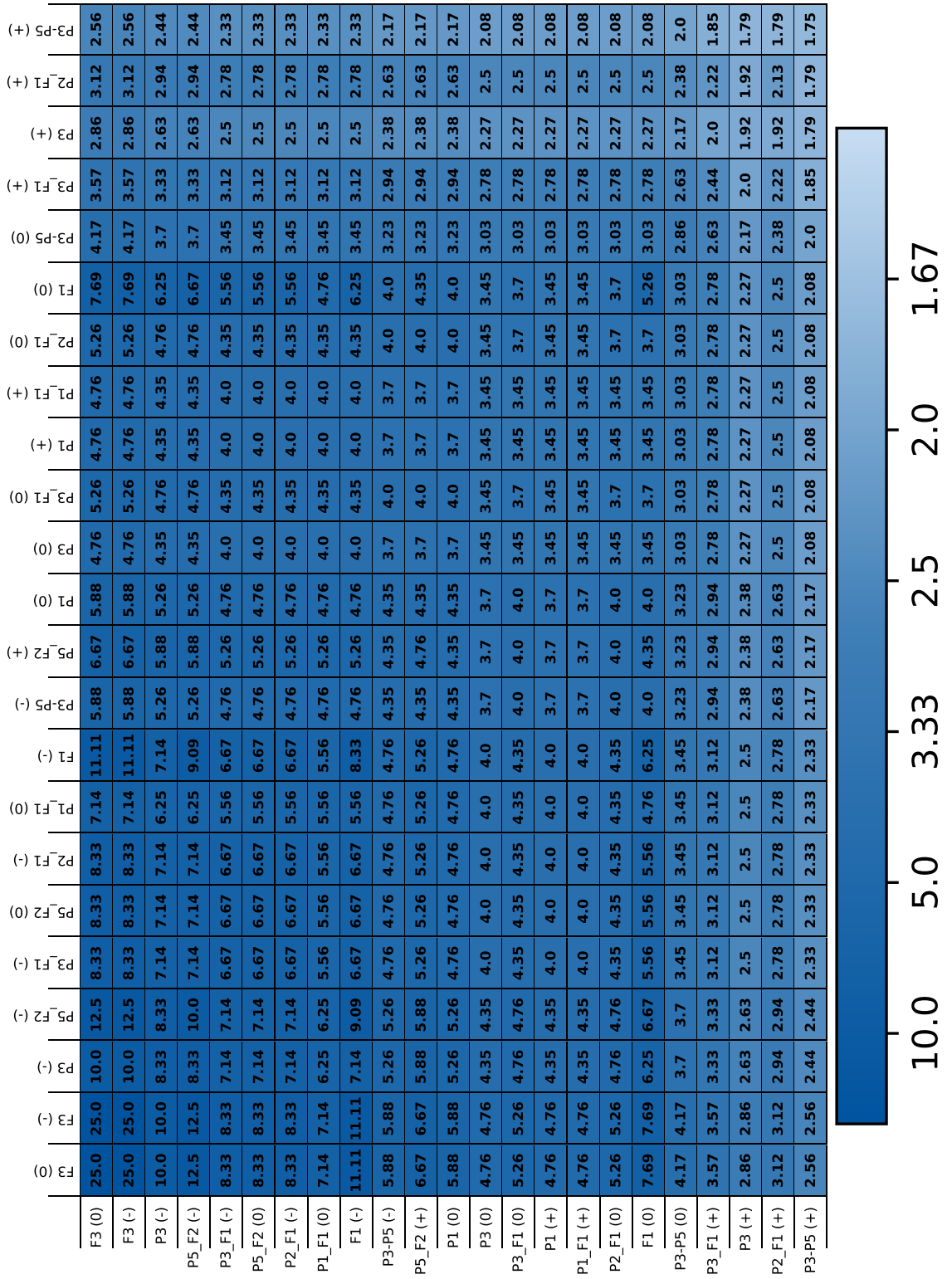


Figure A.7.: Calculated maximum mean service times in minutes for the scenario with an overpass

Appendix A.5: Full Comparison with Capacity Utilization

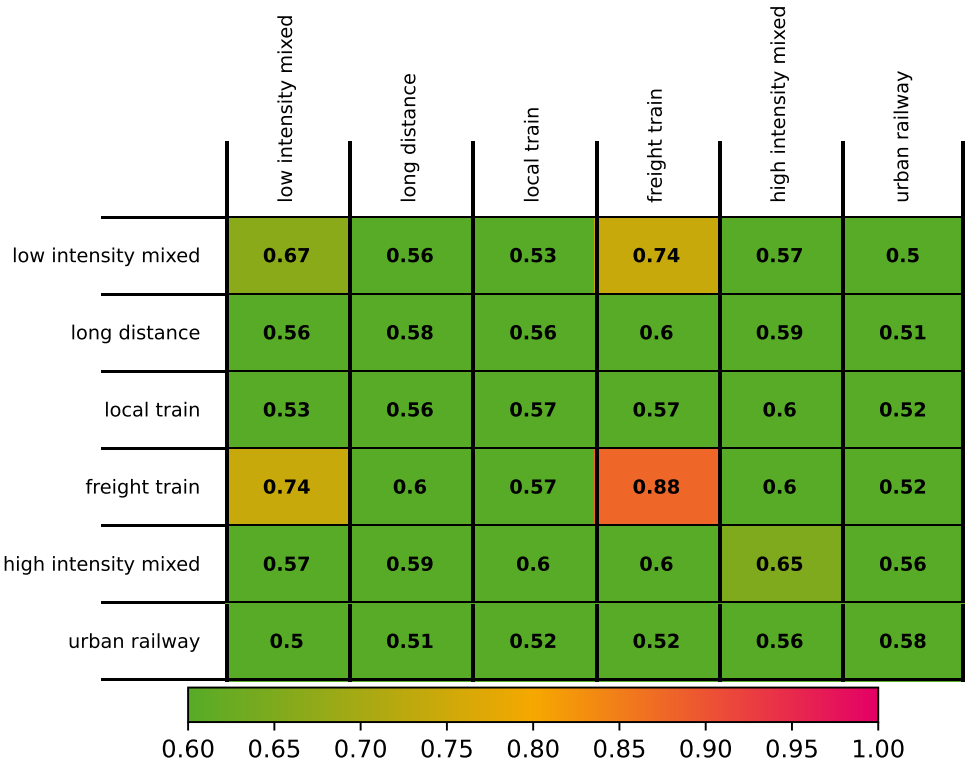


Figure A.8.: Occupancy rates (following equation (3.38)) for the scenario with an overpass, assuming mean service times from Figure 3.15

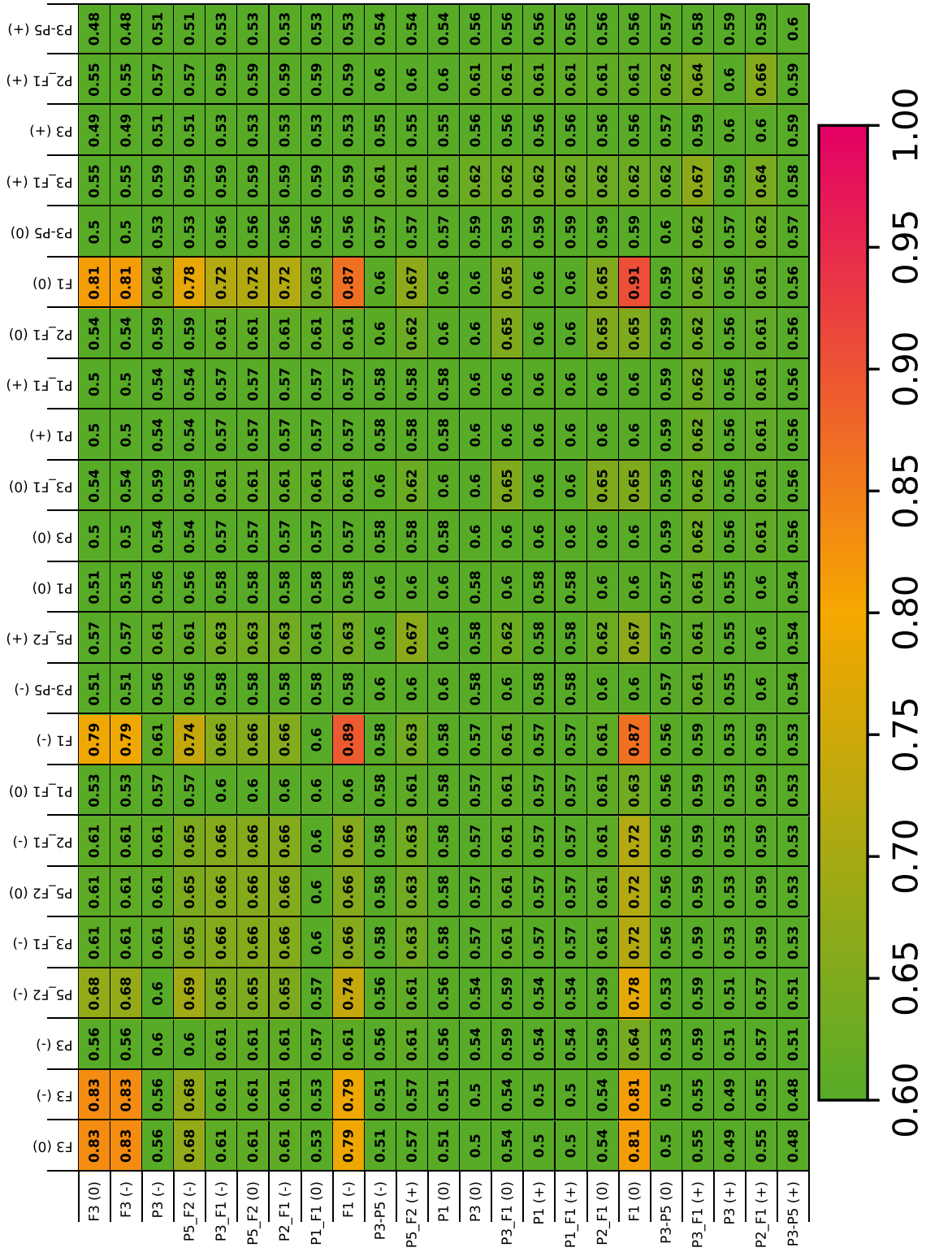


Figure A.9.: Occupancy rates (following equation (3.38)) for the scenario without an overpass, assuming mean service times from Figure A.6

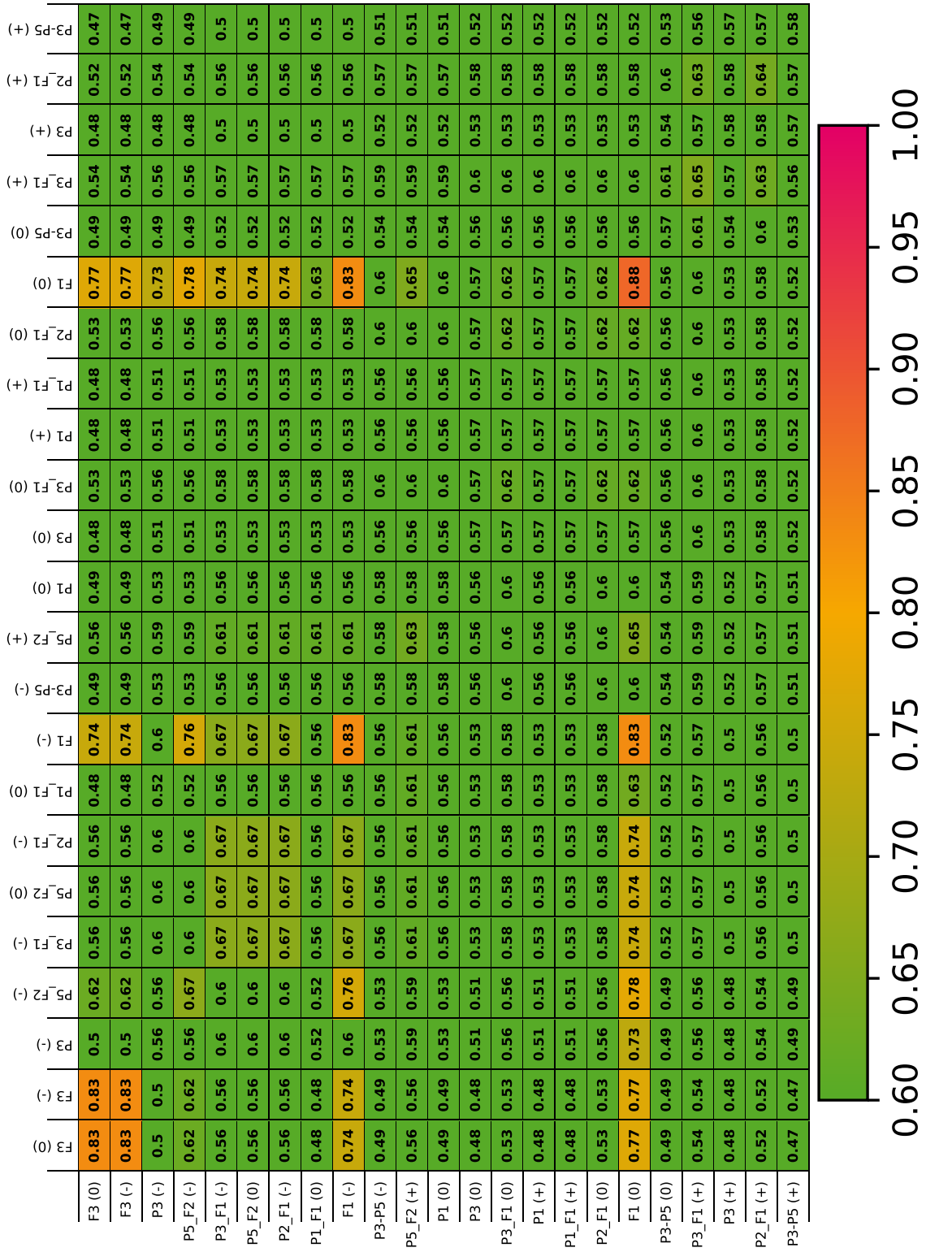


Figure A.10.: Occupancy rates (following equation (3.38)) for the scenario with an overpass, assuming mean service times from Figure A.7

Appendix B.1: Capacity Determination Algorithm

In Section 4.4.6, the root-finding problem has been introduced, which determines the timetable capacity of a railway junction by comparing the estimated queue-length per route L_r to their respective maximum thresholds $L_{r,\text{limit}}$, while varying the total train count n_{total}.

Since some parameters of the railway junction, such as the arrival rates λ_r, service rates μ_r and variation coefficients $v_{A,r}, v_{S,r}$ may depend on the total train count n_{total}, the CTMC model has to be formulated from inside the function call of the root-finding method.

Algorithm 1: Inner Function for Brent's Method

Data: Railway Junction $J = (R, C)$, Minimum Headway Times $h_{(r_i,t_i),(r_j,t_j)}$, Total Number of Trains n_{total}, Request Distribution θ
Result: $\phi\,(n_{\text{total}})$

1 **foreach** *route* $r \in R$ **do**
2 Determine number of trains on the route n_r;
3 Determine arrival rate $\lambda_r = 1/n_r$;
4 Determine share of passenger trains p_{pt};
5 Determine $L_{\text{limit, r}}$;
6 **end**
7 Determine frequencies of train sequences $p_{i,j}$;
8 **foreach** *route* $r \in R$ **do**
9 Determine service times b_r;
10 Determine service rates $\mu_r = 1/b_r$;
11 Determine variation coefficients $v_{S,r}$;
12 Determine phase-type parameters $k_{A,r}, k_{S,r}, \lambda_{r,a}, \lambda_{r,b}, \mu_{r,a}, \mu_{r,a}$;
13 **end**
14 Build the CTMC $MC = (S, T)$;
15 Get $L_r\,(n_{\text{total}})$ with probabilistic model checking;
16 Determine $qf_{\text{max}}\,(n_{\text{total}})$;
17 Determine $\phi\,(n_{\text{total}})$;
18 **return** $\phi\,(n_{total})$

Algorithm 1 therefore describes the complete inner function, that computes all relevant parameters before the calculation of the quality factor and the result $\phi\,(n_{\text{total}})$ for a chosen (n_{total}) in an interval $[n_{\text{total, LB}}, n_{\text{total, UB}}]$.

In the initial `foreach` loop, the number of trains per route, denoted by n_r, along with other computable parameters, is established. Subsequently, the frequencies

of train sequences $p_{i,j}$ are determined, which allows for the setting of the remaining parameters in the following `foreach` loop. The next phase involves formulating the Continuous-Time Markov Chain (CTMC) and calculating the queue lengths L_r. Finally, the difference between the maximum quality factor and the optimal ratio of 1 is computed and returned.

With this information, Brent's method determines whether the set tolerances are met and the timetable capacity n_{max} has been found. Otherwise, the interval $[a, b]$ is shortened, selecting a new bound n_{total} replacing a or b, where ϕ needs to be evaluated. An example of the convergence of this method can be found in Section 4.6.2.

Appendix B.2: Performance Determination for Simulations

In Section 4.5.3, we conduct 100 simulations, each lasting 20 hours, for $p_{\text{main}} \in P_{\text{main}} = \{0.1, \ldots, 0.9\}$ and $n_{\text{total}} \in N_{\text{total}} = \{12.00, 12.04, \ldots, 19.92, 19.96\}$.

This extensive computation was necessary because the algorithm from Section 4.4.6 assumes monotonicity for the function $L_r : N_{\text{total}} \to \mathbb{R}, p_{\text{main}} \mapsto L_r(n_{\text{total}})$.

However, since the simulation is not deterministic, even taking an average of 100 simulation runs does not guarantee a monotonic function approximation. It is therefore possible that there exist $n_{\text{total},1}, n_{\text{total},2} \in N_{\text{total}}$ for a fixed $p'_{\text{main}} \in P_{\text{main}}$, with:

$$L_r(n_{\text{total},1}) \geq L_r(n_{\text{total},2}) \tag{6.1}$$

and $n_{\text{total},1} < n_{\text{total},2}$.

Therefore, it might not be possible to find a single value $n^*_{\text{max}} \in N_{\text{total}}$ that meets the quality threshold requirement:

$$L_r(n^*_{\text{max}}) \leq L_{\text{limit}}, \tag{6.2}$$

where for all larger $n' \in N_{\text{total}}$, with $n^*_{\text{max}} < n'$, the limit

$$L_r(n') > L_{\text{limit}} \tag{6.3}$$

is exceeded **and** for all smaller $n'' \in N_{\text{total}}$, with $n^*_{\text{max}} > n''$, the limit

$$L_r(n'') \leq L_{\text{limit}} \tag{6.4}$$

is not violated.

Hence, we introduce the two bounds to the capacity $n^*_{\text{max, LB}}, n^*_{\text{max, UB}} \in N_{\text{total}}$, where the lower bound $n^*_{\text{max, LB}}$ meets the condition (6.4): For all smaller $n'' \in N_{\text{total}}$, with $n^*_{\text{max, LB}} > n''$, the limit

$$L_r(n'') \leq L_{\text{limit}} \tag{6.5}$$

is not violated.

Similarly, $n^*_{\text{max, UB}}$ meets the condition (6.3): For all larger $n' \in N_{\text{total}}$, with $n^*_{\text{max, UB}} < n'$,

the limit

$$L_r(n') > L_{\text{limit}} \tag{6.6}$$

is exceeded.

These bounds are chosen to be as tight as possible, so that $n^*_{\text{max, LB}}$ represents the highest value and $n^*_{\text{max, UB}}$ represents the lowest value that satisfy the corresponding conditions.

Appendix B.3: Computation Times Case Study

Table B.1 lists the computation times for both models and different p_{main} values from the case study in section 4.6.

Table B.1.: Performance Metrics for Different Model Types at Various p_{main} Values

Model Type	p_{main}	# Iterations	# Calls to Model-Checking	Mean Model-Checking Time (s)	Total Computation Time (s)
M/M	0.1	10	11	0.13	2.89
M/M	0.2	9	10	0.13	2.75
M/M	0.3	9	10	0.12	2.63
M/M	0.4	9	10	0.22	9.75
M/M	0.5	9	10	0.14	2.81
M/M	0.6	9	10	0.13	2.19
M/M	0.7	10	11	0.17	4.15
M/M	0.8	10	11	0.21	9.58
M/M	0.9	10	11	0.13	2.86
PH/PH	0.1	10	11	93.09	1025.04
PH/PH	0.2	9	10	65.20	656.58
PH/PH	0.3	8	9	67.71	610.58
PH/PH	0.4	9	10	60.55	606.76
PH/PH	0.5	9	10	48.74	488.77
PH/PH	0.6	9	10	61.68	617.89
PH/PH	0.7	9	10	43.84	439.48
PH/PH	0.8	9	10	40.65	407.56
PH/PH	0.9	8	9	50.39	454.54

Appendix C.1: Combining Coefficients of Variation

The coefficient of variation of a probability distribution X is described as

$$v_X = \frac{\mathsf{std}(X)}{\mathsf{E}(X)}, \tag{6.7}$$

with the standard deviation $\mathsf{std}(X)$ and expected value $\mathsf{E}(X)$. For multiple stochastic independent distributions $X_1, \ldots, X_n$ the expected value of the combination can be given as

$$\mathsf{E}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathsf{E}(X_i) \tag{6.8}$$

and the Variance as

$$\mathsf{Var}(\sum_{i=1}^{n} X_i) = \sum_{i=1}^{n} \mathsf{Var}(X_i). \tag{6.9}$$

Let the expected value $\mathsf{E}(X_i)$ and the coefficients of variation v_{X_i} be given, then the standard deviation can be calculated with

$$\mathsf{std}(X_i) = v_{X_i} \cdot \mathsf{E}(X_i). \tag{6.10}$$

Using $\mathsf{Var}(X_i) = \mathsf{std}(X_i)^2$, the combined variance can be given as

$$\mathsf{Var}(X) = \sum_{i=1}^{n} (v_{X_i}^2 \cdot \mathsf{E}(X_i)^2). \tag{6.11}$$

Hence, the combined coefficient of variation is obtained by

$$v_X = \frac{\mathsf{std}(X)}{\mathsf{E}(X)} = \frac{\sqrt{\sum_{i=1}^{n} (v_{X_i}^2 \cdot \mathsf{E}(X_i)^2)}}{\sum_{i=1}^{n} \mathsf{E}(X_i)}. \tag{6.12}$$