A Modelling Framework for Aligned Tactical and Operational Timetabling in Conventional and Digital Rail Operations

William Busuttil

Submitted to the Department of Transport & Planning (TP) at
Delft University of Technology
to fulfill the degree of Master of Science in Civil Engineering
30 April 2024

Thesis Committee (7 June 2023 – 30 April 2024)

Chair:	Dr. Ir. E. Quaglietta, Assistant Professor, Technische Universiteit Delft
Supervisors:	Dr. Ir. M. Saeednia, Assistant Professor, Technische Universiteit Delft
	Ir. K. Rigos, PhD Candidate, Technische Universiteit Delft
	Ms. Nadia Hoodbhoy, Principal Engineer, Network Rail Limited

This thesis and related research were performed from 7 June 2023 to 30 April 2024 with support from the Delft University of Technology (TU Delft) and Network Rail Limited, the railway infrastructure manager in England, Scotland, and Wales.

Acknowledgements

Upon reflection, I have many individuals to thank for the opportunity to learn and perform research in a topic of personal interest under the guidance of well-regarded subject matter experts at a renowned university.

I would like to extend my sincere gratitude to Egidio Quaglietta, the Chair of my thesis committee, for sharing his valuable insight and considerable knowledge of railway operations and traffic management. I am greatly appreciative of his sage guidance, humour, and encouragement to persevere, especially when my research proved challenging.

Many thanks to Mahnam Saeednia, Nadia Hoodbhoy, and Kostantinos Rigos for their guidance on this paper, and to Network Rail for supporting my research. My early childhood travels to London first sparked my interest in rail and public transport, so it seems appropriate that I conclude my masters studies by performing research involving Network Rail.

I am also grateful to other members of the Digital Rail Traffic (DRT) Lab for their assistance and camaraderie, and hope to help contribute to its noteworthy achievements.

Finally, words cannot properly convey how thankful I am for my family's love and unwavering support. I am grateful for their advice and encouragement, and for travelling across the Atlantic often to visit me in Delft.

Summary

Background

Railways are increasingly investing in their infrastructure to achieve sustainability objectives and meet expected growth in demand for rail transport services. Since building new physical infrastructure requires large-scale investments and long lead times, many infrastructure managers (IMs) are prioritizing upgrades to their signalling systems to optimize use of their existing networks. By migrating from legacy *multi-aspect* to *distance-to-go* (DTG) signalling, IMs gain the ability to safely run trains closer together, better manage disruptions occurring in normal operations, and implement other compatible technologies that can increase capacity and enhance service.

The key advantages of DTG systems are derived from their ability to provide more precise brake and speed supervision relative to multi-aspect signalling. Existing multi-aspect signalling systems rely on lineside signals to display braking indications to trains and do not consider train braking capabilities when determining the minimum separation between trains. As a result, minimum train separation is based on the worst-case braking distance of any train that could operate on the line. In contrast, brake supervision in DTG signalling systems is performed by the train's onboard computer which calculates a train's braking distance based on that specific train's current speed and braking capability. The ability to determine braking distance based on a train's own characteristics allows more delayed braking relative to multi-aspect signalling. Moreover, migrating the brake supervision from track to train eliminates the need for lineside signalling and related restrictions on block lengths, thereby permitting layouts with shorter blocks than would otherwise be possible with multi-aspect signalling.

To further increase network capacity, IMs are also looking to implement *Connected Driver Advisory Systems* (C-DAS) and *Automatic Train Operation* (ATO) systems. These systems permit greater coordination of train paths at the operational level through the provision of timing windows to trains (referred to as a *Train Path Envelope* or TPE) in real time. This timing information, in turn, helps trains avoid conflicts at the operational level when recovering from small departure delays at stations.

In practice, the benefits of DTG signalling and C-DAS/ATO systems can best be realized if timetabling algorithms consider the actual capabilities of those digital technologies to operate trains closer together in regular operations. It is therefore necessary to align tactical and operational scheduling rules, and to update train planning rules to reflect the capabilities of DTG signalling and C-DAS/ATO systems. However, state-of-the-art-methods for tactical scheduling could result in suboptimal network use as they do not accurately represent operational processes occurring on the railway network, and rely on different levels of detail than those used for real-time traffic management. Furthermore, the abstraction of signalling constraints at the tactical level necessitates additional and more detailed assessments using microscopic methods to achieve a target service plan and timetable feasibility.

Research Objectives

To meet the requirements of IMs, tactical scheduling models should be able to efficiently produce feasible, stable, and capacity-effective timetables that are aligned with traffic management methods. These models need to be detailed enough to realize the available capacity of the infrastructure for the signalling system in use but also simple enough to be tractable for large-scale timetabling problems. To accomplish this, a model should:

- represent the network's signalling layout and constraints as accurately as possible, and align with the IM's traffic management capabilities,
- incorporate the operational ability of C-DAS/ATO systems to guide trains, and
- be computationally simple enough to be used to address the network-level timetabling problems that IMs regularly face.

To address this need, the main research question of this paper is:

How can the tactical and operational planning levels be aligned to produce capacity-effective rail operations for conventional and digital railway operations?

The central questions to be answered in this thesis are:

- What is the current state-of-the-art for train service planning and real-time rescheduling models used for conventional and digital rail operations?
- How can train dynamics and separation constraint be more accurately represented in both tactical and operational planning for both conventional and digital rail operations?
- What method can be used to align tactical and operational train service planning in either conventional or digital rail operations?
- What impact can an aligned tactical and operational train planning framework have on schedule quality?
- What recommendations should IMs consider for aligning scheduling constraints in models for the tactical and operational levels?

At present, railways solve their large-scale timetabling problems using *macroscopic* models which represent the network at the level of stations and major junctions. The minimum headway constraints in these models are based on the schedulers' own experience and expectations of what the infrastructure can handle. The lack of a systematic method for determining minimum headways at the planning level can result in the macro models producing timetables with local conflicts, or incorrectly indicating that a feasible timetable does not exist for the given line plan. This limitation precludes use of existing macro models for areas with very dense traffic, forcing IMs to employ alternative methods to schedule those areas.

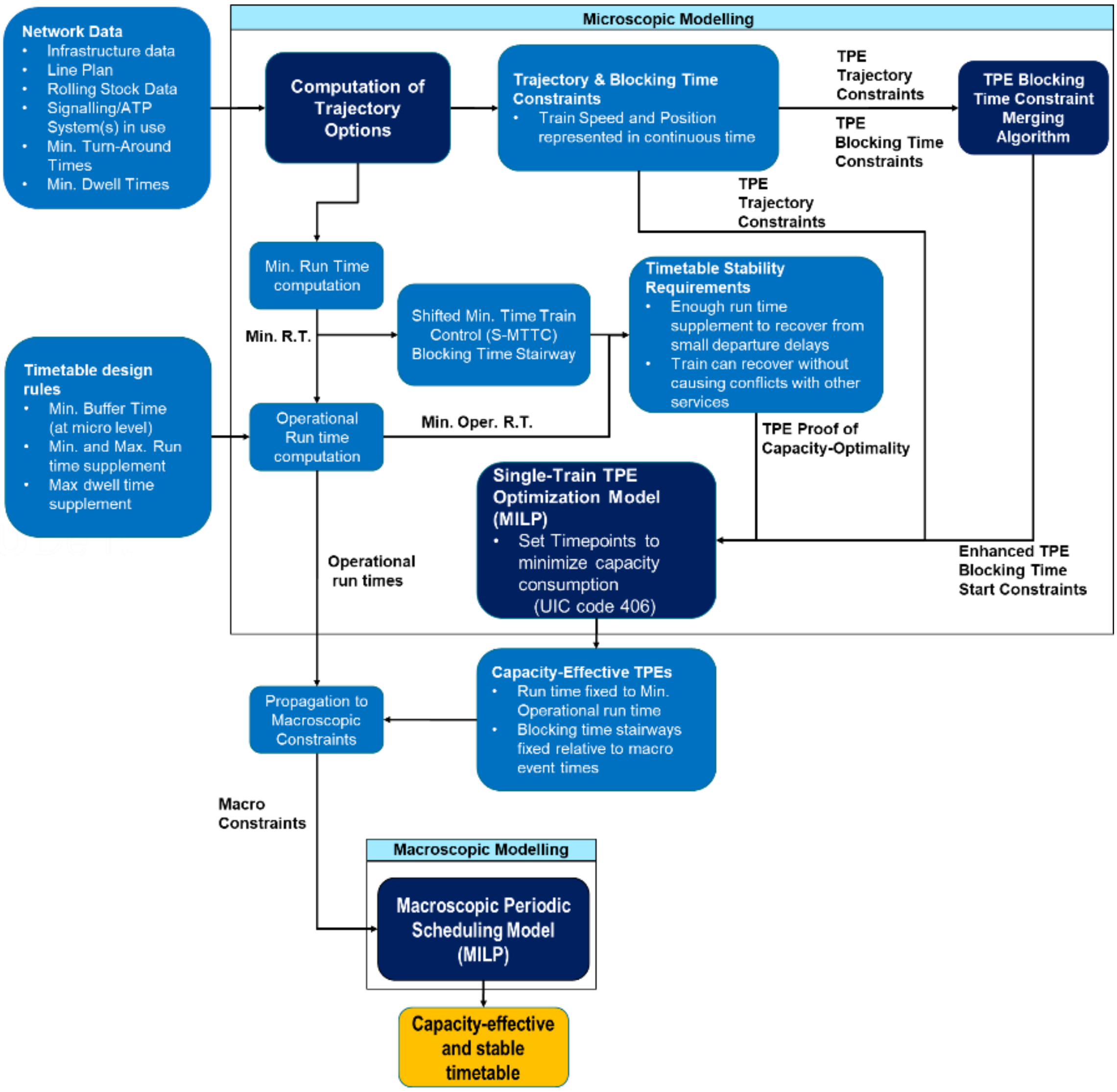
Microscopic models, in turn, are often used to repair the local infeasibilities in timetables generated by macroscopic models. Microscopic models represent the network at the individual operational process level. While guaranteeing that the timetable is feasible, this detailed representation normally increases computationally complexity to such an extent that the microscopic model cannot be scaled to network-level problems. Current microscopic models for conventional and DTG signalling systems use discrete speed levels to model the train's speed-run time and speed-braking distance relationships. Individual train trajectories are represented indirectly though passing times at block start/end locations, necessitating abstraction of the braking distance calculations. The abstraction of the train trajectory creates issues guaranteeing timetable feasibility in areas with very short blocks. To address this problem, the microscopic scheduling model for conventional DTG signalling should be applicable to any signalling layout. This requirement is particularly important for railways with moving block signalling (a form of DTG signalling with zero-length blocks) or European Train Control System (ETCS) Hybrid Level 3 (a DTG signalling system that can have very short virtual sub-sections).

An existing method for timepoint optimization for C-DAS or ATO systems (Wang et al. 2023) generates time windows guaranteeing a conflict-free and sufficiently stable timetable while facilitating energy efficient driving. This method works by identifying the critical bottlenecks and homogenizing the traffic within those sections of the network. While this approach resolves conflicts in macro-level timetables, it requires information about the timetable structure to know where to place the timepoints. Thus, the algorithm cannot be applied until after construction of the macro-level timetable.

A New Modelling Framework for Large-scale Timetabling Problems

To address this research gap, this thesis proposes a novel modelling framework for solving large-scale timetabling problems in networks where a C-DAS or ATO system regulates train speeds in real-time (see Figure 1). A timetable is deemed to be stable if trains have sufficient run time supplement to recover from small departure delays without arriving late at the next station or causing a conflict with another train. This stability criterion is used to identify a train driving strategy that maximizes capacity without having prior knowledge of bottleneck locations, and forms the basis of a timepoint optimization problem that can be decentralized at the individual-train level without sacrificing global optimality. A microscopic TPE optimization is proposed to generate capacity-optimal train planning rules. The TPE optimizer precisely represents the train's speed-distance trajectory through the network. This guarantees feasibility of the trajectory in any DTG signalling layout and allows more accurate representation of the continuous speed-braking distance relationship than current state-of-the-art methods. The results of the TPE optimization problems are used to calibrate the minimum headway constraints of a state-of-the-art macroscopic model.

Figure 1.
Overview of Conceptual Modelling Framework



The results of a case study show that the proposed model delivers the capability to efficiently compute feasible, stable, and capacity-optimal timetables for large networks with areas using ATO/C-DAS speed regulation. The proposed timetabling process is tested on a model of the South West Main Line in the United Kingdom between London and Southampton (~127km long) with ETCS Level 2 signalling (another form of DTG signalling) with short blocks. The EGTrain model reflects the physical track infrastructure in place in 2015 with block lengths optimized for ETCS, so it is not possible to perform a capacity assessment of the whole line with ETCS Level 2 versus the existing multi-aspect signalling. Nevertheless, the single-train TPE optimization problem is solved to optimality within 10 seconds for every service. The results also show that the macroscopic periodic scheduling model is capable of computing optimal timetables for the whole line in under ten hours (256 trains over four periods). When tested on the Main Slow Line within London, migration from 3-aspect signalling to ETCS Level 2 permits a 52% increase in the minimum buffer time between trains (from 27.8s to 41.8s), and a 16s decrease in minimum cycle time by 16s (from 47:22 to 47:06).

Based on these results, it can be concluded that the proposed scheduling framework is suitable for solving large-scale microscopic timetabling problems. The single-train TPE optimization model permits detailed representation of train dynamics and the constraints of any signalling system in use, guaranteeing accurate representation of blocking times under ATO and C-DAS-enabled operations. The TPEs generated are capacity-optimal at the microscopic level, and the associated microscopic minimum headway constraints can be converted into equivalent macroscopic arrive-arrive constraints. The macroscopic constraints can be entered in a state-of-the-art macroscopic model to produce conflict-free, stable, and capacity-effective timetables. Regarding future research, it is recommended that new decompositions be developed for the macroscopic tactical scheduling model, so that it can scale to even larger networks. The methodology should also be verified and validated for real-time traffic management applications so that IMs' operational capabilities are aligned with their tactical scheduling capabilities. Lastly, an algorithm that optimizes departure times and TPEs could further maximize timetable flexibility and provide opportunities for energy-efficient driving within the timetable structures computed by the macroscopic model. This research would prove beneficial because the proposed TPE optimizer prioritizes potential capacity over energy efficiency, and does not provide flexibility beyond the minimum amount needed to comply with train planning rules.

Table of Contents

List of Figures

List of Tables

Glossary of Terms

ATO: Automatic Train Operation

C-DAS: Connected Driver Advisory System

CP: Constraint Programming

Down/Up: UK terminology for the direction the train travelling on a line with respect to the line's major terminal (representing mile/kilometre 0). *Down* means the train is operating in the direction away from the terminal, while *Up* indicates that the train is operating in the direction of the terminal.

DTG: Distance-to-go signalling

EoA: End of Authority

ERTMS: European Rail Traffic Management System

ETCS: European Train Control System

EVC: European Vital Computer

CBTC: Communications-Based Train Control

IM: Infrastructure Manager

LP: Linear Programming

MA: Movement Authority

MILP: Mixed-Integer Linear Program

RBC: Radio Block Centre

RSSB: Railway Safety and Standards Board

SPAD: Signal Passed At Danger

TIMS: Train Integrity Monitoring System

TMS: Traffic Management System

TOC: Train Operating Company

TPE: Train Path Envelope

1. Introduction

1.1. Overview

Many rail infrastructure managers (IMs) are upgrading their railway networks to address expected increases in demand for rail transport and to achieve sustainability objectives. Several research and development (R&D) programmes at the international (e.g., Shift2Rail) and national levels (e.g., Smart-Rail in Switzerland) are exploring options for expanding the capacity of existing rail networks and reducing life-cycle costs. As part of its *Target 190plus (T190+)* programme (Network Rail, n.d.-b), Network Rail, the IM of the United Kingdom's mainline rail network, is seeking to improve sustainability, increase network capacity, and reduce life-cycle signalling costs by migrating from its existing multi-aspect signalling systems to *European Train Control System*[1] (ETCS) Level 2 (RSSB, 2022) and Level 3 Hybrid (Furness et al. 2017). This important initiative also addresses future needs by replacing two-thirds of signalling systems on the network that will reach end-of-life in the next 15 years.

Migration from multi-aspect signalling to ETCS helps Network Rail achieve its T190+ objectives by allowing trains to run closer together and reducing requirements for trackside equipment. In multi-aspect signalling systems, minimum train separation is a function of the number of block sections, which, in turn, depends on the number of available aspects. In contrast, minimum train separation in ETCS Level 2 and Hybrid Level 3 depend on a train's specific speed in real time. These ETCS levels are commonly referred to as *distance-to-go* (DTG) signalling systems because their movement is supervised based on the distance to the next danger point or infrastructure restriction (e.g., a preceding train, an unlocked switch, or a speed limit reduction). The migration of brake supervision from track to train also permits IMs to further increase capacity by deploying DTG signalling systems with very short blocks.

IMs seeking to increase network capacity are also deploying *Connected Driver Advisory Systems* (C-DAS) and *Automatic Train Operation* (ATO) systems in conjunction with migration to ETCS. Both systems guide trains more precisely through the network by using real-time information about the traffic plan. The traffic state is communicated to the train in the form of a *Train Path Envelope* (TPE) (ON-TIME, 2014 and Wang et al, 2023) that specifies conflict-free train paths. In the case of a C-DAS setup, speed advice is provided to a human driver who has control over the train's throttle and brakes. On trains equipped with an ATO system, an onboard computer makes the throttle and brake commands instead. These systems increase network capacity because IMs can reduce the *buffer time* required between trains while still maintaining a stable timetable.

To fully achieve the capacity benefits of multi-aspect and DTG signalling and ATO/C-DAS speed regulation, IMs also require a *tactical scheduling* model that produces capacity-effective timetables reflective of the constraints of these signalling systems, along with an automatic *traffic management system* (TMS) that helps dispatchers resolve conflicts occurring in everyday operations. While a potential issue in conventional signalling systems, addressing the misalignment between tactical and operational planning becomes even more critical to maximizing capacity when migrating to digital

[1] ETCS is the signalling component of the *European Rail Traffic Management System* (ERTMS) (RSSB, 2022), which includes other systems, but these terms are often used interchangeably by the industry to refer to the signalling system. In this document, the term *ETCS* is used to the refer to the signalling system.

operations. Furthermore, since migration to digital signalling systems generally occurs gradually, IMs will face scheduling problems if existing algorithms cannot represent the multiple signalling systems in use across the network. Indeed, maximizing capacity requires that algorithms in-use be capable of represent all signalling systems (and their associated safety constraints) on the network. Since busier network sections are more likely to be prioritized for migration to DTG signalling systems to increase capacity, the model should be able represent the continuous speed-headway relationship present in those systems as accurately as possible as well as reflect the operational constraints in areas with very short blocks. The tactical scheduling model should also consider the capabilities of the ATO/CDAS-TS timepoint configuration algorithm(s) in use, so that it can produce timetables that easily managed in daily operations. In addition, the scheduling model should be computationally simple enough for application to large-scale network timetabling problems.

Most existing research in tactical scheduling is at the *macroscopic* level (Caimi et al, 2011, and Sparing and Goverde, 2017), whereby the network is represented at the level of junctions and open lines, and minimum headway constraints reflect experienced-based scheduling norms. This abstraction of the infrastructure limits the ability of these models to represent signalling constraints at a high level of detail, thereby causing difficulties in finding solutions to large-scale timetabling problems in networks with very dense traffic. Moreover, most existing models that represent minimum headways at the level of individual block sections (known as *microscopic* models) use constraints for multi-aspect signalling (Lamorgese et al. 2017, and Leutwiler and Corman, 2022), such that minimum headway does not depend directly on speed. These models may be unable to produce optimized capacity-effective timetables for lines with ETCS Level 2 or Level 3 Hybrid, where braking distance is a function of train speed in continuous time. Existing research on scheduling for ETCS Level 2, Level 3 Hybrid, and similar DTG systems (such as ETCS Level 3 and Communications-Based Train Control) have headway constraints generated using black-box methods (Schlechte et al, 2022) or do not incorporate information about track geometry when calculating braking curves (Busuttil, 2023). More detailed representation of signalling constraints in microscopic models necessitates the inclusion of additional variables to represent specific operations, limiting their applicability to the large-scale timetabling problems faced by mainline IMs. A review of existing literature did not return any papers which assessed the performance of their models for scheduling traffic for ETCS Level 2 or Level 3 Hybrid signalling.

Most real-time rescheduling research is based on the constraints of three-aspect signalling, where the minimum train separation depends solely on the number and length of blocks between trains, and blocks are at least as long as the worst-case braking distance (D'Ariano et al, 2007 and Pellegrini et. Al, 2015). Current rescheduling methods for DTG signalling systems assume a discrete relationship between fixed speed levels and braking distance (Xu et al, 2017, Long et al. 2021 and Versluis et al, 2023) or no relationship at all (Janssens, 2022). These processes are not reflective of the fact that the DTG braking distance is dependent on the train's position and speed in continuous time.

Thus, the rail industry could benefit from the development of models that increase capacity by aligning tactical planning and real-time rescheduling models for both conventional and DTG signalling. IMs require both tactical and operational scheduling algorithms for areas with DTG signalling, particularly for networks with very short block lengths or with moving block signalling. The models should also exploit the capabilities of ATO and C-DAS systems to operate more precisely in bottleneck areas.

1.2. Problem Statement

A microscopic scheduling model could increase capacity on rail networks with DTG signalling systems for areas with very short blocks or virtual sub-sections. As noted, existing models for solving large-scale timetabling problems (referred to as *macroscopic models*) have minimum headway constraints based on norms derived from experience. These norms are an abstraction of processes at the operational level, and therefore do not guarantee that the timetable output is conflict-free and sufficiently stable for everyday operations. Microscopic scheduling models for multi-aspect signalling systems usually assume that train run time in a particular block is independent of the run time in the previous block. Existing research on scheduling in areas with moving-block or very short blocks (Janssens, 2022 and Versluis et al. 2023) retains this assumption when modelling run times. For this property to hold, it must be possible for a train to physically enter the block at maximum speed, then brake to a stop within the length of the block. This is a reasonable assumption for modelling areas with legacy three-aspect signalling but does not necessarily hold in areas with shorter blocks. Thus, the conclusions of those papers may not always hold, with the implication that research is needed to develop an algorithm that accurately models networks with DTG signalling and very short blocks or virtual sub-sections. Addressing this research gap would help IMs fully realize the capacity benefits of DTG signalling.

IMs could also benefit from a tactical scheduling model that can efficiently produce feasible and stable timetables for large networks, along with real-time rescheduling algorithms that can support signaller decision-making by proposing rescheduling plans that restore timetables created by the tactical scheduling plan. The state-art-methods for tactical scheduling and real-time traffic management contain different levels of detail when representing processes that occur on the railway network. The abstraction of signalling constraints at the tactical level could result in suboptimal use of available capacity in the network. To that end, aligning the network representation used at both the tactical and operational levels ensures that the constraints present in the tactical scheduling problem are reflective of the capabilities of the traffic management system(s) used by the infrastructure manager to manage delays occurring during regular operations.

In summary, this paper makes the following contributions to industry knowledge:

- Proposes a new methodology for constructing capacity-optimal TPEs at the train service planning stage and proves that the timetabling problem for a network with ATO or C-DAS speed regulation is not more complex than the corresponding macroscopic timetabling problem.
- Develops the first microscopic scheduling algorithm applicable to large networks which regulate speed using the C-DAS or ATO system. By aligning the approach for addressing both the scheduling and real-time traffic management problems, this scheduling algorithm would help IMs better design their timetables given constraints faced when managing real-time service.
- Proposes a TPE optimization model that is the first that guarantees conflict-freeness for areas with moving blocking signalling or optimized block layout for ETCS Level 2 or Level 3 Hybrid. This model is also the first that represents the continuous relationship between speed and braking distance, which is defining characteristic of DTG signalling systems.

- Demonstrates that the model can quickly find micro-feasible and stable solutions to large-scale timetabling problems or prove that a given line plan cannot be made sufficiently stable.

1.3. Research Objectives & Questions

As noted, the objective of this paper is to develop a general formulation of the microscopic scheduling problem for DTG signalling that aligns the tactical (medium-term) scheduling of services and real-time speed regulation via C-DAS or ATO systems to produce capacity-effective operations. This work involves extending the continuous speed-headway model proposed in Busuttil, (2023) for use on networks with discretized infrastructure, and consideration of the impact of track geometry (such as grades or curves) on braking distance. The general formulation is immediately applicable to the ETCS Level 2 and Hybrid Level 3 signalling systems that Network Rail seeks to implement, and can be used to approximate the non-discretized infrastructure in moving block signalling systems such as CBTC (Thales, 2020) and ETCS Level 3 Moving Block (Alstom, 2019). The computational complexity is reduced by identifying logical structures that reduce the dependency of headway constraints on knowing the exact trajectory of the train. The model is verified by confirming that the schedules produced are feasible and assessing the output using the simulator EGTrain (Quaglietta, 2014) on the South West Main Line (SWML) in the UK. Accordingly, the main research question is:

How can the tactical and operational planning levels be aligned to produce capacity-effective rail operations for conventional and digital railway operations?

Based on a predetermined line plan, the model can produce microscopically-feasible and stable timetables given infrastructure limitations.

The questions to be answered in this thesis are:

- What is the current state-of-the-art for train service planning and real-time rescheduling models used for conventional and DTG signalling?
- How can tactical and operational planning be consistent with train dynamics and signalling constraints of conventional and digital rail operations?
- What method can be used to align tactical and operational train service planning in either conventional or digital rail operations?
- What impact can an aligned tactical and operational train planning framework have on schedule quality?
- What recommendations should IMs consider for aligning scheduling constraints in models for the tactical and operational levels?

1.4. Thesis Structure

This thesis is organized into six sections, starting with this introduction which provides an overview of the context for this paper, research objectives, and potential contributions. Section 2 provides additional background on the functions of existing multi-aspect signalling, ETCS Level 2 and Hybrid Level 3, and general requirements for both timetabling and traffic management. Section 3 summarizes the state-of-the-art methods for tactical timetabling and real-time rescheduling for both multi-aspect and DTG signalling systems, and identifies research gaps in current literature. Section 4 proposes a methodology for more accurately representing speed-headway relationships in DTG signalling for use in both tactical scheduling and operational models. Application in a case study and the corresponding results are provided in section 5. In section 6, the paper concludes by examining the contribution of this research paper to academic knowledge and identifying potential future avenues for research.

2. Background

This section provides additional background on the function of signalling systems in use on many railway networks (including the UK rail network), and associated timetabling and traffic management requirements. Section 2.1 explains the multi-aspect signalling systems currently in use on most mainline railway networks, and their limitations from capacity and cost perspectives. Section 2.2 outlines the functions of ETCS Level 2 and Hybrid Level 3, and their capacity and cost advantages relative to multi-aspect signalling. Section 2.3 provides an overview of ATO and C-DAS speed regulation. Section 2.4 discusses the role of timetable design and traffic management in providing a high level of service to the *Train Operating Companies* (TOCs) using the network, and the railway's end-customers.

2.1. Conventional Signalling Systems

Multi-aspect signalling systems[2] depend entirely on track-based equipment to detect that the track is clear of trains, and on aspects displayed by lineside signalling to communicate the *Movement Authority* (MA) to the driver. The MA indicates to the driver how far they are authorized to travel. In the UK, three or four-aspect signalling is most common. Under three-aspect signalling, the lineside signals can display one of the following aspects:

1. *Danger* (Red) indicates that the train is not authorized to pass the signal.
2. *Caution* (Single Yellow) indicates that the train is not authorized to pass the next signal. The train needs to reduce its speed so that it can stop if the next signal is displaying *Danger*.
3. *Clear* (Green) indicates that the train is authorized to pass the next two signals and proceed at the line speed.

On lines with three-aspect signalling (Figure 1), trains must always remain at least two blocks apart to travel at line speed. To prevent trains passing a signal at *Danger* (an event known as a Signal Passed At Danger, or *SPAD*), the blocks must be long enough that any train using the line can brake from the line speed to standstill within the length of one block.

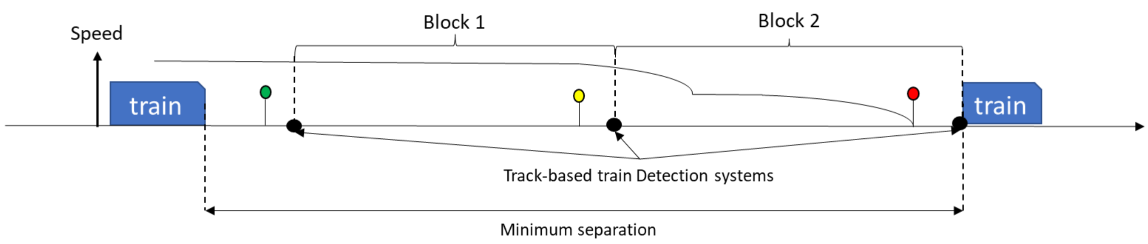


Figure 1: Minimum following distance for three-aspect Signalling (adapted from Busuttil, 2023)

[2] The design of the multi-aspect signalling systems in the UK is governed by the Railway Safety and Standards Board's (RSSB) standard for *Lineside Signals, Indicators and Layout of Signals* (RSSB, 2023b).

Signals on UK lines with four-aspect signalling include an additional *Preliminary Caution* (Double Yellow) aspect between the signals showing *caution* and *clear* (Figure 2). On lines with this standard four-aspect signalling sequence, trains must always remain at least three blocks apart to travel at maximum speed (as shown in Figure 2). To ensure that trains can stop in time for a danger aspect, the blocks must be long enough that any train using that line is able to brake from the line speed to standstill within the distance from the first caution aspect displayed and the signal displaying danger.

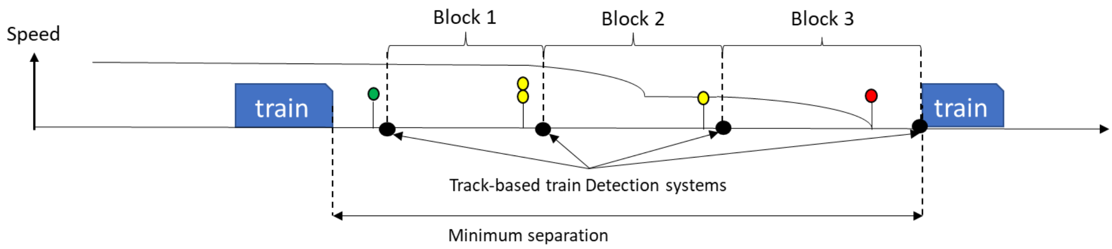


Figure 2: Minimum following distance for standard sequence UK four-aspect signalling (adapted from Busuttil, 2023)

Multi-aspect signalling systems have limitations that can reduce line capacity and increase life-cycle costs. As noted, multi-aspect signalling works through wayside signals that transmit movement authorities to trains as they pass. Since these signals are placed at fixed locations, they can only be read by trains travelling by, so it is not possible to update a train's MA between signals. In situations when trains have minor delays, this update delay can increase the time required to restore the planned schedule.

Furthermore, the limited number of aspects imposes a minimum spacing requirement for signals, with the minimum block length determined based on the worst-case braking distance of any train that could possibly use the line. This minimum block length requirement exists for safety reasons because trains must always have enough distance to brake from the line's speed limit to standstill between the first caution aspect shown and the signal at danger (RSSB, 2023b). Since the signals are track-based, they cannot consider train-specific characteristics when determining which aspect to display. These constraints result in a loss of capacity on lines with mixed rolling stock because the line's block spacing (and therefore minimum following distance for all trains) needs to accommodate the train with the worst braking performance on the line.

2.2. Migration to ETCS Level 2 and Hybrid Level 3

To address the limitations inherent in legacy multi-aspect signalling, some of Network Rail's busier mainline routes are being migrated to ETCS Level 2[3]. This migration involves the movement of the braking distance computation from the track to the train, eliminating the need for lineside signalling. On lines with ETCS Level 2 signalling, the movement authority is communicated to trains via radio messages from the *Radio Block Centre* (RBC) and information about the route (such as track geometry and speed restrictions) from balises mounted between the running rails.

The ETCS movement authority consists of an *End of Authority* (EoA) and information about the train's route, such as speed limits and track geometry. The ETCS EoA represents the furthest location the train is permitted to travel, which is analogous to the location of the signal displaying danger in

[3] A full explanation of the functions of ETCS Level 2 signalling in the UK rail network is available in the RSSB's *ERTMS Handbook* (RSSB, 2022).

multi-aspect signalling. The information contained in the ETCS MA is used by the train's onboard *European Vital Computer* (EVC), which verifies that the train is complying with restrictions imposed by the signalling system. The EVC does this by calculating a *braking curve*, representing the distance required for the train to brake to stop before the EoA. The braking curve is calculated using train-specific characteristics (such as maximum braking ability), route information (such as gradients and speed restrictions), and the distance to the EoA. If the train driver does not keep the train under the braking curve, the EVC will force the train to stop (referred to as *tripping* the train).

An example of how the ETCS braking curve is calculated and enforced is shown in Figure 3.

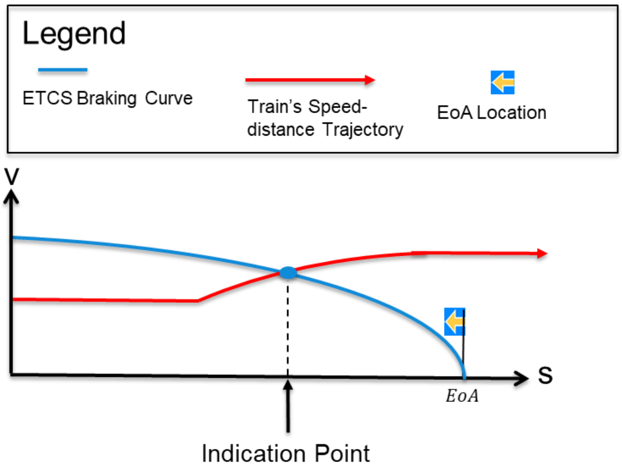


Figure 3: Calculation and Supervision of the braking curve in ETCS

The train's EoA is at the location of a blue-and-yellow ETCS *block marker*. This block marker[4] provides the driver with a visual reference of where to stop if they cannot proceed into the next block. The EVC calculates the braking curve (blue) from the location of the EoA. At the location where the train's trajectory (red) intersects the braking curve (known as the *indication point*), the onboard system notifies the driver that they must start braking. The location of the indication point depends directly on the speed of the train: if the train is travelling at a higher speed, the indication will be further away from the EoA location.

ETCS Level 2 system increases network capacity relative to multi-aspect signalling by permitting the minimum following distance to be calculated based on the speed and particular braking capabilities of the following train, and by eliminating the minimum block length requirements. ETCS Level 2 is classified as a DTG signalling system because the train calculates its *distance-to-go* to the indication point.

Figure 4 shows how minimum separation for two trains is calculated for ETCS Level 2.

[4] Some lines may have EoA locations that are not accompanied by a block marker (RSSB, 2022).

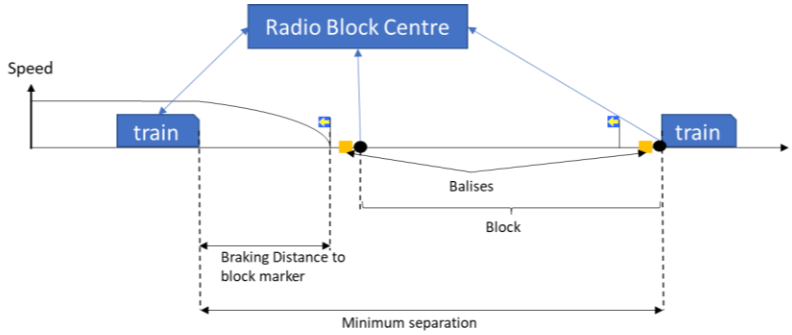


Figure 4: Architecture of ETCS Level 2 Signalling (adapted from Busuttil, 2023)

As shown in Figure 4, the head of following train is at the brake indication point for the block given its current speed. The train position used by the EVC to enforce the braking curve is subject to odometry error, so the block markers are usually offset from where the actual block boundaries (the black circles) are located. As soon as the first train's rear has released the block, the second one can pass the indication point for its current speed.

In the longer term, Network Rail and some other IMs are seeking to migrate lines to ETCS Level 3 Hybrid signalling (proposed by Furness et al. 2017). Relative to ETCS Level 2, the main difference is the addition of a Train Integrity Monitoring System (TIMS) that verifies that the train has not broken apart. If a train's length is known, and its integrity has been confirmed by its TIMS, then it becomes possible to infer the position of the train's rear based on the position of the head of the train. This information, when communicated to the RBC, allows the signalling system to determine which sections of track have been cleared by the train.

The position report-based release is akin to processes currently used in *Communications-Based Train Control* (CBTC) (Thales, 2020) and ETCS Level 3 (Alstom, 2019) signalling to confirm track vacancy. This functionality allows open track to be represented without any discretization (known as *moving-block* signalling), or to be divided into virtual sub-sections that can be released through the confirmation of train integrity plus the position report.

The presence of secondary track-vacancy detection systems in ETCS Level 3 Hybrid also permits operation of trains not fitted with TIMS. These trains (known as *non-integer trains*) are completely reliant on the secondary track vacancy detection for releasing their rear (as in ETCS Level 2). For trains with TIMS (known as *integer trains*), the more permissive of the position report and the secondary detection is used to release the rear.

Network Rail is also investigating application of ETCS Level 3 Hybrid signalling as part of the T190+ programme to reduce life cycle costs below levels possible with ETCS Level 2. As noted, the main advantage of ETCS Level 3 Hybrid relative to Level 2 is that the migration of the vital track-clear detection function from track to train reduces the amount of trackside equipment needed to achieve the desired capacity increases. ETCS Level 3 Hybrid does not increase capacity relative to Level 2 because the onboard braking distance computation process is the same, and both systems can approximate non-discretized infrastructure with very short blocks or sub-sections. Network Rail plans to implement ETCS Hybrid Level 3 using short virtual sub-sections. This is being done to minimize the need for changes to operating rules versus state-of practice.

2.3. ATO/C-DAS Speed Regulation

Railway IMs are looking at implementing C-DAS and ATO systems to improve capacity and punctuality. The essential component of these systems is the provision of real-time traffic information to trains so that they can operate in a more precise and capacity-effective manner during disturbances (ON-TIME, 2014, and Wang et al, 2022). The automated driving function in ATO systems can also increase capacity compared to a human driver using C-DAS because ATO reduces the train's response time to changes in the MA or traffic state. While ATO systems are already widely used in metro systems with homogenous and simple traffic patterns, they are not as common in mainline rail networks (Yin et al, 2017), in part due to their more complex traffic patterns. ATO and C-DAS architectures are not standardized across the industry: the level of integration between the TMS responsible for setting computing the train's trajectory and deciding the driving mode (i.e., whether to accelerate, brake or coast) varies across systems. A full explanation of the possible architectures for ATO and C-DAS Systems is available in ON-TIME, (2014) and Wang et al, (2022).

In ATO-over-ETCS reference architecture (explained in Wang et al, 2022), the TMS is responsible for maintaining a dynamic timetable (known as a *Real-Time Traffic Plan*, or RTTP) with the planned arrival and departure times at each station, and the planned routes and orders of the train at junctions (Quaglietta et al. 2016). This information is fed to the *ATO-Trackside* (ATO-TS) system that sets *timepoints* for the trains to adhere to. The timepoints provided by the TS are imposed as constraints on the trajectory optimization problem that is solved by the *ATO Onboard* (ATO-OB) system (Albrecht et al, 2013). The ATO-OB directly controls the train's throttle and brake.

For a traffic plan to be feasible, the timing windows that make up the TPE provided to trains need to be configured such that each train's ATO-OB can find a feasible trajectory gives its own timing constraints, and no two trains' OB systems compute optimal trajectories which cause a conflict. The TPEs need to be stable to station dwell time variations, so that a train with a small departure delay can still reach the next station on time (modelled as a timing constraint in the trajectory optimization problem) without causing a conflict with another train. For any given inter-station journey, the timepoints provided by the ATO -Trackside to the Onboard system has to contain the *Shifted Min Time Train Control* (S-MTTC) trajectory, and an *Energy-Efficient Train Control* (EETC) Trajectory (Wang et al, 2023)[5]. The S-MTTC trajectory represents the case where the train's departure is delayed up to the run-time supplement, forcing the train to operate with the minimum technical run time between the two stations. The S-MTTC trajectory must be contained within the path envelope to ensure that the train can recover from small departure delays without causing a conflict with another train. If S-MTTC were not included, the timetable would not be sufficiently stable for daily operations. The EETC trajectory is the train's trajectory if it departs on time and arrives at the next station at the scheduled arrival time. The EETC trajectory, calculated by the Onboard system, must respect any timing points specified by the trackside system. This feature gives the trackside system some flexibility to alter the shape of the TPE to better allocate capacity.

[5] Wang et al, (2023) also includes a trajectory called *Restricted Maximum Speed (RMS)*, which has the lowest energy consumption assuming the train cannot coast.

2.4. Timetabling Requirements

Timetable construction is an important element of railway operations and achieving the capacity and service performance improvements offered by ETCS. The *passenger timetable* consists of all revenue passenger services (arrivals and departure times at stations) offered to the travelling public. In addition to ensuring service feasibility, this timetable is used by customers to make informed decisions about their journeys and evaluate service performance. IMs must also be able to effectively construct timetables to accommodate increasing demand for both passenger and freight services on the existing network. This objective must be accomplished while still producing a *working timetable* (WTT) that is feasible (i.e., without conflicts between trains), and stable enough that minor delays can be resolved without significant impact to revenue-generating passenger or freight services (Network Rail, n.d.-c).

Accomplishing these objectives on a network as complex as the UK's requires scheduling tools that accurately represent the constraints of the signalling systems in use, and the processes occuring in everyday operation. Otherwise, the timetable may not make use of all available capacity. The timetable should also be aligned with the capabilities and limitations of systems and processes used to manage revenue service. If the traffic management algorithm(s) used contain more conservative assumptions about minimum headway than those used for scheduling, the tactical-level timetable could be difficult to manage in regular operations.

3. Literature Review on Tactical and Operational Scheduling Models

The literature review covers the state-of-the-art for modelling DTG systems (including ETCS Level 2, Hybrid Level 3, Moving Block, and Virtual coupling) at the service planning and real-time rescheduling levels. This review was performed using keyword searches to find relevant articles, and by investigating the citations of the articles found. The following keywords are used:

Topic	Keywords
Modelling DTG Signalling Constraints	'ETCS Level 2', 'CTCS-3', 'Moving Block', 'CBTC', 'ETCS Level 3', 'Quasi-Moving Block', 'Virtual Coupling', 'Distance-to-go', 'signalling', 'service planning', 'real-time rescheduling', 'traffic management'
State-of-the-art job-shop scheduling (incl. service planning and rescheduling under multi-aspect signalling)	'Job-shop scheduling', 'Alternative Graph', 'scheduling', 'service planning', 'Real-time rescheduling', 'Constraint Propagation', 'disjunctive constraints', 'Periodic Event Scheduling Problem (PESP)', 'Flexible-PESP (F-PESP)'
ATO/DAS Speed Regulation	'Train Path Envelope (TPE)', 'Timing Points', 'Timing Windows', 'Automatic Train Operation (ATO)', 'Driver Advisory System (DAS)', 'Connected Driver Advisory System (C-DAS)'

The literature review is organized into four sections. Section 3.1 outlines state-of-the-art methods for the tactical scheduling problem. Section 3.2 discusses state-of-the-art methods for modelling and solving the real-time rescheduling problem. Section 3.3 reviews state-of-the-art methods for ATO/C-DAS speed regulation. Section 3.4 summarizes the state of existing research and identifies relevant gaps in literature for both the tactical scheduling and real-time rescheduling problems.

3.1. State-of-the-art modelling of tactical scheduling

The literature review returned eight papers that focus on tactical scheduling (Table 1), which are then categorized according to five characteristics:

- **Periodic or Non-Periodic:** Indicates whether the model is designed for periodic scheduling (where services repeat after a certain interval), or non-periodic scheduling.
- **Level of Detail:** Indicates the level of detail used to represent the network. *Macro* indicates that the model only represents arrival and departure events at stations. *Micro* indicates that the amount of detail shown is sufficient to guarantee conflict-freeness. If the model is microscopic, the applicable signalling systems are also listed.
- **Headway modelling:** *Norms* indicate that the minimum headway is based on values derived from experience, as opposed to blocking times. *Fixed Speed* indicates that braking distance does not vary with the train's speed. *Continuous* indicates that the algorithm models the movement of the braking curve in continuous time. In the case of *Discrete Speed Level,* the train can choose from a series of run times and headways corresponding to different cruising speeds. *Black Box* indicates that the process is not specified in detail.
- **Infrastructure discretization:** Indicates how the track release process can be represented by the model. For *Continuous,* the model can represent non-discretized infrastructure without

approximation. If the model cannot represent continuous infrastructure, the minimum block length required for the model to guarantee conflict freeness is indicated. *Black-Box* indicates that the process for generating the constraints in not specified in detail. *N/A* is indicated for macroscopic models.

- **Objective:** Notes the specific objective to be maximized/minimized.
- **Solving Method:** Indicates the method used to solve the model. If the model was created in Mixed-Integer Linear Programming (MILP) form and solves using stock methods, it is considered to use a *Centralized MILP* method. Customizations made to the solving process, if any, are noted.

Most existing research on train service planning and rescheduling problems focus on multi-aspect signalling, where headway is not directly dependent on speed. Existing algorithms represent the problems as a no-wait job shop scheduling problem (Mascis and Pacciarelli, 2002). In this problem, a group of 'jobs' are scheduled on a set of 'machines' subject to constraints which ensure that only one job is assigned to a particular machine at a time. In the railway scheduling problem for multi-aspect signalling, the track vacancy detection segments are the 'machines' that process trains (jobs), subject to single-train minimum running time constraints and disjunctive constraints preventing two trains occupying the same segment at a given time. Typical rescheduling measures include retiming trains, reordering trains at junctions, and rerouting trains in areas with multiple options.

Many railways operate *periodic* timetables, whereby the scheduled services are repeated at regular time intervals (usually a factor or multiple of 60 minutes). For these networks, the tactical planning problem is often formulated as a *periodic event scheduling problem* (PESP) (Serafini and Ukovich, 1989), with discrete train arrival and departure event times for each line selected within the period subject to a macroscopic run time and minimum headway constraints. The idea behind macroscopic modelling to is find a set of departure and arrival events that is either conflict-free (*microscopically feasible*), or that can be modified easily to produce a feasible timetable. Odijk (1996) proposes a methodology for generating MILP constraints for the PESP problem. The flexible-PESP (FPESP) problem (Caimi et al, 2011) specifies event times as a time window (rather than occurring at a single discrete time), improving the likelihood of finding a microscopically feasible train routing, and allowing for better quantification of timetable robustness. The authors also propose a 'flexbox'-based decomposition of the problem, whereby related events (such as a series of timed connections at a station) are grouped together to represent their robustness to initial delays (i.e., the amount of initial delay that can be absorbed by the events in a box). Sparing and Goverde (2017) propose a MILP model of the PESP problem for generating stable timetables to minimize the cycle time of a periodic timetable. This method was extended by Bešinović et al, (2019), which proposes methods for improving the likelihood of finding a robust timetable by reducing the number of services scheduled, relaxing the regularity requirements for lines with multiple services scheduled per period, or relaxing the planning norms for run times. Herrigel et al, (2018) proposes a hierarchical decomposition method, involving categorization of services based on scheduling priority, with each group being scheduled sequentially. To increase the likelihood of finding a feasible solution in later iterations, the scheduled times of higher-priority services can be modified within a prespecified timetable margin.

Some methods for scheduling periodic services have formulations more common in non-periodic timetabling, with additional constraints to ensure that consecutive services from the same line have event times that repeat at regular intervals (or close to regular intervals). Lamorgese et al, (2017) proposes a nonperiodic scheduling problem for lines with both nonperiodic and periodic services. The regularity constraints for periodic line services can also be relaxed to allow for small variation in event times from period to period (which is not possible in the traditional PESP formulation). The allowance for small deviations (which can be penalized in the objective function) increases the likelihood of finding a feasible timetable on lines that also have non-periodic services. Sartor et al. (2023) also proposes a nonperiodic strategic timetabling model for scheduling a line in Norway where the infrastructure manager planned on running a periodic passenger line and some non-periodic freight services. A quasi-periodic version of the model allows the times of services on periodic lines to deviate from the regular event time by a few minutes while remaining compliant with planning rules. Grafe and Schöbel, (2021) proposes a model called *Periodic Timetabling in Aperiodic Network* (PTTA) that is equivalent to PESP. In this formulation, the event times are set at the level of the individual services that make up a single iteration of the period (repeated for as many periods necessary to produce enough of the timetable to guarantee feasibility), subject to macroscopic run time and headway constraints and strict periodicity constraints on the event times for services on the same line service in successive periods. The goal of this formulation is to allow real-time traffic management algorithms (which are designed for a non-periodic problem) to be used for tactical scheduling. Nonperiodic models for periodic timetabling have a structure that more closely resembles the associated real-time rescheduling problem (as train delays are not necessarily periodic), which may be beneficial for creating consistent tactical planning and real-time rescheduling policies.

While railway IMs have shown interest in moving block signalling, there is comparatively little research on optimizing scheduling or rescheduling in this environment. Schlechte et al. (2022) developed a microscopic approach to schedule lines with Moving Block signalling (called *velocity expansion*). In this model, minimum headway depends on both the leader's and follower's speed profiles. While this model can produce feasible schedules, the calculation method for minimum headways depends partly on an 'oracle' to calculate the minimum headway between the two trains at the start and end of each section, and at the time that the following train enters the section. This intermittent formulation of the constraints means that model correctness may depend on the modeller's choice of locations for behavioural nodes. The triple-Big-M formulation of the headway constraints has a weak linear relaxation, possibly affecting computation time.

Busuttil, (2023) developed a continuous-space model for scheduling trains under ETCS Level 2, Moving Block, and Virtual Coupling with the dynamic safety margin proposed in Quaglietta et al. (2022). The paper represents the area of track occupied by a train as a function of its trajectory in continuous time, allowing for calculation of minimum headways using analytical methods. The method was used to evaluate the capacity of Virtual Coupling versus plain Moving Block at the service planning level. While the headway constraints capture the continuous relationship between speed and headway, they have multiple Big-Ms (as in Schlechte et al, 2022) and thus have weak linear relaxations which could cause operational difficulties finding good solutions quickly. The method also requires precise definition of the trajectories (and therefore the running time) to model the release process. The train-centric nature of DTG signalling necessitates more precise trajectory representation than more infra-centric signalling systems, however, the limited number of run time options (compared to state-of the-art speed level models) could result in a thin solution space or infeasibility when solutions do exist in real-life. In addition, the analytical methods for calculating minimum headways do not consider the impact of track geometry (such as grades or curves) on the train's stopping distance. While the proposed model is applicable to scheduling lines with ETCS Level 2 or Hybrid Level 3, the paper did not perform a capacity assessment for either of those signalling systems.

The microscopic train service planning problem is usually decomposed into smaller, easier-to-solve problems, which can be used to modify the original problem to guide it toward feasible solutions. This approach is necessary because the problem grows exponentially with network size and the number of trains to be scheduled, to such an extent that solving times with a commercial solver are unacceptably high. Bešinović et al. (2016) proposes a bi-level scheduling algorithm for constructing robust timetables under fixed-block signalling systems. The algorithm uses a macroscopic model of the infrastructure to produce a rough timetable, that is then checked for feasibility at the microscopic level. If the proposed timetable cannot be made microscopically feasible, the constraints of the macro model are then adjusted to improve the likelihood of finding a micro-feasible timetable. Schlechte et al. (2022)[6] and Sartor et al. (2023) use a row generation approach, such that some or all signalling constraints are initially omitted, and are then added only as needed for feasibility. Methods that do not initially include all signalling constraints exploit the fact that many of them will be non-binding in the optimal solution, so their omission from the initial problem reduces the computational cost of each simplex iteration.

Lamorgese et al. (2017), Leutwiler and Corman, (2022) and Leutwiler and Corman, (2023) use a Benders decomposition (Benders, 1962), whereby solutions to subproblems are used to generate problem-specific feasibility or optimality cuts to the global problem. The generated cuts are from the irreducible infeasible subset of constraints, which is the smallest group of constraints that causes infeasibility in the subproblem. In Lamorgese et al. (2017), the master problem involves setting timings over sections of open track, while the subproblem (at the station levels) involves assigning routes to trains. When a subproblem cannot find feasible routings at a station, Benders cuts are generated and added to the model. Lamorgese et al. (2017) also exploits the fact that the individual station area subproblems are all independent of each other, so they can be solved in parallel. Leutwiler and Corman, (2022) propose a logic Benders decomposition (first proposed by Hooker and Ottosson, 2003) for the nonperiodic scheduling problem that is combined with a geographic decomposition. In the master problem, the complicating variables that connect the different geographic areas are selected, allowing the different geographic areas to be evaluated in parallel. Each of the subproblems is then assessed for feasibility, and an infeasibility in a subproblem is used to generate a Benders cut removing the infeasible partial solution from the solution space. Leutwiler and Corman (2023) propose a logic Benders decomposition for the problem of producing modified schedules in response to short-term operational needs. The addition of logic Benders cuts can result in more efficient information propagation, but the process needs to be tailored to the specific problem.

[6] Schlechte et al. (2022) described their method as a column-and-row generation process, which initially omits the ordering variables, but the initial omission of all signalling constraints already disconnects them from the objective (until a relevant signalling constraint is added).

Table 1: Research on tactical scheduling

Reference	Periodic/ Non-Periodic	Level of Detail	Speed Modelling	Infrastructure Discretization	Objective	Solving Method
Bešinović et al. (2016)	Both	Bi-Level	Fixed speed level	Worst-case braking distance	Optimize trade-off between robustness and short run times	Micro-Macro scheduling algorithm, with robustness evaluated via Monte Carlo simulation
Bešinović et al. (2019)	Both	Micro: 3-aspect	Fixed speed level	Worst-case braking distance	Minimize cycle time	MILP model, with heuristics for cancelling services and relaxing scheduling norms
Busuttil, (2023)	Both	Micro: MB/VC	Continuous	Continuous	Minimize cycle time	Centralized MILP
Caimi et al. (2011)	Periodic	Macro	Norms	N/A	Multi-objective	Centralized MILP
Grafe and Schöbel, (2021)	Periodic	Macro	Norms	N/A	Minimize total run time extension	Centralized MILP
Herrigel et al. (2018)	Periodic	Macro	Norms	N/A	Multi-objective	Hierarchical scheduling of trains based on priority
Lamorgese et al. (2017)	Both	Micro: 3-aspect	Norms	Worst-case braking distance	Minimize total cost	MILP with geographic decomposition and logic Benders decomposition
Leutwiler and Corman, (2022)	Non-Periodic	Micro: 3-aspect	Norms	Worst-case braking distance	Minimize sum of run time delays	MILP with geographic decomposition and logic Benders decomposition
Leutwiler and Corman, (2023)	Non-Periodic	Micro: 3-aspect	Fixed speed	Worst-case braking distance	Minimize planning deviation	MILP with logic Benders decomposition with combinatorial Benders cuts
Odijk, (1996)	Periodic	Macro	Norms	N/A	Find feasible solution	Centralized MILP
Sartor et al. (2023)	Both	Micro: 3-aspect	Fixed speed	Worst-case braking distance	Multi-Objective	MILP solver with logic benders decomposition
Schlechte et al. (2022)	Non-Periodic	Micro: MB	Black Box	Black box	Minimize deviation from desired times	MILP with column- and row-generation
Sparing and Goverde, (2017)	Periodic	Macro	Scheduling Norms	N/A	Minimize Cycle time	Centralized MILP

3.2. State-of-the-art modelling of real-time rescheduling

The literature review returned 12 papers on real-time rescheduling (Table 2), which are then categorized based on the following characteristics:

- **Applicable signalling system:** The signalling system(s) that form the basis for determining the constraints.
- **Rerouting Possible:** Indicates if the model can reroute trains (in additional to retiming and reordering).
- **Speed Modelling:** Indicates if train speeds are a variable in the model. *Fixed* indicates that speed is not an exogenous variable and is only indirectly chosen through the passing times. *Discrete* indicates that the model can choose from a discrete number of trajectory options. *Continuous* indicates that properties of trajectory (such as cruising speed or acceleration) are represented by continuous variables.
- **Objective:** Notes the optimization objective of the solving process. *Max Secondary Delay* refers to the maximum delay experienced by a train when exiting the modelled area. *Weighted delay* refers to objectives which include arrival event times in the modelled area.
- **Solving Method:** The method used to solve to the problem, including the type of problem to be solved (MILP, LP, MINLP, etc.) and any methods used to decompose the problem.

Most existing research on real-time rescheduling is for multi-aspect signalling systems, where headway does not depend directly on speed. D’Ariano et al. (2007) proposed an Alternative Graph (AG) model to re-time and reorder trains after disturbances with the objective of minimizing the maximum secondary delay (equivalent to the makespan in the generic scheduling problem). This model (named ROMA) was extended in D’Ariano et al. (2008) to include local rerouting options in station areas. Implemented in MILP form, this model uses a truncated branch-and-bound procedure that terminates after a pre-specified solving time. ROMA was further adapted in Corman et al. (2010), which proposed a tabu-search method for finding feasible routes before solving the reordering and retiming problems. Lusby et al. (2013) developed a rescheduling model formulated as a set-packing problem, with train trajectories determined outside the model. Since headway constraints are represented as clique constraints, the solving process uses a column-and-row generation process to remove trajectory options that are known to be suboptimal from the model. While the network in the case study (Bern Hauptbahnhof in Switzerland) has multi-aspect signalling, the formulation is flexible enough to accommodate the constraints of DTG signalling systems. Luan et al. (2018) proposed a bi-level solving process for multi-aspect signalling with detailed representation of train trajectory constraints. The bi-level process uses a genetic algorithm to select train orders in the first stage, which is then fed into the second-stage trajectory optimization. The trajectory optimization problem is formulated as a Mixed-Integer-NonLinear-Programming (MINLP) problem with exact specification of the train’s position and speed in continuous time. Luan et al. (2020) propose a distributed optimization method for the real-time rescheduling problem over large networks. The authors propose decomposing the problem into local geographic areas and time intervals, each of which can be solved in parallel. The master problem in the solving process involves fixing the values of complicating variables included in multiple subproblems, after which the local subproblems can be solved in parallel. Lamorgese and Mannino (2019) propose a Benders decomposition method for solving the real-time rescheduling problem for multi-aspect signalling. Pellegrini et al. (2014) develop a MILP-based rescheduling tool for retiming, reordering and rerouting trains with the objective of minimizing total weighted delay. The MILP version of this model (named RECIFE-MILP) was reformulated in Pellegrini et al. (2015), Pellegrini et al. (2017) and Pellegrini et al. (2019) to improve computation time. Most recently, a constraint-programming (CP) formulation of the RECIFE model and a hybrid MILP-CP formulation were proposed by Marlière et al. (2023). The exploitation of constraint propagation to reduce the size of the solution space as variables are fixed could be useful for reducing the values of the Big-M coefficients, whose weak relaxations are a common cause of computational issues.

Nearly all microscopic models for multi-aspect signalling assume that the run time in each block section is independent of the run time in the prior and subsequent blocks on the train's route. This property means that it should always be possible for a train to brake from the maximum line speed (equivalent to achieving the minimum run time in the previous block) to a complete stop within the length of a single block. Although consistent with the safety requirements for three-aspect signalling (where the brake indication is only given 1 block before the EoA), this assumption may not hold in networks with signalling systems that permit block lengths shorter than the worst braking distance. This limits the applicability of these algorithms to resolve conflicts in networks where drivers can be shown four or more aspect sequences on approach to the EoA.

Some research has proposed MILP representations of ETCS Level 2 for the purpose of real-time rescheduling. Xu et al. (2017) propose a real-time rescheduling model for a high-speed line with CTCS-3 (functionally equivalent to ETCS Level 2), with signalling constraints for five different speed levels, and trains only complying with the constraint applicable to their speed level. While extended by Xu et al. (2021) to include the ability to reroute trains, this model still relies on discrete speed levels to represent speed-dependent headways. Liu et al. (2021) propose a bi-level optimization rescheduling process with a headway constraint representation requiring fewer variables and constraints. This approach reduced computation times. However, modelling ETCS Level 2 operations with discrete speed levels prevents it from realizing all available capacity. Long et al. (2019) propose a cell-based rescheduling model for scenarios with a speed restriction on a high-speed line, where the number of cells occupied by a train can be varied depending on its speed. While this approach reduced delays compared to methods that do not consider the relationship between speed and headway, the methodology relies on discrete speed levels to represent signalling constraints. The intermittent checking of signalling constraints could result in sub-optimal capacity utilization. While capable of resolving conflicts, this model may not fully use available capacity because it assumes that the signalling system can only enforce the five defined speed levels, so headway constraints must be based on the worst-case scenario for that level. The lack of precise definition of train trajectories could also cause issues guaranteeing the feasibility of the run times in areas with very short blocks.

The traffic management models proposed by Janssens, (2022) and Versluis et al. (2023) are adapted from models proposed for three-aspect signalling (D'Ariano et al, 2008 and Pellegrini et al, 2015). While both models redefine the braking distance calculation (i.e., approach time) to reflect the train-based brake supervision of DTG signalling, they do not directly represent the train's trajectory, which could result in overestimation of braking distances. In Janssens, (2022), the minimum headway does not vary with speed, so it must be based on the train's braking distance from line speed. If the train is travelling at a slower speed, the braking distance does not decrease as it would in practice. In Versluis et al. (2023), there are only two fixed-speed options to choose from (the maximum speed or scheduled speed). To avoid a thin solution space, the maximum speed headway holds for any speed above the scheduled speed, and the scheduled speed headway is assumed to always hold. Since neither model represents the train's trajectory at a high level of detail, the headway constraints must be based on the worst-case scenario for the given speed options available.

Reference	Applicable Signalling Systems	Rerouting Possible	Speed Modelling	Objective	Solving Method
Corman et al. (2010)	3-aspect	Yes	Fixed	Minimize max secondary delay	Tabu-search for routing, then timings and orderings solved with centralized MILP
D'Ariano et al. (2007)	3-aspect	No	Fixed	Minimize max secondary delay	MILP with branch and bound solving
D'Ariano et al. (2008)	3-aspect	Yes			
Janssens, (2022)	DTG	No	Fixed	Minimize max secondary delay	Centralized MILP
Lamorgese and Mannino (2019)	3-aspect	Yes	Fixed	Minimize weighted delays	MILP with Benders decomposition
Long et al. (2021)	DTG	Yes	Discrete Speed Level	Minimize total delay	Centralized MILP
Liu et al. (2021)	DTG		Discrete speed level	Minimize total delay	Bi-Level solving process
Lusby et al. (2013)	Any	Yes	Supports continuous	Minimize weighted delays	MILP with column-and-row generation process
Luan et al. (2018)	3-aspect	No	Continuous	Minimize weighted delays	Train order selected in first stage (MILP), then trajectories optimized in second stage (MINLP)
Luan et al. (2020)	3-aspect	Yes	Fixed	Minimize (evenly weighted) delays	MILP with geographic and time-based decomposition
Marlière et al. (2023).	3-aspect	Yes	Fixed	Minimize weighted delays	Constraint programming with LP-based warm-start
Pellegrini et al. (2014)	3-aspect	Yes	Fixed	Minimize weighted delays	Centralized MILP solver
Pellegrini et al. (2015), (2017), (2019)	3-aspect	Yes	Fixed	Minimize weighted delays	Centralized MILP with LP-based warm start
Versluis et al. (2023)	DTG	Yes	Discrete speed level	Minimize weighted delay and deviation from scheduled speed	Centralized MILP
Xu et al. (2017)	DTG	No	Discrete speed level	Minimize makespan	Centralized MILP with warm start
Xu et al. (2021)	DTG	Yes			

Table 2: Research on real-time rescheduling

3.3.State-of-the-art ATO/C-DAS Speed Regulation

The literature review returned five papers on ATO/C-DAS speed regulation shown in Table 3. These papers are categorized based on the following characteristics:

- **Topic:** Indicates if the focus of the paper is trajectory optimization, or timepoint configuration.
- **Relevant System:** Indicates the relevant system in the traffic management framework (TMS, Trackside or Onboard).
- **Objective:** Notes the objective to be maximized/minimized.
- **Solution Method:** Indicates the type of algorithm used to solve the problem

Reference	Topic	Relevant System	Objective	Solution Method
Albrecht et al. (2013)	Trajectory Optimization s.t. Timepoint Constraints	Onboard	Minimize energy consumption	Control theory and dynamic programming
Quaglietta et al. (2016)	ATO/C-DAS iteraction with TMS	TMS/Onboard	Propose framework for interaction between TMS and ATO/CDAS trackside	N/A
Wang et al. (2016)	Trajectory Optimization s.t. Timepoint Constraints	Onboard	Minimize energy consumption and delay	Nonlinear Programming
Wang et al. (2017)	Multi-train trajectory optimization	Onboard	Minimize energy consumption and delay	Nonlinear Programming
Wang et al. (2023)	Timepoint Configuration	Trackside	Optimize ATO track-side time points for conflict-free automated train operations	Linear Programming

Table 3: Research on real-time ATO/C-DAS Speed Regulation

Quaglietta et al. (2016) proposes the most commonly used framework for linking the timetable/TMS with ATO/C-DAS speed regulation function. In this framework, the TMS is responsible for setting trains' routes and orders (controlled by the interlocking) and determining scheduled arrival and departure times at stations. This information (known as the Real Time Traffic Plan, or RTTP) forms the basis of the TPE computation, which involves the configuration of timing windows to allow trains to drive in an energy-efficient manner while maintaining conflict-free operations. Each timing window for a given train n consists of a fixed location on the train's route (given the routing selected by the TMS), and functions as a constraint on the train's trajectory (as calculated by the onboard system). The resulting train trajectory optimization problem is either solved at the single-train level (Albrecht et al. (2013) Wang and Goverde (2016)) or as a multi-train problem (Wang and Goverde, 2017).

Wang et al. (2023) propose an algorithm for constructing conflict-free TPEs at the operational level given a macroscopic timetable or RTTP. The algorithm works by identifying the critical bottleneck location between pairs of following trains and placing a timepoint immediately before the bottleneck for the following train. The authors' rationale for this location choice is that homogenizing the following train's Energy-Efficient Train Control (EETC) trajectory with its S-MTTC trajectory reduces the capacity consumed at the bottleneck, possibly resolving the conflict in the macro-level timetable. If the algorithm cannot find timing points that facilitate conflict-free driving, the TMS can be triggered to resolve the conflict by altering the RTTP. The addition of a timepoint results in a trajectory requiring the train to drive more slowly before reaching the critical location (thus consuming more run time supplement), then driving as fast as possible from the intermediate timepoint location to the next scheduled station stop. Wang et al. (2023) propose placing the intermediate timepoint as close as possible to the critical location because doing so provides the greatest possibility for energy-efficient driving. While suitable for operational planning, the algorithm needs information about the overall timetable structure to function. This limits its applicability to situations where the structure of the timetable, as well as the associated bottlenecks, are unknown.

3.4. Conclusion and Literature Gaps

Research found on DTG signalling systems focus on solving only one of the two scheduling problems (tactical or operational), rather than attempting to align planning rules for both levels. Papers on the service planning problem either have signalling constraints generated by an 'oracle' (Schlechte et al, 2022), are based on scheduling norms derived from experience (Caimi et al, (2011), Grafe and Schöbel, (2021)), or use assumptions about the calculation of the braking curve that may not hold in real practice (Busuttil, 2023). Papers focusing on real-time rescheduling tested their algorithms on timetables constructed based on the constraints of the existing multi-aspect signalling system (Janssens, 2022 and Versluis et al, 2023), or on timetables whose construction process was not described (Xu et al, 2017, Long et al, 2019). Thus, research is needed on microscopic models for tactical timetabling and real-time traffic management that accurately represent the brake supervision process in DTG signalling.

There is a need to improve the formulation of DTG signalling constraints, as Multiple-Big-M formulations often have weak relaxations. Although the train service planning model proposed by Busuttil, (2023) can quickly solve problems with homogenous traffic patterns to optimality, it becomes computationally inefficient with line plans where there is variation in stopping patterns. This inefficiency is attributable to the fact that the critical blocks/locations in homogenous service patterns are usually located near stations where speed choices are limited (as trains must reach standstill to service the stop).

Most existing models for real-time rescheduling for multi-aspect signalling discretize the train's journey at the same level as the block sections (or track vacancy detection equipment) in the network, and do not impose maximum run time constraints in the intervals (D'Ariano et al. 2007 and Pellegrini et al, 2015). This assumption is acceptable for three-aspect signalling, where block lengths are required to be long enough for trains to stop within the block. When designing algorithms for signalling systems in cases where this assumption does not hold, such as four-aspect signalling and DTG signalling deployments with very short blocks, it is necessary to check that run times across consecutive intervals are feasible. Existing rescheduling models for multi-aspect signalling also represent headway constraints assuming trains cannot pass signals showing caution aspects, which is inconsistent with the UK rulebook. Thus, there is a need to develop rescheduling algorithms that permit trains to pass signals showing caution aspects, and that can properly represent the run time constraints for four-aspect signalling. While not the focus of this paper, these research gaps need to be addressed to develop a TMS applicable to multi-aspect signalling systems in use today.

There is a clear distinction in the literature between *macroscopic* models, which sacrifice detail for applicability to large-scale problems, and *microscopic* models, which forego computational tractability for feasibility. This trade-off creates issues when scheduling areas with dense traffic, as it may be possible to run trains closer together than otherwise indicated by a macroscopic model. This shortcoming forces IMs to manually schedule trains in these areas, increasing the amount of time required to produce a high-quality timetable. Although previous research solves these issues by implementing a feedback loop between the micro and macro problems (Bešinović et al, 2016), the computational complexity of this method is dependent on the complexity of the macro- and microscopic scheduling subproblems. The bi-level method is expected to become computationally more difficult in the future as IMs migrate from three-aspect signalling to signalling systems with more complex safety constraints at the microscopic level (such as ETCS with blocks shorter than worst-case braking distance). Accordingly, there is a need for microscopic models that are detailed enough to realize the available capacity of the infrastructure, but are simple enough to be tractable for large-scale timetabling problems. Thus, the framework needs to be able to represent the signalling constraints of the DTG signalling systems that IMs are looking to implement in busy areas, but general enough to be applicable to any signalling system in use on their network.

There is comparatively little research on the topic of configuring timepoints for an ATO/C-DAS system over ETCS at the planning level. The literature returned only one paper on the timepoint configuration problem (Wang et al. 2023), which focuses on resolving conflicts in an already-constructed macro timetable. The proposed linear programming algorithm requires knowledge of the timetable structure to determine where to place the timepoints. Thus, there is a need for a timepoint optimization algorithm applicable to tactical and operational scheduling problems, where the timetable structure is not known beforehand.

4. Methodology

To address the noted research gaps, this paper proposes an innovative timetabling framework that integrates run and headway constraints generated at the microscopic level, in a macroscopic timetabling model. In short, this approach seeks to maintain the computational efficiency of macro models, while also more accurately modelling train dynamics and aligning the planning rules at the tactical and operational levels.

The proposed scheduling model is macroscopic, in the sense that the only the arrival and departure events at the station are explicitly represented as variables. The key innovation versus state-of-the-art macroscopic models is that the macroscopic minimum headway constraints are fed by a novel microscopic TPE optimization framework that reflects actual train dynamics and minimum separation rules for any signalling system. In this way, the model can align planning rules at the tactical and operational levels without increasing the computational complexity of large-scale tactical timetabling problems. This capability ensures that the computed timetables are both conflict-free and stable enough for everyday operations.

The tactical scheduling problem is modelled as a Mixed-Integer linear programming (MILP) problem. Given a set of services in a network where a C-DAS/ATO system is used to regulate speed, the model aims to find a capacity-effective and stable periodic timetable by maximizing the minimum buffer time b_{min} present given a fixed cycle time ct, or minimizing the cycle time ct given a minimum buffer time b_{min}. Trains can only be re-ordered or re-timed: routings in station areas are assumed to be fixed. The ability to re-route trains could be added later.

Train trajectories are defined precisely so that the algorithm is applicable to networks with very short blocks/virtual sub-sections, and to enable a more accurate representation of the continuous speed-braking distance relationship in DTG signalling systems. Each train $n \in N$'s route is represented as a series of timepoints TP_n throughout the train's route. The location of each timepoint $tp \in TP_n$ is assumed to be fixed beforehand with its location (in metres from the start of train n's route) represented by the parameter $s_{n,tp}$. Where the train has a station stop, there are two timepoints: one each for the arrival and departure events. For the purposes of notation, timepoints are indexed at the level of the individual service $n \in N$. The timepoint $0 \in TP_n$ denotes the first timepoint on train $n \in N$'s route in the modelled area, and timepoint $fin \in TP_n$ denotes the last one on the route. The set $TP_n^{macro} \subseteq TP_n$ represents the set of macroscopic events on train n's route. This consists of timepoint $0 \in TP_n$ where the train enters the network, timepoint $fin \in TP_n$ where the train exits the network, and all arrival and departure events at stations.

For a given train $n \in N$'s route, it must pass through a series of *track segments* TS_n, which represent the discretized sections of track controlled by the signalling system. To guarantee safety, the signalling system cannot authorize more than one train into a given track section at the same time. In other words, if train $n \in N$ has track segment $ts \in TS_n$ reserved at time t, the signalling system cannot authorize any other train to enter ts at time t. The period where train n is authorized to enter segment ts is its *blocking time.* Blocking times have five components (shown in Figure 5):

- ***Set-up time:*** The time required to confirm that the track segment is clear of other trains, and that any movable elements (such as switches, level crossing gates, or movable bridges) in the track segment are in the correct positions.

- ***Reaction time:*** Once the route has been set-up, there is a delay that occurs while the new MA is communicated to the train. Once the train has received the new MA, its driver or the ATO system reacts to the MA by making a throttle or brake command. This throttle or brake command is subject to train control delay, which is the result of delayed reaction of the driver/ATO to the new MA, and the delay between when the throttle/brake command is made and when the train responds to it.
- ***Approach time:*** The is the period between when the train passes the brake indication point for the track segment, and when its head physically enters the track segment.
- ***Run time:*** The period when the train's head is passing through the track segment.
- ***Clear time:*** The period from when the train's head exits the track segment, to when the train's max safe rear end clears the track segment.
- ***Release time:*** The time required to release the track segment and make it available to be set up for another train.

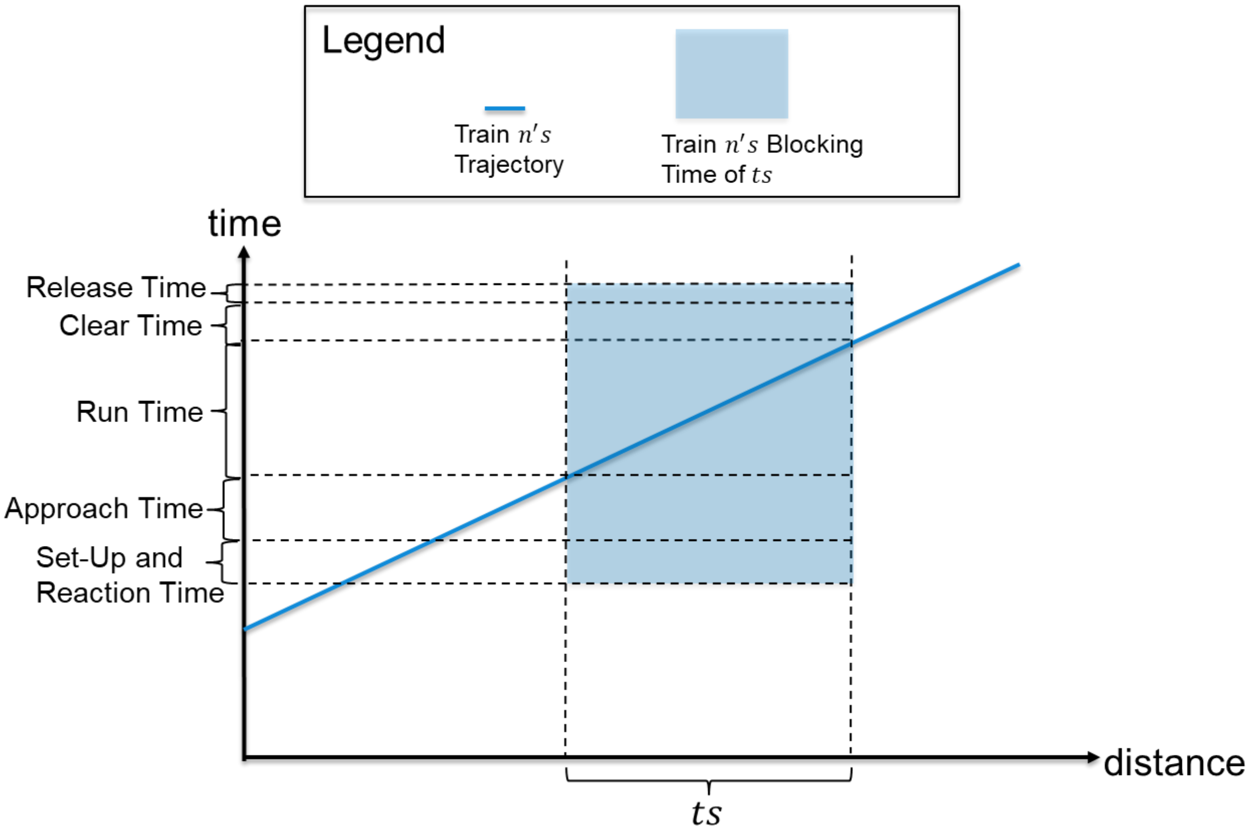


Figure 5: Components of Blocking Time Calculations

It should be noted that the definition of blocking times used assumes that each track segment is set up at the latest time possible without causing a conflict. In practice, a train $n \in N$'s route at track segment $ts \in TS_n$ could be set up earlier, provided that all other trains scheduled to pass through ts before n have released the segment. Blocking time calculations assume set-up occurs as late as possible because it helps to identify conflicts that cannot be avoided by deferring the time a route is set up.

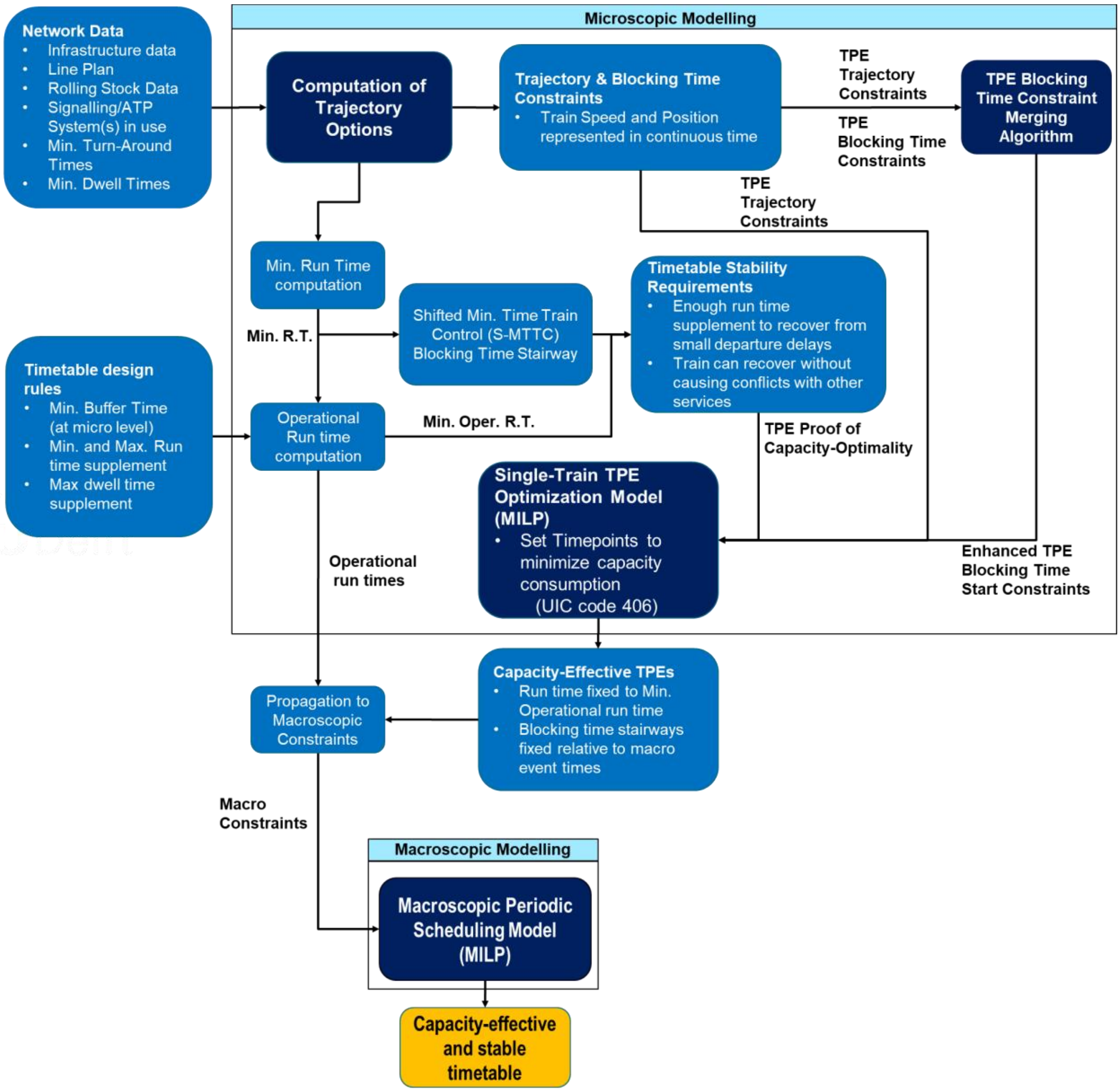


Figure 6: General Modelling Framework

The general modelling framework used is shown in Figure 6 above. Given the set of services (whose stopping patterns and rolling stock allocations are known), simulation is used to obtain the minimum technical running times for each service. These minimum run times (and their associated trajectories) form the basis of the timetable stability requirements (described in section 4.1.2), and the operational run time requirements. The stability requirements provide a proof of capacity-optimality, which forms the basis for the decomposition of the problem into a single-train TPE optimization problem (section 4.1.3) and a macroscopic periodic timetabling problem (section 4.2).

Simulation-based precomputation is also used to generate feasible trajectory options for each train in each timepoint interval $tp \in TP_n$ (section 4.1.1). Information about the train's run times and blocking times are extracted and used to generate the TPE train trajectory constraints (section 4.1.3.1) and TPE blocking time start constraints (section 4.1.3.2). The safety constraints ensure that trains do not need to engage in unacceptably harsh acceleration or braking and that blocking times are modelled accurately. The blocking time start constraints for open track are fed into a constraint merging algorithm (section 4.1.3.3), which propagates information contained in the trajectory and blocking time constraints to reduce the number of constraints required to represent blocking times. The train trajectory constraints and enhanced bocking time constraints are then fed into the TPE optimizer. The single-train TPE optimizer (section 4.1.3) is used to generate capacity-effective TPEs and their associated microscopic run time and headway constraints. The microscopic minimum headway constraints are then propagated into macroscopic headway constraints, which are then used to calibrate the macroscopic timetabling model (4.2). It is shown in section 4.2.2 that the resulting macroscopic minimum headway constraints are equivalent to the micro constraints. The macroscopic model is then used to compute a feasible, stable, and capacity effective timetable.

4.1. Notation For Optimization Framework

The following section lists the sets, parameters, and decision variables for the single-train TPE optimization model, and the macroscopic model.

4.1.1. Sets

N the set of train services
TP_n: The list of timepoints on train $n \in N$'s route
TP_n^{macro}: The list of macroscopic events on train $n \in N$'s route
TS_n: The list of track detection sections (if modelling ETCS Level 2) or virtual sub-sections (if modelling ETCS Hybrid Level 3) that train n occupies over its route
$TS_{n,tp}^{dep}$: The list of track segments train $n \in N$'s route that need to be reserved for station departure event $tp \in TP_n$
$S_{n,tp}$: The list of speed-distance trajectory arcs starting at timepoint tp and ending at timepoint $tp + 1$ on train $n \in T$'s route. Each arc $s_{n,tp}(v_0, v_1) \in S_{n,tp}$ has a starting speed v_0 at tp, and ending speed v_1 at $tp + 1$.
$V_{n,tp}$: The set of speed options for train $n \in N$ at timepoint $tp \in TP_n$. If $v \in V_{n,tp}$, then there exists at least one trajectory $s_{n,tp-1}(v_0, v_1) \in S_{n,tp-1}$ where $v_1 = v$ and at least one trajectory $s_{n,tp}(v_0, v_1) \in S_{n,tp}$ where $v_0 = v$.

4.1.2. Decision Variables

$t_{n,tp}$: the time that train $n \in N$ is scheduled to reach timepoint tp
$w_{n,tp}$: the time that train $n \in N$ waits at location $s_{n,tp}$, if it must stop there
$z_{n,tp}(v_0, v_1)$: Binary variable equal to 1 if speed distance trajectory arc $s_{n,tp}(v_0, v_1) \in S_{n,tp}$ is selected between timepoints tp and $tp + 1$, and 0 otherwise
$q_{n',n,ts}$: Binary variable equal to 1 if train $n' \in N$ reserves track section $ts \in TS_n \cap TS_{n-1}$ before train $n \in N$.
$rs_{n,ts}$: the time that train $n \in N$ begins reservation of track section $ts \in TS_n$
$re_{n,ts}$: the time that train $n \in N$ ends reservation of track section $ts \in TS_n$
ct: The cycle time (in seconds)
sct: The time that the first cycle starts (in seconds)
b_{min}: minimum buffer time (in seconds)

4.1.3. Parameters

$s_{n,tp}$: The location (in metres from the start of train $n \in N$'s route) where timepoint tp is location.
$rt_{n,tp}(v_0, v_1)$: the run time (in seconds) from location tp to $tp + 1$ if arc $s_{n,tp}(v_0, v_1) \in S_{n,tp}$ is selected. These are only defined for arcs on the open line ($s_{n,tp+1} > s_{n,tp}$)
$dwell_{n,tp}$: the minimum dwell time of train n at the station whose arrival event is represented by timepoint $tp \in TP_n$.
$res\_lag_{n,ts,tp}(v_0, v_1)$: If arc $s_{n,tp}(v_0, v_1) \in S_{n,tp}$ is selected, the difference (in seconds) between time $t_{n,tp}$ and the latest time that block $ts \in TS_n$ can be reserved by train $n \in N$ without risk of a blocking time overlap.
$\tau_{setup,ts,n',n}$: combined release and setup time (in seconds) for $ts \in TS_n \cap TS_{n-1}$ if it is released by train $n' \in T$ and subsequently set up for train $n \in N$.
τ_{comm}: Radio Block Centre (RBC) communication delay (in seconds).
$\tau_{ctrl,n}$: the train control delay (in seconds) of train $n \in N$.
$\tau_{n,tp}^{dep}$: the amount of time required to perform station work for train $n \in N$ to depart the start with departure event $tp \in TP_n$
$EoA_{n,ts}$: The furthest location (in metres from the start of train $n's$ route) that train n be authorized to without needing to reserve track section $ts \in TS_n$
l_n: length of train $n \in N$
pos_err_n: position error of train $n \in N$
$timestep$: the desired level of accuracy for run time, reservation lag, and release lag calculations
$\underline{rt}_{n,tp}, \overline{rt}_{n,tp}$ The minimum and maximum process times permitted between macroscopic event $tp \in TP_n^{macro}$ on train n's route and the next macroscopic event on the route.

4.1.4. Parameters specific to macroscopic timetabling

$\tau_{n,ts}^{rel}$: the time difference (in seconds) between when train $n \in N$ clears the release point for track segment $ts \in TS_n$ and the next macroscopic arrival event to occur.
$\tau_{n,ts}^{res,EETC}$: the time difference (in seconds) between when train $n \in N$ begins reservation of track segment $ts \in TS_n$ when using the EETC driving strategy, and the next macroscopic arrival event to occur.
$\tau_{n,ts}^{res,S-MTTC}$: the time difference (in seconds) between when train $n \in N$ begins reservation of track segment $ts \in TS_n$ when using the Shifted MTTC driving strategy, and the next macroscopic arrival event to occur.

4.1. Microscopic Modelling

4.1.1. Computation of Trajectory Options

Each train $n \in N$'s route is represented as a series of microscopic timepoints TP_n along the path travelled. The location of each timepoint $tp \in TP_n$ is assumed to be fixed beforehand with its location (in metres from the start of train n's route) represented by the parameter $s_{n,tp}$. Where the train has a station stop, there are two timepoints: one each for the arrival and departure events. There are no strict requirements as to where timepoints should be placed over the trains' routes to guarantee correctness of the model. The time that train n passes timepoint $tp \in TP_n$ is represented by the variable $t_{n,tp}$. The set $TP_n^{macro} \subseteq TP_n$ is defined as the set of macroscopic events on train $n \in N$'s route.
In areas where the train is braking for a scheduled station stop, additional timepoints are added at the locations where the train would have to start service braking at 20, 40, 60, 80, 100 or 120km/h to stop at the station. If the location of the braking point for a given speed v has a lower speed limit than v, no timepoint is added because the train will never start braking at that location. These braking timepoints were added in areas where driving behaviour is more homogenous: since the train must come to a stop at the station, there is limited ability to choose speed in this area. By creating these braking intervals with few speed options, the headways in station areas can represented with less dependence on the train's speed before it starts braking.

Each train $n \in N$'s passes through a series of track segments TS_n on its route through the modelled network. These track segments can consist of physical track detection sections, or (if applicable) virtual sub-sections that can be released using the train's position report and confirmation of train integrity (if the train has integrity monitoring). For each track segment $ts \in TS_n$, the time that train n begins reserving ts is represented by the variable $rs_{n,ts}$, and the time it ends reservation is represented by variable $re_{n,ts}$.

The minimum run time computation for a service $n \in N$ is performed in two stages. In the first stage, the minimum run time trajectory is computed through simulation-based methods (section 4.1.1). The minimum run times are obtained, and are used to determine the minimum and maximum operational run times $\underline{rt}_{n,tp}, \overline{rt}_{n,tp}$ between consecutive macroscopic events tp' and $tp' + 1 \in TP_n^{macro}$. These parameters are used to calibrate the macroscopic model in section 4.2. The minimum operational run time parameters $\underline{rt}_{n,tp}$ are also used to calibrate the TPE optimizer, so that it produces TPEs with sufficient flexibility to withstand minor delays.

The TPE optimizer requires exact specification of the train's EETC trajectory to guarantee feasibility of the trajectory in areas with short blocks, and to represent speed-dependent signalling constraints as accurately as possible. Between adjacent timepoints tp and $tp + 1 \in TP_n$ on train $n \in N$'s route, (where $s_{n,tp+1} > s_{n,tp}$), a speed-distance trajectory arc must be selected from the set of available arcs $S_{n,tp}$. Each arc $s_{n,tp}(v_0, v_1) \in S_{n,tp}$ is associated with a start speed v_0 at starting timepoint tp, and ending speed v_1 at timepoint $tp + 1$, and a running time $rt_{n,tp}(v_0, v_1)$ that relates the times $t_{n,tp}$ and $t_{n,tp+1}$. Binary variable $z_{n,tp}(v_0, v_1)$ is defined equal to 1 if trajectory $s_{n,tp}(v_0, v_1) \in S_{n,tp}$ is selected. If the trajectory $s_{n,tp}(v_0, v_1)$ is one where train n reaches standstill at $tp + 1$ (i.e., $v_1 = 0$), the train can wait at the location $tp + 1$ if necessary: this extra time is represented by the wait variable $w_{n,tp+1}$, which is constrained to zero if $v_1 \neq 0$. This is necessary to ensure correctness of the run time calculations under any block layout.

The exact specification of the speed-distance trajectory arcs is left up to the modeller, apart from the trajectories being operationally feasible. In the implementation for the case study, the following assumptions were used:

- If the arc $s_{n,tp}(v_0, v_1)$ is one where $v_1 > v_0$, the train immediately accelerates at timepoint tp using maximum acceleration to get from speed v_0 to end speed v_1. Once the train reaches speed v_1, it cruises at that speed until it reaches the next timepoint $tp + 1$. The trajectories for these arcs are computed using simulation-based methods to accurately represent train acceleration at higher speeds over long distances.
- If the $s_{n,tp}(v_0, v_1)$ is one where $v_1 = v_0$, the train cruises at the constant speed from tp to $tp + 1$. If the train is unable to maintain speed v_0 at some point in the interval (e.g., on a portion of track with a steep grade), the arc is considered infeasible, and not added to the model. If the arc $s_{n,tp}(v_0, v_1)$ is one where $v_1 < v_0$, the train brakes at a constant braking rate from speed v_0 at tp to speed v_1 at $tp + 1$.

4.1.2. Timetable Stability Requirements

The minimum run time is used to compute the minimum and maximum operational run times, which depend on the infrastructure manager's standards for allocating run time supplement. The blocking time stairway of the minimum run time trajectory also obtained. This blocking time stairway represents the area that the train will occupy (relative to the arrival events) if it must depart a station with a small delay. The trajectory where the train arrives at a station on time after driving as fast as possible is known as the Shifted-Min. Time Train Control (S-MTTC) trajectory (Wang et al, 2023). The S-MTTC trajectory is dependent only on the performance of the composition operating the service (such as train length, train weight, tractive effort capabilities and service braking rate), and the track geometry (i.e., speed limits, grades and curvature). By definition, the blocking time stairway for the S-MTTC trajectory is fixed relative to the train's scheduled arrival times at stations.

The operational run time requirements and the S-MTTC blocking time stairway are the basis of the stability requirements for the working timetable. For a given timetable to be stable, there needs to be enough run time supplement to allow trains to recover from small departure delays that could occur in everyday operation. Trains must also be able to recover from those delays without causing a conflict with other services. For that reason, an overlap with a train's S-MTTC blocking time stairway indicates that the timetable is not stable.

To satisfy all timetable stability requirements, the proposed scheduling algorithm models trains' paths through the networks through TPEs (Wang et al, 2023). A TPE represents a series of blocking times that permits the train to recover from small departure delays up to the run time supplement (RTS) without causing conflicts with other trains. To achieve this, the TPE contains the blocking times of the *Shifted Min-Time Train Control* (S-MTTC) trajectory, and an *Energy-Efficient Train Control* (EETC) trajectory. The S-MTTC trajectory represents the case where the train departs the station with a delay equal to the run time supplement, and drives as fast as possible to arrive at the next station on time. The EETC trajectory represents the case where the train departs the first station on time and uses the available run time supplement to perform energy efficient driving while arriving at the next station on time and respecting any intermediate timing constraints set (Wang and Goverde, 2016). The EETC trajectory consumes the available running time supplement during the journey through a combination of cruising at a reduced speed, and (if permitted) coasting. Unlike the S-MTTC trajectory, which is fixed given the route and rolling stock used, it is possible to alter the EETC trajectory by altering the intermediate timing point constraints that the train must respect.

A key property of a TPE is that the blocking time ends are always defined by the S-MTTC trajectories, and are independent of the choice of timing points provided to the train (Wang et al, 2023). The requirement for S-MTTC to be feasible also provides partial information about the shape of each train $n \in N$'s TPEs. To That end, the microscopic scheduling problem can be decomposed into a bi-level macro-micro problem. The master problem involves computing event times at the macroscopic level, which is then used to compute the TPEs for all the trains in the network. Given a macroscopically-feasible timetable, the subproblem involves finding an optimal timepoint configuration for a secondary objective (such as minimizing total energy consumption), or finding a feasible TPE as quickly as possible. The subproblem will either return a micro-feasible and stable TPE or return the infeasible core(s) of the TPE computation subproblem(s). Once a given train n's macroscopic event times have been fixed, it is known with certainty when a given train $n \in N$ will release any track segment $ts \in TS_n$. This has the effect of fixing the passing orders of any pair of trains n and $n' \in N$ at every shared track segment $ts \in TS_n \cap TS_{n'}$. With the order of trains n and $n' \in N$ at segment ts known, the permitted reservation start times for second train passing ts can be constrained further in the subproblem. In the case where n is ordered before n' at ts in the macro-timetable, it is known that the time train n' starts to reserve ts must occur after the time train n releases it (which is fixed in the master problem). This property allows for the subproblem to be parallelized at the level of the individual inter-station journey. In any case study, there are at least $|N|$ independent subproblems that can be solved in parallel.

A key observation when configuring timepoints is that it is almost always possible to resolve overlaps between TPEs by configuring the timepoints in a manner that maximizes homogeneity of the EETC trajectory with S-MTTC. This mechanism is employed in Wang et al. (2023), which seeks to resolve conflicts while facilitating energy-efficient driving as much as possible. Their proposed method, while able to resolve conflicts, is dependent on the algorithm having information about the overall timetable structure to determine where to place the intermediate time point(s). If it is not possible to obtain information about the timetable structure, the timepoint with the greatest likelihood of being conflict-free is the one that forces the train to consume all its run time supplement immediately after departing a station, then driving as fast as possible to the next station (with the same trajectory as S-MTTC). This maximum-homogeneity timepoint configuration is the driving strategy that maximizes the possibility of the resulting TPE being conflict-free. In this timetabling methodology, a single-train TPE optimizer is proposed, which generates homogenized TPEs that minimize the possibility of infeasibility in a subproblem. The produced TPE is used to populate the minimum headway constraints of the macro-level problem, which guarantees that this sub-problem will always be feasible. While the homogenized TPEs may not be optimal from an energy consumption perspective, the guarantee of feasibility allows permits large-scale decentralization of the timepoint configuration subproblem. Any additional run time supplement (beyond the run time of the capacity-optimal TPE produced) is converted into additional dwell time supplement allowance when solving the large-scale timetabling problem. This is done so that the microscopic reservation start- and end-times are fixed relative to the macroscopic event times. The result is that the micro-level minimum headway constraints can be converted into macroscopic headway constraints with no loss of detail. These constraints are used to calibrate a state-of-the-art macroscopic model, which produces capacity-effective, feasible and stable timetables.

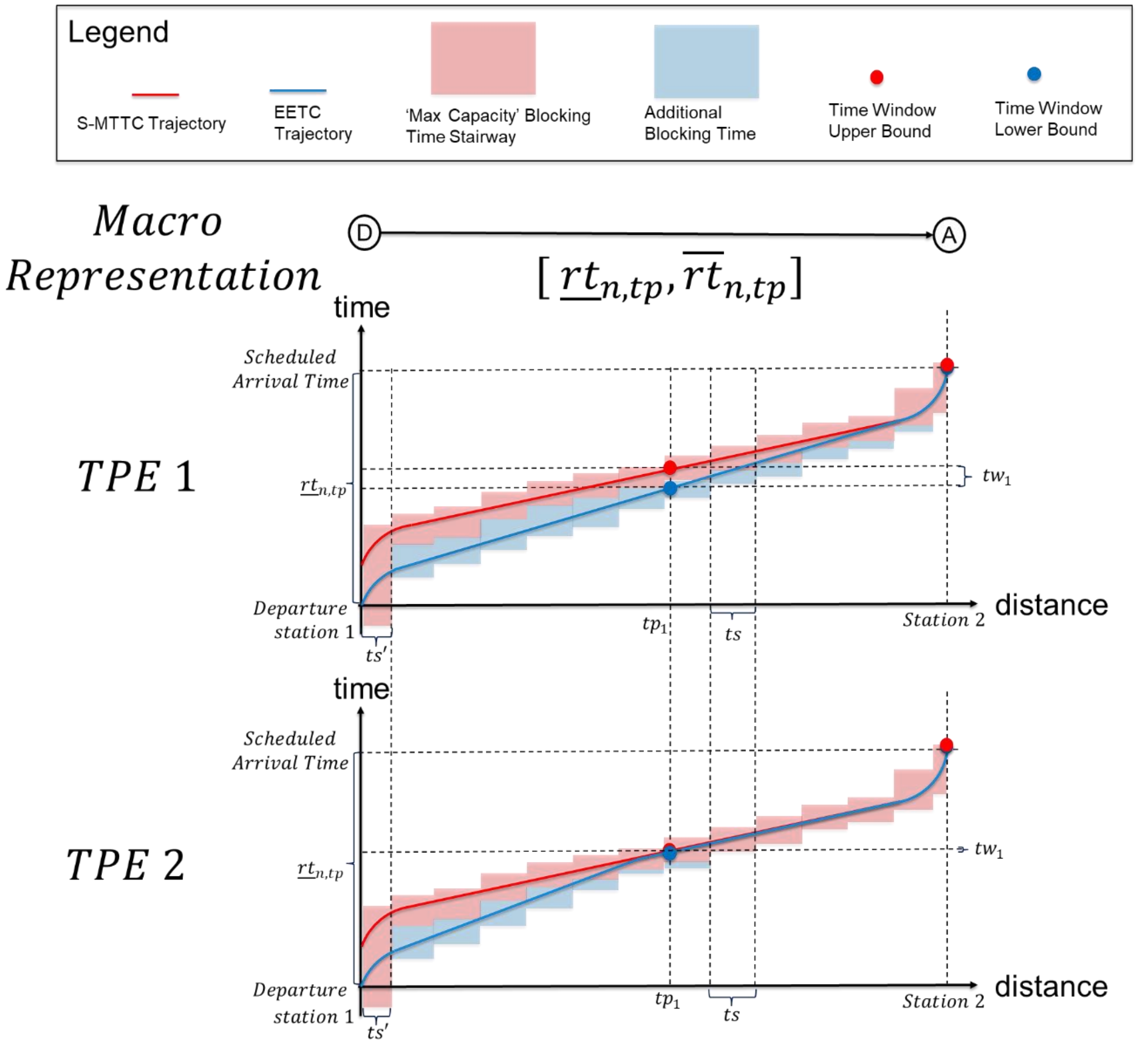


Figure 7: Blocking time diagrams train path envelopes showing the blocking time stairway that maximizes theoretical capacity (red) and additional blocking time (blue) for energy-efficient train control

Figure 7 shows how homogenizing the EETC trajectory with the S-MTTC trajectory increases the possibility of finding a micro-feasible timetable. In the above example, the scheduled run time for the train between station 1 and station 2 must be in the range $[\underline{rt}_{n,tp}, \overline{rt}_{n,tp}]$, given the minimum run time and the infrastructure manager's standards for run time supplement. While running between the two stations, the train traverses track segment ts. The S-MTTC trajectory is shown in red, and the EETC trajectory is shown in Blue. Both TPEs contain an intermediate time window[7] at location tp_1 (which corresponds to time window tw_1) before the entry to track segment ts, and a

[7] The exact number of intermediate timepoints and their locations on the route are up to the scheduler. In this example, only one intermediate timepoint is shown, to demonstrate how changing a time window impacts the shape of the TPE

time window at station *Station 2* (at the scheduled arrival time). The upper and lower bounds of a time window denoted by a red and blue dot, respectively.

The portion of the blocking time stairway that is coloured red is the area that the train must reserve to guarantee feasibility and stability of the timetable. This area includes the space that must be reserved for the S-MTTC trajectory, and the entire blocking time of any segments that need to be reserved for the train to depart $station\ 1$ (in this case, track segment ts') if the operational run time is set to the minimum permitted value $\underline{rt}_{n,tp}$. The portion of the TPE blocking time stairway in blue is the area that needs to be reserved to permit energy-efficient driving given the timepoint configuration.

Overlaps with the red area cannot be resolved by adjusting the timepoints or changing the operational run time. If there is an overlap with the S-MTTC trajectory, it is not possible for the train to recover from a delay without causing a conflict. If there is a blocking time overlap at segment ts' (which must be reserved to exit *Station 1*), reducing the operational run time (lower than $\underline{rt}_{n,tp}$) will reduce the blocking time at ts', but violates operational run time rules. On the other hand, increasing the operational run time will increase the blocking time at ts', which does not resolve the conflict. Thus, setting the run time equal to the minimum value $\underline{rt}_{n,tp}$ maximizes the possibility of finding a micro-feasible timetable.

In a hypothetical timetable with a critical block at ts, the EETC trajectory in the second TPE would consume less capacity (using the compression method from UIC code 406, (2013)), as it reserves ts later. If the minimum headway constraints were based off the first TPE, there is a possibility that an otherwise feasible timetable will be removed from the solution space because of the increased capacity consumption at that location (represented by the blue area). If ts is the critical block with the preceding train, the second TPE would also be capacity-optimal because it does not reserve ts for longer than the minimum time required to guarantee stability. This property is why the using the second TPE provides greater possibility of finding a feasible timetable. This property also holds for any timetable with a critical block between ts and the arrival event at $station\ 2$. For this reason, the objective of the TPE optimizer is to defer the reservation start times as much as safety and operational rules allow.

The driving strategy that provides greatest possibility for maximizing network capacity is one where the train closes its doors at $station\ 1$ as soon as possible (subject to the minimum operational run time constraint) and uses the S-MTTC driving strategy to get to station 2. This driving strategy has no additional reservation time versus the red area in Figure 7, so it is a capacity-optimal TPE regardless of the location of the critical block with the previous train. In practice, this driving strategy may not be permitted because it may be unsafe for the train to wait in $station\ 1$ for extended period with the doors closed and authorization to depart. This could be the case if the infrastructure manager has concerns about persons falling between the platform and the train during the waiting period, or if a level crossing needs to be closed to set up the route to exit $station\ 1$. The shape of the capacity-optimal blocking time stairway may not be the same as the S-MTTC trajectory when these safety constraints are included, so it is necessary to develop a dedicated TPE optimization algorithm for this purpose.

4.1.3. Single-Train Train Path Envelope (TPE) Optimization Model

The TPE optimization model generates timepoint configurations that maximize the possibility of finding a micro-feasible and stable timetable. The optimization process is performed at the level of the individual train $n \in N$ in the network, with the objective of homogenizing the train's EETC and S-MTTC trajectories as much as safety rules allow. Optimization takes place at the single-train level because the blocking time ends of the TPE are independent of the timepoint configuration choice (Wang et al, 2023). The decentralization at the single-train level enables detailed modelling of speed-headway relationships without sacrificing scalability to larger network areas.

The general objective of the TPE optimizer is to defer reservation start times as much as possible. The objective for train $n \in N$'s timepoint configuration problem is:

$$\max \sum_{ts \in TS_n} rs_{n,ts} \qquad (1)$$

Subject to the timing and blocking time constraints:

$$t_{n,tp+1} = t_{n,tp} + w_{n,tp+1} + \sum_{s_{n,tp}(v_0,v_1) \in S_{n,tp}} z_{n,tp}(v_0,v_1) * rt_{n,tp}(v_0,v_1) \; \forall n \in N \; \forall tp \in TP_n : s_{n,tp+1} > s_{n,tp} \qquad (2)$$

$$\sum_{s_{n,tp}(v_0,v_1) \in S_{n,tp}} z_{n,tp}(v_0,v_1) = 1 \; \forall n \in N \; \forall tp \in TP_n : s_{n,tp+1} > s_{n,tp} \qquad (3)$$

$$\sum_{\substack{s_{n,tp}(v_0,v_1) \in S_{n,tp} \\ v_1 = v}} z_{n,tp}(v_0,v_1) = \sum_{\substack{s_{n,tp+1}(v_2,v_3) \in S_{n,tp+1} \\ v_2 = v}} z_{n,tp+1}(v_2,v_3) \;\; \forall n \in N \; \forall tp, tp+1 \in TP_n \forall v \in V_{n,tp+1} : s_{n,tp+2} > s_{n,tp+1} > s_{n,tp} \qquad (4)$$

$$w_{n,tp} \geq 0 \; \forall n \in N \; \forall tp \in TP_n \qquad (5)$$

$$w_{n,tp+1} \leq M \sum_{\substack{s_{n,tp+1}(v_0,v_1) \in S_{n,tp} \\ v_0 = 0}} z_{n,tp}(v_0,v_1) \; \forall n \in N \; \forall tp \in TP_n : s_{n,tp+1} > s_{n,tp} \qquad (6)$$

$$0 \leq t_{n,tp} \leq M \; \forall n \in N \; \forall tp \in TP_n \qquad (7)$$

$$t_{n,tp+1} \geq t_{n,tp} + dwell_{n,tp} \; \forall n \in T \; \forall tp \in TP_n : s_{n,tp} = s_{n,tp+1} \qquad (8)$$

$$w_{n,tp} \leq 0 \; \forall n \in N \; \forall tp \in TP_n : s_{n,tp+1} = s_{n,tp} \cup s_{n,tp} = s_{n,tp-1} \qquad (9)$$

$$rs_{n,ts} \leq t_{n,tp} + \sum_{s_{n,tp}(v_0,v_1) \in S_{n,tp}} z_{n,tp}(v_0,v_1) * res_{lag_{n,ts,tp}}(v_0,v_1) \; \forall n \in T \; \forall ts \in TS_n \; \forall tp \in TP_n \qquad (10)$$

$$rs_{n,ts} \leq t_{n,tp} - \tau^{dep}_{n,tp} \forall n \in T \; \forall tp \in TP_n \; \forall ts \in TS^{dep}_{n,tp} : s_{n,tp} = s_{n,tp-1} \qquad (11)$$

$$t_{n,tp'+1} - t_{n,tp'} = \underline{rt}_{n,tp} \;\; \forall tp', tp'+1 \in TP^{macro}_n, \forall n \in N \qquad (12)$$

Constraints $(2) - (9)$ are individual train trajectory constraints, which are described in section 4.1.3.1. Constraints (10) and (11) define the reservation start times, and are described in section 4.1.3.2. Constraint (12) requires that for every pair of consecutive macroscopic events tp' and $tp' + 1 \in TP^{macro}_n$, the run time must be equal to the minimum process time $\underline{rt}_{n,tp}$ given the infrastructure manager's standards for run time supplement. The TPE optimization model does not need reservation end time constraints because they are independent of the timepoint configuration problem (Wang et al. 2023).

Previous research on the construction of TPEs has observed that the reservation end time $re_{n,ts}$ of any track segment $ts \in TS_n$ by train $n \in N$ depends only on the S-MTTC trajectory (Wang et al. 2023). This means reservation end times are independent of timepoint configuration choices made by the TPE optimizer. Thus, it is not necessary to model reservation end times in this model.

4.1.3.1. TPE Trajectory Constraints

Each speed-distance trajectory arc $s_{n,tp}(v_0, v_1)$ has its trajectory and run time $rt_{n,tp}(v_0, v_1)$ defined precisely in the model. This is done to ensure that the train trajectories are feasible in areas with finely discretized infrastructure, and to model the speed-headway dependencies more accurately. In many previous works (such as Versluis et al, 2023 and Xu et al, 2017), speed is modelled though a series of discrete speed levels. For each of those speed levels, there is an associated minimum run time constraint representing the minimum possible run time for a train using that speed level, but there exists some flexibility to determine the exact run time. Since the speed level options in those models do not correspond to an exact train trajectory, those models must base the headway constraint on the worst-case braking distance that could be realized for the chosen speed level. The requirement for precise definition of the trajectory avoids this capacity loss issue because the exact position and speed of the train is always known to the model. The relationship between the selection of speed-distance trajectory arcs selected and the timepoints is shown in Figure 8 below:

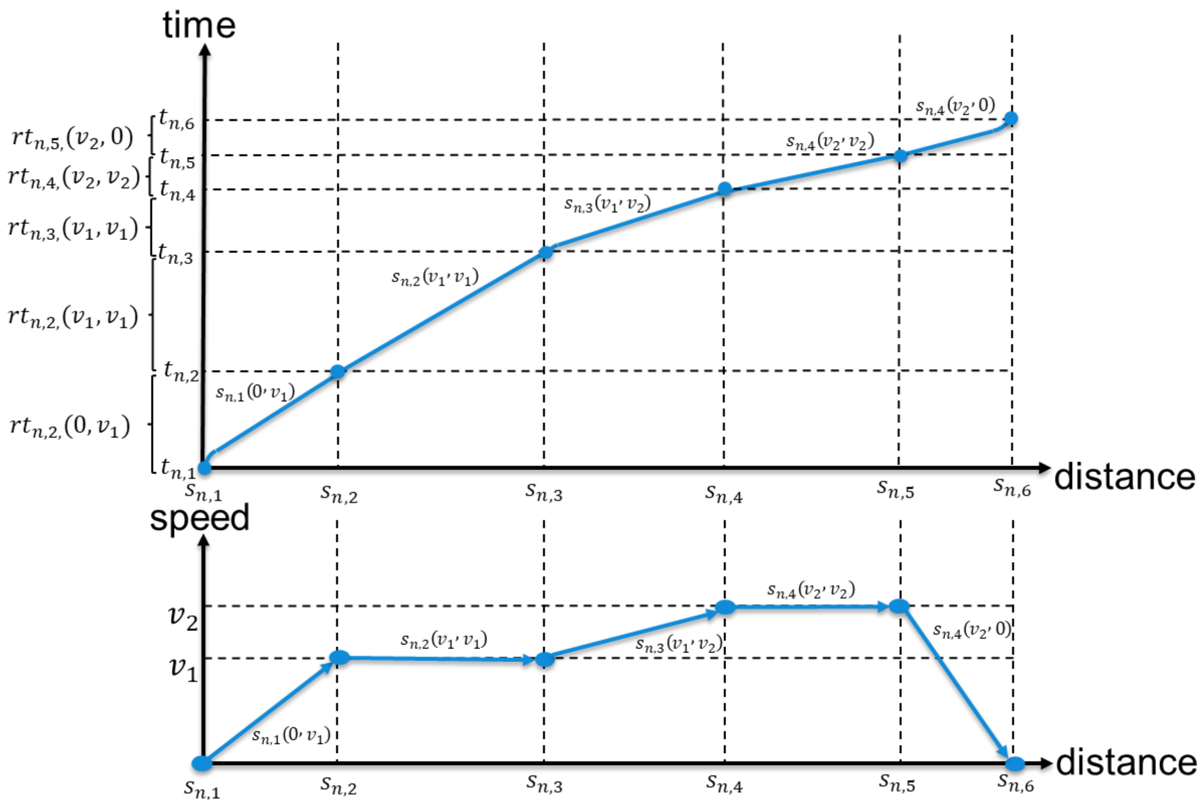


Figure 8: The relationship between speed-distance arc selection and the time-distance trajectory

The run time constraints are expressed as:

$$t_{n,tp+1} = t_{n,tp} + w_{n,tp+1} + \sum_{s_{n,tp}(v_0,v_1)\in S_{n,tp}} z_{n,tp}(v_0,v_1) * rt_{n,tp}(v_0,v_1)\ \forall n \in N\ \forall tp \in TP_n{:}\, s_{n,tp+1} > s_{n,tp} \quad (2)$$

$$\sum_{s_{n,tp}(v_0,v_1)\in S_{n,tp}} z_{n,tp}(v_0,v_1) = 1\ \forall n \in N\ \forall tp \in TP_n{:}\, s_{n,tp+1} > s_{n,tp} \quad (3)$$

$$\sum_{\substack{s_{n,tp}(v_0,v_1)\in S_{n,tp}\\ v_1=v}} z_{n,tp}(v_0,v_1) = \sum_{\substack{s_{n,tp+1}(v_2,v_3)\in S_{n,tp+1}\\ v_2=v}} z_{n,tp+1}(v_2,v_3)\ \ \forall n \in N\ \forall tp, tp+1 \in TP_n \forall v \in V_{n,tp+1}{:}\, s_{n,tp+2} > s_{n,tp+1} > s_{n,tp} \quad (4)$$

$$w_{n,tp} \geq 0\ \forall n \in N\ \forall tp \in TP_n \quad (5)$$

$$w_{n,tp+1} \leq M \sum_{\substack{s_{n,tp+1}(v_0,v_1)\in S_{n,tp}\\ v_0=0}} z_{n,tp}(v_0,v_1)\ \forall n \in N\ \forall tp \in TP_n{:}\, s_{n,tp+1} > s_{n,tp} \quad (6)$$

$$0 \leq t_{n,tp} \leq M\ \forall n \in N\ \forall tp \in TP_n \quad (7)$$

Constraint (2) expresses the relationship between adjacent timepoints $t_{n,tp}$ and $t_{n,tp+1}$ as equal to the run time of the speed-distance trajectory arcs $s_{n,tp}(v_0, v_1) \in S_{n,tp}$ that is selected for the interval and the time $w_{n,tp+1}$ that train n spends at standstill at location $s_{n,tp+1}$ (which is not a scheduled station stop) before starting to move again. Constraint (3) requires that exactly one of the speed-distance trajectories available be selected for the interval between tp and $tp + 1$. Constraint (4) is a speed-conservation constraint, which states at any timepoint location $s_{n,tp}$, the ending speed of the speed-distance trajectory ending at the location must be the same as starting speed of the trajectory in the next interval. This constraint ensures that there are no abrupt speed changes in the train's speed-distance trajectory over the entire journey. Constraints (4), (5) and (6) require that the wait time at any time point on the open line is nonnegative, and equal to zero if the train does not stop at the location. This constraint is necessary to properly represent the ability of trains to stop and wait at location tp until traffic clears down the route. This is for application to real-time rescheduling because it may be necessary for a train to stop and wait at a location outside of a station for the model to find a feasible solution. This could occur in scenarios where fast train gets stuck behind a delayed slow train on the open line. The model may not find any solutions if an allowance is not made for a fast train to stop and wait. Constraint (7) is imposed to ensure all timepoints occur in the time window $[0, M]$, which ensures the validity of any disjunctive minimum headway constraints with a Big-M formulation. An example of how run times are calculated with a train reaching standstill (at timepoint 3) is shown in Figure 9: The relationship between speed-distance arc selection and the time-distance trajectory when a train stops and waits at timepoint 3.

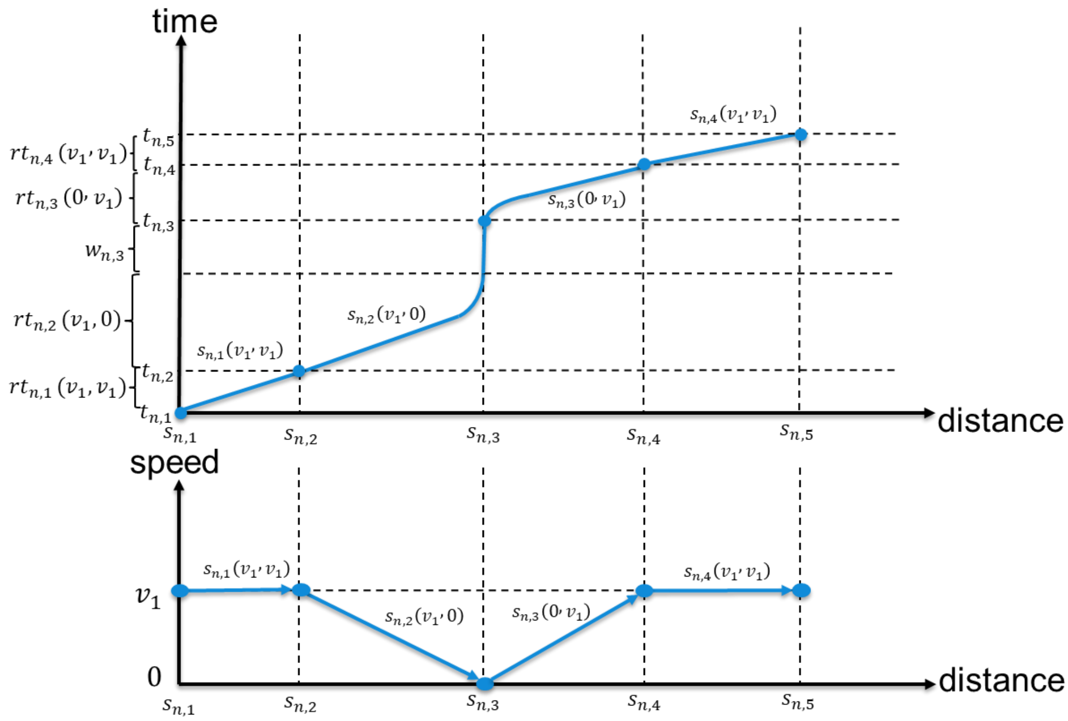


Figure 9: The relationship between speed-distance arc selection and the time-distance trajectory when a train stops and waits at timepoint 3

In Figure 9, train n reaches standstill at the location of timepoint 3, at time $t_{n,2} + rt_{n,2}(v_1, 0)$. At this time, the train can wait for as long as it wants before continuing its journey (which is represented by the next timepoint $t_{n,3}$). Since the train is stationary, it can wait at the location $s_{n,3}$ for any amount of time (represented by variable $w_{n,3}$) that does not result in a conflict with another train.

For timepoints tp and $tp + 1$ representing a scheduled station stop ($s_{n,tp} = s_{n,tp+1}$), the train remains at standstill, so there is no need to specify a speed-distance trajectory between them. It is necessary, however, to ensure that the train complies with the minimum dwell time requirement $dwell_{n,tp}$ for the station. This is accomplished by adding the constraint:

$$t_{n,tp+1} \geq t_{n,tp} + dwell_{n,tp} \;\forall n \in T \;\forall tp \in TP_n\colon s_{n,tp} = s_{n,tp+1} \tag{8}$$

It is assumed that if tp and $tp + 1$ represent the arrival and departure events for a scheduled station stop, all speed-distance trajectories ending at tp have ending speed 0 and all that start at $tp + 1$ have starting speed zero. This assumption is necessary for the overall path to be feasible, since a train cannot safely service a station without coming to standstill. For this reason, it is not necessary to impose constraint (3) for dwell arcs, or to impose the speed-conservation constraint (4) for arrival or departure timepoints (since only one speed can be selected at them). It is also not necessary to create the wait time variable $w_{n,tp}$ if $t_{n,tp}$ represents an arrival or departure event. To simplify the presentation of the constraints, it is assumed that $w_{n,tp}$ is created for every timing location $tp \in TP_n$, and the constraint:

$$w_{n,tp} \leq 0 \;\forall n \in N \;\forall tp \in TP_n\colon s_{n,tp+1} = s_{n,tp} \cup s_{n,tp} = s_{n,tp-1} \tag{9}$$

Is added to the model if tp is an arrival or departure event.

4.1.3.2. TPE Blocking Time Start Constraints

To produce a conflict-free timetable or traffic plan, the model needs to represent the speed-headway relationship characterizing the DTG signalling systems as accurately as possible. For each train $n \in N$, the set TS_n is defined as the set of all track segments (whether physical detection sections or virtual sub-sections) that train n will need to reserve to complete its trip through the modelled area. For modelling lines with ETCS Level 2, TS_n is the set of physical track detection sections on the line. For modelling with ETCS Hybrid Level 3, TS_n should be defined as the set of virtual sub-sections that the train passes through. For modelling of moving-block systems (such as CBTC or ETCS Level 3), sections of open track should be discretized into sufficiently small virtual sub-sections to approximate the continuous release process.

Each track segment $ts \in TS_n$ has an associated End-of-Authority $EoA_{n,ts}$ representing the furthest location that that train n can be authorized to without needing to reserve the segment. The location of $EoA_{n,ts}$ corresponds to the furthest location that the signalling system can authorize train n to, which may not be the same location where train n would physically enter the segment. This distinction is made to correctly model the route set-up rules in interlockings with sectional route release. An example of how the parameter $EoA_{n,ts}$ is set in for a diverging route in an interlocking is shown in Figure 10.

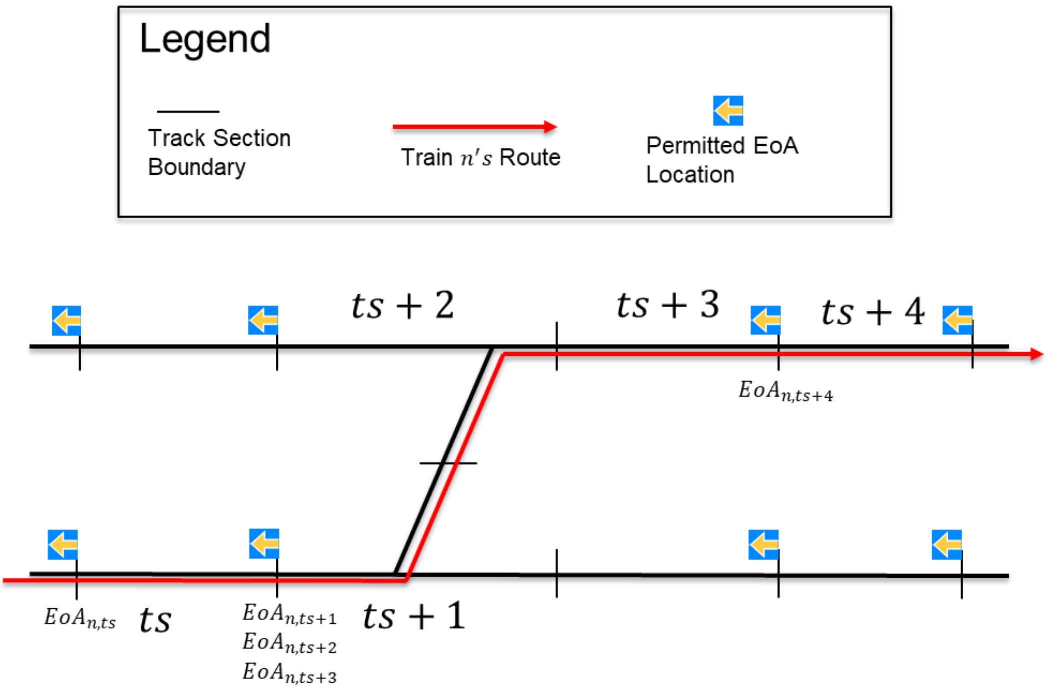


Figure 10: Placement of reservation start locations in Interlockings with Sectional Route Release

In the example in Figure 10, the train can be authorized to the physical start locations of segments ts, $ts + 1$ and $ts + 4$, so their EoA locations are all located at those positions. For the diverging route at the crossover, train n needs to have all of segments $ts + 1$, $ts + 2$ and $ts + 3$ set up. Since it is not possible to authorize the train to the physical start of segment $ts + 2$ in between the two diverging switches, the location of $EoA_{n,ts+2}$ is placed at the same location as $EoA_{n,ts+1}$. Similarly, it is not possible to authorize trains to the start of segment $ts+3$, so $EoA_{n,ts+3}$ is placed at the same location as $EoA_{n,ts+2}$ (which happens to be located at the same position as $EoA_{n,ts+1}$).

To model the relationship between a train's speed distance trajectory and its blocking times, the variable $rs_{n,ts}$ is introduced to represent the latest time that train $n \in N$ can begin reservation of section $ts \in TS_n$ without risk of a blocking time overlap with another train. Similarly, the variable $re_{n,ts}$ is introduced to represent the earliest time that train $n \in N$ can end reservation of $ts \in TS_n$ without risk of a blocking time overlap. The definition of $rs_{n,ts}$ assume that there is no set-up, reaction or release time required. This is done because the exact amount of time required to release the route of a train $n' \in N$ and set up the route of another train $n \in T$ at a common track segment $ts \in TS_n \cap TS_{n'}$ could vary across segments.

For a given trajectory followed by train n, where the train's position and speed at time t are represented by $s_n(t)$ and $v_n(t)$, respectively, the area that train n needs to have occupied to guarantee safety is represented as the job-starting frontier $JS_n(t)$ of the train. This is the area in front of the train's head plus the distance required for the train to come to a complete stop using service braking, as shown in Figure 11.

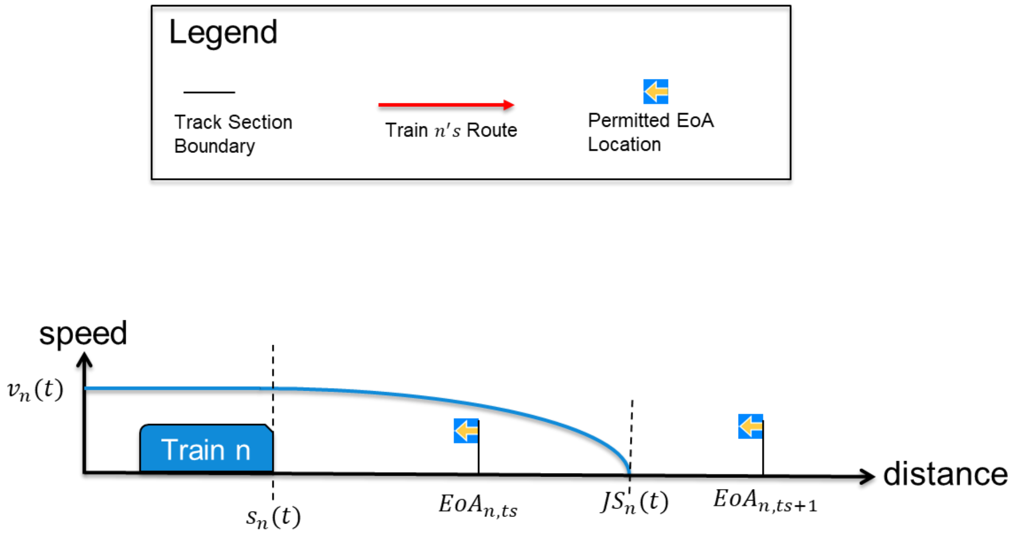


Figure 11: Visualization of the Job-start frontier of train n at time t in relation to the position of its head

In Figure 11, the head of train n has not yet reached the start of track segment ts at location $EoA_{n,ts}$, but its job-start frontier $JS_n(t)$ given its current position and speed has already passed $EoA_{n,ts}$. This means that track segment ts must be reserved at or before time t. At the same time, the job-start frontier has not yet passed the start of track segment $ts+1$ at location $EoA_{n,ts+1}$. This means that $ts+1$ does not need to be reserved by train n yet, and another train may reserve that track segment at time t without causing a conflict.

The definition of the job-start frontier $JS_n(t)$ should be defined as accurately as possible, to maximize the capacity gains that can be realized by the model while guaranteeing feasibility. A train's braking distance depends on the following factors:

- *The train's service braking rate:* A train with a higher service braking rate will be able to stop in a shorter distance, and thus does not need to occupy as much track in front of it.

- *The train's speed at time t:* Braking distance typically increases the faster a train is travelling. There may also be an interaction between speed and the train's braking ability, where the service rate improves (or worsens) when the train's speed crosses a certain threshold.
- *The gradient of the track:* On track with a negative gradient, gravity cause the train to accelerate (compared to travelling on flat track), reducing the train's effective braking rate to below the normal service rate. On track with a positive gradient, the train will be travelling upward, so gravity will assist in slowing the train down. This results in a higher effective braking rate compared to flat track. The gradient may not be constant over the entire braking distance.
- *The curvature of the track:* When a train goes through a curve, the forces imparted by the rails on the train to change its direction results in additional resistance that can cause the train to slow down.
- *Air Resistance:* Air resistance is a product of the train's body moving through space. Higher amounts of air resistance require the train to use more power when cruising or accelerating, and reduce the distance required for the train to stop.
- *Wheel-Rail Adhesion:* In conditions with low wheel-rail adhesion, it will take longer for a train to brake to standstill. Under these conditions, the train will need to reserve a greater distance of the track ahead of it to guarantee safety.
- *Rotating Mass Factor:* for trains with high rotating mass factor, more force is required to make the wheels stop rotating, which increases the required braking distance.

For this case study, the job-starting frontier $JS_n(t)$ for operations with ETCS is calculated as a piecewise summation of the train n's braking curve over the track ahead of $S_n(t)$, considering how changes in gradient affect its braking ability as it moves forward. If train n is at position where the track has gradient G, it's effective braking rate b_n^{eff} over that section of track is:

$$b_n^{eff} = b_n + g * G$$

Where b_n is the train's service braking rate (m/s^2) on flat track (with zero gradient), g is the gravitation acceleration (assumed to be 9.81 m/s^2), and G is the track gradient (%). When the train crosses into a section of track with a different grade, the effective braking rate will change, as shown in Figure 12.

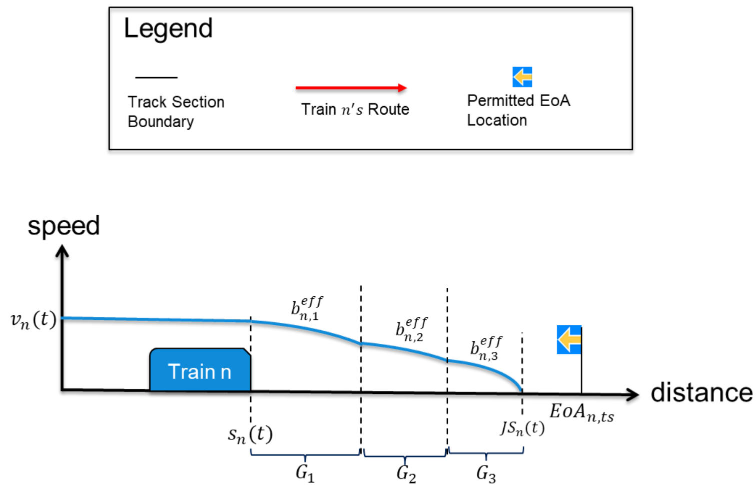


Figure 12: The impact of changes in track grade on the effective braking rate and the job-start frontier location

The calculation of the reservation start time $rs_{n,ts}$ for track segment ts is calculated by finding the location on train $n's$ route where its speed-distance trajectory intersects with the service braking curve for location $EoA_{n,ts}$. This is the furthest that train n can be authorized to on its route without needing to reserve ts. When the train's speed-distance trajectory intersections, it will receive a braking indication if ts is not yet reserved at the time, so it will need to reserve ts by that time for the path to be conflict-free. Since train n will need to pass the location $EoA_{n,ts}$ to complete its journey through the modelled area, it is guaranteed that the train's speed-distance trajectory will intersect the brake curve for ts at some location. An example of how the indication point is derived from the train's trajectory is shown below in Figure 13.

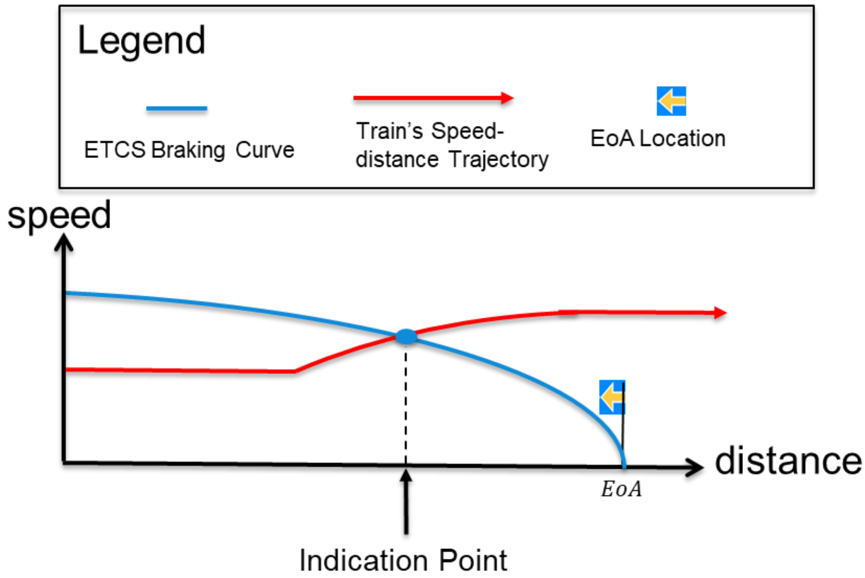


Figure 13: Derivation of ETCS full supervision indication point for a track segment for a fixed trajectory.

If train n is operating under multi-aspect signalling, the job-start frontier $JS_n(t)$ is always at the same position as the train's head $S_n(t)$. The equivalent braking curves for three- and four-aspect signalling is shown in Figure 14 below, if yellow and double-yellow passes are not permitted. The brake indication point for train segment ts is the first signal to show a brake indication (i.e., a non-green aspect) if track segment ts is not occupied. The EoA location for ts is assumed to the last signal train n passes before physically entering ts, which would show a red signal if ts were occupied. If there is a three-aspect sequence, the indication point is at the location of the previous signal (regardless of speed). In a four-aspect sequence, the indication point is two signals back from the EoA.

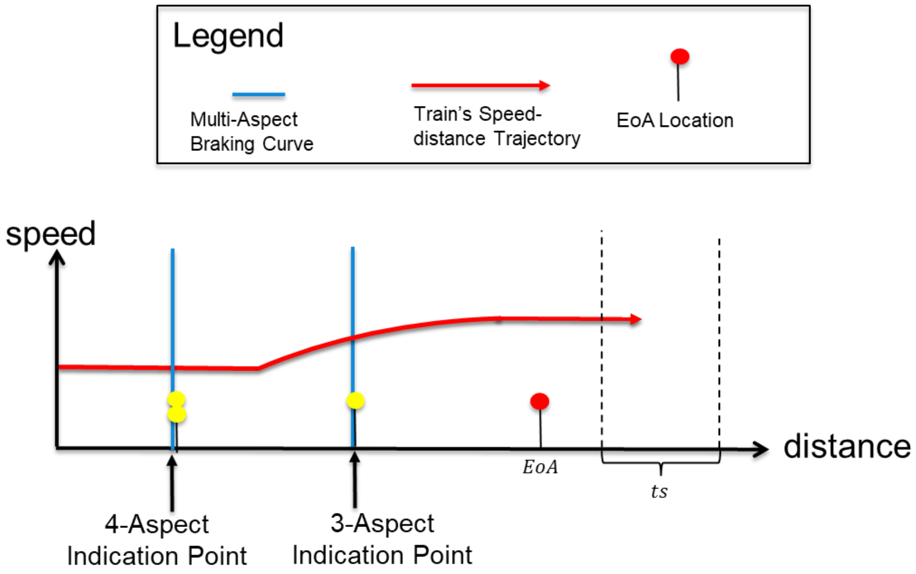


Figure 14: Derivation of three- and four-aspect indication points for a track segment for a fixed trajectory.

While it is guaranteed that the train's speed-distance trajectory will intersect the braking curve, the exact location where it occurs may depend on the speed-distance trajectory chosen by the train. In general, the faster a train is travelling, the further away the indication point will be from the EoA location. The timepoint interval that the brake indication occurs in may be different for different speeds, as shown in Figure 15 for ETCS full supervision. In that example, the indication point occurs between timepoints 1 and 2 if the train chooses a speed-distance trajectory with speed v_2 at $t_{n,1}$, whereas if the train remains at speed v_1 at $t_{n,1}$, the indication occurs between timepoints 2 and 3. The exact location where the indication occurs varies depending on the exact speed-distance arcs chosen.

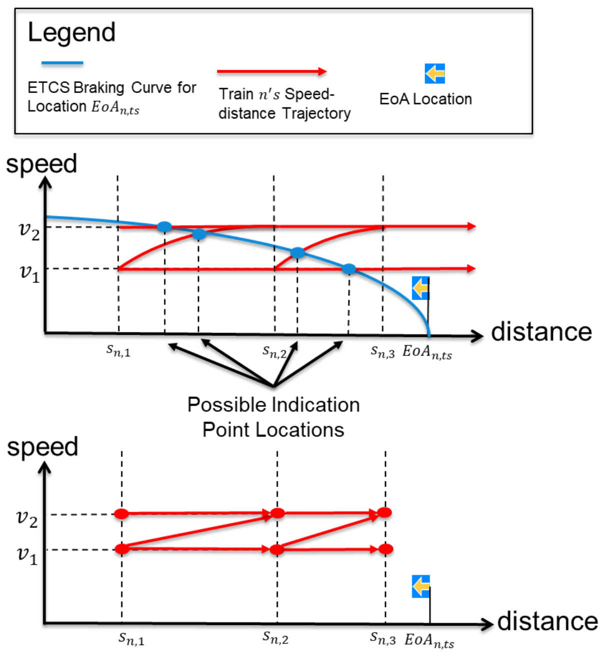


Figure 15: Impact of Speed-distance arc chosen on the interval the indication point occurs in.

To obtain all the possible scenarios for a brake indication to occur for track segment $ts \in TS_n$, it is necessary to check every speed-distance arc $s_{n,tp}(v_0, v_1) \in S_{n,tp}$ in every interval $tp \in TP_n$: $s_{n,tp} + pos\_err_n \leq EoA_{n,ts}$. For each of these arcs, the *reservation lag* parameter $res\_lag_{n,ts,tp}(v_0, v_1)$ is defined as the time difference between $t_{n,tp}$ and the latest time reservation of ts can safely begin.

The reservation lag times are calculated using the following process:

Algorithm 2: *Generation of Reservation Lag Weight* $res\_lag_{n,ts,tp}(v_0, v_1)$ *for arc* $s_{n,tp}(v_0, v_1) \in S_{n,tp}$	
Input:	*Train position* $s_n(t)$*, speed* $v_n(t)$*, track segment EoA location* $EoA_{n,ts}$*, arc run time* $rt_{n,tp'}(v_0, v_1)$ *of simulation-based process used to generate arcs*
	$z_{n,tp}(v_0, v_1) \leftarrow 1$ $t \leftarrow t_{n,tp}$
	If $JS_n(t_{n,tp}) > EoA_{n,ts}$ ***Then***
	Return $-timestep$
	While $t < t_{n,tp} + rt_{n,tp'}(v_0, v_1)$ ***do***
	If $JS_n(t) \leq EoA_{n,ts} < JS_n(t + timestep)$ ***Then***
	Return $t - t_{n,tp}$
	Else
	$t \leftarrow t + timestep$
	Return M
Output:	*Release Lag time Value* $res\_lag_{n,ts,tp}(v_0, v_1)$*, or indication to use Big-M value*

If the algorithm returns the value M, the train can run through the entire arc $s_{n,tp}(v_0, v_1)$ without needing to reserve track segment ts, so a Big-M value should be assigned to avoid cutting off the optimal solution. If the algorithm returns $-timestep$, then the train must already have the track segment ts reserved when starting the arc $s_{n,tp}(v_0, v_1)$, so it can be inferred that the reservation start must occur before time $t_{n,tp}$. While this information does not indicate the exact time that the reservation of ts by train n needs to start, it adds an additional constraint on $rs_{n,ts}$, reducing the dependence of the headway constraints on knowing the exact train trajectory. These additional constraints make it more difficult for the solver to achieve unrealistically low headways when integrality constraints on the binary variables $z_{n,tp}(v_0, v_1)$ representing the trajectory options are relaxed. If the algorithm returns any value for $res\_lag_{n,ts,tp}(v_0, v_1)$ other than M or $-timestep$, then the train needs to reserve track segment ts within $res\_lag_{n,ts,tp}(v_0, v_1)$ of when it passes timepoint tp.

While the reservation lag values vary depending on which arcs are selected, only one constraint is needed per interval $tp \in TP_n$ to represent all lag possibilities in that interval. This can be done because of constraint (2), which requires that exactly one arc be selected in each interval. This means that the summation:

$$\sum_{s_{n,tp}(v_0,v_1) \in S_{n,tp}} z_{n,tp}(v_0, v_1) * res\_lag_{n,ts,tp}(v_0, v_1)$$

Will always be equal to the exact reservation lag $res\_lag_{n,ts,tp}(v_0, v_1)$ of the arc $s_{n,tp}(v_0, v_1) \in S_{n,tp}$ that is selected for the interval.

The resulting constraint representing reservation start time is:

$$rs_{n,ts} \leq t_{n,tp} + \sum_{s_{n,tp}(v_0,v_1) \in S_{n,tp}} z_{n,tp}(v_0, v_1) * res_{lag_{n,ts,tp}}(v_0, v_1) \; \forall n \in T \; \forall ts \in TS_n \; \forall tp \in TP_n \qquad (10)$$

Constraint (10) requires that the reservation start time $rs_{n,ts}$ occur at or before the time train n passes timepoint $tp \in TP_n$ plus the reservation lag of the arc selected. This constraint ensures that reservation will start at or before train n passes the brake indication point for segment ts. If this constraint were not included for the model, it would be possible for train n to pass the indication point for the track segment without first reserving it. If this were to occur in real operations, the train would be forced to brake, and it would be infeasible to follow the speed distance trajectory selected by the model. The reservation start time is expressed considering communication and control delay, which affects the ability of the train to react to the new movement authorization. The formulation of (10) does not consider set-up time because the timepoint optimization process is performed at the single-train level, whereas the amount of time required to set up the route may depend on the route of the previous train to pass through segment ts. Set-up time is instead incorporated at the macroscopic timetabling model described in section 4.2.

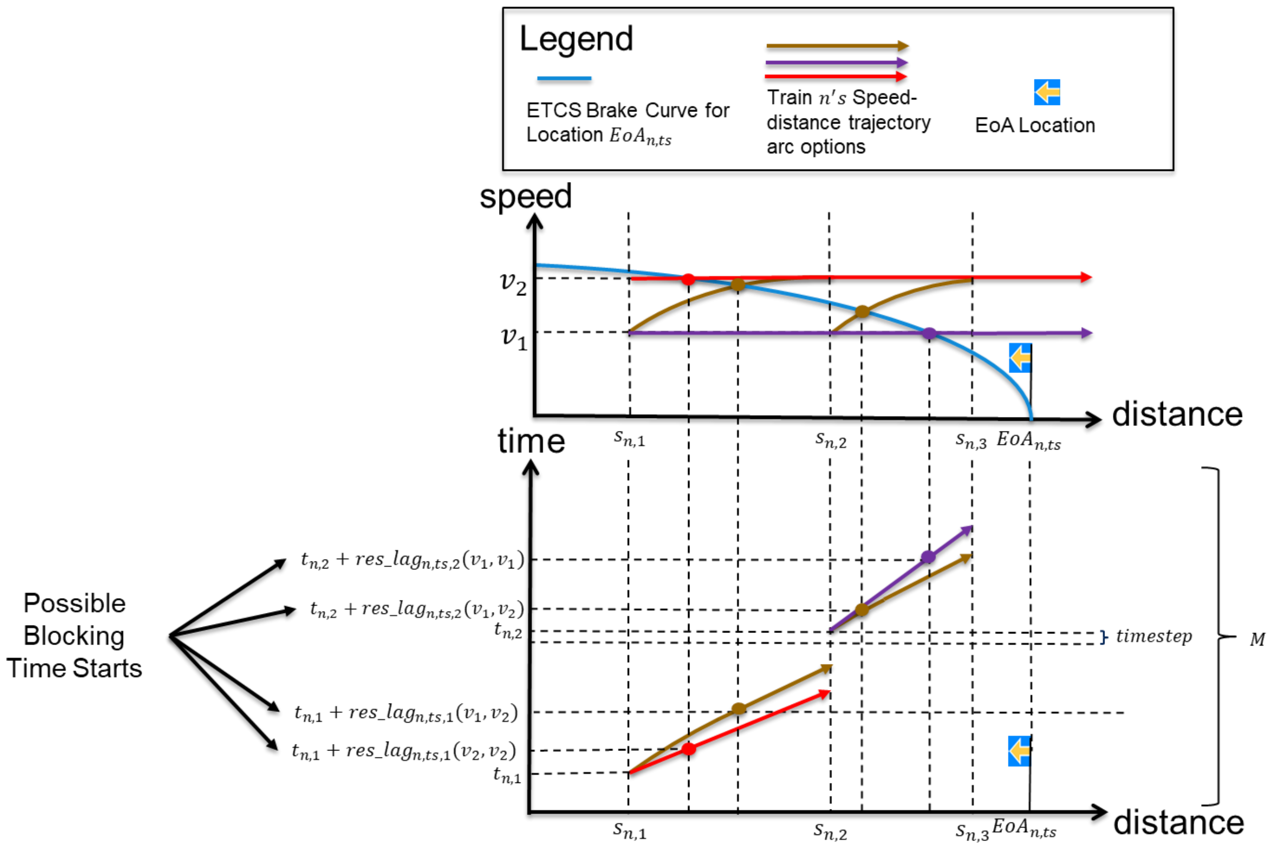


Figure 16 : Relationship between speed-distance trajectory selection and exact reservation start times, with M and $timestep$ shown for scale. Only arcs with a reservation lag other than M or $-timestep$ are shown in the time-distance diagram.

An example of how exact reservation lags are calculated and grouped by interval is shown in Figure 16, with the values of $timestep$ and a sufficiently large M shown for scale. In interval 1, there are two arcs $s_{n,1}(v_1, v_2)$ (in brown) and $s_{n,1}(v_2, v_2)$ (in red) that intersect the braking curve for ts, and one arc $s_{n,1}(v_1, v_1)$ (in purple on the speed-distance diagram) that is entirely to the left of the braking curve. The reservation lag values of $s_{n,1}(v_1, v_2)$ and $s_{n,1}(v_2, v_2)$ correspond to the time difference between $t_{n,1}$ and the latest time that reservation of ts can safely start if those arcs are chosen. The reservation lag value $s_{n,1}(v_1, v_1)$ is set equal to M, which is sufficiently high that the constraint:

$$rs_{n,ts} \leq t_{n,1} + M$$

is not more restrictive than any of the constraints that are imposed in later intervals.
In interval 2, there are two arcs $s_{n,2}(v_1, v_1)$ (in purple) and $s_{n,2}(v_1, v_2)$ (in brown) that intersect the braking curve for ts, and one arc $s_{n,2}(v_2, v_2)$ (in red on the speed-distance diagram) that is entirely to the right of the braking curve. The reservation lag values of $s_{n,2}(v_1, v_1)$ and $s_{n,2}(v_1, v_2)$ represent the time difference between $t_{n,2}$ and the latest time that reservation of ts can safely start if those arcs are chosen. For arc $s_{n,1}(v_2, v_2)$, which needs ts to be reserved before time $t_{n,2}$, the constraint:

$$rs_{n,ts} \leq t_{n,2} - timestep$$

is imposed if the arc is selected. This is not the most restrictive constraint that would be imposed $rs_{n,ts}$, since arc $s_{n,2}(v_2, v_2)$ is necessarily preceded by an arc whose trajectory intersects the braking curve (in this case, by $s_{n,1}(v_1, v_2)$ or $s_{n,1}(v_2, v_2)$). If the reservation lags of the preceding arcs are not known, it cannot be proven that:

$$rs_{n,ts} \leq t_{n,2} - timestep$$

is not the most restrictive constraint on $rs_{n,ts}$. For that reason, it is impossible to assign $s_{n,2}(v_2, v_2)$ a more restrictive reservation lag than $-timestep$ without incorporating the information contained in reservation lags from the previous interval.

From the formulation of the constraint (10) and the reservation lags shown in Figure 16, it can be concluded that:

$$rs_{n,ts} \leq t_{n,2} + res\_lag_{n,ts,2}(v_1, v_1)$$

without knowing anything about the trajectory selected for train n. It is possible to conclude this because:

- The exact reservation lag times associated with arcs $s_{n,2}(v_1, v_1)$ and $s_{n,2}(v_1, v_2)$ are known.
- While the exact reservation start time associated with arc $s_{n,2}(v_2, v_2)$ is not known, it is known that reservation will start before $t_{n,2} - timestep$ if the arc is selected.

If this information provided by $s_{n,2}(v_2, v_2)$ about the reservation start time were not included into the constraint, the solver could only conclude that:

$$rs_{n,ts} \leq t_{n,2} + M$$

without knowing anything about the trajectory. The incorporation of the information provided by arc $s_{n,1}(v_2, v_2)$, while not exact, helps the solving algorithm know what kinds of solutions will not be feasible. By knowing where not to look, it is easier for the solver to find good quality solutions.

Most IMs also have regulations requiring a train stopped at a station to obtain a movement authority before the departure process can start. In the UK, this process is governed RSSB's *Station Duties and Train Dispatch* rules (RSSB, 2023c) that determine when a train's doors can be closed, and when the train can depart. For the departure procedure to commence, the driver must have a movement authority that allows them to clear the entire platform. Once the MA has been received, the train's doors can be shut. Once the train's doors are shut, and it has been confirmed that it is safe to depart, the train can begin to move. For station departure timepoint $tp \in TP_n$ on train $n \in N$'s route, the blocks that need to be reserved to depart are represented by the set $TS_{n,tp}^{dep} \subseteq TS_n$. The time required to perform station duties is represented by the parameter $\tau_{n,tp}^{dep}$. The train dispatch constraint is:

$$rs_{n,ts} \leq t_{n,tp} - \tau_{n,tp}^{dep} \forall n \in T\ \forall tp \in TP_n\ \forall ts \in TS_{n,tp}^{dep} : s_{n,tp} = s_{n,tp-1} \qquad (11)$$

The constraint starts the track segments $ts \in TS_{n,tp}^{dep}$ that need to be reserved for departure event $tp \in TP_n$ all need to be reserved at least $\tau_{n,tp}^{dep}$ before the time the train starts to move ($t_{n,tp}$).

The exact list of track sections $TS_{n,tp}^{dep}$ that need to be reserved to dispatch train n at $t_{n,tp}$ should be obtained from the information about the interlockings in station areas, and the exact location of the end of the platform (which may not be where the head of the train is stopped). For the implementation in EGTrain, where the station is represented as a single point on the route at location $s_{n,tp} = s_{n,tp-1}$, any track segment $\forall ts \in TS_n$ where:

$$s_{n,tp} \leq EoA_{n,ts} \leq s_{n,tp} + l_n$$

is included in the set $TS_{n,tp}^{dep}$ for departure event $t_{n,tp}$.

4.1.3.3. TPE Blocking Time Constraint Merging Algorithm

While constraint (10) as written is a correct representation of trains' blocking time starts in the single-train TPE optimization problem, it is possible to include more information in the constraints relating $t_{n,tp}, tp \in TP_n$ and $rs_{n,ts}$ by considering how the selection of the speed-distance trajectory in interval tp affects the trajectory options available in subsequent timepoint intervals. Algorithm 3 shows the iterative process used to add more information to reservation start constraints.

Algorithm 3: *Strengthening process for constraint for train $n \in N$ relating reservation start of track segment $ts \in TS_n$ with passing time of timepoint $tp \in TP_n$*

Input:
Original Constraint $rs_{n,ts} \leq t_{n,tp} + \sum_{s_{n,tp}(v_0,v_1)\in S_{n,tp}} z_{n,tp}(v_0,v_1) * res\_lag_{n,ts,tp}(v_0,v_1)$,
Run time $rt_{n,tp_i}(v_1,v_2)$ and reservation lag $res\_lag_{n,ts,tp_i}(v_1,v_2)$ for every speed-distance trajectory $s_{n,tp_i}(v_1,v_2) \in S_{n,tp_i} \forall tp_i \in TP_n: tp_i \geq tp$,
Location s_{n,tp_i} for every timepoint node $tp_i \in TP_n: tp_i \geq tp$
Speed options V_{n,tp_i} for every timepoint node $tp_i \in TP_n: tp_i \geq tp$

$tp_i \leftarrow tp + 1$
$DonePropagation \leftarrow False$
1 $InitializeConstr\left(rs_{n,ts} \leq t_{n,tp} + \sum_{s_{n,tp}(v_0,v_1)\in S_{n,tp}} z_{n,tp}(v_0,v_1) * res\_lag_{n,ts,tp}(v_0,v_1)\right)$
2 **While not** $DonePropagation$ ***and*** $s_{n,tp_i} < s_{n,tp_i+1}$ **do**
3 $\quad DonePropagation \leftarrow True$
4 $\quad$ **For** speed $v \in V_{n,tp_i}$ **do**
5 $\quad\quad NotBigMpresent \leftarrow False$
6 $\quad\quad$ **For** $s_{n,tp_i-1}(v_1,v_2) \in S_{n,tp_i-1}: v_2 = v$ **do**
7 $\quad\quad\quad$ **If** $res\_lag_{n,ts,tp_i-1}(v_1,v_2) \neq M$ **then**
8 $\quad\quad\quad\quad NotBigMpresent \leftarrow True$
9 $\quad\quad\quad\quad$ **break**
10 $\quad\quad$ **If not** $NotBigMpresent$ **then**
11 $\quad\quad\quad DonePropagation \leftarrow False$
12 $\quad\quad\quad$ **For** $s_{n,tp_i-1}(v_1,v_2) \in S_{n,tp_i-1}: v_2 = v$ **do**
13 $\quad\quad\quad\quad ChangeRHSVarWt(Var = z_{n,tp_i-1}(v_1,v_2), Weight = rt_{n,tp_i-1}(v_1,v_2))$
14 $\quad\quad\quad$ **If** $v = 0$ **then**
15 $\quad\quad\quad\quad AddNewRHSVarWt(Var = w_{n,tp_i}, Weight = 1)$
16 $\quad\quad\quad$ **For** $s_{n,tp_i}(v_1,v_2) \in S_{n,tp_i}: v_1 = v$ **do**
17 $\quad\quad\quad\quad AddNewRHSVarWt(Var = z_{n,tp_i}(v_1,v_2), Weight = res\_lag_{n,ts,tp_i}(v_1,v_2))$
18 $\quad tp_i \leftarrow tp_i + 1$

Output: *Strengthened constraint (12) for train $n \in N$ relating reservation start of track segment $ts \in TS_n$ with passing time of timepoint $tp \in TP_n$*

The first step is to generate the basic constraint relating reservation start time $rs_{n,ts}$ with the interval timepoint interval $tp \in TP_n$ (line *1*). For each speed $v \in V_{n,tp_i}$ that the train could take at $tp_i \in TP_n: tp_i > tp$, the speed-distance trajectories $s_{n,tp_i-1}(v_1,v_2) \in S_{n,tp_i-1}$ that end with speed v are checked to see if their reservation lag is equal to M (Lines *6-9*). If all those trajectories have reservation lag M, it is known that reservation start will occur after time t_{n,tp_i} if the train has speed

v at timepoint tp_i. This information is added to the constraint by changing the weights associated with the trajectories $s_{n,tp_i-1}(v_1,v_2) \in S_{n,tp_i-1}: v_2 = v$ on the right-hand-side of the constraint from M to their run times (lines *12-13*). If the speed v being checked is also zero (standstill), the wait time w_{n,tp_i} of the train at tp_i is also added to the right-hand side of the constraint (lines *14-15*). The variable w_{n,tp_i} is added with weight *1* because it is part of the run time for train n between the nodes if it stops and waits at the location of tp_i, but is otherwise equal to zero by constraint (5). These two steps cause the right-hand side to be equal the passing time t_{n,tp_i} if the indication point does not occur before the location of tp_i. Finally, the reservation lags of the speed-distance trajectories $s_{n,tp_i}(v_1,v_2) \in S_{n,tp_i}$ that start at tp_i with speed v are added to the right-hand side (lines *16-17*). This process iterates for every interval $tp_i \in TP_n: tp_i > tp$ until the train either reaches a scheduled station stop ($s_{n,tp_i} = s_{n,tp_i+1}$) or if it is no longer possible to add more information about the exact reservation start time to the constraint. The algorithm is guaranteed to terminate by the end of the modelled route. This will occur because the train must pass the EoA location $EoA_{n,ts}$ to complete its journey through the modelled area, so the indication point is guaranteed to occur in at least one of the speed-distance trajectory arcs in the complete trajectory of train n.

An example of how constraint (10) is strengthened by algorithm 3 is shown in Figure 17 for train n and track segment ts. In the diagram above, the time that train n needs to start reserving segment ts could occur in interval 1 or 2, depending on the exact trajectory selected. Two basic constraints of type (10) need to be created (one for each of these intervals) to properly model $rs_{n,ts}$.

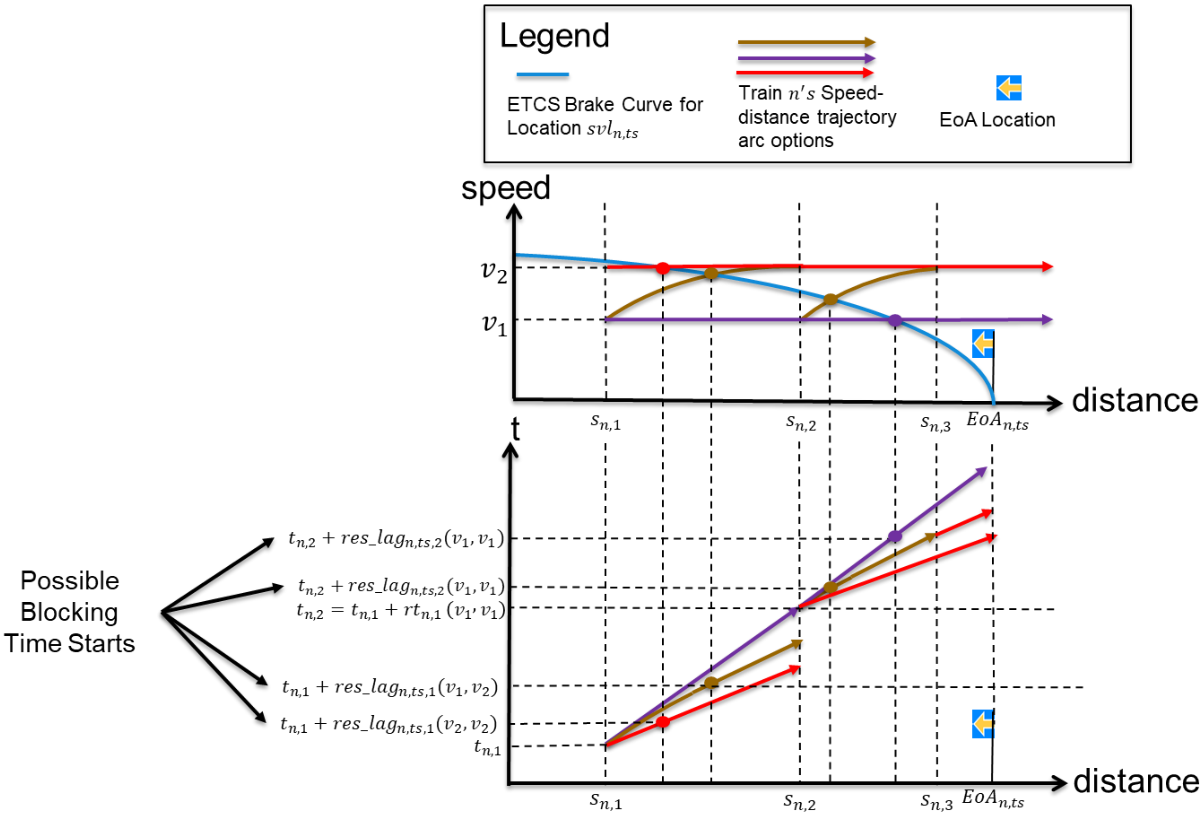


Figure 17 : Relationship between speed-distance trajectory selection and exact reservation start times across intervals with mutual exclusivity property shown.

The initial reservation lag constraint of type (11) for track segment ts for the trajectories in timepoint interval 1 is:

$$rs_{n,ts} \leq t_{n,1} + M * z_{n,1}(v_1, v_1) + res\_lag_{n,ts,1}(v_1, v_2) * z_{n,1}(v_1, v_2) + res\_lag_{n,ts,1}(v_2, v_2) * z_{n,1}(v_2, v_2)$$

This constraint says:

- if speed-distance trajectory option $s_{n,1}(v_1, v_1)$ is selected $(z_{n,1}(v_1, v_1) = 1)$, then it cannot be inferred when $rs_{n,ts}$ will occur relative to time $t_{n,1}$ (hence $res\_lag_{n,ts,1}(v_1, v_1) = M$).
- If $s_{n,1}(v_1, v_2)$ or $s_{n,1}(v_2, v_2)$ are selected, then $rs_{n,ts}$ occurs between timepoints 1 and 2 (so the appropriate reservation lag values are used).

This constraint formulation for timepoint 1, while correct, does not consider that selecting speed-distance trajectory $s_{n,1}(v_1, v_1)$ provides information about the run time in interval 1, and the speed-distance trajectories that can selected in the following interval 2. If speed-distance trajectory $s_{n,1}(v_1, v_1)$ is selected, then it is known that:

- $t_{n,2} = t_{n,1} + rt_{n,1}(v_1, v_1)$ (from constraint (1)), and
- The speed-distance trajectory selected in the next interval must be $s_{n,2}(v_1, v_1)$ or $s_{n,2}(v_1, v_2)$ (from constraint (3)).

This information can be included in the constraint for timepoint interval 1 by rewriting it as:

$$\begin{aligned} rs_{n,ts} \leq t_{n,1} &+ res\_lag_{n,ts,1}(v_1, v_2) * z_{n,1}(v_1, v_2) + res\_lag_{n,ts,1}(v_2, v_2) * z_{n,1}(v_2, v_2) \\ &+ rt_{n,1}(v_1, v_1) * z_{n,1}(v_1, v_1) + res\_lag_{n,ts,2}(v_1, v_2) * z_{n,2}(v_1, v_2) \\ &+ res\_lag_{n,ts,2}(v_1, v_1) * z_{n,2}(v_1, v_1) \end{aligned}$$

This rewritten constraint states:

- If $s_{n,1}(v_1, v_2)$ is selected, then $rs_{n,ts} \leq t_{n,1} + res\_lag_{n,ts,1}(v_1, v_2)$, or
- If $s_{n,1}(v_2, v_2)$ is selected, then $rs_{n,ts} \leq t_{n,1} + res\_lag_{n,ts,1}(v_2, v_2)$, or
- If $s_{n,2}(v_1, v_2)$ is selected, then $rs_{n,ts} \leq t_{n,1} + rt_{n,1}(v_1, v_1) + res\_lag_{n,ts,2}(v_1, v_2)$, or
- If $s_{n,2}(v_1, v_1)$ is selected, then $rs_{n,ts} \leq t_{n,1} + rt_{n,1}(v_1, v_1) + res\_lag_{n,ts,2}(v_1, v_1)$

The last two statements hold because constraint (3) states that $s_{n,1}(v_1, v_1)$ must be selected for interval 1 if either $s_{n,2}(v_1, v_1)$ or $s_{n,2}(v_1, v_2)$ are selected for interval 2. The four cases in the constraint are mutually exclusive by constraint (2) (in both intervals 1 and 2, exactly one speed-distance trajectory must be chosen), so this constraint is a correct representation.

With this new constraint formulation, it is now known from Figure 17 that:

$$rs_{n,ts} \leq t_{n,1} + rt_{n,1}(v_1, v_1) + res\_lag_{n,ts,2}(v_1, v_1)$$

without knowing anything about trajectory selected for train n. In the original formulation of the constraint, it could only be inferred that:

$$rs_{n,ts} \leq t_{n,1} + M$$

with the same information. Incorporating this information in constraint (1) makes it easier for the solving algorithm to know if a solution is infeasible, allowing it to find good-quality solutions faster.

In the implementation of the model for TPE optimization, all constraints of type (10) are implemented with the enhancements made by *algorithm 3* to the formulation. Any constraints that are made redundant by the algorithm are excluded from the model in the precomputation phase.

4.2. Macroscopic Periodic Scheduling Model

The macroscopic periodic timetabling model seeks to either maximize the minimum buffer time:

$$Max\ b_{min} \qquad (12)$$

with a fixed cycle time ct, or to minimize the cycle time ct:

$$Min\ ct \qquad (13)$$

With a buffer time $b_{min} = 0$, subject to:

$$t_{n,\mathrm{tp}'+1} - t_{n,\mathrm{tp}'} \geq \underline{rt}_{n,tp} \ \ \forall tp', tp' + 1 \in TP_n^{macro}, \forall n \in N \qquad (14)$$

$$t_{n,\mathrm{tp}'+1} - t_{n,\mathrm{tp}'} \leq \overline{rt}_{n,tp} \ \ \forall tp', tp' + 1 \in TP_n^{macro}, \forall n \in N \qquad (15)$$

$$re_{n,ts} = t_{n,tp_{n,ts}^{rel}} - \tau_{n,ts}^{rel,S-MTTC} + b_{min} \ \ \forall ts \in TS_n \forall n \in N \qquad (16)$$

$$rs_{n,ts} = t_{n,tp_{n,ts}^{res}} - \max\left(\tau_{n,ts}^{res,S-MTTC}, \tau_{n,ts}^{res,EETC}\right) - \tau_{comm} - \tau_{ctrl,n} \ \ \forall ts \in TS_n \forall n \in N \qquad (17)$$

$$q_{n,n\prime,ts} + q_{n\prime,n,ts} = 1 \forall n, n' \in T \ \forall ts \in TS_n \cap TS_{n\prime} \qquad (18)$$

$$rs_{n,ts,tp} \geq re_{n',ts,tp} + \tau_{setup,ts,n',n} - M\left(1 - q_{n',n,ts}\right) \forall n, n' \in T \ \forall ts \in TS_n \cap TS_{n'} \qquad (19)$$

$$sct + (p-1) * ct \leq \ t_{n_p,0} \leq sct + p * ct - timestep \ \forall n_p \in N_p \ \forall p \in P \qquad (20)$$

$$t_{n_{p+1},tp} = t_{n_p,tp} + ct \ \forall tp \in TP_{n_p} \ \forall tp \in TP_n^{macro} \ \forall n_p \in N_p \forall p \in P \qquad (21)$$

4.2.1. Run Time Constraints

The results of the micro-level TPE optimization algorithm are used to create the blocking time stairways for the trains in the macro-level timetabling problem. For each pair of consecutive macroscopic events tp' and $tp' + 1 \in TP_n^{macro}$, the $\underline{rt}_{n,tp}$ and $\overline{rt}_{n,tp}$ are the minimum and maximum process times permitted between the two events. For each pair of macro events, the constraints:

$$t_{n,\mathrm{tp}'+1} - t_{n,\mathrm{tp}'} \geq \underline{rt}_{n,tp} \ \ \forall tp', tp' + 1 \in TP_n^{macro}, \forall n \in N \qquad (14)$$

$$t_{n,\mathrm{tp}'+1} - t_{n,\mathrm{tp}'} \leq \overline{rt}_{n,tp} \ \ \forall tp', tp' + 1 \in TP_n^{macro}, \forall n \in N \qquad (15)$$

ensure that run, dwell, and turnaround times comply with planning rules. To simplify the formulation, it is assumed that in running arcs, any run time supplement provided above the minimum operational running time $\underline{rt}_{n,tp}$ is converted into dwell time supplement at the departing station. Future research will include developing algorithms to covert this dwell time supplement into run time supplement to add more flexibility and to facilitate more energy efficient driving.

4.2.2. Headway Constraints

The computed trajectories in section 4.1.1 are used to define the reservation start and end times relative to the macroscopic arrival events. For each track segment $ts \in TS_n$, the macroscopic release reference event $tp_{n,ts}^{rel} \in TP_n^{macro}$ for $ts \in TS_n$ is defined as the next event to occur after train $n \in N$ clears the release point for track segment ts. $tp_{n,ts}^{rel}$ will either be a station arrival event in the network, or the event where the train exits the modelled area. The time that train n releases tracks segment ts is governed by the blocking time stairway produced by the S-MTTC trajectory, which is fixed relative to $t_{n,tp_{n,ts}^{rel}}$. The time difference between the arrival is represented by the *macroscopic release offset* parameter $\tau_{n,ts}^{rel,S-MTTC}$, which depends only on the minimum run time trajectory (Wang et al, 2023). The constraint representing the reservation end time is:

$$re_{n,ts} = t_{n,tp_{n,ts}^{rel}} - \tau_{n,ts}^{rel,S-MTTC} + b_{min} \quad \forall ts \in TS_n \forall n \in N \tag{16}$$

where variable b_{min} is minimum permissible buffer time between trains' blocking time stairways.

The macroscopic reservation reference event $tp_{n,ts}^{res} \in TP_n^{macro}$ is the next arrival event to occur after train n begins reservation of segment ts. If train n uses a S-MTTC driving strategy, the time difference between the next arrival event and the reservation start time is represented by the *macroscopic S-MTTC reservation offset* parameter $\tau_{n,ts}^{res,S-MTTC}$ obtained from the minimum run time computation step. If train n instead follows the EETC driving strategy computed by the TPE optimizer, the time difference is represented by the *macroscopic EETC reservation offset* parameter $\tau_{n,ts}^{res,EETC}$. The constraint representing the reservation start time is:

$$rs_{n,ts} = t_{n,tp_{n,ts}^{res}} - \max\left(\tau_{n,ts}^{res,S-MTTC}, \tau_{n,ts}^{res,EETC}\right) - \tau_{comm} - \tau_{ctrl,n} \quad \forall ts \in TS_n \forall n \in N \tag{17}$$

The constraint states that train n must begin reservation of segment ts before it receives a brake indication for either driving strategy (S-MTTC or EETC). The times where the train would receive a brake indication are both fixed relative to $t_{n,tp_{n,ts}^{res}}$, so reservation must start before the earlier of the two possible events.

The time required to release the route of train n' and set up the route of train n at shared track segment ts $TS_n \cap TS_{n'}$ depends on where the two trains' routes differ at the segment. More time is required for set-up if the route of train n at ts differs from that of train n' because some switches would need to be moved to properly set-up the second route. The combined release and set-up time is represented by the parameter $\tau_{setup,ts,n',n}$. The RBC communication delay, a component of reaction time that is assumed to be constant across the network, is represented by parameter τ_{comm}. The train control delay for train $n \in N$ is represented by parameter $\tau_{ctrl,n}$.

For each track segment $ts \in TS_n \cap TS_{n'}$, which both trains n and n' must reserve to complete their journey, the binary variable $q_{n',n,ts}$ is set equal to 1 if train n' reserves it before train n. To simplify the model presentation, another binary variable $q_{n,n',ts}$ is created to represent the opposite order of reservation. The constraints preventing blocking time overlap are:

$$q_{n,n',ts} + q_{n',n,ts} = 1 \forall n, n' \in T \; \forall ts \in TS_n \cap TS_{n'} \tag{18}$$

$$rs_{n,ts,tp} \geq re_{n',ts,tp} + \tau_{setup,ts,n',n} - M\left(1 - q_{n',n,ts}\right) \forall n, n' \in T \; \forall ts \in TS_n \cap TS_{n'} \tag{19}$$

Constraint (18) requires that an order be specified between trains $n, n' \in T$ at any track segment ts that they both must reserve. Constraint (19) prevents blocking times from overlapping, ensuring that the timetable is conflict-free when considering set-up, reaction, and release time. These constraints have a microscopic form, but can be converted into macroscopic constraints.

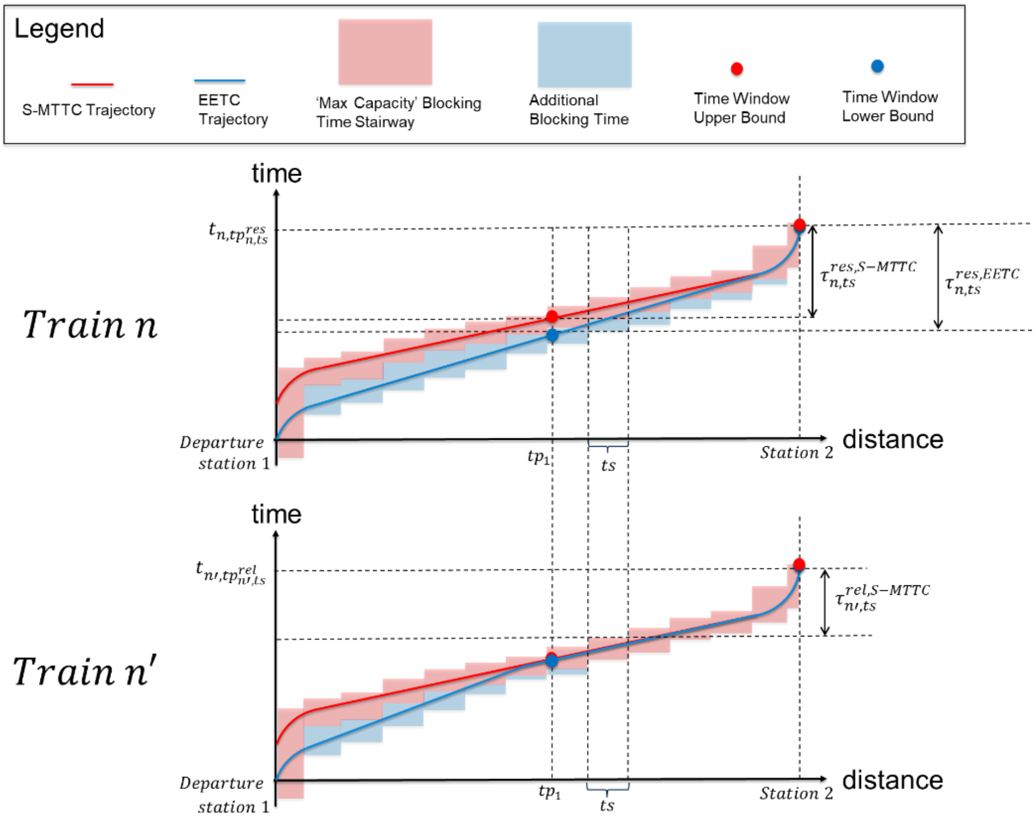


Figure 18 : Representation of Macroscopic Reservation and Release Offsets

An example of how a microscopic constraint for track segment $ts \in TS_n \cap TS_{n'}$ is converted into macroscopic constraints is shown in Figure 18 above. The first train diagram shows the TPE for train $n \in N$'s journey from *Station 1* to *Station 2,* and for train $n \in N$'s journey from *Station 1* to *Station 2.*

For train n', the next macroscopic event that occurs after it releases ts is the arrival event at *Station 2.* The exact time $re_{n',ts,tp}$ that train n' releases ts can therefore be expressed as a fixed offset (constraint (16)) from the time that train n' arrives at *Station 2:*

$$re_{n',ts,tp} = t_{n\prime,tp^{rel}_{n\prime,ts}} - \tau^{rel,S-MTTC}_{n\prime,ts} + b_{min}$$

Where $t_{n\prime,tp^{rel}_{n\prime,ts}}$ is the train n''s arrival event time at *Station 2,* and $\tau^{rel,S-MTTC}_{n',ts}$ is the time difference (in seconds) between that arrival event and the TPE reservation end time.

At the same time, the next macroscopic event to occur after train n begins reservation of ts is its own arrival event at *Station 2.* Its reservation start time for ts can therefore be represented as a fixed offset from the arrival event at *Station 2* (Constraint (17)):

$$rs_{n,ts} = t_{n,tp^{res}_{n,ts}} - \max\left(\tau^{res,S-MTTC}_{n,ts}, \tau^{res,EETC}_{n,ts}\right) - \tau_{comm} - \tau_{ctrl,n}$$

Where $t_{n,tp^{res}_{n,ts}}$ is the train n''s arrival event time at *Station 2,* $\tau^{res,S-MTTC}_{n,ts}$ is the time difference (in seconds) between that arrival event and the reservation start time for the S-MTTC trajectory, and $\tau^{res,EETC}_{n,ts}$ is the time difference (in seconds) between that arrival event and the reservation start time for the EETC trajectory. The larger of the two offsets is used in the model to preserve flexibility.

If train n' is routed through track segment ts before train ts is, the constraint of type (19) that prevents a blocking times overlap between the two train at ts is:

$$rs_{n,ts,tp} \geq re_{n',ts,tp} + \tau_{setup,ts,n',n} - M\left(1 - q_{n',n,ts}\right)$$

If the constraints defining variables $re_{n',ts,tp}$ and $rs_{n,ts}$ are substituted into this constraint, it produces the constraint:

$$t_{n,tp^{res}_{n,ts}} - \max\left(\tau^{res,S-MTTC}_{n,ts}, \tau^{res,EETC}_{n,ts}\right) - \tau_{comm} - \tau_{ctrl,n} \geq t_{n\prime,tp^{rel}_{n\prime,ts}} - \tau^{rel,S-MTTC}_{n\,\prime\,,ts} + b_{min} + \tau_{setup,ts,n',n} - M\left(1 - q_{n',n,ts}\right)$$

Which simplifies to:

$$t_{n,tp^{res}_{n,ts}} - t_{n',tp^{rel}_{n',ts}} \geq \tau_{setup,ts,n',n} + \tau_{comm} + \tau_{ctrl,n} + \max\left(\tau^{res,S-MTTC}_{n,ts}, \tau^{res,EETC}_{n,ts}\right) + b_{min} - \tau^{rel,S-MTTC}_{n',ts} - M\left(1 - q_{n',n,ts}\right)$$

Relating the arrival event of train n' at *Station 2* with the arrival event of train n at *Station 2.* This process can be used to convert every microscopic minimum headway constraint (19) into an equivalent macroscopic arrive-arrive constraint. The model thus has macroscopic computational complexity.

4.2.3. Periodicity Constraints

If a periodic timetable is being modelled, additional constraints are needed to ensure that the services repeat at regular intervals. For a given problem, it is assumed that services for P consecutive cycles are being modelled. Given that periodic tactical scheduling problem needs to be align with the nonperiodic real-time traffic management problem, it is necessary to model the tactical problem in a nonperiodic environment (as in Grafe and Schöbel, 2021). It is assumed that P cycles are sufficient to verify that conflict-freeness of the resulting timetable.

For each individual period $p \in P = \{0,1,2, \dots, p_{end}\}$, the set of services in the period is represented by set $N_p \subseteq N$. In this scenario, service $n_p \in N_p$ has the same characteristics (including route, rolling stock and scheduled station stops) as train $n_{p+1} \in N_{p+1}$ of the following cycle $p \in P$. The period's cycle time is represented by the variable ct. The cycle time represents the headway between any two services n_p and n_{p+1} in adjacent periods.

The periodicity constraints are:

$$sct + (p-1) * ct \leq\ t_{n_p,0} \leq sct + p * ct - timestep\ \forall n_p \in N_p\ \forall p \in P \tag{20}$$

$$t_{n_{p+1},tp} = t_{n_p,tp} + ct\ \forall tp \in TP_{n_p}\ \forall tp \in TP^{macro}_n\ \forall n_p \in N_p \forall p \in P \tag{21}$$

Constraint (20) requires that for every train $n_p \in N_p$ part of period $\forall p \in P$, the time that the train enters the network (event $t_{n_p,0}$) must occur within the period interval

$$[sct + (p-1) * ct, sct + p * ct - timestep]$$

Where sct is the time that the first period starts. Since $timestep$ is the smallest run time increment in the model, this constraint says that:

$$t_{n_p,0} \in [sct + (p-1) * ct, sct + p * ct)$$

This constraint ensures that all trains of period $p \in P$ start their journey in the modelled area in a common time window. Depending on the area being modelled, this constraint can be added at a different timepoint of the modeller's choice. Constraint (21) requires that the timing of $n_{p+1} \in N_{p+1}$ at every timepoint $tp \in TP_n^{macro}$ occur one cycle time after the timings of train n_p of the previous period. Constraints (20) and (21) are identical to ones used by Grafe and Schöbel, (2021) for modelling a periodic timetable in a non-periodic environment.

5. Case Study

To assess the effectiveness of the proposed model, a case study is conducted based on the South West Main Line (SWML) in the UK. This line, which runs from London to Weymouth, carries a mix of intercity services (including cities such as Exeter, Bournemouth, and Portsmouth), local services within Metro London, and regional services to London's commuter belt. The case is performed for ETCS Level 2 signalling on a model with 250 metre blocks in the microscopic simulator EGTrain (Quaglietta, 2014). The principal train operating company on the line is the *South Western Railway,* which operates all passenger services into the London Waterloo Terminal. Between Basingstoke and Southampton, the line also supports long-distance services operated by *CrossCountry* as well as some freight services. Waterloo station is the reference location for direction-keeping on the line. Trains in the direction of Waterloo are operating in the *Up* direction; trains in the opposite direction are operating in the *Down* direction. The case study focuses on the portion of the line between London Waterloo and Southampton Central. From Waterloo to Basingstoke, the line has four tracks. This section of the line consists of the *Main Slow Line* (MSL) that has platforms at all stops, and the *Main Fast Line* (MFL) that is used by services making a few stops in the section. Down from Basingstoke, the line is mostly double-track. The case study area is shown in red in Figure 19 below.

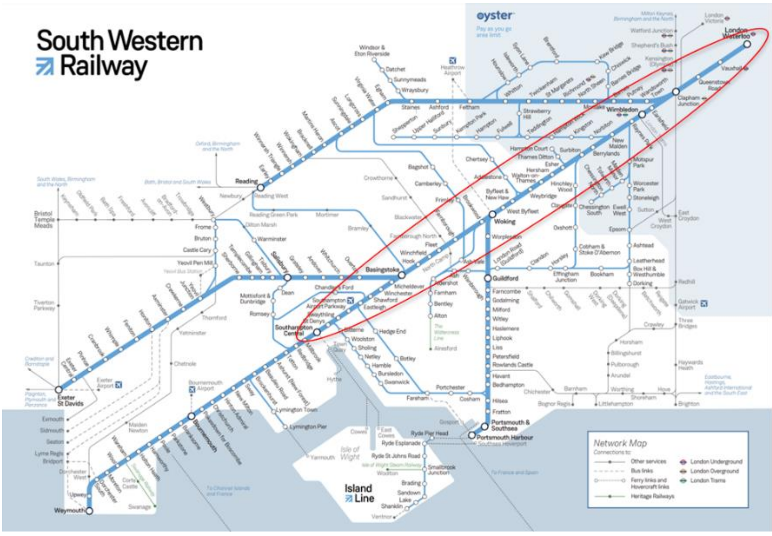


Figure 19 : Case Study area (Adapted from South Western Railway, 2022)

Services departing Waterloo on the Slow Line are typically operated by the British Railways Class 455 electric multiple units (EMUs) in 4- or 8-carriage compositions. These services operate within Metro London (e.g., the Chessington South, Hampton Court and Kingston Loop services), and provide slower services to London's commuter belt. The Fast line services consist of a combination of intercity services to outlying areas and commuter services to stops in the commuter belt. The commuter services typically run fast to Surbiton (the last stop within Metro London), where they switch to the Slow Line, making some (but not all) local stops between Surbiton and Woking. Commuter services are typically run by the Class 450 EMUs in 4-,8- or 12 carriage compositions. The intercity services either operate nonstop from Waterloo to Woking, or call only at Clapham Junction Station. Services operating past Basingstoke to Salisbury, Yeoville Junction and Exeter St. Davids are operated by Class 159 diesel multiple units (DMUs) (which can form 3-,6- or 9-carriage compositions), while services to other destinations are typically operated by Class 450s.

CrossCountry also operates services on the line between Basingstoke and Southampton using the Class 220 *Voyager* DMU. The line also sees some freight services, which are typically hauled by a Class 66 Diesel Locomotive. The performance characteristics of the trains are shown in Table 4.

Train Type	**Length**	**Total mass (kg)**	**Adhesion mass (kg)**	**Max Speed (mph)**	**Service Braking rate** (m/s^2)	**Cross-Section Area** (m^2)	**Sauthoff Resistance Coefficient** r_0	**Jerk** (m/s^3)
Class 455	81.60m (4 cars)	33,988	33,988	75	0.6	1.45	0.004	0.75
Class 450	80.92m (4 cars)	42,500	42,500	100	0.7	1.45	0.004	0.75
Class 159	69.63m (3 cars)	37,800	37,800	90	0.6	1.45	0.004	0.75
Class 220	93.34m (4 cars)	46,400	46,400	200	0.8	1.45	0.004	0.75
Class 66	21.39m	129,600	129,600	65 or 75	0.6	1.45	0.004	0.75

Table 4: Performance characteristics of passenger trains used on South West Main Line

The sets $S_{n,tp}, tp \in TP_n, n \in N$ of speed-distance trajectory options are computed using simulation-based methods. For the set-up of the case study, the following assumptions were made when generating speed-distance trajectories:

- If timepoint $tp \in TP_n$ is the departure event for a scheduled stop, the maximum permitted start speed v_1 of trajectory $s_{n,tp}(v_1, v_2)$ must be zero for the trajectory to be deemed feasible. If timepoint $tp + 1 \in TP_n$ corresponds to an arrival event, only trajectories $s_{n,tp}(v_1, v_2)$ where the end speed is zero (i.e., $v_2 = 0$) can be feasible.
- Trains are otherwise permitted to cruise at speeds that are multiples of 30km/h, at 40km/h, the maximum permitted speed they can reach (which is dependent on the train's own acceleration ability, maximum speed, and any speed restrictions on the line), and 5km/h below the maximum attainable speed.
- For each train, timepoints are never more than 500 metres apart. Timepoints are also placed at any location where there is a speed limit increase or decrease, and at the braking points (for each scheduled station stop) corresponding to the permitted cruising speeds.
- Trains may change between the permitted cruising speeds by accelerating at their maximum capacity, or by using maximum braking (with a constant braking rate in each interval).
- If a train is cruising at speed v_1 at timepoint tp, but cannot accelerate/brake to cruising speed v_2 within that interval alone, the trajectory option $s_{n,tp}(v_1, v_3)$ is created, where v_3 is the ending speed at $tp + 1$ with maximum acceleration/braking. In the following interval $tp + 1$, an additional speed distance trajectory $s_{n,tp+1}(v_3, v_4)$ is created to continue modelling maximum acceleration/braking. This continues until cruising speed v_2 is reached.
- All coefficients in the model (including run times) are rounded with $timestep = 0.1s$.

Once all trajectory options have been generated for service $n \in N$, the technical minimum run time is obtained by constructing a model for that train's journey and minimizing the total end-to-end journey time. This problem is solved to optimality using IBM CPLEX, and the results are then used to determine the run time supplements for the tactical scheduling problem. The timepoint optimization problems were solved to optimality within 10 seconds for every service in precomputation.

The resulting technical minimum run times interstation pairs in the down direction on the Slow Line with an 8-carriage Class 455 composition are shown in Table 5. The first and second columns list the arrival and departure stations. The technical minimum run times are compared against the scheduled run times from the May 2023-December 2023 working timetable WH01 for Monday-Friday service (Network Rail, n.d.-d), which covers the line from Waterloo to Surbiton. At some minor stations, the same time is indicated for both arrival and departure. The table indicates whether the dwell time at the arrival station is included in the calculation. Some interstation pairs also have run times which vary among trains. The minimum scheduled time is used if this occurs. The reported times in Table 5 exclude time allowances.

Depart Station	Arrive Station	Minimum Run Time (mm:ss)	Reference Service	Scheduled Run Time (From Working Timetable)
Waterloo	Vauxhall	03:30	2J43	05:00 incl. Dwell at Vauxhall
Vauxhall	Clapham Jn	03:56	2J43	04:00
Clapham Jn	Earlsfield	02:44	2J43	03:30 incl. Dwell at Earlsfield
Earlsfield	Wimbledon	02:46	2J43	03:00
Wimbledon	Raynes Park	02:28	2J43	03:00 incl. Dwell at Raynes Park
Wimbledon	Surbiton	06:12	2G43	06:30
Raynes Park	New Malden	02:10	2J43	02:30 incl. Dwell at New Malden
New Malden	Berrylands	02:14	2J43	02:30 incl. Dwell at New Malden
Berrylands	Surbiton	02:03	2J43	03:00
Surbiton	Berrylands	01:59	2J46	02:30 incl. Dwell at Berrylands
Surbiton	Wimbledon	06:18	2G44	07:30
Berrylands	New Malden	02:10	2J46	03:00 Incl. Dwell at New Malden
New Malden	Raynes Park	02:07	2J46	02:30 Incl. Dwell at Raynes Park
Raynes Park	Wimbledon	02:23	2J46	03:00
Wimbledon	Earlsfield	02:01	2J46	03:30 Incl. Dwell at Earlsfield
Earlsfield	Clapham Jn	02:48	2J46	03:00
Clapham Jn	Vauxhall	02:39	2J46	05:00 Incl. Dwell at Vauxhall

Table 5: Technical Minimum run times from Model compared to scheduled run times in the May-December 2023 Timetable

The technical minimum run times are largely aligned with the run times shown in the schedule. Differences between the minimum times provided by the model and Network Rail's working timetable are likely attributable to the fact that the UK working timetable is specified in 30-second increments. The technical minimum run times are expected to be slightly higher than actual, as the construction of trajectories with braking assumes a constant braking rate (which may not be reflective of actual practice). In any case, the run times provided are close enough to be used to schedule services.

The initial verification of the tactical scheduling model is for a possible hourly pattern containing the following services. The first two columns show the outer destinations of the trains, and their stopping pattern on the SWML from Waterloo to Surbiton. The sample timetable (which is meant to fit within an hourly cycle) is shown in Table 6 below.

Service	Rolling Stock Type	Stopping Pattern	Frequency (Tph)	Regularity Requirements
Kingston Loop/Shepperton	Class 455 (8 coaches)	All Stops to New Malden	4	Even Headways
Dorking/Guildford Via Epsom	Class 455 (8 coaches)	All Stops to Raynes Park	4	Even headways
Chessington South	Class 455 (8 coaches)	All Stops to Raynes Park	2	Even Headways
Hampton Court	Class 455 (8 coaches)	All Stops to Surbiton	2	Even Headways
Woking (via Slow line)	Class 455 (8 carriages)	All Stops to Wimbledon, Fast to Surbiton	2	Even Headways
Guildford via Cobham	Class 455 (8 carriages)	All Stops to Wimbledon, Fast to Surbiton	2	Even Headways

Table 6: Sample Timetable for Slow Line

Tests were performed with the regularity requirements for services with multiple trains per cycle. The tests were done for cases with all trains using ETCS Level 2 signalling, with 250m blocks, and for trains with three-aspect signalling with 500m block lengths. The 500m block length used for the three-aspect test is the minimum acceptable length given the Slow Line's speed limit and the performance characteristics of the Class 455. The goal of the initial verification is to test that the model works for both signalling types, and to assess the capacity impact of ETCS level 2 relative to the best-case layout for three-aspect signalling within Metro London.

An additional test is performed on all services on the line from London to Southampton Central (~127km in length). The goal of this test is to verify that the macro scheduling model is scalable to large areas, and to assess the capacity impact of ETCS Level 2 versus the current timetable. An additional test is performed on the entire line to demonstrate the algorithm's ability to support operations with mixed signalling. In this test, the Class 455s operating with three-aspect signalling with 500m blocks, and all other rolling stock types using ETCS Level 2 with 250m block lengths. Each three-aspect block over open track contains two physical track detection sections (each corresponding to a single ETCS Level 2 block), and route release is sectionalized at the level of the individual track detection section.

Tests for the Fast Line (MFL) include the services shown in Table 7 as well as the Slow Line (MSL) services shown in Table 6 above. This timetable is meant to be representative of a typical off-peak service pattern for the South West Main Line. The timetable includes one freight service per hour operating non-stop between Basingstoke and Eastleigh. It is pathed for a freight train of up to 700m in length, with a maximum speed of 65mph and a total weight of up to 579,600 kg, that is hauled by a single Class 66 diesel locomotive.

Service	Composition	Stations Called at (line)	Frequency (Tph)	Regularity Requirements
London – Weymouth	Class 450 (12 coaches)	Waterloo (MFL), Woking (MFL), Winchester, Southampton Airport Parkway, Southampton Central	2	Even headways
London – Salisbury	Class 159 (9 coaches)	Waterloo (MFL), Basingstoke (MFL)	2	Even headways
London – Alton	Class 450 (12 coaches)	Waterloo (MFL), Surbiton (MSL), West Byfleet (MSL), Woking (MSL), Brookwood (MSL)	2	Even headways
London – Basingstoke	Class 450 (12 coaches)	Waterloo (MFL), Woking (MFL), Brookwood (MSL), Farnborough (Main) (MSL), Fleet (MSL), Winchfield (MSL), Hook (MSL), Basingstoke (MSL)	2	Even headways
London – Portsmouth via Eastleigh	Class 450 (12 coaches)	Waterloo (MFL), Woking (MFL), Basingstoke (MFL), Micheldever, Winchester, Eastleigh	2	Even headways
London – Portsmouth via Direct	Class 450 (12 coaches)	Waterloo (MFL), Woking (MFL)	2	Even headways
Birmingham – Bournemouth	Class 220 (4 coaches)	Basingstoke (MFL), Winchester Southampton Airport Parkway Southampton Central	1	N/A
Brighton – Bournemouth	Class 450 (8 coaches)	Southampton Central	2	Even headways
Reading – Portsmouth Freight	Class 66, 65mph, 700m length, 579,600 kg	Non-Stop From Basingstoke to Eastleigh	1	N/A

Table 7: Sample Timetable for Fast Line

Tests for the entire line are intended to assess the applicability of the proposed scheduling algorithm to large-scale timetabling problems. A complete assessment of the line's capacity cannot be performed because the model used is based on the infrastructure in place in 2015. It excludes the Hampton Court Reversible (a third track at Surbiton Station in the down direction), and a pair of switches near Hampton Court Junction which are used in the current timetable. These pieces of infrastructure are designed to reduce the number of conflicts between trains moving from the Fast to Slow lines and services branching off from the Slow Line to Hampton Court or the New Guildford line. The validity of the three-aspect block lengths (i.e., whether the layout complies with safety rules for three-aspect signalling) is also not verified past Surbiton because of time limitations. For these reasons, no information on the capacity of the South West Main Line can be derived from the tests involving the entire line.

The minimum cycle time tests (with objective (15)) involve compressing the trains' stable TPEs to obtain the minimum cycle time. This process differs from the UIC Code 406 (2013) guidelines for calculating capacity consumption which involves compressing the precise EETC train trajectories.

Thus, capacity consumption using the proposed framework is expected to be higher than if the UIC Code 406 method were followed. For this reason, the results of the computational experiments cannot be assessed using UIC Code 406 guidelines for timetable stability.

The following assumptions are made in all cases examined:

- All train timings and order are selected by a centralized agent.
- In tests with objective (14) (maximizing minimum buffer time), the cycle time is constrained equal to 1 hour (3600s).
- Combined Release- and Set-up time is 2s on open track, and 10s in interlockings if route set-up requires a switch to be moved.
- Radio Block Centre Communication Delay τ_{comm} is 2s.
- The train control delay (due to the delayed reaction of the driver/ATO system and the train) $\tau_{ctrl,n}$ is 4s.
- The time required to perform train dispatch process $\tau_{n,tp}^{dis}$ is 12s.
- Train position error for calculating the braking curve pos_err_n is 20 metres.
- In problems where the objective is to maximize the minimum buffer time b_{min}, solutions with b_{min} under 20s are deemed infeasible.
- For each train $n \in N$, the timepoint event used for constraint (20) is the arrival event time at Waterloo Station. If train $n \in N$ does not serve Waterloo station, the event where it enters the network is used instead.
- The minimum turnaround time at Waterloo station (if turnaround constraints are included) is 8 minutes. No maximum turn-around time is specified.
- If trains n and n' share the same rolling stock (if turnaround constraints are included), all constraints of type (18) between the two trains are omitted from the model.
- Run time supplement for each interstation pair must be between 5% and 7% of the technical minimum running time obtained in pre-processing.
- A 10-hour (36,000s) solving time limit was used for all cases.

5.1. Results and Discussion

The computational test results for the eight cases are shown in Table 8 below. For each case, the table shows the number of cycles and individual services modelled, the cycle time obtained after 1 hour along with the capacity consumed (assuming a 1-hour period), and the optimality gap. The solving time is shown in parentheses. In some cases, the exact solving time exceeds 1 hour. This occurred on some tests because the solving algorithm was in the middle of performing a solving procedure at the time when solving time reached 1 hour. The train diagrams for each test are shown in Appendix A.

Case	Signalling in use	No. Hourly Cycles	No. Trains	Max Min Buffer Time		Min Cycle Time	
				Objective Value	Gap (time)	Objective Value (Capacity)	Gap (time)
MSL only, with Turnaround	ETCS Level 2	4	128	41.8s	0% (8:34)	47:06 (78.5%)	0% (6:09)
MSL only, no turnaround	ETCS Level 2	4	128	41.8s	0% (18:54)	47:06 (78.5%)	0% (1:26:06)
MSL only, with turnaround	3-aspect	4	128	27.4s	0% (8:29)	47:22 (78.9%)	0% (33:49)
MSL only, no turnaround	3-aspect	4	128	27.8s	0% (11:35)	47:20 (78.9%)	0% (1:40:07)
Whole line, with turnaround	ETCS Level 2	4	256	41.8s	0% (29:26)	50:09 (83.5%)	0% (14:44)
Whole line, no turnaround	ETCS Level 2	4	256	41.8s	0% (40:11)	47:06 (78.5%)	0% (9:33:27)
Whole line, with turnaround	Class 455: 3-aspect, otherwise ETCS L2	4	256	27.4s	0% (39:23)	50:09 (83.5%)	0% (19:18)
Whole line, no turnaround	Class 455: 3-aspect, otherwise ETCS L2	4	256	27.8s	0% (1:03:38)	47:20 (78.8%)	3.49% (10:00:02)

Table 8: Results of Computational experiments

Results show that the macroscopic scheduling model can produce micro-feasible timetables within 1 hour for the entire line from London Waterloo for Southampton Central. In seven of the eight cases with the objective of maximizing minimum buffer time, the solver was able to prove optimality within the 1-hour time limit. The results for the maximum-minimum buffer time indicate that the minimum cycle time is less than 1 hour (3600s) in all eight scenarios examined. When the objective of minimizing the cycle time is used, the solver could only prove optimality for the whole line within 1 hour when turnaround constraints are added. The inclusion of turnaround constraints simplifies the problem by creating strict precedence constraints among services that share the same rolling stock. This logic reduces the number of binary ordering variables in the problem. Without turnaround constraints, the model was able to find the minimum cycle time within 10 hours in 3 of 4 cases.

As compared to the existing three-aspect signalling on the Main Slow Line, migration to ETCS Level 2 with optimized block lengths reduces capacity consumption by 16 seconds. The minimum buffer time in the timetable can also be increased by 52% (from 27.4s to 41.8s) to improve timetable stability. A similar increase in minimum buffer time is observed in cases without the minimum turnaround time constraints at London Waterloo. The increased associated with ETCS Level 2 is attributable in part to the train-based braking supervision that allows approach times to be computed based on the train's actual trajectory instead of the worst-case braking distance. The ability of ETCS Level 2 to accommodate shorter block lengths also allows trains to reserve the track segments containing station platforms faster than would be possible with three-aspect signalling. The finer discretization of the infrastructure enabled by DTG signalling further reduces the approach time for some track detection sections on the line.

The relatively small difference in minimum cycle time for ETCS L2 vs. 3-aspect signalling on the Slow Line is attributable to that fact that the Waterloo station throat is the main bottleneck on the line. Trains are limited to 15mph in the throat, and there are conflicts between trains entering and leaving. It is also not possible to homogenize the TPEs for trains in the down direction because the routes through the throat need to be set up to begin the departure process at Waterloo station. Trains in the Up direction are also travelling as fast as possible through the throat. This maximizes theoretical capacity by homogenizing the EETC trajectory with the S-MTTC trajectory, but limits the degree to which the train-based braking supervision can reduce approach time versus 3-aspect signalling.

The comparatively worse performance of the tests aimed at minimizing cycle time (objective (15)) in proving optimality is likely due to the larger feasible solution space. In tests on the whole line, the solver was unable to find feasible solutions for the desired 1-hour period within the solving time limit. The cycle time constraint used in cases with objective (14) has the effect of cutting off train orderings whose cycle time must exceed 1 hour. The absence of this constraint in the minimum cycle time cases makes it is more difficult for the solver to determine whether to prune that set of train orders from the branching tree. Thus, it may be beneficial to impose a maximum cycle time constraint for the purpose of cutting off solutions which are already known to be suboptimal, or which have a cycle time that exceeds the maximum acceptable value for the line plan. Future research should explore methods for improving computation time of the minimum cycle time problem.

6. Conclusion

This thesis seeks to determine whether alignment of the tactical and operational planning levels could produce capacity-effective rail operations on lines with conventional or digital signalling. A new method for generating capacity-optimal TPEs is proposed for scenarios where the timetable structure (i.e., the orders of trains) is unknown beforehand. Special properties of the TPE allow the TPE optimization problem to be decentralized at the level of the individual train in the network. This permits very detailed modelling of any signalling system's safety constraints without sacrificing applicability to the large-scale timetabling problems faced by IMs. The optimized TPEs are then used to generate the run time and minimum headway constraints for a state-of-the-art macroscopic model for tactical timetabling. Results from tests on the South West Main Line (SWML) in the UK show that the algorithm can produce conflict-free timetables over large networks. In addition to guaranteeing a conflict-free plan (a necessity for microscopic models), the resulting timetabling model is not more complex than state-of-the-art macroscopic methods for solving large-scale timetabling problems.

6.1. Answers to Research Questions

This section outlines the conclusions for each of the research sub-questions.

6.1.1. What is the current state-of-the-art for train service planning and real-time rescheduling models used for conventional and DTG signalling?

Most existing *microscopic* models are designed around the safety rules for three-aspect signalling. The calculation of braking distances (i.e., approach time) is based on the lengths of the blocks on the line and the number of aspects shown to the driver. Run time constraints assume that blocks are at least as long as the worst-case braking distance for the line. These assumptions limit the applicability of those algorithms to signalling systems with train-based braking distance computation (the essential characteristic of DTG signalling) or to networks where blocks are shorter than the worst-case braking distance. This assumption precludes their use by IMs who are seeking to increase line capacity by leveraging the ability to discretize the infrastructure into shorter blocks.

To date, research on modelling for DTG signalling has largely focused on redefining the braking distance computation from processes applicable to multi-aspect signalling to those relevant to DTG signalling. Methods proposed in published research either calculate braking distance based on the maximum speed (e.g., Janssens, 2022), or use discrete speed levels (e.g., Versluis et al, 2023) to represent the relationships between speed and run time, and speed and braking distance. The simplification of the continuous speed-braking distance relationship can also fail to utilize available capacity to the extent that braking distance is overestimated.

Current methods for solving large-scale timetabling problems are *macroscopic,* in that they represent the network at the level of the stations/interlockings on the route. Minimum headway constraints are set by the schedulers based on their experience and judgement of what the infrastructure can handle, as opposed to representations of all the individual processes taking place in regular operations. This process creates a risk that the computed timetable will contain conflicts in some areas (needing to be resolved through alternative methods), or result in sub-optimal use of available capacity. This practice may hinder the ability of IMs to run more trains on the existing infrastructure and require additional methods to produce feasible timetables in areas with dense traffic.

6.1.2. How can tactical and operational planning be consistent with train dynamics and signalling constraints of conventional and digital rail operations?

A novel microscopic scheduling model is proposed that precisely represents the train's trajectory in continuous time and space. The precise specification permits more accurate calculation of run times and blocking times relative to models with discrete speed levels or no speed modelling, thus making the model suitable for scheduling in networks with conventional or digital signalling with blocks of any (or no) length. The more precise trajectory representation also permits more detailed modelling of the continuous speed-minimum separation relationships in areas where trains are accelerating or braking. This model is used to created capacity-effective TPEs which can be used by ATO or C-DAS to guide the train through the network. Special properties of the TPEs (identified by Wang et al, 2023) are used to decentralize the TPE optimization process. This approach allows for faster computation time relative to processes where the TPEs are calculated centrally. The capacity-effective TPEs are then successfully used to generate the run time and minimum headway constraints for a state-of-the-art macroscopic model.

6.1.3. What method can be used to align tactical and operational train service planning in either conventional or digital rail operations?

Trains' blocking times are modelled using TPEs that enable conflict-free ATO/C-DAS operations, and provide enough flexibility for trains to recover from small delays (up to the run time supplement) without causing a conflict. The TPE consists of a S-MTTC trajectory (representing the case where the train departs late and drives as fast as possible to arrive at the next station on time) and an EETC trajectory (representing the case where the train departs on time, and consumes the run time supplement between the two stations). The requirement that the S-MTTC trajectory is contained in the TPE means that all block release times are fixed once the trains' macroscopic arrival events are fixed (Wang et al, 2023). This property is exploited through a macro-micro decomposition, where the macroscopic arrival events are fixed in the initial problem, and a timepoint configurations are chosen in the subproblem. It is shown that the timepoint configuration subproblem can be parallelized at the level of the individual train's inter-station journeys. It is also shown that TPEs that maximize the homogeneity of the S-MTTC and EETC trajectories have the greatest likelihood of being feasible in the timepoint configuration problem. This principle is used to construct capacity-effective TPEs offline, which are used to generate microscopic minimum headway constraints. These microscopic constraints are then propagated into macroscopic minimum headway constraints for a large-scale macroscopic timetabling model. The result is that the macroscopic timetabling model can produce feasible, stable, and capacity-effective timetables for large areas.

6.1.4. What impact can an aligned tactical and operational train planning framework have on schedule quality?

Compared to state-of-the-art methods, the run time and headway constraints in the proposed timetabling model are based on actual track occupation information, and do not require simplification of the processes occurring in regular operation. Thus, timetables produced by the proposed model are guaranteed to be feasible and make use of all available network capacity. This model could also assist the line planning process by verifying whether a given line plan can be made stable enough for regular operation.

On the Slow Line on the South West Main line, using the peak frequencies from the Fall 2023 schedule, migration to ETCS Level 2 (with 250m blocks) allows the minimum buffer time between trains to increase by 52% versus 3-aspect signalling with 500m blocks (from 27.4s to 41.8s). A similar increase (from 27.8s to 41.8s) is observed if turnaround constraints at Waterloo terminal are omitted. These increases are attributable to the ability of ETCS L2 to tailor brake supervision to the train's actual speed and performance characteristics, and to support shorter blocks than are possible with 3-aspect signalling. The minimum cycle time for ETCS L2 (48:08) is 16 seconds lower than that of 3-aspect signalling (48:22) with minimum turnaround constraints included. The relatively small difference in minimum cycle times is attributable to that fact that the Waterloo station throat is the main bottleneck on the line. Trains are limited to 15mph in the throat, and there are conflicts between trains entering and leaving. It is also not possible to homogenize the TPEs for trains in the down direction because the routes through the throat need to be set up to begin the departure process at Waterloo station. The TPE for trains in the Up direction includes the S-MTTC trajectory to guarantee stability, so there is limited ability to reduce approach time in that area.

6.1.5. What recommendations should IMs consider for aligning scheduling constraints in models for the tactical and operational levels?

To realize the available capacity of their networks, IMs should use microscopic representations of train dynamics and signalling processes when generating macroscopic train planning rules. The process should include the IM's processes for configuring ATO/C-DAS timepoints at the traffic management phases. The inclusion of the ATO/C-DAS timepoint construction process is especially important, as it ensures that modelled train dynamics are reflective of the processes that occur in regular operation. Inclusion of ATO/C-DAS timepoint configuration process also increases the tractability of the scheduling algorithm by reducing the computational complexity to that of a state-of-the-art macroscopic model. IMs should strive to maintain detailed models of their infrastructure so they can generate accurate minimum headway constraints that are reflective of the train's actual capabilities.

6.2. Recommendations For Future Research

The proposed novel algorithms for optimizing TPEs, and tactical and operational scheduling represent an important first step towards achieving an integrated planning and operational process. Future research could continue to build on this work by:

- Developing methods to decompose the tactical problems, so the framework can be applied to even larger networks. This could involve creating novel geographic or temporal decomposition methods for the macroscopic problem described in section 4.2.
- Verifying and validating the model for real-time rescheduling applications.
- Designing an algorithm to optimize departure times and TPEs to maximize timetable flexibility and provide more opportunities for energy-efficient driving. The model proposed in section 4.2 assumes that run times between stations are set at the minimum permitted

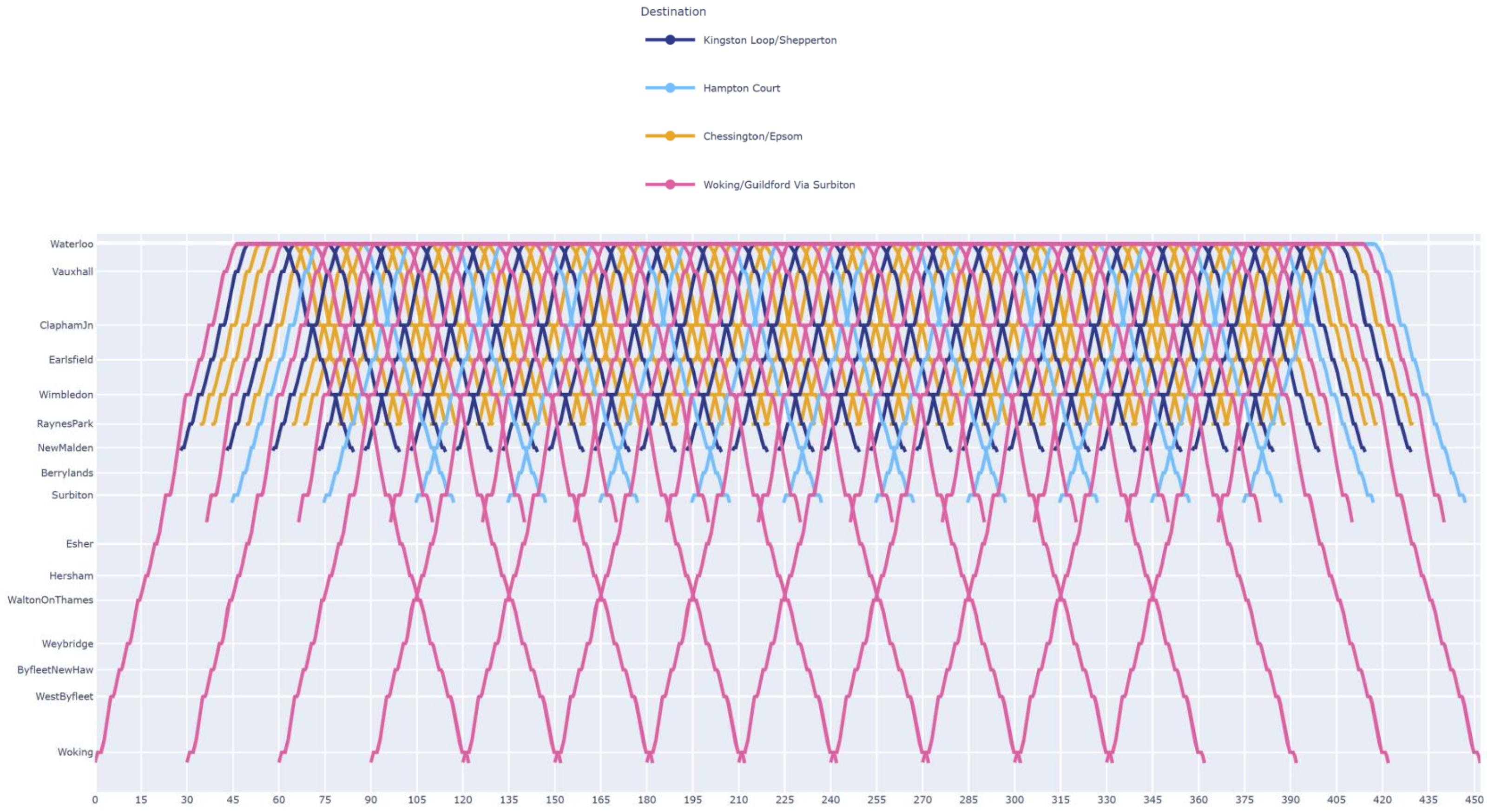
time given stability requirements. Although this guarantees that the timetable is acceptable for regular operation, it may not make use of all the flexibility available in the computed timetable structure.

- Adding the capability to choose routings in complex station and junction areas to increase the likelihood of finding a feasible timetable.

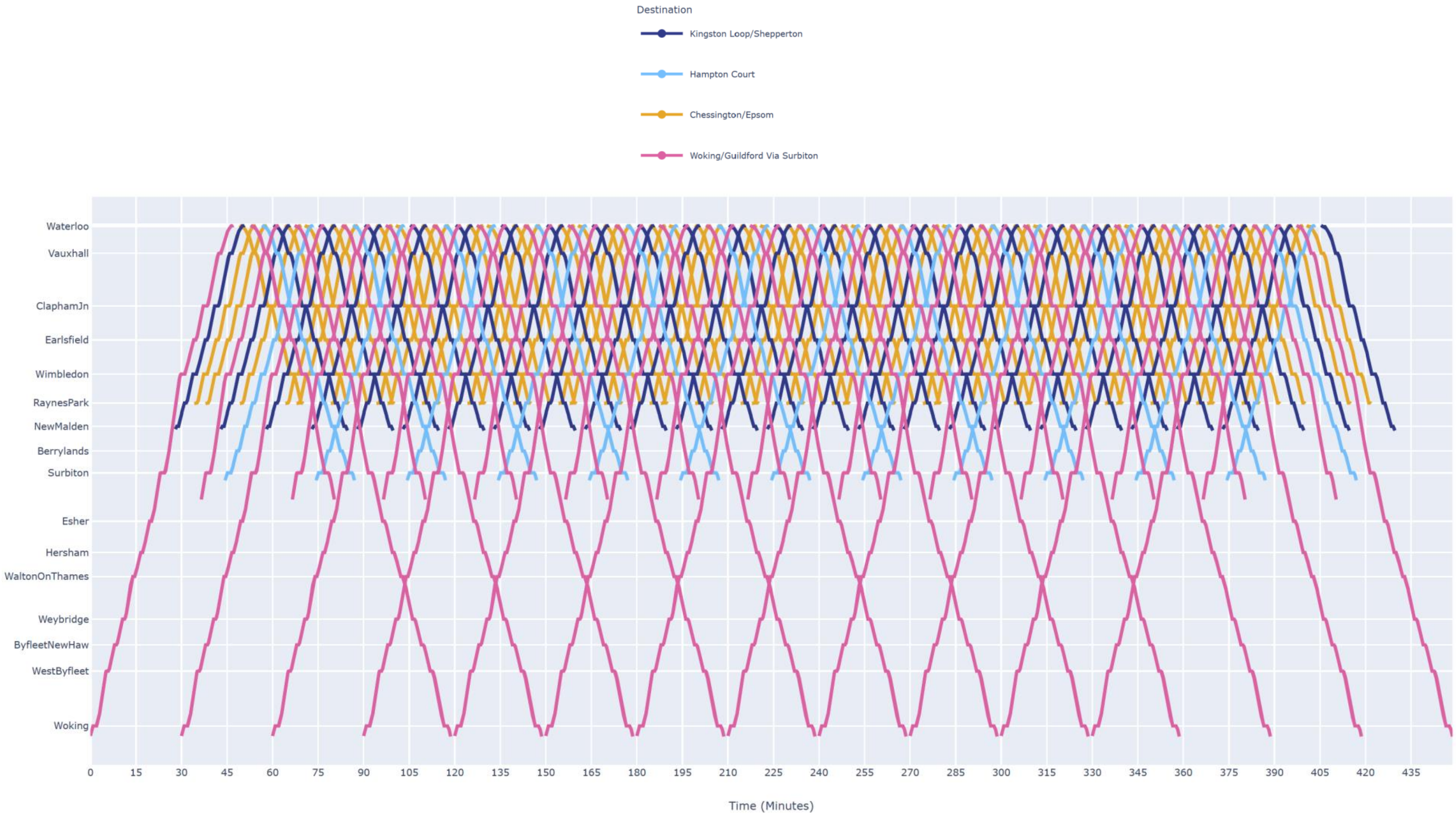
A. Appendices

A.1 Results for Main Slow Line with ETCS Level 2 signalling (6 cycles shown)

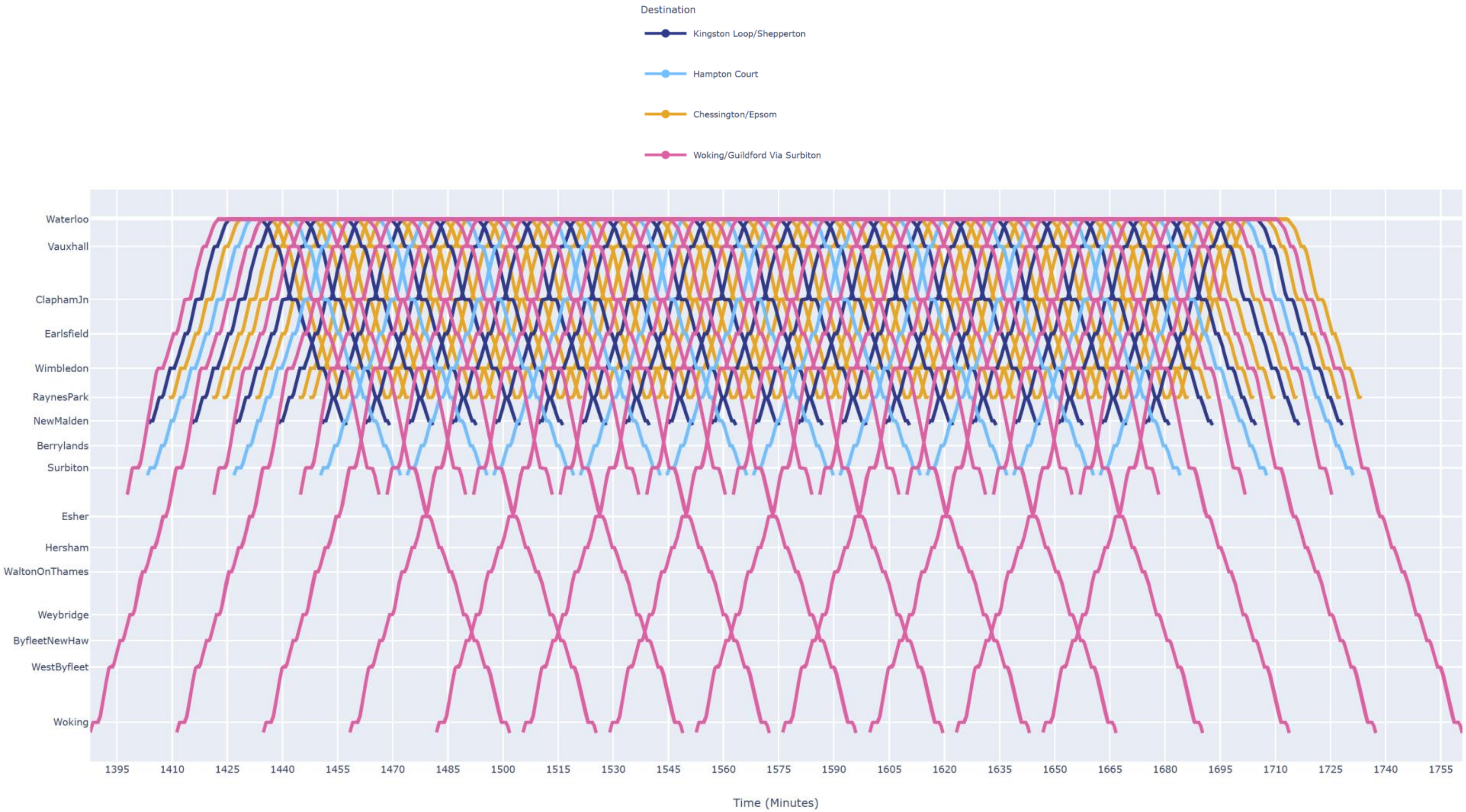
A.1.1 Maximized Buffer Time including Turnaround Constraints



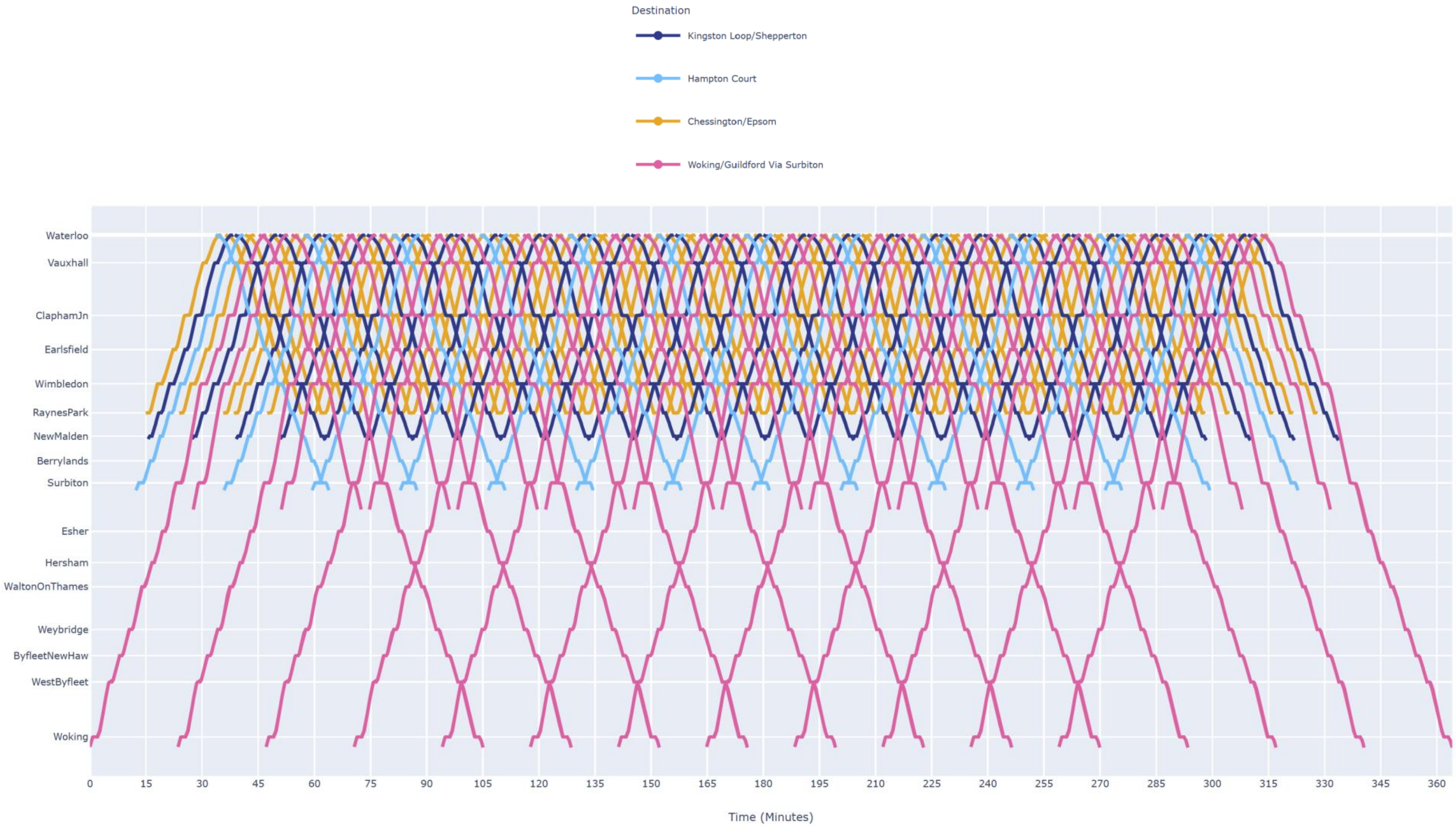
A.1.2 Maximized Buffer Time with No Turnaround Constraints



A.1.3 Minimized Cycle Time with Turnaround Constraints

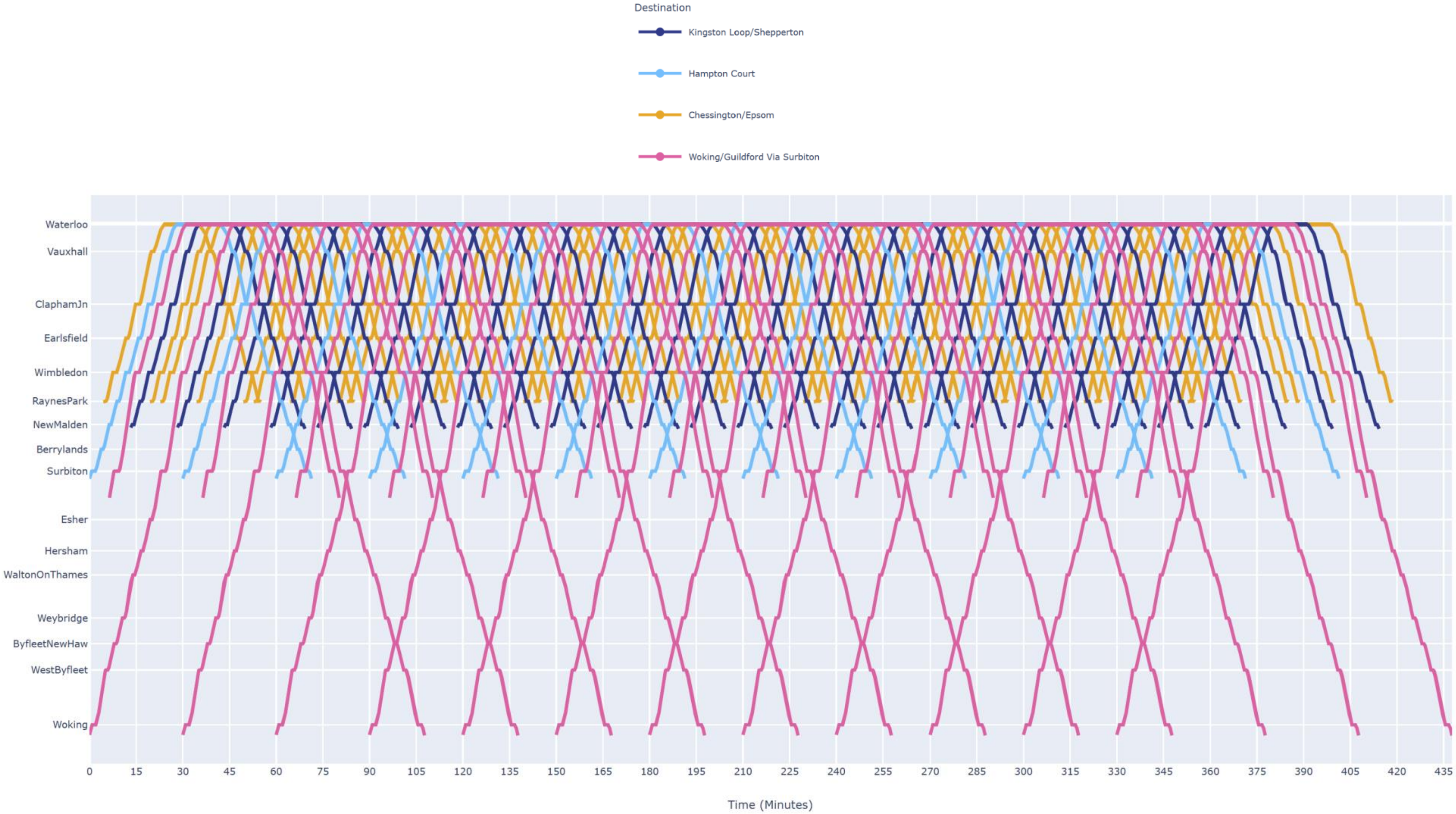


A.1.4 Minimized Cycle Time with No Turnaround Constraints

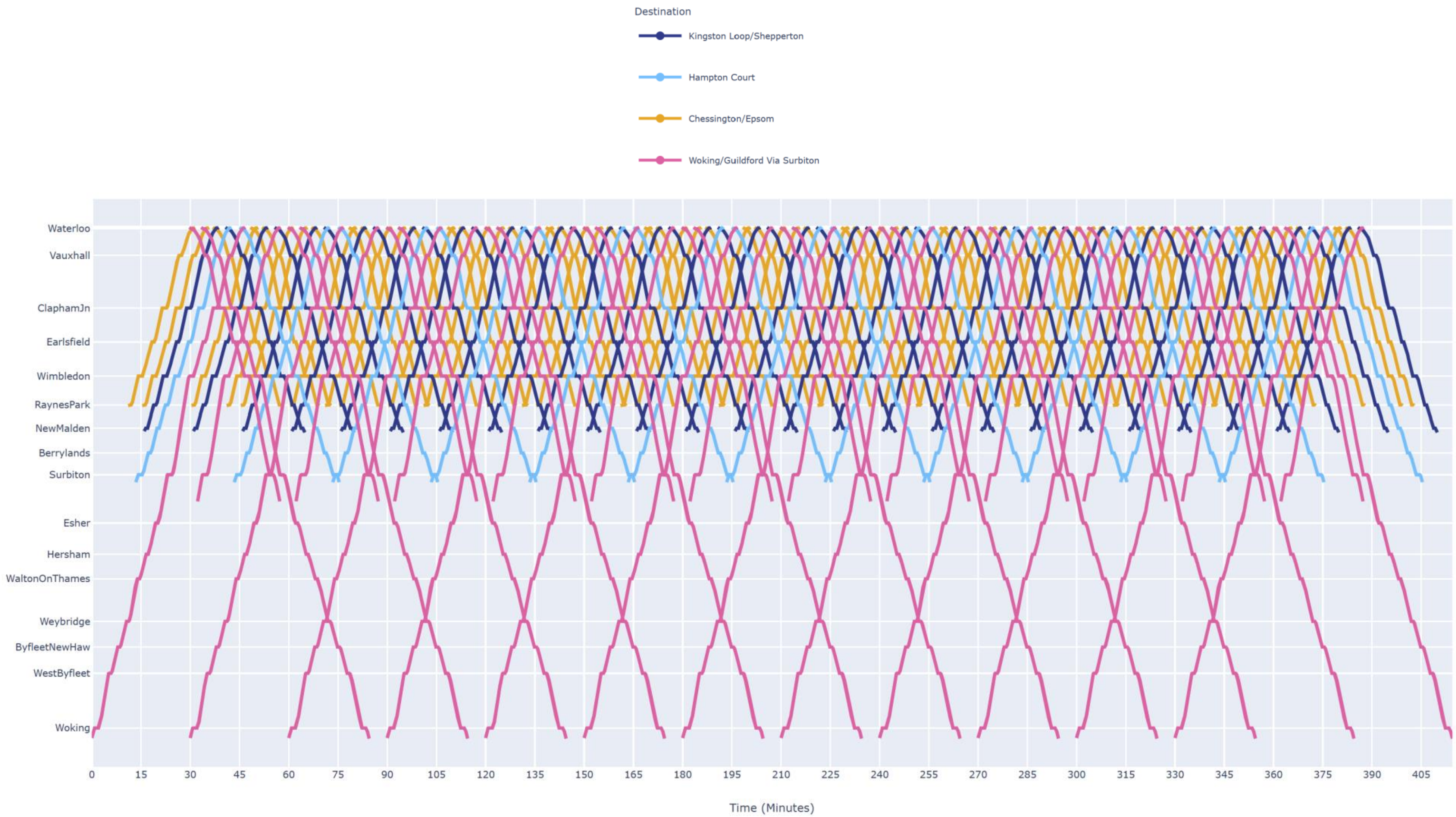


A.2 Results for Main Slow Line with three-aspect Signalling (6 Cycles Shown)

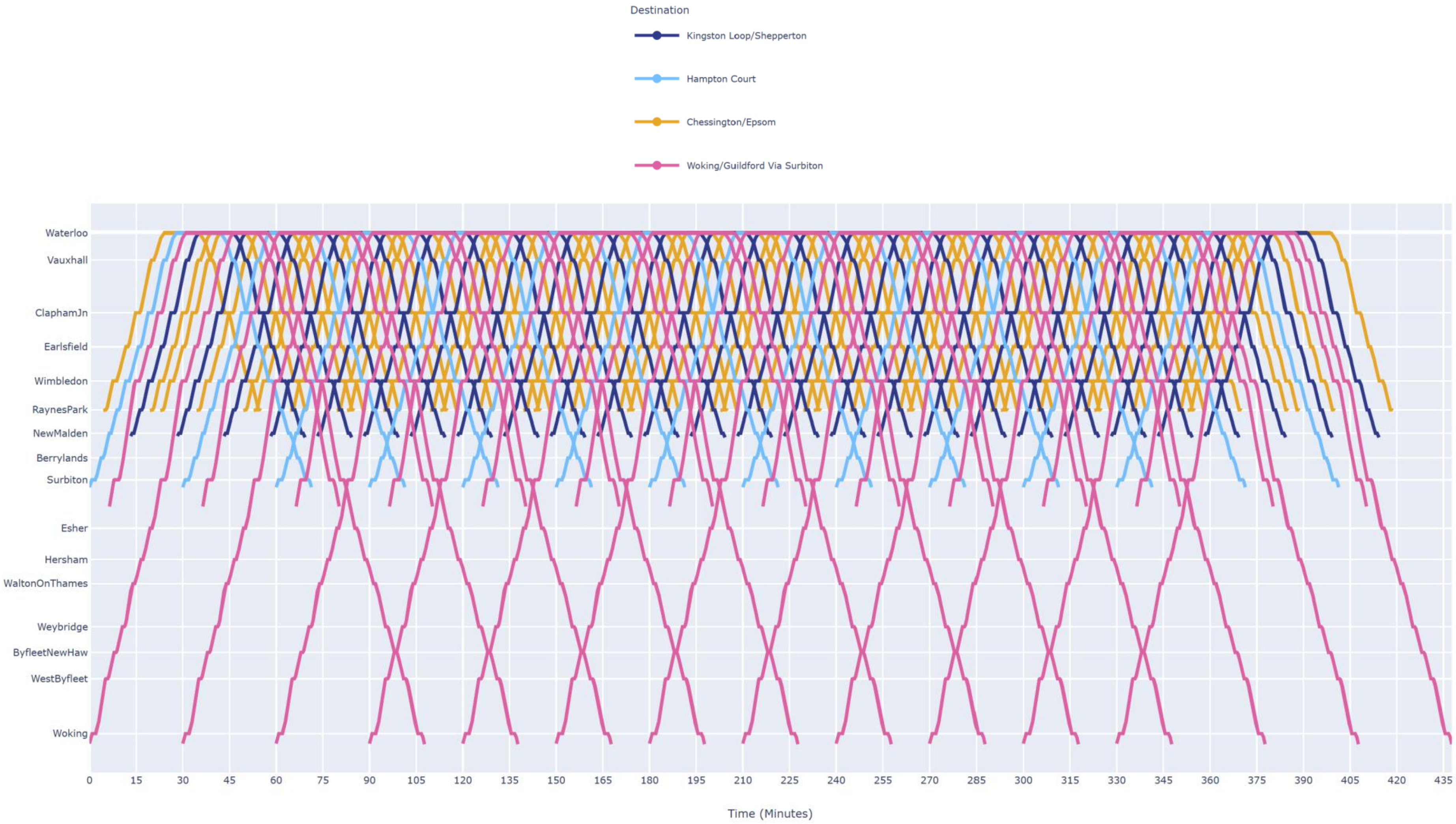
A.2.1 Maximized Buffer Time with Turnaround Constraints



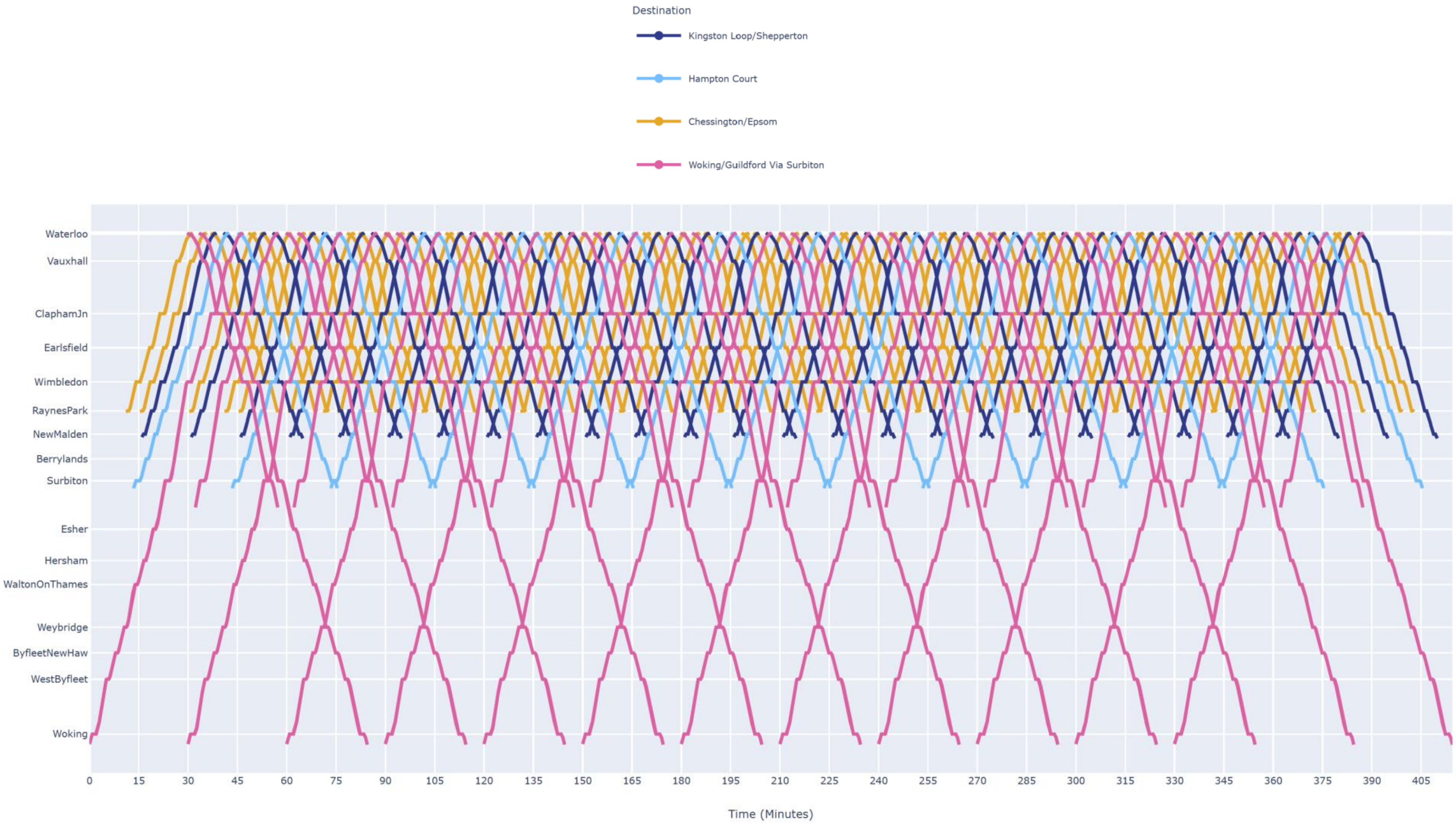
A.2.2 Maximized Buffer Time with No Turnaround Constraints



A.2.3 Minimized Cycle Time with Turnaround Constraints

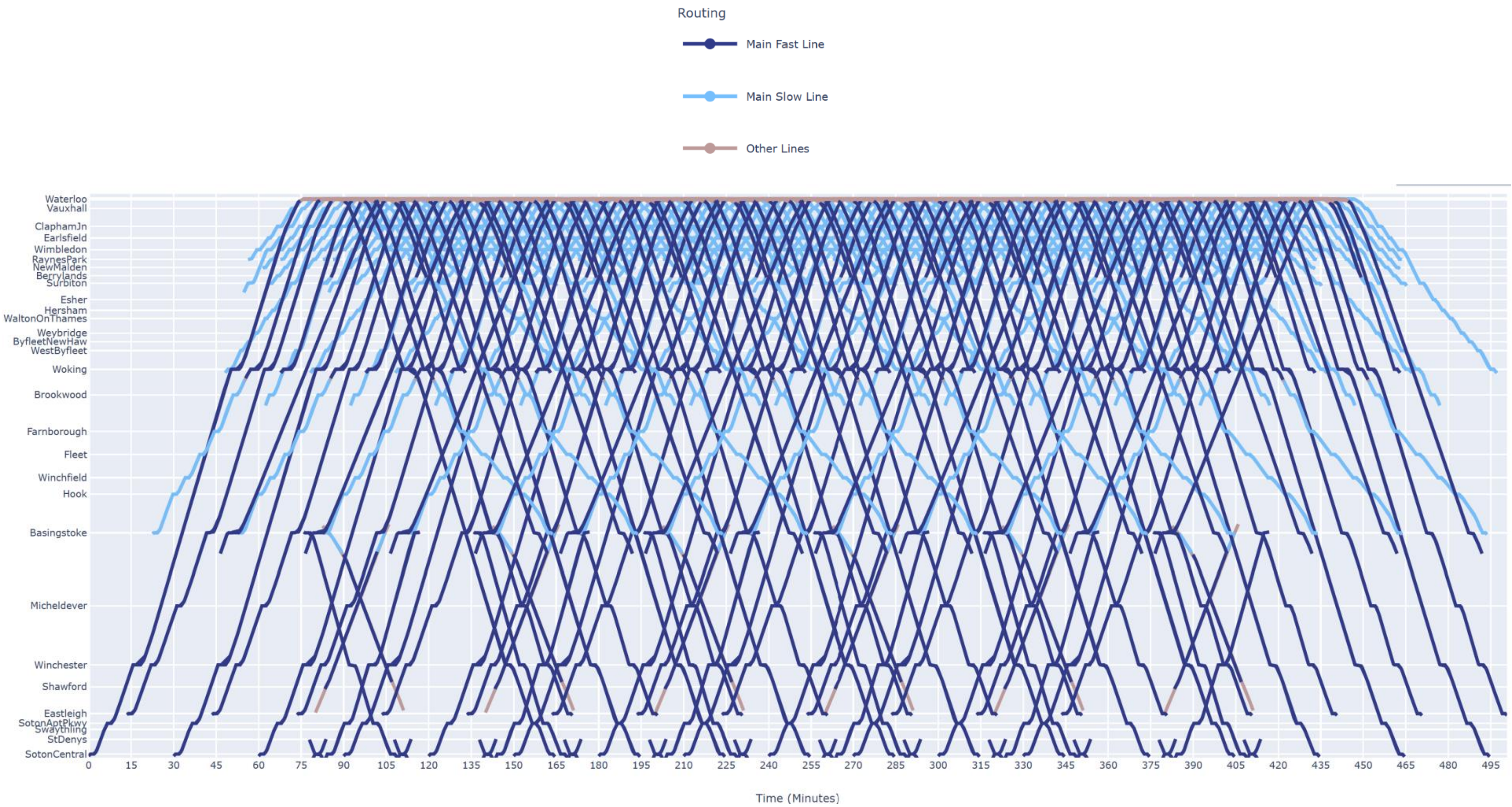


A.2.4 Minimized Cycle Time with No Turnaround Constraints

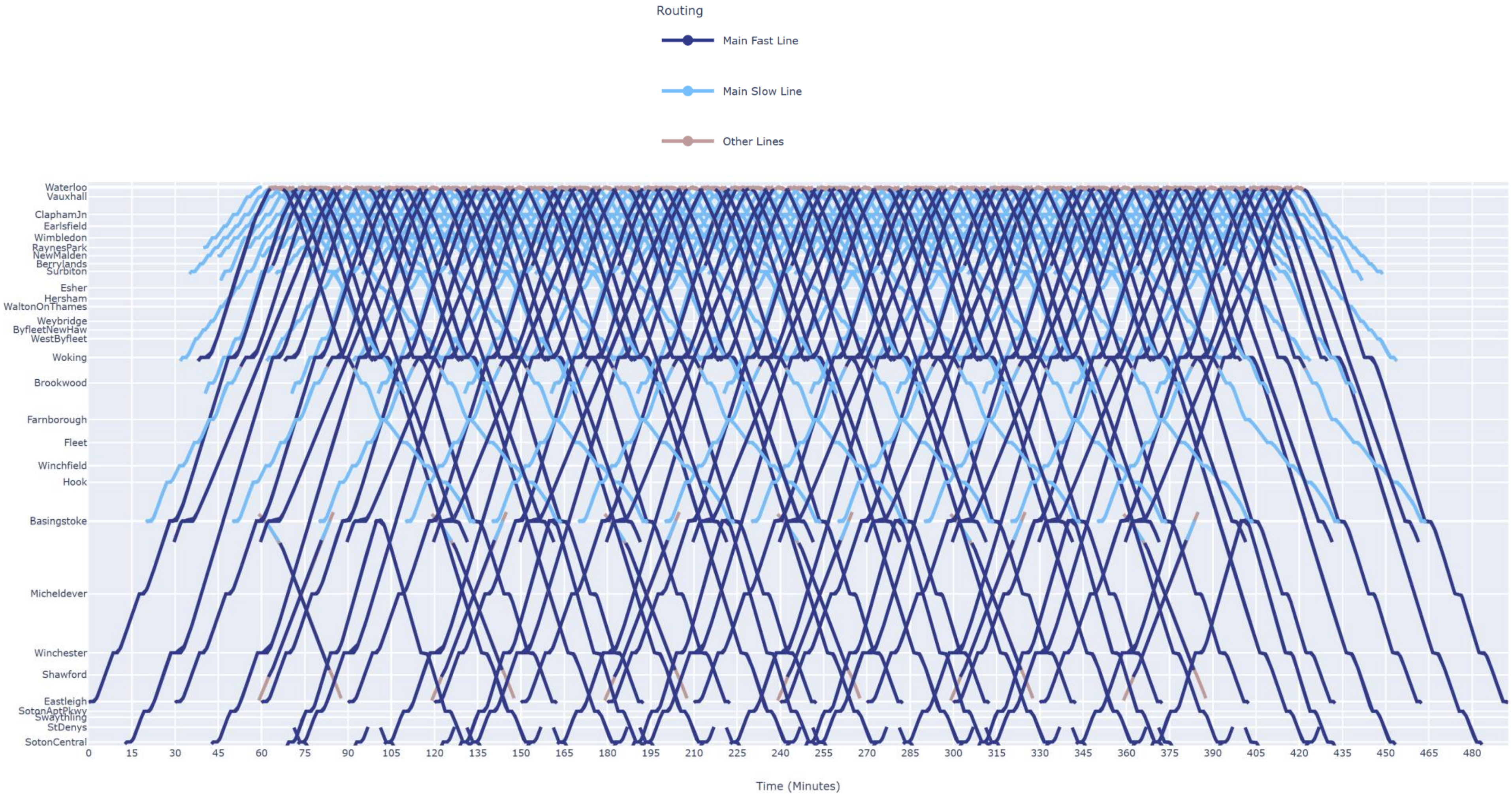


A.3 Results for Whole South West Main Line with ETCS Level 2 signalling (6 cycles shown)

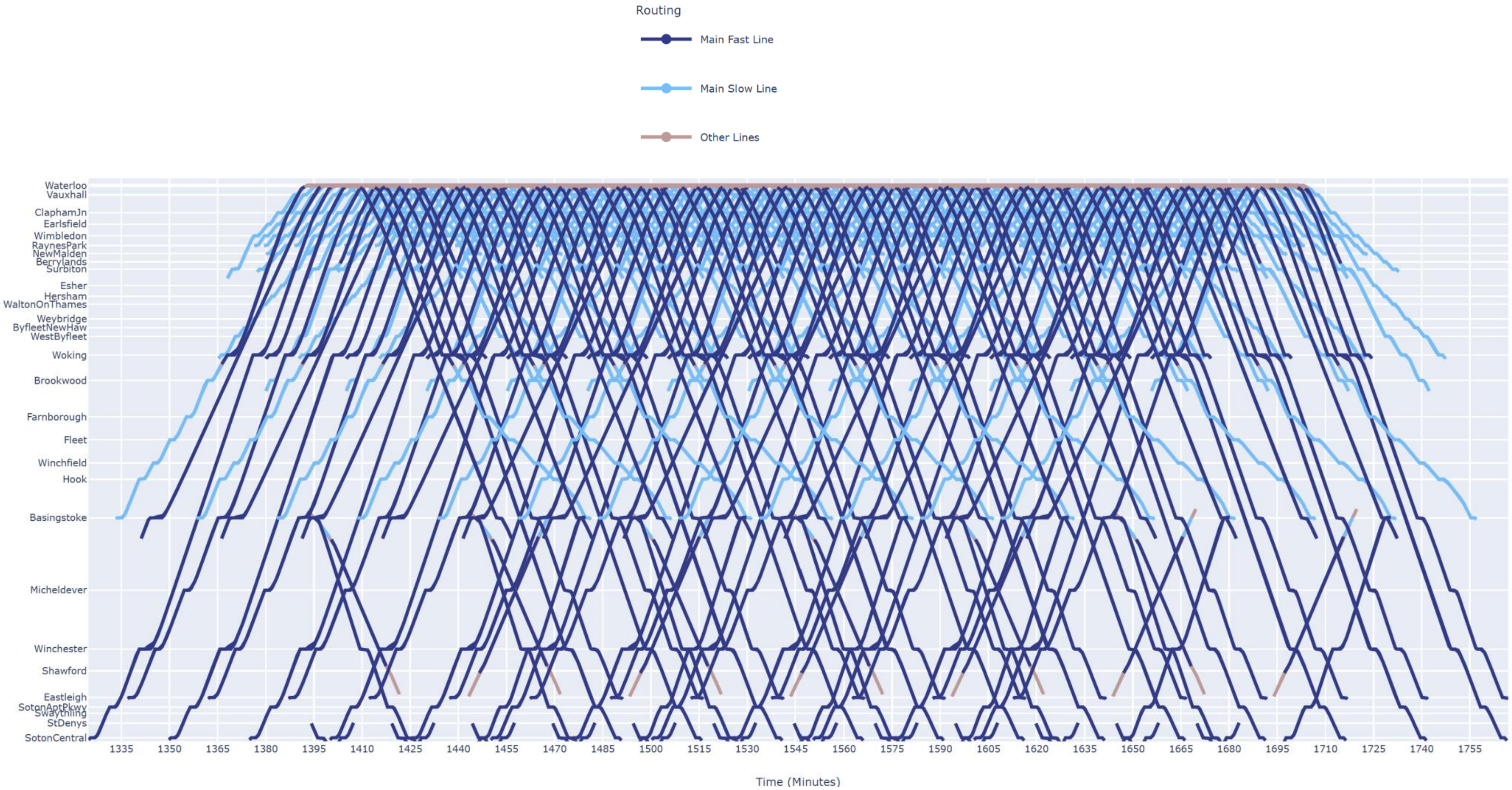
A.3.1 Maximized Buffer Time with Turnaround Constraints



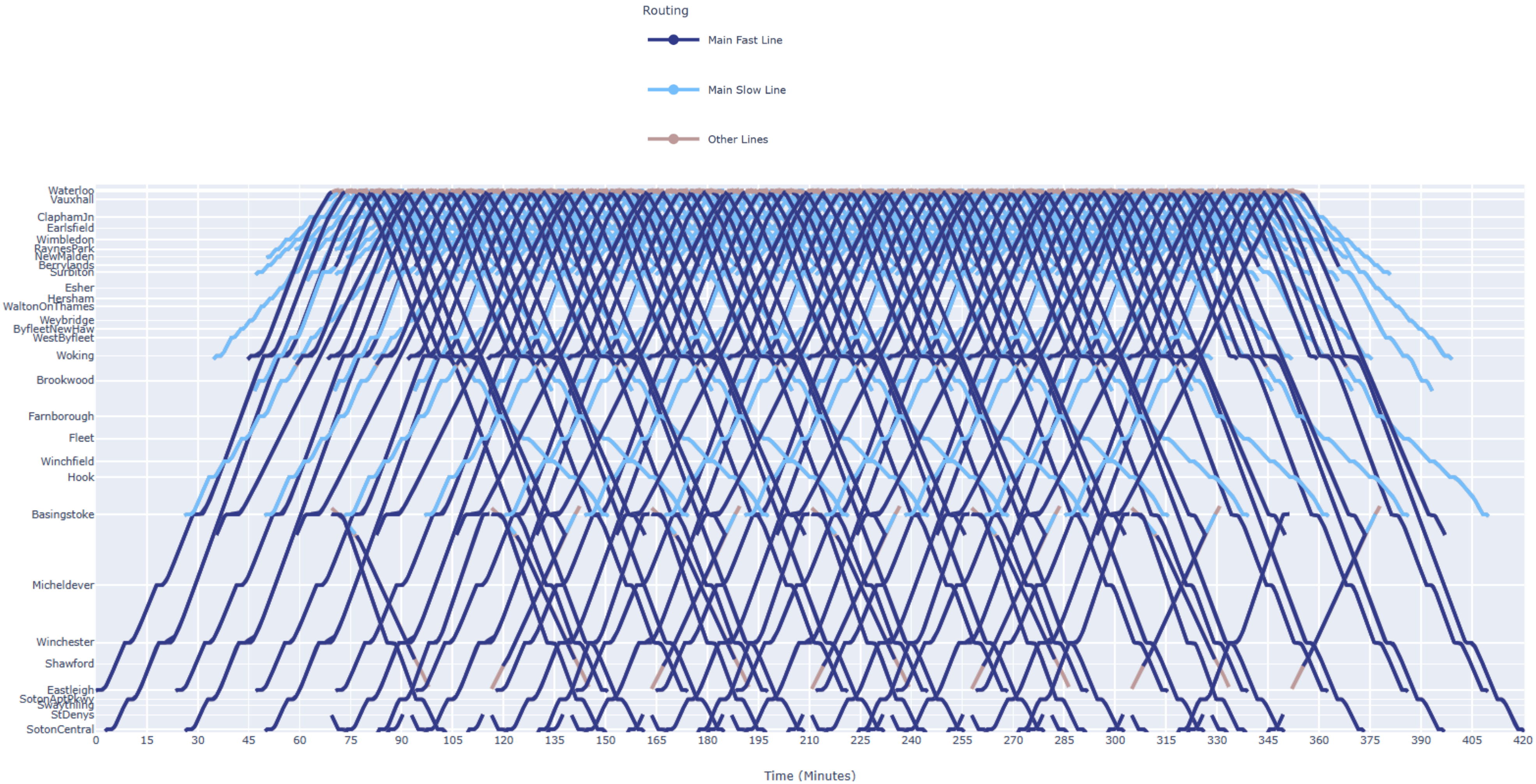
A.3.2 Maximized Buffer Time with No Turnaround Constraints



A.3.3 Minimized Cycle Time with Turnaround Constraints

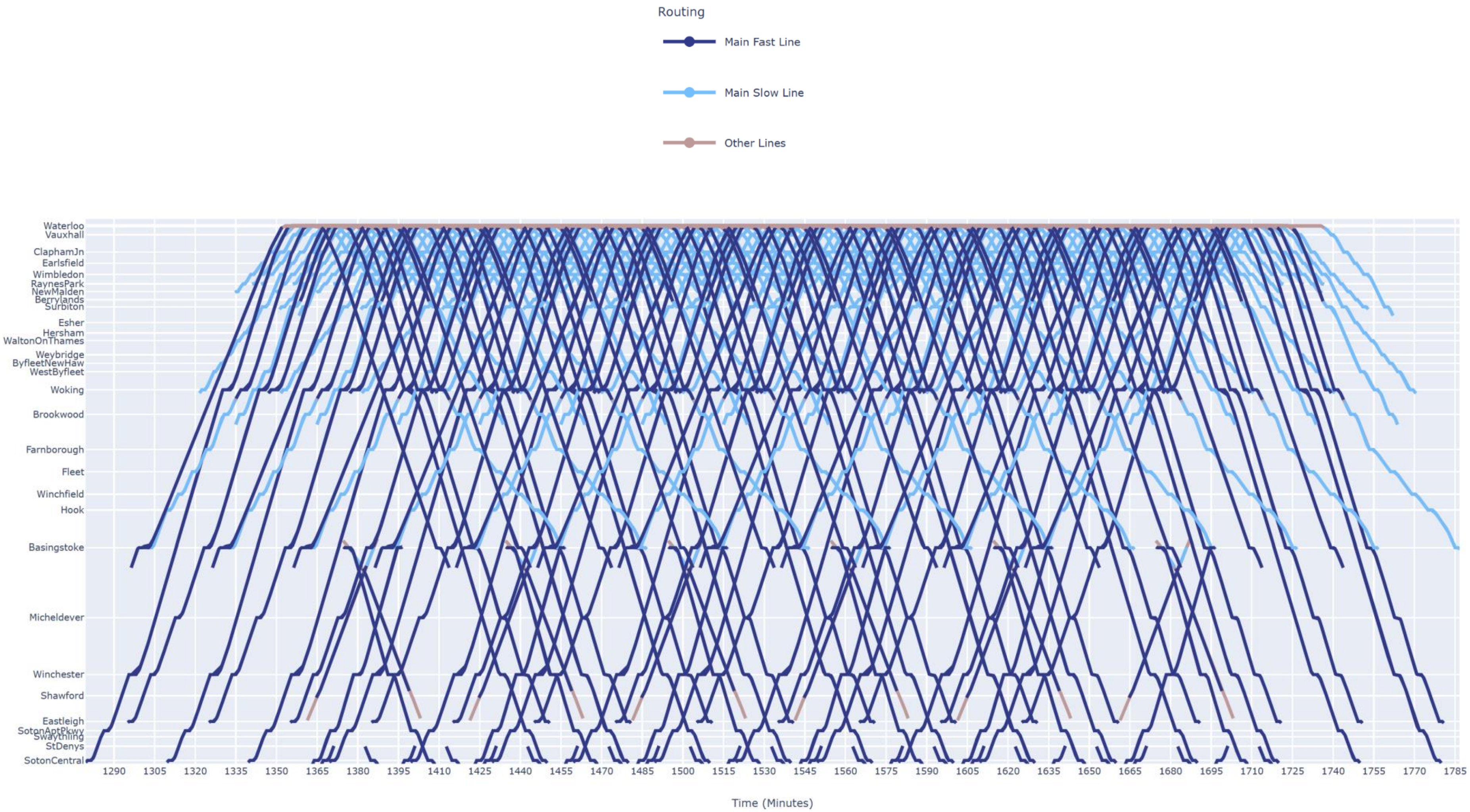


A.3.4 Minimized Cycle Time with No Turnaround Constraints

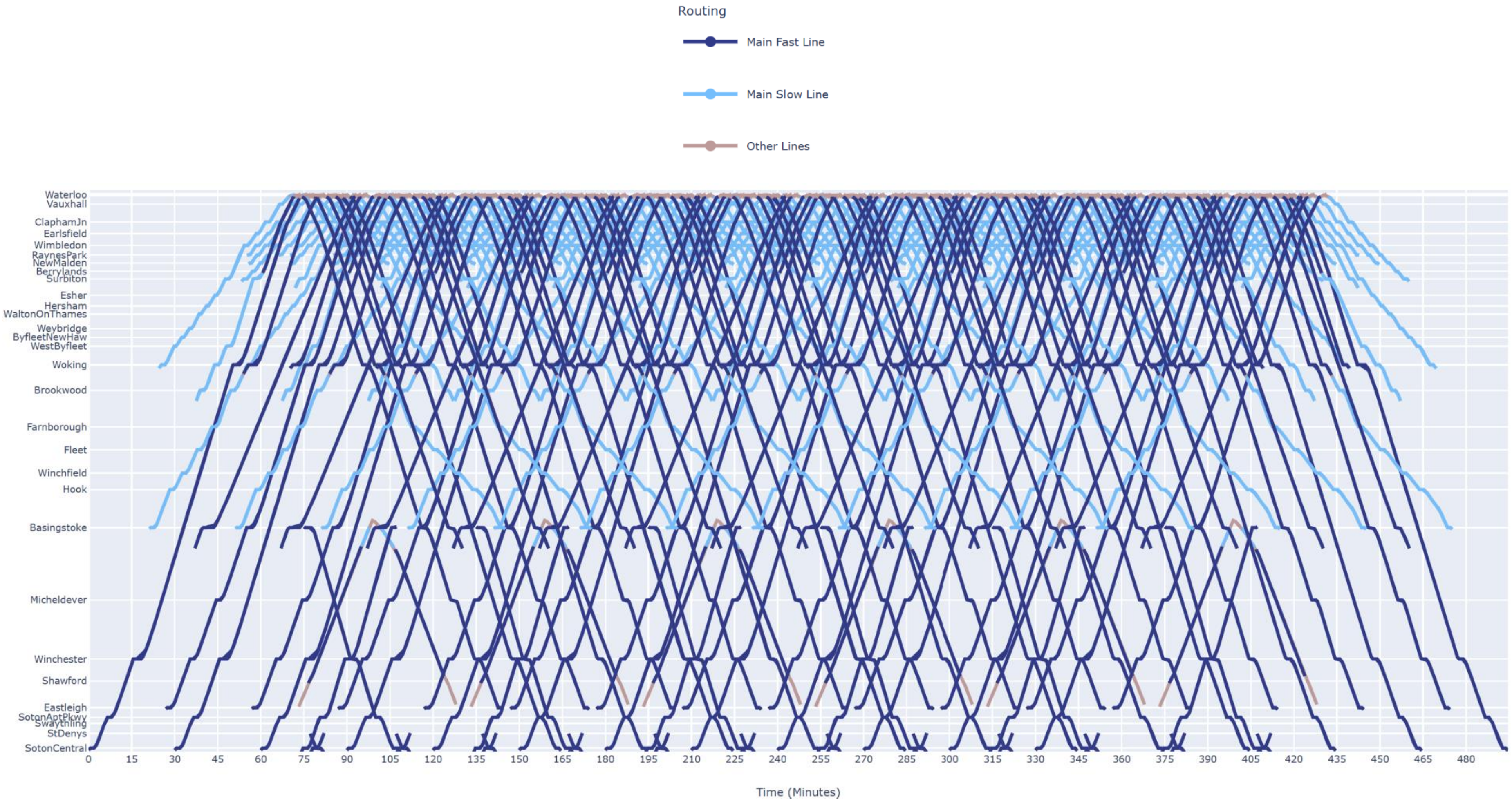


A.4 Results for Whole South West Main Line with Mixed signalling (6 cycles shown)

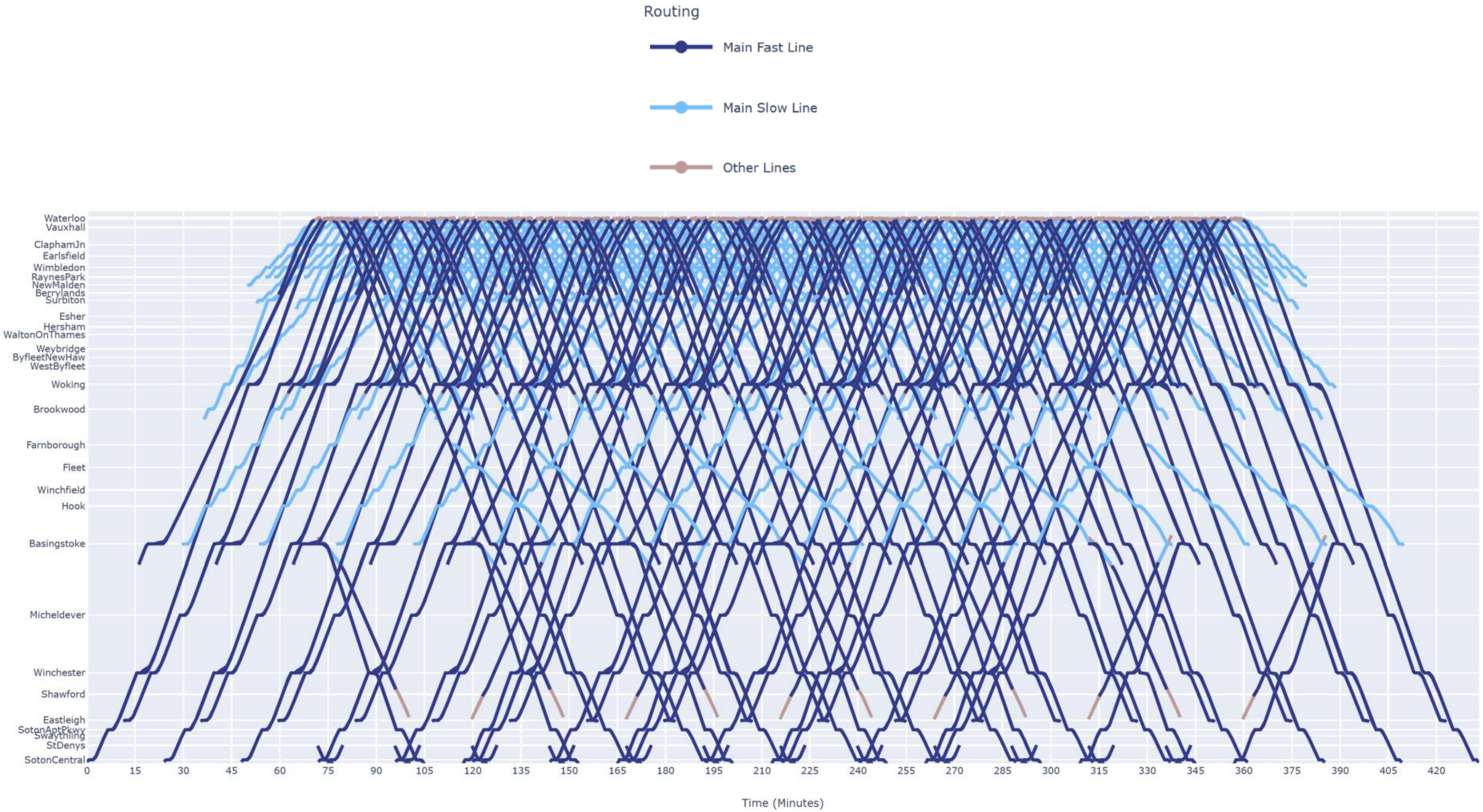
A.4.1 Maximized Buffer Time with Turnaround Constraints



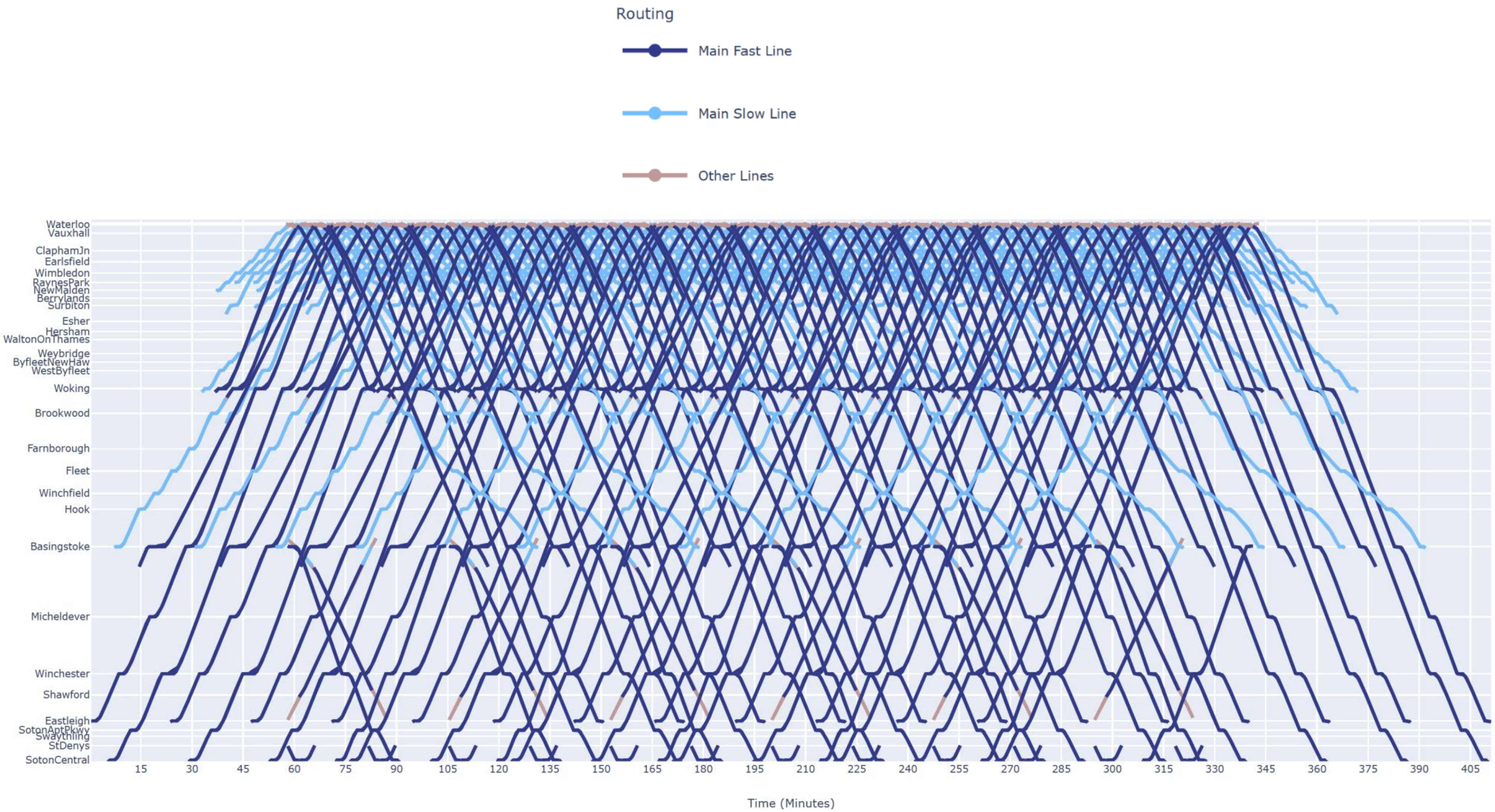
A.4.2 Maximized Buffer Time with No Turnaround Constraints



A.4.3 Minimized Cycle Time with Turnaround Constraints



A.4.4 Minimized Cycle Time with No Turnaround Constraints



References

Albrecht, T., Binder, A., Gassel, C. (2013). Applications of real-time speed control in rail-bound public transportation systems. *IET Intelligent Transport Systems*, 7(3), 305–314. https://doi.org/10.1049/iet-its.2011.0187

Alstom (2019). Alstom digital train control system enters service on Wuppertal Suspension Railway. https://www.alstom.com/press-releases-news/2019/9/alstom-digital-train-control-system-enters-service-wuppertal-suspension

Benders, J.F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik, 4(1), 238-252.* doi:10.1007/BF01386316

Bešinović, N., Goverde, R. M. P., Quaglietta, E., & Roberti, R. (2016). An integrated micro-macro approach to robust railway timetabling. *Transportation Research Part B: Methodological, 87*, 14-32. doi:10.1016/j.trb.2016.02.004

Bešinović, N., Quaglietta, E., & Goverde, R. M. P. (2019). Resolving instability in railway timetabling problems. *EURO Journal on Transportation and Logistics, 8*(5), 833-861. doi:10.1007/s13676-019-00148-3

Busuttil, W. (2023). *An alternative graph formulation for train service planning under ETCS Level 2, Moving Block and Virtual Coupling signalling systems with dynamic speed supervision and continuous infrastructure representation* *https://repository.tudelft.nl/islandora/object/uuid:1f8da4af-0483-4c93-8e11-3ec5aa246ad0?collection=education*

Caimi, G., Fuchsberger, M., Laumanns, M., & Schüpbach, K. (2011). Periodic railway timetabling with event flexibility. *Networks, 57(1), 3-18. doi:10.1002/net.20379*

Corman, F., D'Ariano, A., Pacciarelli, D., & Pranzo, M. (2010). A tabu search algorithm for rerouting trains during rail operations. *Transportation Research Part B: Methodological, 44*(1), 175-192. doi:10.1016/j.trb.2009.05.004

D'Ariano, A., Pacciarelli, D., & Pranzo, M. (2007). A branch and bound algorithm for scheduling trains in a railway network. *European Journal of Operational Research, 183*(2), 643-657. doi:10.1016/j.ejor.2006.10.034

D'Ariano, A., Corman, F., Pacciarelli, D., & Pranzo, M. (2008). Reordering and local rerouting strategies to manage train traffic in real time. *Transportation Science, 42*(4), 405-419. doi:10.1287/trsc.1080.0247

Furness, N., Van Houten, H., Arenas, L., Bartholomeus, M. (2017). ERTMS Level 3: the Game-Changer. *IRSE News*, 232, 2-9.

Grafe, V., & Schöbel, A. (2021). Solving the periodic scheduling problem: An assignment approach in non-periodic networks. Paper presented at the *OpenAccess Series in Informatics, 96* doi:10.4230/OASIcs.ATMOS.2021.9

Herrigel, S., Laumanns, M., Szabo, J., & Weidmann, U. (2018). Periodic railway timetabling with sequential decomposition in the PESP model. *Journal of Rail Transport Planning and Management, 8*(3-4), 167-183. doi:10.1016/j.jrtpm.2018.09.003

Hooker, J. N., & Ottosson, G. (2003). Logic-based benders decomposition. *Mathematical Programming, Series B, 96*(1), 33-60. doi:10.1007/s10107-003-0375-9

M. Janssens (2022). *Multi machine approaches for conflict resolution under moving block signalling* https://repository.tudelft.nl/islandora/object/uuid:ee472a92-c634-47b8-bcee-d828ee29b49e?collection=education

Lamorgese, L., Mannino, C., & Natvig, E. (2017). An exact micro–macro approach to cyclic and non-cyclic train timetabling. *Omega (United Kingdom), 72*, 59-70. doi:10.1016/j.omega.2016.11.004

Lamorgese, L., & Mannino, C. (2019). A noncompact formulation for job-shop scheduling problems in traffic management. *Operations Research, 67*(6), 1586-1609. doi:10.1287/opre.2018.1837

Leutwiler, F., & 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

Leutwiler, F. & Corman, F. (2023). Set covering heuristics in a Benders decomposition for railway timetabling. *Computers & Operations Research, 159, 106339.* doi:https://doi.org/10.1016/j.cor.2023.106339

F. Liu, J. Xun, H. Dong, Z. Zhang and Y. Li. (2021). "A Bi-level Real-time Rescheduling Approach for Train Operation in High-speed Railways," *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021, pp. 2417-2422, doi: 10.1109/ITSC48978.2021.9564832.

Luan, X., Wang, Y., De Schutter, B., Meng, L., Lodewijks, G., & Corman, F. (2018). Integration of real-time traffic management and train control for rail networks - part 1: Optimization problems and solution approaches. *Transportation Research Part B: Methodological, 115*, 41-71. doi:10.1016/j.trb.2018.06.006

Luan, X., De Schutter, B., Meng, L., & Corman, F. (2020). Decomposition and distributed optimization of real-time traffic management for large-scale railway networks. *Transportation Research Part B: Methodological*, *141*, 72–97. https://doi.org/10.1016/j.trb.2020.09.004

Lusby, R. M., Larsen, J., Ehrgott, M., & Ryan, D. M. (2013). A set packing inspired method for real-time junction train routing. *Computers and Operations Research, 40*(3), 713-724. doi:10.1016/j.cor.2011.12.004

Long, S. Meng, L. Wang, Y. Miao, J. Li, T & Corman, F. (2019). "A Train Rescheduling Optimization model with considering the Train Control for A High-Speed Railway Line under Temporary Speed Restriction," *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pp. 2809-2816, doi: 10.1109/ITSC.2019.8917173.

Mascis, A., & Pacciarelli, D. (2002). Job-shop scheduling with blocking and no-wait constraints. *European Journal of Operational Research, 143*(3), 498-517. doi:10.1016/S0377-2217(01)00338-1

Marlière, G., Sobieraj Richard, S., Pellegrini, P., & Rodriguez, J. (2023). A conditional time-intervals formulation of the real-time railway traffic management problem. *Control Engineering Practice, 133* doi:10.1016/j.conengprac.2022.105430

Network Rail (2020, June 12): *New West Midlands training facilities help develop future railway signallers* https://www.networkrailmediacentre.co.uk/news/new-west-midlands-training-facilities-helps-develop-railway-signallers-for-the-future

Network Rail (n.d.-a): *Target 190plus: Sustainable Signalling Renewals.* Retrieved September 26, 2023, from https://www.networkrail.co.uk/industry-and-commercial/research-development-and-technology/research-and-development-programme/target-190plus-sustainable-signalling-renewals/

Network Rail (n.d.-b) *Target 190plus Project Explainers - How we are making signalling sustainable in the GB Railway.* Retrieved September 26, 2023, from https://sacuksprodnrdigital0001.blob.core.windows.net/t190/Find%20out%20more%20about%20the%20Target%20190plus%20projects/Target%20190plus%20Project%20Explainers%20-%20How%20we%20are%20making%20signalling%20sustainable%20in%20the%20GB%20Railway.ppsx

Network Rail (n.d.-c) *The Timetable*. Retrieved October 1, 2023 from https://www.networkrail.co.uk/running-the-railway/the-timetable/

Odijk, M. A. (1996). A constraint generation algorithm for the construction of periodic railway timetables. *Transportation Research Part B: Methodological, 30*(6), 455-464. doi:10.1016/0191-2615(96)00005-7

ON-TIME, (2014). *Optimal Networks for Train Integration Management across Europe*. https://cordis.europa.eu/project/id/285243/reporting

Pellegrini, P., Marlière, G., & 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

Pellegrini, P., Marlière, G., Pesenti, R., & Rodriguez, J. (2015). RECIFE-MILP: An effective MILP-based heuristic for the real-time railway traffic management problem. *IEEE Transactions on Intelligent Transportation Systems, 16*(5), 2609-2619. doi:10.1109/TITS.2015.2414294

Pellegrini, P., Marlière, G., Pesenti, R., & Rodriguez, J. (2017). A MILP reformulation for train routing and scheduling in case of perturbation. Paper presented at the *Springer Proceedings in Mathematics and Statistics, , 217* 495-503. doi:10.1007/978-3-319-67308-0_50

Pellegrini, P., Pesenti, R., & Rodriguez, J. (2019). Efficient train re-routing and rescheduling: Valid inequalities and reformulation of RECIFE-MILP. *Transportation Research Part B: Methodological, 120*, 33-48. doi:10.1016/j.trb.2018.12.008

Quaglietta, E. (2014). A simulation-based approach for the optimal design of signalling block layout in railway networks. *Simulation Modelling Practice and Theory, 46,* 4-24. doi:10.1016/j.simpat.2013.11.006

Quaglietta, E., Pellegrini, P., Goverde, R.M.P., Albrecht, T., Jaekel, B., Marlière, G. , Rodriguez, J., Dollevoet, T. , Ambrogio, B., Carcasole, D. Giaroli, M. Nicholson, G. (2016). The ON-TIME real-time railway traffic management framework: A proof-of-concept using a scalable standardised data communication architecture. *Transportation Research Part C: Emerging Technologies, 63*, 23-50. doi:10.1016/j.trc.2015.11.014

Quaglietta, E., Spartalis, P., Wang, M., Goverde, R. M. P., & van Koningsbruggen, P. (2022). Modelling and analysis of virtual coupling with dynamic safety margin considering risk factors in railway operations. *Journal of Rail Transport Planning and Management, 22* doi:10.1016/j.jrtpm.2022.100313

Railway Safety and Standards Board. (2018, December). *ERTMS level 2 train signalling regulations.* Retrieved from https://www.rssb.co.uk/standards-catalogue/CatalogueItem/GERT8000-TS10-Iss-4

Railway Safety and Standards Board. (2022, December). *ERTMS Handbook*. Retrieved from https://www.rssb.co.uk/standards-catalogue/CatalogueItem/rs525-iss-1

Railway Safety and Standards Board. (2023a, December). *General signalling regulations.* Retrieved from https://www.rssb.co.uk/standards-catalogue/CatalogueItem/ris-0703-ccs-iss-2\

Railway Safety and Standards Board. (2023b, September). *Signalling Layout and Signal Aspect Sequence Requirements.* Retrieved from https://www.rssb.co.uk/standards-catalogue/CatalogueItem/ris-0703-ccs-iss-2

Railway Safety and Standards Board. (2023c, December). *Station Duties and Train Dispatch*. Retrieved from https://www.rssb.co.uk/standards-catalogue/CatalogueItem/gert8000-ss1-iss-9

Sartor, G., Mannino, C., Nygreen, T., & Bach, L. (2023). A MILP model for quasi-periodic strategic train timetabling. *Omega (United Kingdom), 116* doi:10.1016/j.omega.2022.102798

Serafini, P., Ukovich, W. (1989). A mathematical model for periodic scheduling problems. *SIAM Journal on Discrete Mathematics 2.* https://doi.org/10.1137/0402049

Schlechte, T., Borndörfer, R., Denißen, J., Heller, S., Klug, T., Küpper, M., Lindner, N., Reuther, M., Söhlke, A., & Steadman, W. (2022). Timetable optimization for a moving block system. *Journal of Rail Transport Planning & Management*, *22*, 100315. https://doi.org/10.1016/j.jrtpm.2022.100315

South Western Railway Network Map. South Western Railway. (2022). Retrieved from https://www.southwesternrailway.com/plan-my-journey/our-network

Sparing, D., & Goverde, R. M. P. (2017). A cycle time optimization model for generating stable periodic railway timetables. *Transportation Research Part B: Methodological, 98*, 198-223. doi:10.1016/j.trb.2016.12.020.

Thales (2020, September 3). New contract for Thales to provide CBTC Signalling to Vancouver's Broadway subway project. https://www.thalesgroup.com/en/group/journalist/press-release/new-contract-thales-provide-cbtc-signalling-vancouvers-broadway

UIC, C. (2013). 406: capacity. *International Union of Railways, Paris*.

Versluis, N. D., Pellegrini, P., Quaglietta, E., Goverde, R. M. P., & Rodriguez, J. (2023). An Approximate Conflict Detection and Resolution Model for Moving-Block Signalling by Enhancing RECIFE-MILP. *RailBelgrade 2023: 10th International Conference on Railway Operations Modelling and Analysis.*

Wang, P., Goverde, R.M.P. (2016). Multiple-phase train trajectory optimization with signalling and operational constraints. *Transportation Research Part C: Emerging Technologies*, 69, 255–275. https://doi.org/10.1016/j.trc.2016.06.008

Wang, P., Goverde, R.M.P. (2017). Multi-train trajectory optimization for energy efficiency and delay recovery on single-track railway lines. *Transportation Research Part B: Methodological*, 105, 340–361 https://doi.org/10.1016/j.trb.2017.09.012

Wang, Z., Quaglietta, E., Bartholomeus, M. G. P., & Goverde, R. M. P. (2022). Assessment of architectures for Automatic Train Operation driving functions. *Journal of Rail Transport Planning & Management*, *24*, 100352. https://doi.org/10.1016/J.JRTPM.2022.100352

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

Xu, P., Corman, F., Peng, Q., & Luan, X. (2017). A train rescheduling model integrating speed management during disruptions of high-speed traffic under a quasi-moving block system. *Transportation Research. Part B: Methodological*, *104*, 638-666. https://doi.org/10.1016/j.trb.2017.05.008

Xu, P., Zhang, D., Guo, J., Liu, D., & Peng, H. (2021). Integrated Train Rescheduling and rerouting during Multidisturbances under a quasi-moving block system. *Journal of Advanced Transportation*, *2021*, 1–15. https://doi.org/10.1155/2021/6652531

Yin, J., Tang, T., Yang, L., Xun, J., Huang, Y., & Gao, Z. (2017). Research and development of Automatic Train Operation For Railway Transportation Systems: A survey. *Transportation Research Part C: Emerging Technologies*, *85*, 548–572. https://doi.org/10.1016/j.trc.2017.09.009

Yin, J., Tang, T., Yang, L., Xun, J., Huang, Y., & Gao, Z. (2017). Research and development of Automatic Train Operation For Railway Transportation Systems: A survey. Transportation Research Part C: Emerging Technologies, 85, 548–572. https://doi.org/10.1016/j.trc.2017.09.009