Evaluating the Robustness of Crew Schedules for
Rail Transit Systems

by
Harshavardhan Ravichandran
B.Tech., M.Tech., Civil Engineering
Indian Institute of Technology Madras (2011)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation
at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2013

© Massachusetts Institute of Technology 2013. All rights reserved.
Evaluating the Robustness of Crew Schedules for Rail Transit Systems

by

Harshavardhan Ravichandran

Submitted to the Department of Civil and Environmental Engineering on May 23, 2013, in partial fulfillment of the requirements for the degree of Master of Science in Transportation

Abstract

Crew scheduling has traditionally been the last step in the process of service planning, and it has traditionally aimed at minimizing manpower costs because it was assumed that the crew schedule does not directly impact service quality. There has been a growing recognition that crew schedules do in fact affect performance, especially during disruptions, and in that context different crew schedules perform differently.

The primary objective of this research is to evaluate the robustness of a crew schedule, or in other words, the performance of a crew schedule under a range of commonly observed disrupted conditions. While the thesis focuses on the Piccadilly Line on the London Underground, the concepts and methods developed are intended to be applicable to a range of metro lines and systems.

The thesis has four components: first, a description of the types of incidents that take place on a line, the type of service control interventions used to respond to incidents, and the intimate relationship between the crew schedule and service control policies; second, a comparative analysis of the structure of two Piccadilly Line crew schedules to demonstrate how two crew schedules with similar underlying timetables can have very different structures, and how those structural differences can affect performance; third, the development of a simulation-based framework for evaluating the robustness of crew schedules, and a simulation model of the Piccadilly Line; and finally, the application of the simulation model to evaluate and compare the performance of the same two Piccadilly Line crew schedules under a range of disrupted conditions.

Based on observation and understanding the relationship between the crew schedule and service control, a service control module has been implemented in the simulation. This module mimics the actions of a service controller when the operations plan is disrupted and is a key contribution of this research. This allows for the simulation of incidents on the line which leads to an understanding of how the structure of a crew schedule affects its performance during disruptions. Elements of a crew schedule such as slack time and relief locations are identified as key drivers of robustness. The effect of these elements on performance is demonstrated by comparing the simulated performance of two crew schedules that differ in the distribution of these elements.
The simulation also allows the testing of hypothetical crew schedules, such as those corresponding to different labour agreements. Therefore, it can be used to test the performance impact of changes in labour agreements. This is demonstrated for the case of the Piccadilly Line through the simulation of two hypothetical scenarios where two different labour constraints are relaxed.

The thesis concludes with recommendations to Piccadilly Line and London Underground management and staff regarding crew scheduling, service control and data collection. Results suggest that the slack time in Piccadilly Line crew schedules could be redistributed in a way that improves performance by relaxing other constraints, and that service controllers currently make control decisions without easy access to critical crew information and control decisions could be improved by providing them with better information on crew activities.

Thesis Supervisor: Nigel H.M. Wilson
Title: Professor of Civil and Environmental Engineering

Thesis Supervisor: Harilaos N. Koutsopoulos
Title: Visiting Professor of Civil and Environmental Engineering
To my parents,
Srilatha Ravichandran & N. Ravichandran,
for their unconditional love and support.
Acknowledgments

I wish to wholeheartedly thank Professors Nigel Wilson and Haris Koutsopoulos for their time and patience. They have been a constant source of support, and without their knowledge, advice and dedication, this thesis would not have been possible. I cannot thank them enough for all that they have done. Others at MIT, particularly John Attanucci and Ginny Siggia, also played an important role in guiding my research and helping this document come to fruition.

I am indebted to the staff at Transport for London for their generous help in supporting this research. I would like to thank, among others, Chris Taggart and Phil O’Hare from the Piccadilly Line; Charles Horsey, Stuart Wilson, Mark Godfrey and Paul Western from Scheduling Services; and Caroline Harper, Lauren Sager-Weinstein, Madhuri Shah and Duncan Horne from TfL for all their help and feedback throughout the process.

It is my friends and roommates that made life in Boston so much fun. I am eternally thankful to Naveen Kartik, Rama Krishna Simhadri, Ajit Kamath and Vivek Sivathanu for their companionship; and specially to Varun Pattabhiraman for sharing so much of his life with me. I have to thank Aditi Balachandar and Vaishnavi Surendra for making my stays in London so much fun. Finally, I would like to thank all my friends at the Transit Lab for getting me through these two years.
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-1</td>
<td>Transit service delivery as a business process [2].</td>
<td>17</td>
</tr>
<tr>
<td>2-1</td>
<td>London’s public transport network. [17]</td>
<td>24</td>
</tr>
<tr>
<td>2-2</td>
<td>Change in number of trips since 1993, by mode. [18]</td>
<td>25</td>
</tr>
<tr>
<td>2-3</td>
<td>The geometry and service pattern of the Piccadilly Line</td>
<td>26</td>
</tr>
<tr>
<td>2-4</td>
<td>Peak thirty minute (AM) flow on the Piccadilly Line - Eastbound [15]</td>
<td>28</td>
</tr>
<tr>
<td>2-5</td>
<td>Peak thirty minute (AM) flow on the Piccadilly Line - Westbound [15]</td>
<td>29</td>
</tr>
<tr>
<td>2-6</td>
<td>Example of a renumbering operation</td>
<td>44</td>
</tr>
<tr>
<td>3-1</td>
<td>A sample duty from the Acton Town weekday crew schedule</td>
<td>54</td>
</tr>
<tr>
<td>3-2</td>
<td>An example of a single-spell duty</td>
<td>55</td>
</tr>
<tr>
<td>3-3</td>
<td>An example of a typical two-spell, two run duty</td>
<td>56</td>
</tr>
<tr>
<td>3-4</td>
<td>An example of a night duty</td>
<td>56</td>
</tr>
<tr>
<td>3-5</td>
<td>Some of the cover duties scheduled at Acton Town</td>
<td>56</td>
</tr>
<tr>
<td>3-6</td>
<td>A sample TDA</td>
<td>58</td>
</tr>
<tr>
<td>3-7</td>
<td>An excerpt from the Acton Town roster</td>
<td>59</td>
</tr>
<tr>
<td>3-8</td>
<td>Distribution of driving time in the two crew schedules</td>
<td>61</td>
</tr>
<tr>
<td>3-9</td>
<td>Meal break slack vs Lateness at ACT-EB</td>
<td>63</td>
</tr>
<tr>
<td>3-10</td>
<td>Meal break slack vs Lateness at ACT-WB</td>
<td>63</td>
</tr>
<tr>
<td>3-11</td>
<td>Meal break slack vs Lateness at AGR-EB</td>
<td>64</td>
</tr>
<tr>
<td>3-12</td>
<td>Meal break slack vs Lateness at AGR-WB</td>
<td>64</td>
</tr>
<tr>
<td>3-13</td>
<td>EoD Late running margin vs. Time to hit a parameter</td>
<td>67</td>
</tr>
<tr>
<td>3-14</td>
<td>Cumulative distribution of duty lengths</td>
<td>68</td>
</tr>
<tr>
<td>4-1</td>
<td>Framework for simulation-based evaluation of robustness</td>
<td>75</td>
</tr>
<tr>
<td>4-2</td>
<td>Simulation Methodology</td>
<td>78</td>
</tr>
<tr>
<td>4-3</td>
<td>Processing a Projected Arrival event</td>
<td>80</td>
</tr>
<tr>
<td>4-4</td>
<td>Processing an Actual Arrival event</td>
<td>82</td>
</tr>
<tr>
<td>4-5</td>
<td>Processing a Projected Departure event</td>
<td>84</td>
</tr>
<tr>
<td>4-6</td>
<td>Processing an Actual Departure event</td>
<td>85</td>
</tr>
<tr>
<td>4-7</td>
<td>Scatter plot of running times of successive trains (ACT-AGR)</td>
<td>89</td>
</tr>
<tr>
<td>4-8</td>
<td>Scatter plot of running time deviations of successive trains (ACT-AGR)</td>
<td>89</td>
</tr>
<tr>
<td>4-9</td>
<td>Waterfall diagram showing train movement around an incident</td>
<td>92</td>
</tr>
<tr>
<td>4-10</td>
<td>Reduced route structure of the Piccadilly Line used in the simulation</td>
<td>95</td>
</tr>
<tr>
<td>4-11</td>
<td>Histogram of running times from Acton Town to Arnos Grove</td>
<td>97</td>
</tr>
<tr>
<td>4-12</td>
<td>Actual and scheduled running time (ACT-AGR) by time of day</td>
<td>98</td>
</tr>
</tbody>
</table>
4-13 Scatter plot of dwell times at Acton Town eastbound................. 100
4-14 Histogram of dwell times at Acton Town eastbound.................. 101
4-15 Service control strategy................................................. 107

5-1 Running times of successive trains between ACT and AGR (Schedule 48, undisrupted)................................................................. 112
5-2 Average lateness of trains arriving at crew change locations (Schedule 48, undisrupted)................................................................. 114
5-3 Average headway between trains departing AGR-EB (Schedule 48, undisrupted)............................................................. 115
5-4 Average lateness with and without service control for Scenario 1 .... 117
5-5 Average lateness for Scenario 2................................................ 120
5-6 Slack versus Lateness at ACT-EB................................................ 124
5-7 Slack versus Lateness at ACT-WB................................................ 125
5-8 Spatial distribution of Missed pickups........................................ 126
5-9 Temporal distribution of Missed pickups...................................... 126
List of Tables

3.1 Overall statistics for the two crew schedules ........................................... 58
3.2 Comparison of Relief Directions ................................................................. 60
3.3 Comparison of Meal break Slack time ......................................................... 62
3.4 Comparison of intermediate reliefs ............................................................... 65
3.5 Comparison of EoD reliefs ......................................................................... 66

5.1 Performance measures for Scenario 1 ........................................................... 116
5.2 Distribution of incident duration by period (in minutes) ................................. 121
5.3 Spatial distribution of incidents on the Piccadilly Line .................................. 122
5.4 Performance measures for schedules 48 and 49 ........................................ 123
5.5 Meal break slack across missed pickups ..................................................... 127
5.6 EoD slack across early reliefs ................................................................. 127
5.7 Performance measures for schedule 49 with meal break length constraint relaxed ........................................................................................................ 128
5.8 Performance measures for schedule 49 with reduced walking time at Arnos Grove ........................................................................................................... 129
Chapter 1

Introduction

This thesis presents research on the evaluation of the robustness of the crew schedule for rail transit service. It highlights the importance of the structure of the crew schedule in improving robustness and making operations more resilient to disruptions. Two different approaches to measuring this robustness — structural and simulation — are discussed. A simulation based model is developed and validated for the Piccadilly Line, and is used to perform a comparative evaluation of the robustness of two different crew schedules.

This research discusses ways in which a crew schedule can be made more robust. The simulation model that is developed here can be used by schedulers and management to predict the impact of changes in the crew schedule on robustness, caused either by self-directed changes, or changes in labour agreements or service patterns. This will also enable the identification of best practices for robust crew scheduling.

The main contribution of this research is the development of the simulation model to evaluate robustness. The model simulates the line’s operations under a spectrum of real operating conditions with varying degrees of disruption. Simulating operations during and after a disruption requires that real-time service control also be modelled, and this has been done for the Piccadilly Line. While this kind of approach has been taken in airline systems, it is novel in rail systems.

1.1 Background and motivation

The passenger experience on a high frequency metro is principally determined by short and regular headways, which is not necessarily the same as running to timetable. However, the timetable is still important for operations because late train arrivals can mean that train operators are late for their meal breaks which can lead to missed
connections, or are late in finishing their duties which can lead to overtime costs. Secondly, the timetable is an operations plan that is understood by all the parties involved (including service controllers, crew managers, station managers etc.) and deviations from the plan, such as result from real-time control interventions, need to be effectively communicated to all parties. Doing this in real-time requires advanced information systems that continuously and accurately report the state of the system, including real-time changes being made to the system in response to disruptions, and new actions that are needed as a consequence of those changes. Large amounts of capital are required to upgrade old communication systems and to provide such advanced capabilities. In the absence of such systems, disruptions cannot be handled entirely in real-time by service control – there is a need to plan for them.

Traditionally, delays are absorbed in the timetable by a combination of recovery time at the terminus at the end of a trip and padded running times. These strategies do not work for a high-frequency metro service because providing recovery times requires extra platform capacity at the termini, and padded running times can result in blocking back of trains. Late running trains are often renumbered and “put back on time” at the time of a crew change, or they may be short-tripped or cancelled, or some combination of the above, which can have adverse effects on the customer experience. The primary motivation for doing so is to bring drivers back to their crew bases in time for their meal break, or next train, or the end of their duty. This may be better achieved through the combination of a robust crew schedule and intelligent service control.

The robustness of a schedule is often characterized by two mutually dependent properties: stability, or the ability to absorb small disruptions and still remain feasible, and flexibility, or the ability to recover quickly from large disruptions. A crew schedule is made robust by the presence of slack or buffer times, providing spare operators, placing crew changes in locations and at times that maximize flexibility, and other such methods.

Clearly, the price of robustness is increased crew costs, and to make a sensible trade-off it is necessary to evaluate the utility of that robustness. Robustness can also be improved by deploying manpower more effectively: deploying slack time and spare operators where they are needed, providing the correct mix of relief locations, directions and times etc. Therefore, it is necessary to identify best practices for robust crew scheduling. One way of doing this is to perform a comparative evaluation of different crew schedules that systematically vary the above characteristics.

It is not possible to evaluate the robustness of a crew schedule without an understanding of service control because it is the service controllers who make use of the flexibility offered by a crew schedule to effect real time interventions. Therefore, it is necessary to evaluate the robustness of a crew schedule in the context of service
control strategies.

1.2 The service delivery process

One can imagine a transit service as a business process, as shown in Figure 1-1. The overall service policies such as span of service, frequency and routing are determined at the management levels of the transit agency. Service policies are usually based on expected or actual demand, network connectivity considerations, financial constraints and political considerations. These policies are then used by the planning department, which is responsible for developing an operations plan.

The operations plan is the detailed plan which describes the utilization of the agency's resources – rolling stock, personnel and infrastructure – in order to meet the service policies. The most important components of the operations plan are the timetable and the crew schedule. The timetable describes all train movements reflecting where and when transportation service should be provided to customers. And the crew schedule assigns drivers to trips, ensuring that all trips in the timetable are manned, while respecting labour constraints.

The last piece under the transit agency's control is at the operational level, where the operations plan is implemented. It includes all front-line staff (train operators,
station managers) as well as service controllers, crew managers, vehicle and infrastructure maintenance engineers and operational support personnel. Service control (also, operations control or service management), which is an essential component at the operational level, oversees and coordinates the implementation of the operations plan and modifies it to deal with unforeseen events and short-term infeasibilities: It is the central, real-time control function of schedule-based operations. Its centralized nature, in which controllers are provided with information on the state of the entire system, sets it apart from local dispatching techniques, where a supervisor is positioned at a terminal or a station along the line.

The service delivery process results in the daily operations which are provided to passengers. It is important to note that in daily operations, passengers do not experience how the service was planned to be operated, they experience the actual operations. This difference between planned and actual operations comes from the natural variability in operations, and from externalities (i.e. incidents) which cause disruptions.

A cost-optimized operations plan that aims to provide the most service at the least cost might work very well under ideal operating conditions, but may stutter under minor disruptions and fail completely under major disruptions. On the other hand, a more robust operations plan may provide lower service levels and/or use more manpower. But it is better able to continue functioning when minor disruptions strike, and provide reasonable service with a quick recovery during a major disruption. Therefore, it is possible that a robust schedule can provide better performance (i.e. passenger service) on average, when considered over the full spectrum of operating conditions.

### 1.3 Why study robust crew scheduling

Crew scheduling is typically the last stage of the planning process at most transit agencies. It is performed after timetabling, and is usually done in a way that minimizes manpower cost while satisfying all the labour constraints. The problem with this approach is that a minimum cost crew schedule may not be the best crew schedule when things don't go as planned. As Carrel[2] and Rahbee[14] observe from their study of the service control process on the London Underground, many service control decisions during recovery after an incident are motivated by the need to get drivers back to their crew bases on time, to the detriment of the passenger experience. A robust crew schedule can reduce the need for such actions, and thereby improve average performance and/or reduce average operations costs.

Secondly, whenever changes are made to labour agreements or other crew-related matters, it is often assumed that the only impact will be lower or higher manpower costs. But it is often the case that such changes also have an impact on the performance
of the line. For instance, "fixed-link" was an agreement that was introduced at the Arnos Grove depot of the Piccadilly Line in 2009 which imposed certain constraints on the crew rostering process. In the process of satisfying these new constraints, the mix of crew relief locations and directions changed in the subsequent crew schedule. This unintended consequence had a negative effect on performance because the new crew schedule was found to be less flexible during disruptions. Had the management known this before the negotiation process started, they might have negotiated differently. Labour agreements are reviewed periodically, and it is in the interest of the management to know beforehand how any proposed changes might impact the line’s performance.

1.4 Research objectives

This research has three objectives:

1. To analyse the structure of the crew schedule, identify the elements that make the schedule robust and compare different schedules on the basis of these elements. Such an analysis would deepen our understanding of what makes a schedule robust, and give schedulers, service controllers and the line management a way of predicting schedule performance.

2. To develop a simulation-based framework to evaluate the performance of a crew schedule, and develop and validate a working model for the Piccadilly Line. The model would simulate the performance of the schedule over a range of operating conditions, and provide performance measures which are a measure of schedule robustness.

3. To use the simulation model to perform a comparative evaluation of the performance of two different Piccadilly Line crew schedules, and combine this with the insights obtained from the structural comparison of the two crew schedules, so as to provide insights into best practices for robust crew scheduling.

1.5 Research approach

This research aims to be of value to anyone interested in crew scheduling for a rail transit line, specifically a high-frequency metro. While the research is motivated by the desire of the London Underground to improve operational reliability on the Piccadilly Line without expensive capital upgrades, the research is just as applicable to any other rail transit system. The simulation model is first developed in a general manner, and then the implementation of the model for the Piccadilly Line is presented. Some of the details of the model, such as service control strategies, are necessarily specific to the geometry, labour agreements and operating conditions of
the line and cannot be generalized. However, a deliberate effort is made to keep the discussion of the model as general as possible.

The objectives outlined above were achieved through the analysis of data available from the London Underground, and by extended interactions with planners, duty schedulers, service controllers, train operators, crew managers and the management of the Piccadilly Line. The first objective involved a thorough analysis of the crew schedules of the Piccadilly Line. This analysis led to insights on what made the schedule robust, and also to the realization that a structural analysis would not be sufficient to measure robustness, and that a simulation based model was needed.

Based on detailed observations during interactions with the Piccadilly Line staff and analysis of train operations data from the London Underground, a simulation based framework for the evaluation of the crew schedule’s performance was formulated, and a specific instance of the model was developed and validated for the Piccadilly Line. This model is then used to perform a comparative evaluation of two crew schedules.

1.6 Organization of the thesis

Chapter 2 gives a detailed introduction to the operations plan, including how it is formulated, and showing how in daily operations, train service can deviate from the plan. It then reviews past research and literature on robust scheduling and points out the gaps which this research is intended to fill.

Chapter 3 analyses the structure of the crew schedule and identifies the elements that make it robust. This analysis is done by comparing Piccadilly Line crew schedules from two years that have the same underlying timetable but starkly different crew schedules. While the analysis is performed on two Piccadilly Line crew schedules, the inferences are kept general.

Chapter 4 discusses the shortcomings of the structural analysis approach to evaluate robustness, and proposes an alternate framework based on simulation. This framework is presented first in the context of a general rail transit line. Thereafter, details of a specific implementation of the model for the Piccadilly Line are explained, including detailed discussion on the inputs, methodology and outputs.

Chapter 5 validates the simulation model of the Piccadilly Line under a range of operating conditions. The model is then used to perform a comparative analysis of the performance of the same two crew schedules that were compared in Chapter 3 on the
basis of their structures. The results of the simulation-based approach are combined with that of the structural approach to provide some insights into best practices for robust crew scheduling for the Piccadilly Line.

Finally, Chapter 6 summarizes the findings of the research, highlights some of the shortcomings and suggests directions for future research.
Chapter 2

Background & Literature Review

This chapter presents the background information that provides the context of this research. Section 2.1 provides an introduction to Transport for London and the London Underground, Section 2.2, the Piccadilly Line. Section 2.3 provides an overview of the Crew Scheduling process as it is performed in most transit agencies, and then discusses the specific case of the London Underground and the Piccadilly Line. Section 2.4 thoroughly discusses the function of real-time service control which is essential to an understanding of robustness. Section 2.5 establishes the link between service control and robust crew scheduling. Finally, Section 2.6 provides an overview of the literature in this area.

2.1 Transport for London & the London Underground

The public transport network of greater London is under the responsibility of Transport for London (TfL), a local government body created in 2000 as part of the Greater London Authority (GLA), and overseen by the office of the Mayor of London. TfL has oversight and planning authority over all modes of public transport and the urban arterial street network within London, with the exception of most National Rail suburban rail services. TfL manages approximately 700 bus routes, 11 heavy rail lines which constitute the London Underground network, the London Overground heavy rail lines, the Docklands Light Rail (DLR), a small tram network (Tramlink), the Barclay’s Bike Share network and the Emirates Airline cable car.

Operations are generally contracted out to private companies, with the exception of the London Underground, which is operated by London Underground Ltd. (LUL),
a subsidiary of TfL. The London Underground network is one of the largest metro systems in the world carrying about 3.7 million passengers daily on a total length of 402 line-km, approximately half of which is below ground. It is comprised of 207 stations on 11 lines, with many of the lines having multiple branches.

Figure 2-1 illustrates TfL’s rail network, which consists primarily of Underground lines. The Overground network is shown as dotted orange lines; Tramlink, which serves the southern part of the region, is shown as a dotted green line; and the DLR is shown as a dotted aqua line. The National Rail network, not managed by TfL, is shown as a series of gray lines. The Piccadilly Line, which is the focus of this research, is shown in dark blue. Altogether, the network is large, complex, and highly interconnected.

![Image of London's public transport network](image)

**Figure 2-1:** London’s public transport network. [17]

Under the leadership of the Mayor of London, TfL has substantially expanded and improved the region’s public transportation network. Since 2000, several Underground lines have been upgraded, the Overground network has been created, and the DLR has been substantially extended and improved. Bus services have been expanded with
increased operating hours and more frequent services. At the same time, the regional government, working with local councils, has expanded the number of available bike lanes, improved the walking environment, and introduced a congestion charge in central London.

These initiatives have contributed to significant changes in the commute patterns of residents of Greater London. As documented in the Travel in London report, the public transport mode share of all journeys in the region has increased from 24% in 1993 to 35% in 2011, while that of private motorized vehicles (motor cycles, automobiles, and taxis) declined from 51% to 40% over the same period, while the share of non-motorized modes (walking, bicycling) remained steady. In absolute terms, the number of trips by various public transportation modes has increased by 55-95% in this time period, while motorized trips have declined by 10%, and non-motorized trips have increased by 20% (Figure 2-2). This is remarkable given that the region’s population has increased by about 1 million in this time period.

![Figure 2-2: Change in number of trips since 1993, by mode. [18]](image)

The increased use of the public transport network can also be quantified in terms of passenger entries into the system. In 2011, the TfL system provided more than 3.5 billion rides, of which 63% were provided by bus and 31% by the Underground. 2011 was a record year for the Underground in particular, with more than 1.1 billion rides. Current estimates are that annual ridership will increase to 1.3 billion by 2015 [19]. London’s continued population growth (it is expected to increase by a further 1.25 million by 2030) and increasing use of the public transport network demonstrates the vital importance of providing efficient, reliable and comprehensive transport services throughout the region.
2.2 The Piccadilly Line

Though the simulation-based framework for evaluating the robustness of a crew schedule that is developed in this thesis is general and can be applied to any rail transit line, it has been motivated by and is applied to the Piccadilly Line. Therefore, this section describes some of the salient features of the LU Piccadilly Line.

The Piccadilly Line is a heavy rail transit line that of 71 kilometers, the second longest in the LU system (after the Central Line), and serves 52 stations. Figure 2-3 (adapted from [5]), shows the geometry of the line along with the service pattern. The line runs from Cockfosters in North London, through the center of London, and then to two western branches—one to Heathrow Airport and the other to Uxbridge. Along the way, the line serves many important stations, including Arsenal, King’s Cross-St. Pancras, Piccadilly Circus and the theatre district, and the museum district in South Kensington. The central section of the line is in a deep tube that first opened in 1869, but most of the branch sections run on the surface or on embankments.

![Figure 2-3: The geometry and service pattern of the Piccadilly Line](image)

As Figure 2-3 shows, LU provides a range of service at rush hour to different parts
of the line. The central section, from Acton Town to Arnos Grove, is scheduled for a peak service of 24 trains per hour. The Heathrow branch also receives more service than the Uxbridge branch, both because of the importance of connecting to the airport and also because the Rayner's Lane—Uxbridge section of the line is shared with the Metropolitan Line, which offers more direct and more frequent service to central London.

The four-track section of the line between Ealing Common and Earl's Court is shared with the District Line, which operates as a “local” service to the “express” Piccadilly Line (the two services are on separate, but adjacent tracks), which serves fewer stations.

Service is generally operated from 5:30 to 0:30, with 894 scheduled one-way trips on the line. Not all the trips operate from terminus to terminus; many trips terminate at Arnos Grove and at Rayner's Lane.

The Piccadilly Line has two depots: at Northfields, and at Cockfosters. In addition there are a number of sidings (Acton Town, Arnos Grove, South Harrow and Uxbridge) where small numbers of trains are stabled overnight. The Piccadilly Line has two crew bases— at Acton Town and at Arnos Grove—and these are the locations where crew changes take place. Finally, the service control facility is located at Earl's Court.

The Piccadilly Line, which runs through some of London's most popular residential districts as well as many of its major destinations, serves more than 210 million annual journeys, making it the fourth busiest line on the LU network. Congestion, as on the TfL system in general, is heavily concentrated at peak times and in the central area. Figures 2-4 and 2-5 show the flows on the Piccadilly Line vis-à-vis the capacity during the peak thirty minutes of the morning peak, for the eastbound and westbound directions respectively [15].

It is observed that westbound trains operate at capacity (5 standing passengers per square meter) between King's Cross and Holborn, and eastbound trains are nearly as crowded coming into Acton Town from the Heathrow branch. This implies that passengers are often uncomfortably packed into trains or left waiting on platforms. It also implies that even a single missed train during the peaks can have significant passenger impacts.

This thesis focuses on the Piccadilly Line in part because the service is one of TfL's most important. Its ridership is likely to increase over the next decade as London's population continues to increase and Heathrow Airport and King's Cross-St. Pancras are expected to become even more popular than they are today. But just as important
Figure 2-4: Peak thirty minute (AM) flow on the Piccadilly Line – Eastbound [15]
Figure 2-5: Peak thirty minute (AM) flow on the Piccadilly Line – Westbound [15]
to the selection of the Piccadilly Line as the subject of this analysis is the fact that the line is years away from a major technology upgrade. Though the service has been well maintained for decades, it is handicapped by having some of the oldest trains in the system (dating from 1973), an antiquated signaling technology, and decaying capital infrastructure resulting from the line's age.

The TfL business plan suggests that capital improvements are many years off, though the Mayor’s Transport Strategy notes that there is an effort to provide the Piccadilly Line “new trains, more capacity and quicker journeys” [10]. Upgrades for the Piccadilly Line are expected in the “2020-30s” period. Moreover, any alternative investments that will lead to reduced stress on the line are unlikely in the medium term. The Victoria Line, which parallels the Piccadilly Line for much of its route, is often even more crowded. And the proposed “Crossrail 2”—a new service mentioned in the Mayor’s Transport Strategy that would run between Chelsea and Hackney (and could parallel the Piccadilly Line from Wood Green and King’s Cross to Piccadilly Circus)—has yet to be funded, let alone begun construction [9]. These conditions make the Piccadilly Line an ideal choice for this important case study.

2.3 The crew scheduling process

2.3.1 The service planning process

The operations plan is, generally speaking, the set of plans which fully describe the utilization of transit agency resources in daily operations under ideal conditions. It is designed to meet the service policy requirements set forth by management while complying with crew work rules, vehicle management and infrastructure maintenance requirements. Typically the operations plan is built around the service plan. While the service plan is focused on customer services, that is, all train movements in passenger service, the operations plan includes everything needed to produce the service plan, including a working timetable (including train movements which are not in passenger service), a crew schedule, and a crew roster.

The crew schedule develops driver shifts to cover all parts of the working timetable, and the crew roster then links individual employees to those shifts. The crew roster development process is agency-specific and depends heavily on agreements between management and unions in terms of work rules. The published timetable is the most important component of the operations plan. In a sense, it is a promise by the agency to the customer. However, the timetable is also very important for asset management and strategic planning on a metro system.
The operations planning process is described in detail by Ceder[3]. It consists of four steps: network route design, timetable development, vehicle scheduling and crew scheduling. In an existing metro system, network design will not be of great importance, leaving the three other steps to be repeated whenever a new operations plan is needed.

The primary scheduling parameter for high-frequency rail services is the service frequency per line section by time-of-day, which is a function of passenger demand, maximum and minimum headway constraints and infrastructure characteristics. It is usually expressed in trains per hour (tph). The timetable development process builds on the above mentioned factors and service frequency policies. Furthermore, it requires running times, dwell times, layover times and capacity limitations as input.

Historically, these variables were primarily derived from models and assumptions since operational data were difficult to collect and often involved significant uncertainties. However, as many transit agencies have installed digital signaling and train control systems in recent years, the availability, accessibility and quality of operational data have greatly improved and the roles of data analysis and modeling in the operations planning process are shifting. Where possible, the analysis of operational data can provide information on many of the variables mentioned above without a direct need for models. The increased availability of data can also assist in the development and calibration of better predictive models to determine the effect of future changes.

The result of the timetable development is a set of individual end-to-end train trips on the line which constitute the timetable for passenger service. The last step in the operations planning process is crew scheduling, which aims to define crew duty pieces such that all vehicle movements have a driver assigned to them, typically at minimal overall crew cost, while satisfying all the crew constraints. Aside from the scheduled crews, a number of spare operators are allocated to crew depots to cover for absences and unforeseen needs. In this thesis, the terms crew, operator and driver are used synonymously, as are the terms timetable and schedule.

2.3.2 Crew scheduling constraints on the Piccadilly Line

The previous section described the service planning process as it is generally carried out in a transit agency. This section focuses on one aspect of service planning, namely crew scheduling, and with specific reference to the Piccadilly Line. This is not a discussion of the general crew scheduling process, which is discussed at length in [3]. Rather, the subject of interest here is the specific constraints on crew scheduling that are present in the case of the Piccadilly Line. In Chapter 3, the structure of Piccadilly Line crew schedules is analyzed; this section merely discusses the constraints
that result in that structure.

Crew scheduling (or duty scheduling as it is referred to in the London Underground) refers to the process of assigning crews to scheduled trains runs in the vehicle schedule (or the timetable). A train's crew might consist of just a driver or train operator, as is the case with the London Underground, or might have more than one person, which is common in the case of inter-city trains. A crew has to be assigned to every single scheduled train movement, which includes not just train movements that carry passengers, but other train movements such as depot moves and deadheading.

Furthermore, this assignment must be performed in a manner that does not violate any agreements that govern how much and what kind of work a person can perform. These agreements, typically signed between the management of a transit service and labour unions place restrictions on what drivers can or cannot do, and therefore constrain the process of crew scheduling. The conventional approach to crew scheduling has been to build a schedule that satisfies the operations plan without violating any crew constraints at minimum cost, which usually means, using the least amount of manpower. The problems with a purely cost-minimizing approach to crew scheduling are well known, and many leading transit agencies today try to build in some resiliency or robustness into their crew schedules at the cost of more manpower.

In the case of the Piccadilly Line, there are a number of constraints on the drivers. Drivers cannot be scheduled to drive for more than four hours and fifteen minutes without a meal break, and their duty cannot be longer than eight hours (excluding the meal break). The meal break must be at least thirty minutes long, in addition to an agreed upon time to walk to and from the meal-room. Similarly, there are agreed upon allowances for the time required to check-in, to walk (from the crew base to the platform or siding or depot), to prepare or stable a train etc. There are restrictions on how early duties can start, or how late they can finish. A certain number of duties can start very early or finish very late to allow for early pull-outs and late pull-ins at the depots. Not all duties start at the crew bases; some duties start at the depots, though there are restrictions on how many drivers are allowed to check-in directly at the depot.

In addition to these hard constraints, there are also soft constraints which are usually a result of driver or service controller preferences. For instance, meal breaks are not scheduled at certain periods of the day because of driver preference. Eastbound reliefs at Arnos Grove and Westbound reliefs at Acton Town are prioritized because service controllers find that they have more flexibility with these reliefs. There are also minimum late running margins which are incorporated into the crew schedule as hard or soft constraints to increase schedule robustness.
In addition to the above constraints on driver duties, there are also constraints in the rostering process. The average duty length across all the duties in a roster must equal 36 hours per week. There is a single roster for all Acton Town duties, and six rosters for Arnos Grove duties. These six rosters correspond to six links or times of the day when duties start (e.g., Dead early, intermediate early, dead late etc.). Therefore, the average duty length constraint applies individually to each of these six rosters. This is known as the fixed-shift link constraint, and its effects on the crew schedule are discussed in Chapter 3. Drivers are allowed to take a fixed number of leaves in a year, and the roster includes spare operators to cover for drivers who are on vacation. There is also a required number of spare drivers who must be present to cover for sickness, casual absenteeism etc. In addition to this, some more spare drivers can be scheduled to increase schedule robustness.

### 2.4 Service control

In daily operations, the train service can deviate from the service plan for a number of reasons including variability in passenger demand, variability in driver behavior, unforeseen operating conditions, or incidents. Due to these events, the actual operations may deviate from the schedule, and certain parts of the operations plan, such as scheduled crew movements may even become infeasible.

Service control can be described as the work of modifying the operations plan in real-time to deal with the aforementioned unforeseen events. It is a process which is both proactive and reactive. Some constraints are known before they become immediately relevant for service delivery; for instance, the availability of rolling stock at the beginning of service can be limited due to maintenance requirements or defective trains, or certain track sections might be unavailable due to engineering work. In that case, service controllers can plan ahead and modify the schedule in order to allocate the remaining resources optimally and avoid conflicts.

However, more commonly, service control must deal with disruptions and train delays as they occur, causing deviations from the schedule. Section 2.4.1 examines the common types of disruptions and schedule deviations in more detail. From the service controller's point of view, these events immediately affect service quality or make the schedule infeasible, thus requiring controllers to reschedule service in order to maintain it at the best possible level, subject to the momentary constraints, and to eventually restore the full level of service.

The real-time control of a transit line functions much like a control loop, in which the two elements are the service control center and the operations on the transit line.
The service controllers constantly monitor the state of the system and compare indicators of the level of service to the service plan and to other service quality objectives. Deviations will cause controllers to perform corrective interventions (a list of possible interventions will be given in Section 2.4.2).

The choice of intervention is informed by the service controller’s knowledge of the system, the momentary constraints, the target state (which is often the timetable) and a projection of the effect of the intervention on the system; this process is explored in Section 2.4.3. For a more comprehensive treatment of service control for rail transit lines, the reader is referred to [2] from which some of this material is adapted.

2.4.1 Disruptions and schedule deviations

A disruption (or incident) is defined as a single, unforeseen event which causes one (or more) trains to be unable to complete their trips as scheduled. A disruption has a beginning and an end in time and a location at which its effects are felt. This is in contrast to congestion from demand peaks and routine variability, which can generally cause different running times and dwell times than the schedule sets out. Uniman [20] refers to the result of congestion (and small regularly occurring anomalies) as recurrent unreliability in contrast to the aforementioned disruption-related unreliability and shows that the effects in terms of passenger travel times can be distinctly different.

The consequence of disruptions and delays from congestion is a train service that deviates from the service plan, which will henceforth be referred to as a disrupted service. Although delayed trains are the most common form of disrupted service, there can also be early trains or trains which are completely missing from service (e.g., they become defective and are withdrawn). An understanding of the different types of disruptions and service deviations is essential for studying robustness since robustness measures the resilience of a schedule to disruptions. The rest of this subsection discusses the typical causes and effects of disruptions in a rail transit line.

Causes of disruptions

The cause of a service disruption can be endogenous or exogenous to the transit system. Exogenous causes can be, for example, a passenger operating the emergency alarm, an object on the tracks or weather (on open track sections.) These causes are basically beyond the direct control of the transit agency. The occurrence of disruptions with endogenous causes, on the other hand, can be influenced by the transit agency’s maintenance procedures, accountability structures, and employee discipline policies. Examples of endogenous causes are defective trains, infrastructure problems (such as signal failures), staff communication errors or the unavailability of a driver.
Despite their varied causes, disruptions usually manifest themselves on the train service in a limited number of ways:

- **Stationary line blockage**: A blocked train is not able to leave the station it is berthed at or proceed beyond a certain point on the line. Since it is generally not possible for trains in rapid transit systems to pass each other, a train which is blocked on the line will not only be delayed, but will also cause following trains to queue behind it and thus be delayed.

- **Slow-moving line blockage**: This might be caused by a defective train which is able to move only at a reduced speed. Unlike the stationary blockage, service from affected stations does not come to a standstill, but travel times are increased, not only for the defective train but also (as a function of the service frequency) for those following it.

- **Single train delay**: This type of delay is experienced by one single train without directly affecting other trains in the depot or on the line. It could, for example, be a train which pulls out of its depot late because no train operator was available at its scheduled pull-out.

- **Train blocked in terminal**: A train might not be able to depart a terminal for various reasons. The effect on other trains depends very much on the specific terminal configuration, but in a typical stub-end terminal with more than one reversing track, this will not result in a complete blockage. Instead, the capacity of the terminal is reduced and following trains are only able to reverse at a reduced frequency.

- **Reduced infrastructure capacity**: Unlike the aforementioned disruptions, this phenomenon does not necessarily involve a train. Reduced infrastructure capacity can be caused by track, switch or trackside equipment failures, resulting either in temporary slow zones or sections of the line becoming unusable. For instance, a track failure may cause a certain section of a line to have only one usable track for both directions, resulting in reduced capacity.

**Effect on the service**

The effect on the service can be defined from either the agency’s or the passenger’s perspective. This difference is especially important when considering high-frequency metro lines where passengers can be expected to arrive randomly, without referring to a published timetable. In this case, from the passenger’s perspective, important variables of service quality are platform waiting time (which is a function of expected headways and headway variance) and on-train travel time, which is directly related to service speed, but not necessarily schedule adherence.
However, from the agency's point of view, the degree of adherence to the operations plan (and thus to the crew and vehicle schedules) is of critical importance. Disruptions can cause both a deviation from the schedule (i.e., train or driver lateness) and a deviation from service quality standards. The two effects, although correlated, are not necessarily in a direct cause-and-effect relationship.

For instance, it may be possible to maintain regular headways as scheduled on a line section despite trains running late due to an earlier disruption. That would be a deviation from the schedule, but a passenger waiting for a train would experience service at the expected headways, and there would be no deviation in service quality with regard to regular headways. In such a situation, the drivers in the different trains may be running late. While this does not affect the passenger experience directly, it is possible that service control interventions will be performed to reduce driver lateness (so as to maintain the feasibility of the crew schedule or reduce overtime costs), and these interventions may ultimately affect the passenger experience.

Fundamentally, a disruption can cause a gap in the service (possibly followed by bunched trains), an incorrect sequence of trains, general lateness or a combination of these effects. The possible impacts on passengers vary depending on the situation.

A gap caused by a line blockage (whether stationary or slow-moving) will often be followed by a group of delayed trains with short headways, depending on the duration of the blockage and the scheduled service frequency on the line. The primary impact on passengers is through long waiting times during the gap and crowding in the first train or series of trains after the gap. The first train or few trains after the gap may also experience longer than scheduled running times due to passenger congestion. Once these effects have "passed", passengers essentially experience regular headways and might classify the service as "good". However, from the agency’s point of view, there will be a series of trains and drivers on the line which are delayed with respect to the schedule, and this may be undesirable for various reasons.

A single train delay, on the other hand, can cause a trip not to be covered, resulting in a gap of two headways between two other trains which are running on time. The delayed train might enter service later, out of sequence. To passengers, the gap has the same effect as described above (increased waiting times, potential crowding on the first train after the gap), and the first train after the gap may experience increased running times. However, from the point of view of service control, the number of trains which are off schedule is smaller than in the above case (generally only one or two), and the spacing between trains after the gap is generally more consistent with the timetable than after a blockage.

As mentioned above, a single train delay can cause a train to enter service out of
sequence. On lines with branches, delays on one branch do not immediately affect the other branches, but, a different train sequence from that in the timetable can result when trains then merge onto the common trunk section. While this deviation from the schedule may not matter to passengers as long as regular headways are maintained, it may be problematic for the agency.

General lateness with respect to the timetable can occur if the scheduled running times are inadequate, for example during peak hours with high demand and passenger congestion. It can also result from disruptions where the throughput capacity of the line or a terminal is reduced. Therefore, passengers may experience longer overall travel times as a result, but general lateness need not be accompanied by any gaps in the service or changes in train sequence.

2.4.2 Service control interventions

During a disruption, the main service control tasks are to coordinate the responses to the disruption, maintain an adequate level of service on line sections not directly affected by it and avoid conflicts between trains. This task is known as disruption management. Once a disruption has cleared or peak demand has abated, service controllers move into the phase of service recovery, also known as service restoration, where they work with real-time information and a toolbox of changes they can make to the system to achieve an ultimate target state (typically the service plan).

Collectively these changes will be referred to as service control interventions. The controller’s choice of intervention is driven by a set of constraints, objectives and priorities, which are either defined by agency policy or informally by the controllers. An intervention usually applies to an individual train. Together, the set the interventions which are performed in order to restore service to its target state constitute the service recovery strategy. It is important to note that when analyzing service recovery, the strategy must be considered as a whole – the purpose of individual interventions may not be clear unless one understands how they fit into the overall strategy.

Another point to bear in mind is the strong linkage between the train schedule and the crew schedule. Although in the planning process the crew schedule is developed given the train schedule, in daily operations the two constantly need to be coordinated to deal with variations in service patterns and resource availability, in order to ensure that each train trip has a crew. A change to a train’s trajectory (e.g., a diversion to a different destination) will always change a driver’s trajectory and his or her original schedule.

The remainder of this section presents a comprehensive list of train-related and crew-
related interventions which service controllers can perform. Crew-related interventions affect crew assignments but not train routings. Train-related interventions, on the other hand, affect trains and crews alike. In looking at the list, the reader will realize that, in order to achieve a certain outcome (e.g., put a late train back on schedule), there are many different possible interventions from which a controller can choose. Also, many of these interventions often need to be performed in tandem to achieve a desired outcome.

Train-related interventions

- **Short-turn:** A train which is short-turned is reversed before it reaches its scheduled destination. This can be done either to fill a gap in the opposite direction or to reduce its lateness by shortening its cycle time. Trains can also be short-turned upstream of a blockage in order to prevent a queue from forming. On-board passengers with destinations beyond the short-turn point need to alight and wait for the next train to their destination.

- **Withdrawal:** A train which is prematurely withdrawn from passenger service into a siding or depot is another form of intervention. This can happen at the terminal or at a siding along the line. The effect on passengers is the same as in a short-turn. This may be done because the train becomes defective, or because the train driver needs to be relieved but no relief driver is available. In addition, withdrawing trains can be used as a disruption management technique in times of reduced capacity.

- **Diversion:** On lines with branches, trains with a destination on one of the branches can be diverted to serve another branch instead. The objective might be to fill a gap on the branch to which the train is being diverted or to withdraw it to a depot located at the end of that branch. However, on branches with different cycle times a diversion can also be performed in order to shorten the cycle time of a late train or to lengthen the cycle time of an early train. The impact on passengers to the original branch is the same as if the train were short-turned. If a diversion takes effect after the train departs a terminal, passengers traveling to stations on the original branch will need to alight and wait for the next train serving their branch.

- **Trip Cancellation:** Unlike a short-turn, withdrawal or diversion, in which a train does not operate part of the trip it was scheduled for, a canceled trip is not operated in its entirety. Service controllers can cancel a trip if there are not sufficient resources (trains and drivers) available to operate it. However, cancellations may also serve as a disruption management or service restoration technique. Controllers may temporarily remove a train from service and cancel one or more round trips, then insert it back into service at a later point. This can serve as a technique for putting late trains back on schedule, or it can
be done to reduce congestion and facilitate real-time train management during disruptions. On high-frequency line sections where passengers arrive randomly, they will most likely not be aware of the missing train, but they will experience longer waiting times or more crowded trains.

- **Renumbering**: When a service controller renumbers a train, the assignment between the train (or rolling stock) and trip changes. This means that the train is now associated with a different vehicle block, and thus a different set of scheduled trips to be operated. Renumbering trains is typically associated with one of the aforementioned changes in routing (a diversion, an extension or a short-turn). In the event that a train is renumbered to a train/trip number with the same original destination, the controller's intent most likely is to assign it to a scheduled trip for which it is on time. It is important to note that renumbering trains in such a case without any other intervention only affects the lateness of a train with respect to the timetable; the driver's lateness does not change.

- **Holding at stations**: This intervention consists of delaying the departure of a train from a station beyond the end of its normal dwell time. Trains may be held to even out headways, to ensure on-time departures if they are running early or for connections with other trains. During blockages, trains can be held downstream of the blocked train in order to counteract the formation of a gap and they can be held upstream to avoid the formation of a queue.

- **Expressing**: Expressing is a technique in which a train in passenger service skips one or more stops it was scheduled to serve. Passengers on board that train must alight if their destination is one of the skipped stops, and passengers waiting at those stops experience an increase in waiting time as they wait for the next stopping train.

- **Extension**: An extended train travels past its scheduled destination to a station located beyond it. There is no negative impact on passengers on board the train as all scheduled stops are served. The degree to which passengers to the “new” destination(s) benefit depends on when the extension is announced. An extension can be performed to fill a gap, withdraw a train to a depot or to lengthen its cycle time.

- **Adding service**: Passenger service can be added in the form of unscheduled train trips which were originally not part of the service plan. This may be achieved by using a spare train or a train which was withdrawn from another part of the line. Train service can be added either to supplement scheduled services if they prove to provide insufficient capacity to serve passenger demand or to fill gaps caused by disruptions. Trains which need to be moved from one depot to another can be run in passenger service if they do not have any major defects, thus providing a benefit to passengers at little extra cost.

- **Adding an out-of-service trip**: This intervention is effectively an unplanned deadhead, which refers to any train movement on the line which is not in passenger service. There is a large variety of reasons for such trips, but they are
generally related to moving trains or drivers from a point in the line where they are no longer needed to their new site of operation or to the depot.

- **Train priority at junctions and terminals:** At any junction where branches merge, and at terminals, service control must establish a train priority scheme. The planned form of this scheme may be embedded in the automatic signaling system and can be as simple as "first come, first served". Trains can also be held, either to their scheduled sequence or as an intervention to establish a desired service pattern downstream.

**Crew-related interventions**

- **Substituting a spare driver:** Replacing a rostered driver with a spare driver is one form of crew-related intervention. For example, such an intervention can be used to relieve a late-running driver who needs to step off but has not yet completed his/her driving assignment.

- **Dropping back:** Dropping back in a planned form is a crew scheduling technique which allows the train layover time at a terminal to be shortened without compromising the driver layover time. Specifically, every driver steps off his/her train at the terminal, takes a break and then departs on the following train. Dropping back can also be imposed by service controllers as an ad-hoc intervention to speed up the reversal process with the help of a spare driver.

- **Jumping up:** On a line where drivers are planned to drop back, the intervention called "jumping up" refers to a change back to every driver departing with the same train that he or she arrived on, i.e., a discontinuation of the policy of dropping back.

- **Switching drivers:** If trains need to be resequenced, this can easily be done by renumbering them, but more importantly the drivers may need to be resequenced – one driver needs to move to a train ahead of his/her train and the other needs to move to a train behind. If the two drivers are at a terminal at the same time or meet each other on the way, a cross-platform driver change is possible. This might be problematic since it compromises the layover time of one of the drivers. To avoid this, controllers can work with a spare driver, for example as follows. The drivers of trains 1 and 2, both for the same destination and where train 2 is following train 1, need to be switched. The two drivers shall be called driver 10 (on train 1) and driver 20 (on train 2). A spare driver steps onto train 2 a few stations short of the terminal. Driver 20 waits at that station and steps onto train 1 after train 1 has reversed. Driver 10 then steps off and waits for train 2, where he or she replaces the spare driver. Using this technique, driver 10 had a layover at the terminal, whereas driver 20 had a layover at the station where the spare driver stepped on.

- **Stock and crew:** Stock and crew functions similarly to switching drivers; it
is a technique for moving late trains (and late drivers) forward in the timetable without short-turning the train. However, it only works if there is a spare train and driver available at the terminal. While the late train is traveling towards the terminal in passenger service, the spare driver with the spare train departs the terminal at the time the late train was scheduled. The late train and the spare train meet at a station and the two drivers switch trains across the platform. The spare train then receives the number of the scheduled train and continues on its path whereas the train heading towards the terminal becomes the new spare train and is stabled by the spare driver. As with the crew changeover described above, a problem might arise because the original driver does not get a layover at the terminal.

2.4.3 The service control process

Service control needs a set of simple, real-time performance measures through which the operations on a line are evaluated in real-time and compared to their target state defined by the service or operations plan. These performance measures largely define the decision rules and priorities for service control.

Common supply-centric measures can be headway regularity, lateness of service or total missed trips. The last two measures define schedule adherence, and headway regularity is featured separately because it can be independent from schedule adherence. Passenger-centric measures might include total passenger delays or travel time reliability metrics, but they are generally much more difficult to calculate in real-time and to relate to the service variables which controllers can influence. Adherence to the crew schedules is currently not explicitly tracked by the London Underground, though as we shall see, crew management is a major driver of service control decisions.

Based on these measures, the state of the system is monitored in real time. When a disruption occurs (as described in Section 2.4.1), the state of the system changes and it may become necessary for the service controller to intervene, in ways that were described in the previous section. In this section, the primary drivers behind service control are examined, as well as the service strategies that are employed.

The overarching objective of the service controllers of the Piccadilly Line (and likely for most rail transit lines) is to operate as many trips as possible, and to operate those trips as close to schedule as possible. This objective is based on the premise that the timetable is the optimal service output, and that achieving this objective automatically achieves the ultimate objective of providing the best service for the passenger.

It is important to note that headway regularity is not explicitly embedded in the
above objective, even though it is one of the (if not the) most important determinant of service quality in a high-frequency metro. One reason for this could be that service controllers are not directly provided with information on headways, as their information systems only report train lateness. It is also not possible to observe headways from the on-screen displays because they are often not to scale. Furthermore, the job of maintaining regular headways is often relegated to signal controllers who may decide to hold trains to regulate headways.

Although, schedule adherence is treated as the main objective, it is possible during some disruptions that the target system state is not the operations plan, but some other feasible plan. This is especially the case during very severe disruptions or partial line suspensions when it is realized that the best course of action is often to operate a reduced and simplified service.

The primary driver for controller intervention is train lateness, which is often but not always the same as driver lateness. In the case of the Piccadilly Line, ten minutes is often used as a threshold for train lateness when making a decision on whether or not to intervene. This is because the service controllers feel that trains that are late by ten minutes or less have a fairly good chance of making up (some of) that lateness in their recovery time.

It is important to note that service controllers are not explicitly made aware of driver lateness or how much slack time a driver has. While it is possible for them to calculate this from the crew schedule, the demanding nature of the job especially during disruptions, means that this is impractical. As a result, service controllers use the minimum slack as a proxy for deciding when to intervene. As will be discussed in detail in Chapter 3, the crew schedule ensures that most drivers have at least ten minutes of slack time, and therefore, it is not essential to intervene and put those trains back on time.

But in general, it would be correct to say that crew constraints are the biggest driver of controller actions. Section 2.3.2 discussed many of the constraints that drivers need to adhere to, such as a maximum continuous driving time of four hours and thirty minutes etc. While the crew schedule is built to ensure that these constraints are satisfied, that may not be true during a disruption. A driver who has exceeded the maximum permitted driving time is entitled to park the train at a siding and wait for his relief. However, several controllers indicated that in the case of serious disruptions, drivers will generally be understanding if the driving time regulations are breached, as long as they can assume that controllers are working to minimize the excess driving time.

Another reason that lateness of more than ten minutes is actively corrected is that
a driver who gets off a train late might be late for their next run. This might not always be the case depending on the late running margin (or slack time) that the driver has, but this is not something that the service controllers are actively aware of. Similarly, it might not be advisable for the driver getting on to the train to use up all the slack right at the beginning of the run in case there is a disruption later in the day.

Finally, trains are put back on time because service controllers wish to avoid drivers starting their runs late. This is partly because controllers do not have a means of directly talking to drivers in the depots, and need to rely on the Duty Managers in the depots to inform a driver about which train to board. Any confusion in this operation might result in a line blockage because the driver is not at the right place at the right time. Because of the scope for misinformation in this process, controllers prefer to have drivers start their runs with minimal lateness.

Thus, in most cases it is okay for drivers to step off their train a little late, but not to step on late. Therefore, service controllers use a combination of renumbering and rerouting delayed trains in order to put drivers back in time for their reliefs. This strategy is illustrated in the following example (Figure 2-6).

Consider the case of two trains 240 and 241 that are scheduled as described in the table below. They travel past AGR (this could be any station) in the Eastbound (EB) direction to CFS, where they reverse, and then travel through AGR Westbound (WB). Also assume that both trains are scheduled to have a crew change at AGR-WB.

Now consider the case when both of these trains are running 15 minutes late because of a disruption on the line. In the absence of controller intervention the trains would arrive at their relief locations about 15 minutes late (they might make up some time at the terminus, if there is recovery time available). Therefore, the drivers would get off 15 minutes late, which might be acceptable if they have 15 minutes of slack time, or don’t have any other runs afterwards. The drivers who are getting on to the trains would then start their runs 15 minutes late, and may be using up all their slack time at the beginning of their run.

In order to avoid this situation, the following operation is possible:

- Short-turn train 241 at AGR.
- Renumber train 240 at 241, and train 241 as 240 at AGR-WB.

If the above interventions were performed, the table below describes what would happen. Train 240 would arrive at AGR-WB 15 minutes late, but would leave on time as train 241. Train 241 would arrive at AGR 15 minutes early for the relief, and would depart AGR-WB on time as train 240. The driver of train 240 gets off 15 minutes late, and the driver of train 241 gets off 15 minutes early. But the two new drivers...
Original Departure Times

<table>
<thead>
<tr>
<th>Train #</th>
<th>AGR-EB</th>
<th>CFS</th>
<th>AGR-WB</th>
</tr>
</thead>
<tbody>
<tr>
<td>240</td>
<td>10:00</td>
<td>10:15</td>
<td>10:30</td>
</tr>
<tr>
<td>241</td>
<td>10:15</td>
<td>10:30</td>
<td>10:45</td>
</tr>
</tbody>
</table>

Modified Departure Times

<table>
<thead>
<tr>
<th>Train #</th>
<th>AGR-EB</th>
<th>CFS</th>
<th>AGR-WB</th>
</tr>
</thead>
<tbody>
<tr>
<td>240</td>
<td>10:15</td>
<td>10:30</td>
<td>10:45 as 241</td>
</tr>
<tr>
<td>241</td>
<td>10:30</td>
<td>--</td>
<td>10:30 as 240</td>
</tr>
</tbody>
</table>

Figure 2-6: Example of a renumbering operation

both start their runs on time, which was the desired end result.

The negative consequence of this strategy is the fact that one AGR-CFS-AGR round trip was cancelled. In the case of the Piccadilly Line, this might be an acceptable compromise because AGR is very close to CFS, and the demand in that section is extremely low.

This strategy can be varied in many different ways depending on the exact configuration of the crew reliefs, the availability of spare drivers, the necessity of maintaining service to the end of the line and other factors. But the above example captures the spirit of the strategy of renumbering and rerouting trains in order to reduce driver lateness.

In general, controllers frequently employ recovery strategies which hinge on crew reliefs. Crew reliefs act as "fixed points" in the timetable, and controllers often reroute and renumber late trains to meet these departures as scheduled, because in doing so, they are simultaneously restoring service to the timetable and meeting crew management constraints.
Controllers have the option of using spare drivers to operate some trips. This may become necessary when the scheduled driver is close to his maximum driving time, or when a driver who is supposed to step onto a train is unavailable for some reason (such as a late arrival from the previous run), or in many other scenarios. Based on conversations with multiple Piccadilly Line controllers, controllers try not to rely heavily on spare drivers because the availability of spare drivers is not always known reliably beforehand.

This concludes our discussion of service control. In this section, we have explored common disruptions that take place in a rail transit line, what service control interventions are possible to deal with such disruptions, and the service control process that is followed to determine what interventions are actually used in resolving the problem. The next section explores the link between service control and robust crew scheduling.

### 2.5 Service control and robust crew scheduling

Figure 1-1 showed a flow of decisions from the management level to the planning & scheduling teams and on to the operations level, but at the same time, a flow of information is needed from the operational level back to planning & scheduling and management. What these connections describe is in fact a very strong bidirectional link between service control and operations planning & scheduling. This interaction can best be understood by revisiting the operations planning process. The inputs used for scheduling are usually data on running times and dwell times, and assumptions or models of scheduling variables for which insufficient (or) no data are available, such as layover times, throughput capacity and junction and terminal capacity.

The validity of the assumptions or models can only be assessed with the help of data gathered during operations. Planners can also use operational data to identify bottlenecks, for example by analyzing the origin of delays, the variance of certain variables in daily operations or levels of train impedance.

However, it is very important to recognize that the system being analyzed is not a system of autonomous actors (i.e., trains, drivers and passengers), but rather a system which is controlled by a central, intelligent entity (service control) which has an understanding of the state of the entire network and can influence individual actors within its control (trains and drivers) to change that state. That means that the data being observed (e.g., running times, dwell times) may tell an incomplete story without being linked to service control interventions, and that this, in turn, can distort the models used for determining scheduling variables.
An excellent example of this relationship comes from the process of crew scheduling. A certain amount of slack time is included in every single duty in order to mitigate the effects of lateness and ensure that the duty does not become infeasible the moment a train is late. The amount of slack to include usually depends on the typical train lateness at the time of relief. This lateness can be observed from train tracking (AVL) data, but the lateness thus calculated is the lateness that was present after any service control interventions. Thus we see that service control actions influence the data used in operations planning.

Another example of this reverse feedback is the mix of crew relief directions. When crew reliefs take place at a non-terminal station (as is the case with the Piccadilly Line), it is possible to schedule reliefs in either direction at that relief location. Clearly there are a large number of relief configurations that are possible by varying the time, location and direction of individual reliefs. If the service is running as planned, there is little to choose from between the reliefs, and the best configuration of reliefs would be one that satisfies all crew constraints at minimum cost. However, when a disruption occurs, this is not the case. Service controllers are acutely aware that different types of reliefs have different advantages and disadvantages, and offer varying amounts of flexibility during service recovery. Therefore, the choice of an ideal set of crew reliefs depends on the ability of service controllers to work with those reliefs when dealing with the type of disruptions that typically affect the line.

The manner in which service control manages disruptions also provides valuable input to service planning. By understanding how disruption management strategies are implemented as a function of its parameters (e.g., train frequency, scheduled recovery time), by investigating where spare network capacity such as reversing tracks or crossovers is used and what resources (e.g., spare trains and drivers) are required for disruption management, planners can try to accommodate service control better in the planning process. For instance, they can restrict the usage of these resources in the operations plan or make them easy to reallocate in the event of a disruption.

Furthermore, understanding service control can also help an analyst understand which parts of the operations plan are most vulnerable to disruptions. For instance, one can analyze which scheduled trips are most frequently changed and which stations are most frequently affected. This may lead to some surprises, as the effects of disruptions may be felt much less in the line sections where the disruptions occur than in the line sections where trains and personnel are removed by service controllers in order to deal with the disruption.

This directly leads to the issue of robust scheduling. Robustness, as we have discussed before, is the characteristic of a schedule that allows operations to remain feasible during a disruption, and permits easy service recovery. It is clear that the robustness of a schedule can only be discussed in the context of service control strategies, as it is
service control that uses the flexibility embedded in a schedule to handle disruptions. The aim of this thesis is to develop tools to evaluate the robustness of crew schedules, and such an evaluation can only be performed with a thorough understanding of how service control works. This can then be applied to provide feedback to planners and schedulers about how planning and scheduling decisions affect the robustness of a schedule.

2.6 Literature review

This section presents a review of literature that is relevant to this research. This is organized into three subjects: simulation, reliability & robustness, and service recovery.

2.6.1 Simulation

Hoogheimstra and Teunisse [7] describe their research on robustness in timetables in the Dutch railway network. A program called DONS (Designer Of Network Schedules) is used to generate timetables, and a DONS-simulator is developed. The simulation tool enables the authors to study the effect of small disturbances on the punctuality of trains in the entire network. The simulator can also be used to evaluate how investments in infrastructure affects.

Middelkoop and Bouwman [12] describe the architecture and features of the simulation program Simone (Simulation Model Network). Simone is a simulation environment developed with the purpose of determining the robustness of a timetable and the stability of a railway network, and thereby improve the quality and stability of the timetables from a set of different criteria. It does this by determining bottlenecks in the network, by examining the number of delayed departures for all the stations in the network. Simone can also be used for analyzing delays and exploring causes and effects of delays for different layouts of railway infrastructures and timetables. Unlike many simulation models, Simone can simulate an entire railway network.

Simone is used to compare the quality of different timetables. Quality in timetables depends on network properties such as correspondences between trains and use of shared capacity. When there are no disturbances all trains run according to schedule. When disturbances occur Simone inspects the different types of delays (primary and secondary) and the user gets extensive information on the delays and delay propagation in a specific simulation. This makes it useful for comparing the robustness and punctuality of different timetables. This is similar to the approach used in this thesis to evaluate the performance of a schedule, though the key difference is that in
Simone there is no representation of any crew constraints and crew related service control interventions.

Sandblad et al. [16] describe various concepts within simulation of train traffic for use in both planning and training, and the development of a new simulation system which can contribute to improved methods for train traffic planning, experiments for developing new systems and training of operators.

The paper presents thorough description of the various uses of simulation as a planning tool or as a learning tool, e.g. understanding the behaviour of the system, as a base for difficult decisions, or for controlling the system. In addition the report describes the different phases in the planning and implementation of a simulation project. These include problem specification, construction of the model, validation of the model, programming, verification of the program, planning the experiments, realization of experiments, evaluation of results and conclusions. The real contribution of this paper is the thorough treatment of the methodology of simulation, and its applications.

### 2.6.2 Reliability and robustness

Carey [1] describes different heuristic measures of stability of a schedule. The reason for using these measures is that analytical methods are practical only for simple systems, and simulation methods are time consuming and involved, so in practice the most widely used measures are heuristic.

The author proposes performance measures which can be used in advance, such as in the design phase or to estimate the reliability of a proposed schedule. It should be noted that even though the measures are meant to be used in advance some past information is needed to determine the distribution of the occurrence of delays.

Initially the author states a measure of reliability which assumes that no secondary delays occur (i.e. the slack time absorbs all secondary delays), except secondary delays caused to immediately following trains.

As a second type of measures, the author proposes a number of measures which build on the expected size of the secondary delays, instead of on the probability of occurrence of secondary delays. Furthermore the author proposes some heuristic measures of reliability which do not use probabilities. These measures are not based on information on the previous occurrence of delays and are based on the number, size and spread of minimum headways.
While this work represents an interesting treatment of delays, its biggest shortcoming is the simplified handling of secondary delays. This shortcoming is especially pertinent to high-frequency rail lines, such as the Piccadilly Line, where secondary delays can be even more significant than the primary delays. The purpose of service control is largely to mitigate the effect of these secondary delays, and this is best captured through simulation given the complex interactions that take place.

Mattson [11] reviews various methods of deriving causal relationships between disruptions and train delays. The article presents three methods for deriving relationships for train delays: analytical methods, statistical analysis and simulation approaches. The focus of the paper is on secondary delays and especially on how the amount of secondary delays can be related to the amount of primary delays and the capacity utilization. Clearly, higher capacity utilization needs lower buffer times which causes more unreliability, and the author reviews various theories connecting these two quantities. The conclusion is that analytical methods and statistical analysis do not require as much input data or computational effort as simulation models. But they are mathematically demanding, and depending on the context may provide significantly inferior results to simulation. Simulation methods offer the most detailed representation of a railway system and the various interactions that take place, and are therefore more reliable when it comes to studying delays.

Klabjan et al. [8] discuss the problem of robust airline crew scheduling in airlines. The approach taken for evaluating the robustness of a crew schedule is the same as in this thesis. A simulation tool (SimAir) is used to simulate the operations of an airline network in the presence of disruptions, including the effects of crew constraints and real-time operations control, and the performance of the schedule is measured.

This approach differs from this thesis in the way disruptions are handled. SimAir uses the complete distribution of block time (time from door closing to door opening) etc., which includes observations from disrupted days, and the simulation draws from these distributions. In this thesis however, the distributions of running time (equivalent to block time) are drawn from undisrupted conditions, and disruptions are specifically super-imposed by the simulator. This approach has the advantage of being able to simulate any kind of disruption. The types of recovery strategies used by airlines are also completely different from those used in a high-frequency rail service.

Chiraphadhanukul and Eggenberg [4] also discuss the evaluation of airline schedule robustness. This paper discusses the merits and demerits of a priori methods (i.e. methods that examine the structure of the schedule analytically) versus simulation based methods. They also provide a useful discussion of the effects of different kinds of performance measures and data sources on the measurement of robustness.
2.6.3 Service recovery

Goodman and Takagi [6] discuss the problem of dynamic rescheduling of trains following a disruption. The paper reviews some of the applications of computers to the problem of recovering from disturbances. When recovering from disruptions different criteria are considered; regaining the scheduled departures as quickly as possible, aiming for regular headways or maximizing capacity utilization. The authors recognize that different types of networks (such as metro or inter-city) need different recovery strategies, and suggest that for metro systems the recovery strategy of regaining regular headways is often more effective than regaining the original schedule. Two approaches to recovery are discussed: One where a known set of rules is used to recover when disturbances occur, or performing service interventions based on developing an objective function and a search procedure and iteratively finding the optimal recovery strategy.

Finally, they also discuss the differences between computer and human controlled recovery processes. The paper ends with a literature review on scheduling and recovery strategies.

Puong and Wilson[13] describe the development of, and experiments with, a train holding model. The goal of using the model is to limit the negative impacts smaller disturbances have on a train network. The holding model in the article is formulated as a Mixed Integer Problem (MIP). The objective in the model is to minimize the total passenger waiting time at stations and the extra passenger riding time due to a train holding. The objective function is minimized with respect to constraints on track capacity behind and ahead of the blockage, minimum safe headways, maximum deviation from schedule and queuing situations. The MIP is solved with a two-step procedure, starting with finding a worst case but feasible solution and then improving this solution. This two-step technique makes the execution time fast enough to solve problems in real time.

The authors discuss how the developed holding model can be used if larger disruptions occur, and also how to model short-turning. They conclude that buffer time at end stations and even headways have a positive effect in the reduction of secondary delays.

Carrel [2] presents a comprehensive description of service control and recovery strategies for high-frequency rail lines using the London Underground’s Central Line as an example. The author discusses the different types of interventions, and what drives their use. There is also an excellent discussion of the environment in which service control decisions are made and the different types of objectives that compete during service recovery. This work also devotes substantial attention to the effect of crew constraints on service recovery.
2.6.4 Conclusion

While there has been substantial research on reliability and robustness in recent years, much of it focuses on the robustness of the timetable, and not so much the crew schedule. There is very little literature on what makes a crew schedule robust, and how the robustness of crew schedules is evaluated. While there is well-developed literature related to robust crew scheduling for airlines, that is not the case for rail systems, especially high-frequency rail systems. This thesis will contribute to filling this gap in the literature.
Chapter 3

Structural Evaluation of Piccadilly Line Crew Schedules

Chapter 2 described the service planning process, and one of its outputs – the crew schedule. In Section 2.3, the process of crew scheduling was discussed, along with the specific crew scheduling constraints that apply to the Piccadilly Line. The crew schedule can affect the performance of a system, and Section 2.5 explored the link between service control and robust crew scheduling.

In this chapter, the structure of the Piccadilly Line’s crew schedules is analyzed, keeping in mind the ultimate objective of evaluating its robustness. The elements of the crew schedule that affect its robustness are identified, and the link between these elements and the performance of the schedule is established. This link is further explored in Chapter 5, when the robustness of different crew schedules is tested by simulating them under different operating conditions.

Section 3.1 qualitatively describes the overall structure of three different documents – the crew schedule, the train & duty analysis, and the crew roster. These three documents together provide a comprehensive overview of the assignment of crews to trains. Section 3.2 delves deeper into the crew schedule by identifying the elements of the crew schedule that affect robustness, and comparing the distribution of these elements across two different crew schedules – schedule 48 (operated in 2012) and schedule 49 (operated in 2013). This comparison gives the reader a sense of how two crew schedules that serve very similar timetables can still have very different structures.
3.1 Data sources

In this section, we look at three different sources of information: the crew schedule, the train & duty analysis and the crew roster. While the specifics of this discussion are based on the Piccadilly Line, the overall structure of these documents is similar across many transit systems.

3.1.1 The crew schedule

The Piccadilly Line crew schedule is specified separately for each of its crew bases: Acton Town and Arnos Grove. It is also specified separately for Weekdays, Saturdays and Sundays. All the discussion in this chapter is based on the weekday schedule, though it is equally applicable to the other two days also. Figure 3-1 is an excerpt from the 2012 Acton Town weekday crew schedule, and it is used to illustrate the way a duty is specified.

<table>
<thead>
<tr>
<th>Duty</th>
<th>Report At</th>
<th>No.</th>
<th>From</th>
<th>Time</th>
<th>Place</th>
<th>To</th>
<th>Finish At</th>
<th>Duty</th>
</tr>
</thead>
<tbody>
<tr>
<td>427</td>
<td>07:02 ACT</td>
<td>348</td>
<td>07:19 ACT E</td>
<td>08:12</td>
<td>AGR</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12:05</td>
<td>ACT</td>
<td>267</td>
<td>12:11 ACT W</td>
<td>13:22</td>
<td>ACT</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>254</td>
<td>12:45 ACT W</td>
<td>14:24</td>
<td>ACT E</td>
<td>14:41</td>
<td>ACT</td>
<td></td>
<td>07:08</td>
<td></td>
</tr>
</tbody>
</table>

Figure 3-1: A sample duty from the Acton Town weekday crew schedule

This is a single duty, to be performed by a driver. We note the following details about this duty: The duty starts at 07:03 and ends at 14:41. It consists of two distinct spells: the first spell starts at 07:03 and ends at 11:35, and is 4:32 long. The second spell begins at 12:05 and ends at 14:41, and is 2:36 long. The duty itself is 7:08 long which is within the eight hour limit. The first spell has 3:19 of driving time, and the second spell has 2:00 of driving time. Both of these are within the 4:15 limit on continuous driving time.

The two spells are separated by a thirty minute meal break. This is the standard time allotted for a meal break. Not every duty has two spells – many duties have just a single spell of work (see Figure 3-2 for example), though they still have a meal break at the end of the spell. The overall length of the duty excluding the meal break must be less than eight hours, and the driver may not be scheduled to drive a train for more than four hours and fifteen minutes in any spell.

Notice that the duty starts at 07:03, but the first train run starts at 07:19. This sixteen minute gap ensures that the driver has enough time to check-in at the crew base, and then walk to the platform. There is a gap between the two runs within a spell. This slack time is a buffer against delays, so that even if the first train is a
little late, the driver would not miss the second train.

There is also a large gap of forty six minutes between when the driver gets off train 345, and when the spell officially ends. Some part of this goes towards the time required to walk from the platform to the crew base (four minutes in the case of Acton Town). The majority of the time is a buffer or margin against late running, to ensure that even if the train is late, the driver has enough time to have the thirty minute meal break and get onto the next train on time. In this thesis, this buffer time is referred to as meal break slack time. There is very little slack time after the meal break – a driver usually has just enough time to walk to the platform. There is also some slack time at the end of the duty, i.e. between the time a driver gets off the last train and when the spell ends. This is referred to as EoD (End of Duty) slack time.

This particular duty begins and ends at Acton Town. Most duties start and end at one of Acton Town or Arnos Grove. There are also some duties that begin at the depots (Cockfosters and Northfields) rather than the crew bases. This is the case when the first train run in that duty is a pull-out from a depot, and such a duty is said to have a “remote book-on”. As mentioned in Section 2.3.2, there is a limit on the maximum number of drivers who are allowed to book-on remotely. For duties where the driver books on at the crew base and needs to travel to the depot to pick the train up, or conversely, where the driver stables a train at the depot, and needs to travel back to the crew base to sign out, sufficient travel time is provided. If the drivers cannot use other Piccadilly Line trains to travel to/from the depot (as is the case with the very early and late duties), a special staff taxi is provided to transport the drivers. In the case of the duty in Figure 3-1, the meal break also takes place at the driver’s home base, Acton Town. While this is preferred and fairly common, it is not always the case.

Let us now turn our attention to the train runs in the duty shown in Figure 3-1. While this duty has two spells and two train assignments per spell, this is fairly uncommon. Most duties have only one train assignment per spell, though both single-spell and double-spell duties are common (see Figure 3-3 for example). For each train assignment, there is a start time and place, and an end time and place, and these are tied to the underlying timetable. Except for depot moves, all train runs start and end at the two crew bases, i.e. crew changes do not take place at the termini of the line. In addition to just denoting the start/end location of the run, the direction is also
specified (ACT E, ACT W, AGR E, AGR W, AGR). This tells us the direction of the train when the relief takes place, and this plays an important role in determining the robustness of the schedule. When the relief location is specified simply as AGR without any directional modifier (E or W), it means that the train reverses at Arnos Grove.

![Table]

**Figure 3-3:** An example of a typical two-spell, two run duty.

It should also be noted that not all duties start and end in the same day. There are overnight duties that start in the night and end the following morning (see Figure 3-4 for example). These duties are necessary because some depot pull-outs and pull-ins are too early/late to be performed by normal daytime duties.

![Table]

**Figure 3-4:** An example of a night duty

In addition to daytime duties and overnight duties, the third important class of duties is cover duties or spares. These duties do not have any train assignments, though they do have a start and an end time, and a crew base. The drivers assigned to cover duties act as spare drivers and wait in the crew base. They are used as required to cover for casual absenteeism, sickness, and in the case of disruptions, to cover for drivers who miss their reliefs or have hit their parameters. See Figure 3-5 for an example.

![Table]

**Figure 3-5:** Some of the cover duties scheduled at Acton Town

### 3.1.2 The train and duty analysis (TDA)

The TDA, as it is commonly known in the London Underground, provides the same information as the crew schedule, but from the train’s perspective. It lists every trip
that a train is scheduled to make, as well the duty that each trip is assigned to. At crew change locations, it also describes what the outgoing driver does next and what the incoming driver did previously. While there is no new information in the TDA, it is a very useful way of visualizing the information present in the crew schedule. Figure 3-6 shows the TDA for Train 343.

3.1.3 The crew roster

The roster links duties together to form weekly rotations for drivers. Figure 3-7 shows an excerpt from the Acton Town roster.

Note that rotas 1 and 2 are assigned to leave cover. These are intended to cover drivers' annual vacation leave. The remaining rotas define the weekly driver assignment. Rota 3, for instance, shows what duties that driver will be performing. The policy for the assignment of drivers to specific rotas depends on each crew base, and may depend on seniority. Note that the roster also mentions the total number of hours per week and per fortnight for each rota. This is important because the average weekly duty length across the roster cannot exceed thirty-six hours, and a particular rota cannot exceed thirty nine hours. There are similar rules for the fortnightly duty length also. As alluded to in Section 2.3.2, Acton Town has only a single roster for the entire crew base, whereas Arnos Grove has six rosters for six different times of the day (Dead Early, Late Early, Inside Early, Inside Late, Early Late and Dead Late).

This concludes our qualitative description of the structure of the crew schedule. In the next section, we break the crew schedule down to individual elements, and compare these elements across two different crew schedules.

3.2 Comparing the structure of two crew schedules

We begin the comparison of crew schedules 48 (2012) and 49 (2013) by first looking at some overall statistics which are presented in Table 3.1. The underlying timetable was modified only slightly between the two duty schedules. In the 2013 timetable, trains no longer reverse at Ruislip during peak hours, and running times in general have been refined to better reflect actual performance. This is reflected in a decrease in the total driving time.

Total driving time and average duty length are only calculated for regular duties (excludes overnight duties, cover duties, spare days on the rota). Also, the average
### Figure 3-6: A sample TDA

<table>
<thead>
<tr>
<th>DUTY</th>
<th>FROM</th>
<th>TO</th>
<th>CREW</th>
<th>CHANGE</th>
<th>TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>05 15 AGW</td>
<td>05 20 AGW</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>05 28 AGW</td>
<td>06 17 ACT W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>06 18 HRT W</td>
<td>06 41 NRT F</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>06 47 HRF</td>
<td>07 19 ACT E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 27 ACT E</td>
<td>08 12 AGW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 48 AGW</td>
<td>11 40 AGW W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11 40 AGW W</td>
<td>15 11 AGW W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 29 AGW W</td>
<td>18 29 AGR W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 12 AGW</td>
<td>22 44 ACT W</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 46 ACT W</td>
<td>23 56 ACT E</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 54 ACT E</td>
<td>00 42 AGW</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>00 42 AGW</td>
<td>00 51 CFS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Schedule

<table>
<thead>
<tr>
<th>Schedule</th>
<th>Schedule 48</th>
<th>Schedule 49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of duties (ALL duties)</td>
<td>308</td>
<td>310</td>
</tr>
<tr>
<td>Reporting at ACP</td>
<td>152</td>
<td>154</td>
</tr>
<tr>
<td>Reporting at AGW</td>
<td>156</td>
<td>156</td>
</tr>
<tr>
<td>Total driving time</td>
<td>1953h 54m</td>
<td>1928h 24m</td>
</tr>
<tr>
<td>Average duty length</td>
<td>6h 47m</td>
<td>6h 45m</td>
</tr>
</tbody>
</table>

Table 3.1: Overall statistics for the two crew schedules
duty length does not include the thirty minute meal break time, but includes walking time.

### 3.2.1 Relief directions

One of the major objectives while creating crew schedule 49 was to increase the number of reliefs in the direction preferred by service controllers, i.e. AGR-EB and ACT-WB. The majority of the reliefs were in the preferred direction in duty schedules prior to schedule 47. The fixed link constraint at AGR that was first introduced in schedule 47 resulted in fewer reliefs in the preferred direction. That has been reversed in schedule 49. Table 3.2 summarizes some key information about relief directions and types.

Note that AGR-EB and ACP-WB are the preferred relief directions. When the train is scheduled to reverse at AGR, the relief is denoted as just AGR without any directional suffix. These reliefs are functionally equivalent to AGR-EB reliefs because drivers do not have to go past their relief point, like they do with AGR-WB reliefs. Ending in depots/sidings refers to those duties which end at AGR/ACP/SHR sidings or NFD/CFS depots.
It is clear from the table that the scheduling team succeeded in its objective of increasing the number of reliefs in the preferred directions. Another important change to note is the decrease in the number of intra-spell reliefs from 38 to 33. These reliefs occur when a driver has more than one train run per in a spell.

### 3.2.2 Driving time: The 4h 15m parameter

A driver is allowed to drive continuously without a meal break (this can include multiple trains) for a maximum of four hours and fifteen minutes. Figure 3-8 shows the histogram of continuous driving time for reliefs where the driving time is over three and a half hours.

We observe that in duty schedule 48 only 2 duties had a scheduled continuous driving time of over 4h 5m. In schedule 49, this number has increased to 10, and 4 of these duties have a driving time of over 4h 10m. These drivers are almost certain to violate their parameters in the event of even a slight delay. Also, a number of these duties have larger slack values, which means that the slack cannot be used fully because doing so would violate the driver’s parameters. It must be noted that while in theory

<table>
<thead>
<tr>
<th>Duty Schedule</th>
<th>48</th>
<th>49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of reliefs</td>
<td>469</td>
<td>463</td>
</tr>
</tbody>
</table>

**By DIRECTION**
- AGR/AGR-EB/ACT-WB | 175 | 277 |
- AGR-WB/ACT-EB | 212 | 105 |
- Ending in depots/sidings | 82 | 81 |

**By TYPE**
- End of duty | 229 | 229 |
- Meal break | 202 | 201 |
- Intra-spell | 38 | 33 |

<table>
<thead>
<tr>
<th>End of duty reliefs by DIRECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGR/AGR-EB/ACT-WB</td>
</tr>
<tr>
<td>AGR-WB/ACT-EB</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Meal reliefs by DIRECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGR/AGR-EB/ACT-WB</td>
</tr>
<tr>
<td>AGR-WB/ACT-EB</td>
</tr>
</tbody>
</table>

Table 3.2: Comparison of Relief Directions
drivers are entitled to stop their trains if they violate their parameters, the reality is different. Most drivers are usually cooperative as long as they know that service controllers are actively trying to put them back on time.

This increase in driving times was necessary to achieve the objective of improving the relief direction mix in schedule 49. It is with trade-offs such as this that simulation comes in handy. It is not immediately obvious if this is a good trade-off to make, and simulation would allow such a trade-off to be evaluated without actually putting in into operation.
3.2.3 Meal reliefs

Meal reliefs are very important to analyse because they represent a potential missed connection. When a train arrives late for a meal relief, not only is the incoming driver late, the incoming driver has to drive another train after the meal break, and could potentially be late for that pickup.

In this context, the meal break slack is defined as the lesser of: (i) the time between when an operator gets off the train and when the meal break officially starts, less walking time; and (ii) the time required for the driver to hit the 4h15m driving time parameter. The meal break slack allows an operator to be late for a relief, and yet not miss the next train, while ensuring that the 4h15m driving parameter is not violated. It should be noted that in the majority (close to 95%) of cases, it is criterion (i) that is binding, i.e. drivers are far more likely to be late for their next train than violate their 4h15m parameter. The exceptions are duties with exceptionally high driving times as highlighted in the previous section. Therefore, it is necessary to view meal break slacks in the context of train lateness at the relief location at that time. Figures 3-9-3-12 show the meal break slack for every relief at each relief location/direction. The 95th percentile of train lateness at that location at that time of day is superimposed on this.

The X-axis represents the time when the operator is scheduled to get off the train. The lateness at each location was calculated from the March-May 2012 NetMIS data for that location. The day is divided into 15 minute intervals, and for reliefs that are scheduled in each 15-minute interval a distribution of lateness is constructed, and the 95th percentile of this distribution is used. There are a number of sharp kinks in the lateness graphs because of the poor quality of the NetMIS data. Table 3.3 summarizes some key statistics about the meal break slacks:

<table>
<thead>
<tr>
<th></th>
<th>48</th>
<th>49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average meal break slack (minutes)</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Min. meal break slack (minutes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- Considering the 4h15m parameter</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>- Not considering the 4h15m parameter</td>
<td>11</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 3.3: Comparison of Meal break Slack time

It is interesting to see that both the minimum and average slack times have decreased. This is partly the result of having low slacks when lateness is low, and high slacks when lateness is high, rather than set a uniform minimum for the whole day. But there are still some reliefs that are below the 95th percentile of lateness, such as the evening time reliefs at ACP-EB or late morning reliefs at AGR-EB. The simulation
Figure 3-9: Meal break slack vs Lateness at ACT-EB

Figure 3-10: Meal break slack vs Lateness at ACT-WB
Figure 3-11: Meal break slack vs Lateness at AGR-EB

Figure 3-12: Meal break slack vs Lateness at AGR-WB
model would enable us to find “critical” reliefs which are missed more often than others.

It is interesting to note that there are many duties with very high slack times of more than thirty minutes. That much of slack time is unlikely to be used as service controllers would not let a train arrive 30 minutes late merely because the driver has enough slack time; they actively work to put trains back on time. Therefore, a crew schedule that had a tighter distribution of slack time with the same average slack time but a higher minimum (and consequently, a lower maximum) slack time would potentially be more robust. Again, if such a hypothetical crew schedule were created, its performance could be tested through simulation.

3.2.4 Intermediate reliefs

This refers to reliefs that happen in between a spell that has two train runs. A majority of these are step-backs at AGR, which are not traditionally considered reliefs. However, every step-back represents a potential missed connection in that a driver may arrive late at AGR while the train that he/she is scheduled to step back into is on time. This is especially possible on the Piccadilly Line if there is a delay on one of the branches which only delays trains from that branch.

There is a slack time associated with these reliefs which is simply defined as the time between the arrival of the incoming train and the departure of the following train. This does not account for the minimal amount of walking that may be required (such as between platforms). Table 3.4 summarizes some key information about intermediate reliefs:

<table>
<thead>
<tr>
<th></th>
<th>48</th>
<th>49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of intermediate reliefs</td>
<td>38</td>
<td>33</td>
</tr>
<tr>
<td>Average slack (minutes)</td>
<td>15</td>
<td>17</td>
</tr>
<tr>
<td>- Step backs (minutes)</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>- Other intermediate reliefs (minute)</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Minimum slack (minutes)</td>
<td>8</td>
<td>9</td>
</tr>
</tbody>
</table>

Table 3.4: Comparison of intermediate reliefs

The number of intermediate reliefs has decreased which leads to simpler operations. It is notable that the average slack time for step-backs is ten minutes, which is essentially the time between successive AGR-reversing trains. In contrast, other intermediate
slacks have an average slack of twenty five minutes.

### 3.2.5 End of Duty (EoD) reliefs

This refers to reliefs where the driver finishes his duty after being relieved from his train. EoD reliefs are different from meal reliefs for a number of reasons. Firstly, the issue of a driver being late for the next train does not arise in the case of EoD reliefs, so driver lateness is less of an issue. The more important issue here is potential driver refusal to go past, say AGR-EB, to CFS and back to AGR-WB for an EoD relief when the train is running 20 minutes late. This is in fact the primary reason for preferring AGR-EB and ACT-WB reliefs, and this is modelled in the simulation. Secondly, there are two parameters to be concerned with here: the 4h15m driving time parameter, and the 8h duty length (excluding meal break) parameter.

Therefore, the EoD slack is defined as the minimum of: (i) the late running margin (i.e. the time between the relief and end of duty, less travel/walking time, stabling time, check off time etc.) and (ii) the time needed to hit the closer of the 4h15m driving time or 8h duty length parameter. Table 3.5 summarizes some key information about EoD reliefs:

<table>
<thead>
<tr>
<th></th>
<th>48</th>
<th>49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of EoD reliefs</td>
<td>229</td>
<td>229</td>
</tr>
<tr>
<td>Average slack (minutes)</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Reliefs with unusable slack</td>
<td>50</td>
<td>43</td>
</tr>
</tbody>
</table>

Table 3.5: Comparison of EoD reliefs

Both the average and minimum slacks are far lower for EoD reliefs compared to meal relief which makes sense given the difference in their impacts on reliability. In fact there are a number of EoD reliefs with no slack time at all.

Figure 3-13 shows the amount of late running margin that EoD reliefs have versus the amount of time required to hit a parameter. There are reliefs where the late running margin at the end of the duty might not be used fully because a driver might hit one of their parameters before being able to use the margin fully. Such duties represent potentially wasted slack, and their number has decreased from 50 in schedule 48 to 43 in schedule 49. They constitute the duties below the dark line in figure 3-13.
3.2.6 Duty length

Figure 3-14 shows the cumulative distribution of duty lengths in the two crew schedules. The two schedules have very similar distributions of duty length. What is interesting, however, is the fact that close to 40% of the duties have a duty length less than seven hours, with many duties that are just a little over four hours long, when in fact the maximum permissible duty length is eight hours (excluding a thirty minute meal break). The mean duty length is about six hours and forty five minutes.
The reason for this varied distribution is two-fold:

Firstly, there is a thirty six hour limit on the weekly driving time for a driver, which over a five day week, translates to an average of seven hours and twelve minutes. As many drivers do not prefer to work on weekends, the schedulers deliberately have longer than average duties on weekends so that fewer drivers are scheduled on weekends. Therefore, the average weekday duty length is less than 7h 12m, and is in fact equal to 6h 45m.

Secondly, Section 2.3.2 mentioned the fixed-link constraint at Arnos Grove which divided the day into six periods. Each period has its own roster, and each of these rosters has to maintain an average weekly duty length of thirty six hours. The problem with this is that, if the scheduling software were left to its own devices, very early duties and very late duties would be shorter than other duties. But because of the fixed-link constraint, these duties also have to have the same average length as other duties. The scheduler achieves this by forcing other duties to be short, and making these early/late duties longer by padding them with extra slack time. This is one of the reasons for this varied distribution of duty length. Another side-effect of this fixed-link constraint is that slack time is often allocated to duties to make them longer, when that slack time could be more usefully deployed elsewhere.

Figure 3-14: Cumulative distribution of duty lengths
3.2.7 Conclusion

This first part of the chapter qualitatively described the crew schedule and the roster. The second part of the chapter delved deeper into the crew schedule, by looking at individual elements of the crew schedule that impact robustness and comparing the distribution of those elements across two different crew schedules.

Some of these elements include cover duties, relief mix, slack time, number of two-run spells and continuous driving time. Given finite resources it is not possible to make improvements in every one of these areas. In moving from schedule 48 to 49, there was an overarching goal to improve the mix of relief directions. At the same time, the number of two-run spells also decreased, which should be good for robustness. However, the tradeoff was a slight increase in the number of long duties and a slight decrease in the average amount of slack time.

It is not possible to say if this tradeoff is good or bad, or if the change should have been more or less drastic, by merely looking at the structure of the crew schedule as we have in this chapter. This is because there are a number of complex interactions that come together to determine the performance of a schedule and it is not possible to capture all of these interactions by merely looking at the structure of the schedule as we have done here. This is the primary motivation for the simulation approach that is described in the next chapter.

Using simulation, it is possible to evaluate these, and other hypothetical tradeoffs and identify best practices in scheduling. In chapter 5, the simulation approach is used to perform a comparative evaluation of the performance of the same two crew schedules (48 and 49), and in doing so, we will refer to many of the structural elements (such as slack time, relief directions etc.) that were highlighted in this chapter.
Chapter 4

Simulation-based Framework for Evaluating Crew Schedules

4.1 Introduction

The previous chapter attempted to evaluate the robustness of a crew schedule by analyzing its structure. There are various characteristics of the crew schedule that make it robust, for instance, the presence of slack time and spare operators. However, there are also other characteristics whose effect on robustness is less clear. For instance, most rail transit lines have a number of locations where crew reliefs may take place. While it is generally acknowledged that it is necessary to have a mix of various kinds of reliefs both kinds of reliefs for operational flexibility, it is not clear what the optimal mix of reliefs is. Even in cases where the robustness benefits are obvious, such as the addition of slack time and spare operators, the utility of the additional resources is unknown. For instance, it is possible that adding two duties will result in a significant increase in reliability, whereas subsequent additions will have decreasing marginal returns.

The primary problem with this structural approach to evaluating robustness is the difficulty in relating the elements of the crew schedule to the actual performance that is observed under disrupted conditions. While it is possible to make general statements such as “more slack is better” it is much harder to isolate the impact of a specific element of robustness by just looking at the structure of the crew schedule, because of the inter-dependence of the different elements defining the crew schedule. For instance, it is possible that a particular relief has very little slack time associated with it. But that relief may offer many options for substitution, and therefore, if that train is late arriving for that relief, it can easily be renumbered. The different elements of the crew schedule such as slack, relief location, relief time and spare operator availability interact with the actions of service controllers in the context of commonly observed disruptions to produce the performance that is observed over a
period of time. Therefore, it is difficult to isolate the effect of an element of the crew schedule just by examining the structural differences between different crew schedules.

Secondly, even if we were able to completely identify the elements that have positive and negative impacts on the performance of the crew schedule, it is difficult to quantify their impact. However, some kind of quantification is needed for taking decisions on whether to increase or decrease manpower, or whether to negotiate labour agreements. In order to make these kinds of decisions, it is necessary to predict the impact of those changes on the performance. Similarly, if it were possible to predict the impact of varying certain elements of the crew schedule, say relief directions, then it would be possible to generate and evaluate multiple crew schedules with minor variations of those elements and thereby identify best practices for robust crew scheduling. Simulation is a commonly used tool to build predictive models because it naturally captures interactions, it can easily use empirical/observed data, and it is adaptable.

Depending on the level of detail of the simulation, it can capture many levels of interaction. Simulation models store the state of every object in the system, and capture the reactions of each object to the state of the system, which includes every other object. Vehicular micro-simulation models, for instance, capture the physical interaction between vehicles that are close to one another. TSM (London Underground Ltd., 2006) is an example of a micro-simulation model that is employed by the London Underground to analyse the capacity of their underground lines by performing detailed simulations of train movements (including acceleration and deceleration), loads and dwell times at stations. It is also possible to build models that simulate disruptions and recovery strategies. While it may be possible to write analytical expressions that capture some of these effects, it is difficult to retain mathematical tractability while also capturing detail.

Simulation models go hand in hand with data-driven analysis. They allow the use of large datasets directly without questionable assumptions on their distributions, unlike analytical models which are based on theoretical distributions. Most modern transit systems use some combination of AVL (Automatic Vehicle Location or train tracking), AFC (Automatic Fare Collection) and APC (Automatic Passenger Count) data which provide an excellent source of disaggregate data in large amounts. It may be difficult to fit these data to theoretical distributions with reasonable goodness of fit. With simulation, however, empirical distributions can be based directly on observations, and they can take on any shape. The tradeoff with using empirical distributions is greater difficulty in interpreting the results and the need for larger sample sizes. It is generally easier to identify the cause of changes with analytical models; this is more difficult with arbitrary distributions and with simulation. Moreover, a large sample size should be used to build empirical distributions if the model is to capture the effects of low-probability events.
Adaptability refers to the relative ease with which we can make changes to the model. For instance, we may wish to test different service recovery strategies. Doing so is easier with a simulation model than an analytical model because the behavior of the simulation can be defined in modules. For instance, there may be a module for recovery in the absence of decision-support systems, and another for recovery in the presence of decision-support systems, and we can choose which module to use in a particular simulation experiment. There is no dependent analytical expression that loses validity in response to which recovery strategy is in place.

Clearly, simulation is a powerful tool. However, there is a need to balance complexity with necessity. While developing a simulation model, it is necessary to keep in mind the ultimate purpose of the model, and capture a level of detail that is sufficient for that purpose. It is necessary to remember the cliché that a model must be made as simple as necessary (or as complex as necessary), but no more. For instance, a model for the pedestrian flow in a busy transit interchange station might find it necessary to simulate the movement of individuals, but the same might be unnecessary for a simulation of the capacity of the transit lines at that station. The more complex a model, the more difficult it is to calibrate, validate, and the longer it takes to run. Therefore, the level of detail of a simulation model must be tailored to the problem at hand.

The rest of the chapter describes a simulation model of the daily operations of a high-frequency rail transit line that is used to evaluate the robustness of the line's crew schedule. While the model has been developed with the London Underground in mind, and specifically the Piccadilly Line, it has consciously been developed in a manner that allows it to be adapted to other contexts. Section 4.2 discusses the general framework that is used for the simulation-based evaluation of a schedule. Section 4.3 describes the details of the simulation model, including the methods, inputs and outputs. The evaluation framework and simulation model are general and can be applied to any rail transit service. Section 4.4 details the implementation of the model for the Piccadilly Line, and discusses the running time models, dwell time models and service control strategies for the Piccadilly Line.

4.2 Framework for simulation-based evaluation of robustness

This section introduces the basic framework for the simulation-based evaluation of the robustness of a crew schedule. Qualitatively, robustness is the resilience of a schedule to disruptions - the ability to maintain operational feasibility in the event of minor incidents, and the ability to recover easily from major incidents. The approach that is taken to measure the robustness of a schedule, is to simulate the operations of the line with that schedule under different kinds of operating conditions (normal, and varying
degrees of disruption), and to measure the performance under those conditions.

The simulation model simulates a day's operations of the transit line given the route structure, operating plan (timetable and crew schedule), observed running time and dwell time distributions, and service control strategies. The above elements comprise the "normal" operating conditions of the line. In addition, the model also allows the simulation of incidents, the consequent disruption, and the operations under those disrupted conditions.

During the course of a schedule's life (typically a year at the London Underground), it is operated under a wide range of conditions. There will be days when nothing goes wrong and the operations plan is closely followed, days when there are minor incidents (such as a passenger emergency alarm or a defective train) that cause slight deviations from the operations plan, and yet other days when severe incidents (such as a signal or track failure, or a fatal accident) causes widespread disruption.

Therefore, the crew schedule must also be simulated under a variety of conditions that are representative of real operating conditions in order to evaluate its robustness. But in doing so it is necessary to exercise some judgment about the severity of the disruptions that we wish to simulate. While it is no doubt good if a schedule is resilient to even the gravest of incidents, such a schedule would be needlessly expensive to operate, because it would have to incorporate massive amounts of slack resources to achieve that level of resilience. Therefore, a schedule can only be designed to be resilient to a certain range of incidents and in evaluating its robustness, it is sufficient to simulate that range.

Therefore, the approach that is taken in this thesis is to simulate the operations of the transit line for a reasonable range of incidents, and we discuss later what that means. Each run of the simulation gives us the performance of the line for that day under those specific conditions. After having simulated the entire range of operating conditions, we have performance measures for multiple days of operations, with each day representing a specific operating condition. These daily performance measures are then aggregated to evaluate the performance of the crew schedule over the full range of reasonable disruptions, which is ultimately a measure of robustness. Figure 4-1 presents this framework graphically.

It is clear that it is possible to evaluate the performance of not just an existing crew schedule, but a hypothetical crew schedule that, say, relaxes a key labour constraint. That would allow the user to predict the impact of that change by examining the performance of the new crew schedule compared with the baseline case. It would also be possible to create multiple different crew schedules with, for instance, varying mixes of relief directions to try and find the best combination. Similar exercises in system-
Performance of the crew schedule over all days, i.e. over the full range of reasonable disruptions

Figure 4-1: Framework for simulation-based evaluation of robustness

atically varying other elements of the crew schedule would allow the identification of best practices in robust crew scheduling.

4.3 Simulation architecture

This section introduces the framework for the simulation and describes the inputs, methods and outputs of the simulation. While the simulation model was developed with the Piccadilly Line as the intended application, the framework is presented in more general terms which allows it to also be applied to other transit lines. Occasionally, examples taken from the Piccadilly Line are used to illustrate finer details, or to state principles in a less abstract and more concrete fashion.
4.3.1 Simulation inputs

The following are the inputs to the simulation listed in Figure 4-1:

- **Route structure:** The route structure is a collection of locations that are connected to one other in a certain way. It is the network on which the simulation takes place, i.e. trains only move between the locations identified in the route structure. The route structure can simply be the network of the transit line being studied, or it could be simplified. For example, as will be explained later, the route structure of the Piccadilly Line that is used in the simulation is a simplification of the actual network where only crew bases and termini are included. While simplifying the route structure care must be taken to ensure that all train movements can be represented, and all important details are captured.

- **Timetable:** The timetable describes the movement of trains or rolling stock over the route. For each train, it lays out the locations where it is expected to travel and the time when it is expected to be at that location.

- **Crew schedule:** It is an assignment of drivers to trains that ensures that every train movement is manned by a driver. It is subject to crew constraints as described in section 2.x.

- **Running time & dwell time distributions:** A running time distribution is specified for each segment, i.e. for every pair of adjacent locations in the route structure. The running time distribution models the variability in the running time of trains, in the absence of disruptions. This variability is caused by a number of factors: each driver drives differently, even the same driver does not always drive in exactly the same manner, and trains sometimes hit amber or red signals which slow or stop the train temporarily. If the segment in question has stations within it, then the running time also includes the dwell times at those intermediate stations, which clearly have their own variability.

  Dwell time distributions are specified for each station in the route structure. They describe the dwell time for the train whenever it arrives at a location in the route structure. Please note that the route structure may not contain all the stations in the real network (as is the case with the model in this thesis), in which case the dwell time at stations that are not explicitly included in the route structure would be incorporated into the running times of the segments of which they are a part. Dwell times vary due to a number of factors such as varying passenger alightings and boardings, driver differences etc.

  These distributions can be analytical or empirical, univariate or multivariate or conditional on some system state variable. Because of the availability of large amounts of AVL data at most modern transit agencies, these distributions can be easily estimated.

- **Service control strategies:** As seen above, there is clearly variability in the movement of trains. This variability might be the result of normal processes,
or due to incidents, which lead to disruptions. Hence, the operations plan as laid out by the timetable and the crew schedule is not always followed. In such cases, it sometimes becomes necessary for the controller to intervene in real-time and override the operations plan in order to meet some passenger objectives (such as headway control) or crew objectives or constraints (such as reducing driver lateness to avoid parameter violations). It is also possible that the deviation from the operations plan will result in operational infeasibility. For instance, a driver who is on a train that is 20 minutes late might be required to pick up another train, and might not be able to make it to that pickup on time. In such cases, real-time intervention is necessary to maintain operational feasibility. These real-time interventions are effected by service controllers who act to meet those objectives and constraints. A simulation of train movements, especially during disrupted conditions, must necessarily include control actions, and this is implemented in the form of service control strategies which are clearly defined rules that take into account the current system state, and modify the operations plan in a manner that ensures feasibility while meeting any passenger or operational objectives.

- **Range of incidents:** This describes the distributions of incidents that are to be simulated. For the purpose of this simulator, an incident is defined as an event that stops train movements at a location, at a certain time, and for a certain duration. Section 4.2 contained a brief discussion of the types of incidents that are to be simulated, and this will be followed up in Chapter 5. For those incidents that are to be simulated, it is necessary to know the distributions of their frequency, location, time of occurrence and duration. This model will allow more than one incident may take place per day or replication.

Section 4.4 discusses how the inputs were obtained for the specific case of the Piccadilly Line.

### 4.3.2 Simulation methodology

The previous section explained the inputs needed by the simulation. In this section, the simulation methodology is examined. Figure 4-2 presents this simulation methodology in the form of a flowchart.

**Preprocessing**

As explained in the discussion about the route structure in the previous section, the route structure may contain only a subset of all stations on the line. In that case, it is necessary to extract the information that is relevant to the route structure from the timetable and the crew schedule. In other words, trains now move between the locations on the route structure, and not the original locations on the line, and the
Preprocessing:
- Adjust timetable and crew schedule to conform to the route structure
- Create an initial set of events corresponding to the projected arrival of every train at a location in the route structure (after pulling out of the depot/siding) and add these events to a queue

Event Heap
(ordered by Event time)

Event Type

Simulating Train Movement:
While the heap of is not empty, extract & process the topmost event

Performance Measurement:
- Calculate performance measures such as headways, train lateness, trip cancellations, short-turns, spare operator utilization etc.

Figure 4-2: Simulation Methodology
timetable and crew schedule must reflect this reality. If the route structure is chosen judiciously, this should not result in a loss of correctness or detail.

The simulation is event-driven, and as shown in Figure 4-2 there are four types of events: Projected Arrival, Actual Arrival, Projected Departure and Actual Departure. These events are now discussed in detail.

**Projected Arrival Event**

This event is created at the time of departure of a train from a station, and corresponds to the projected time of arrival of that train at its next station. A train may not actually arrive at the station at the projected time of arrival for a number of reasons. The platform(s) that the train is supposed to use may be occupied and there may be other trains in queue to use that (those) platform(s). Or the projected arrival time may be too close to the departure time of the previous train from that platform, resulting in a violation of the minimum time separation requirements between trains. These occurrences cannot be predicted when the train departs from the previous station, and therefore, a distinction is made between the projected arrival and the actual arrival of a train at a station. Figure 4-3 describes the handling of a Projected Arrival event.

When processing a Projected Arrival event, the first step is to check if the platform(s) that can be used by that train is (are) available. To perform this check, an arrival queue is maintained. Trains are entered in this queue in chronological order of their original projected arrival time at that station. Therefore, the arrival queue represents the order in which trains queue up in the approach to the station.

In the case of a simple through station with only one platform for available for through movement in each direction, each platform has one arrival queue. In the case of a terminus with multiple platforms where trains reverse, it would make sense to maintain a single arrival queue for all the platforms. This is because trains waiting for a platform outside a terminus are normally assigned to the first platform that becomes available. In the case of a station where two branches merge, the choice of arrival queue configuration depends on how the merger occurs. If the merger occurs before the station, and the merged lines use a common set of platforms (as is the case with Acton Town, for instance), we would maintain a common arrival queue for trains from both branches. On the other hand, if the merger happens after the station, and both branches have their own dedicated platforms, then separate arrival queues would have to be maintained for the two branches.

In this simulation, the trains in the arrival queue are processed in a first-come-first-serve basis, though this does not always have to be the case. For instance, in the case
Processing a Projected Arrival type Event

1. Projected Arrival

Is the target platform(s) free?

YES

The train actually arrives at the station at the projected time.
Create ACTUAL ARRIVAL event
Add event to heap

NO

The train does not actually arrive at the station at the projected time.
Calculate new projected arrival time based on station arrival queue
Create PROJECTED ARRIVAL event
Add event to heap

Figure 4-3: Processing a Projected Arrival event
of a station such as Acton Town where two branches merger just before the station, it is possible that one branch is accorded higher priority over the other.

The arrival queue is used to determine if the target platform(s) is available. If it is available, then the train is said to actually arrive at the projected arrival time. Therefore, an Actual Arrival event is created for that train with the projected arrival time being the event time, and this event is added to the event queue to be processed next.

If the platform(s) is (are) not available, then the train does not arrive at the projected arrival time. In that case, a new projected arrival time is estimated, and a new Projected Arrival event is created for the train. To calculate the new projected arrival time the following procedure is employed:

- Examine the arrival queue.
- If there are no trains ahead of this train that have not yet arrived, then the new projected arrival time is estimated as the departure time of the last train plus the minimum separation $h_{\text{min}}$ of sixty seconds.
- If there are trains ahead of this train in the arrival queue that have not yet arrived, then the new projected arrival time is estimated as the projected arrival time of the last such train plus the minimum separation $h_{\text{min}}$ of sixty seconds. This is a conservative estimate of the projected arrival time in that the actual arrival can never take place before this.

**Actual Arrival Event**

This event corresponds to the Actual Arrival of a train at a station. Figure 4-4 describes the steps involved in processing this type of event.

First, the trains is removed from the station arrival queue. The operations plan is used to decide where (i.e. which of the adjacent stations on the inputted route structure) the train will go to next, and which driver will operate the train. The operations plan could either be the scheduled operations plan, or one that has been modified in real-time by the service control logic. The train’s next location could be the depot in case the train is being stabled.

The departure time function is used to determine the projected departure time of the train. This function, which is described later in this section, determines the departure time based on factors such as the dwell time distribution at that station, the type of move (reversing or through) and the occurrence of an incident at that station. Corresponding to the estimated departure time, a Projected Departure event is created
Processing an Actual Arrival type Event

- Event type
- Event time
- Train#
- Station → 2 Actual Arrival

Remove train from station arrival queue

Determine Next Location

Determine Projected Departure Time

Create PROJECTED DEPARTURE event and add to heap

Add train to station departure queue

Current Operations Plan (as modified by service control)

Departure Time Function (depends on dwell times, type of move: through or reversing, incidents)

Figure 4-4: Processing an Actual Arrival event
for that train and added to the event queue.

**Projected Departure Event**

This type of event is spawned by an Actual Arrival event at a station. A train may not be able to depart at the projected departure time because the projected departure time may be too close to the departure of the previous train, violating the minimum separation requirement. Figure 4-5 describes the steps involved in processing a Projected Departure event.

The first step is to check if the train is able to depart at the projected departure time. To perform this check, we maintain a departure queue for every platform (or set of platforms, in case multiple platforms feed into the same track, as in the case of a terminus). This departure queue represents demands from trains to depart from that platform(s). Trains are ordered in the departure queue on the basis of their original projected departure times, and are served in a first-come-first-served basis.

A train is able to depart from the station at the projected departure time if there are no trains ahead of that train in the departure queue, and if the previous train to depart departed at least sixty seconds ($h_{min}$) ago. In that case, an Actual Departure event is created corresponding to the projected departure time, and is added to the event queue to be processed next.

If the train is unable to depart at the projected time, an updated departure time is estimated, and a corresponding Projected Departure event is created. The new projected departure time is estimated as follows:

- If there are no trains ahead of this train in the departure queue, then the new projected departure time is estimated as the actual departure time of the previous train plus the minimum separation ($h_{min}$) of sixty seconds.
- If there are trains ahead of this train in the departure queue, then the new projected departure time is estimated as the projected departure time of the last such train plus the minimum separation ($h_{min}$) of sixty seconds.

**Actual Departure event**

This event corresponds to the actual departure of a train from a station. Figure 4-6 describes the steps involved in processing this type of event.
Processing a Projected Departure type Event

Event type
Event time
Train#
Station

3 Projected Departure

Is the train able to depart?
(i.e. have there been no other departures in that direction in the last $h_{\text{min}}$ seconds?)

YES

The train actually departs from the station at the projected time.

Create ACTUAL DEPARTURE event
Add event to heap

NO

The train does not actually depart from the station at the projected time.

Calculate the new projected departure time based on the station departure queue
Create PROJECTED DEPARTURE event
Add event to heap

Figure 4-5: Processing a Projected Departure event
Processing an Actual Departure type Event

- Event type
- Event time
- Train#
- Station

4 Actual Departure

Remove train from station departure queue

Determine Projected Arrival Time at next station

Arrival Time Function (depends on running times, incidents)

Create PROJECTED ARRIVAL event and add to heap

Add train to next station arrival queue

Figure 4-6: Processing an Actual Departure event
The first step is to remove the train from the station departure queue. The train's next location has already been determined while processing the train's arrival at that station. The projected arrival time at the next station is calculated using the Arrival Time function, which will be described later in the section. This function calculates the arrival time at the next station, given the departure time from the previous station based on the distribution of running times between the two stations and the presence of disruptions. A Projected Arrival event is then created corresponding to the projected arrival time of this train at the next station, and this event is added to the event queue.

**Processing events**

The four types of events used in this simulation have been described. Events are stored in a queue chronologically and processed accordingly. At the start of the simulation, a Projected Arrival event is added for each train corresponding to its first arrival at a station in the route. A Projected Arrival event spawns an Actual Arrival event, which spawns a Projected Departure event and in turn an Actual Departure event. When a train arrives at its last station for the day, no more events are created for that train. The event queue is processed till it is empty, at which point every train has been stabled, and the operations end.

This way, an entire day's operations are simulated from its start (around 5am in the case of the Piccadilly Line) to its close (at around 1.30am on the following day in the case of the Piccadilly Line). We do not have to deal with the issue of specifying boundary conditions. All trains start and end at the depots, and all drivers start and end their duties at their crew bases.

**4.3.3 Departure time function**

The departure time function is called whenever an event (i.e train arrival) is being processed and it determines when the projected departure time of the train from that station. The departure time function clearly depends on the distribution of dwell times for that location as well as the type of location. A location could either be a through-station, or a terminus (i.e. reversing point) or a depot.

Clearly, the departure time function will be very different for each of these locations. For a through station, the dwell time is determined by passenger alightings and boardings (which vary by time of day). At a terminus or reversing location, the departure time might depend on the train's lateness. A train with lateness greater than the recovery time might not be held beyond the minimum reversing time, whereas a train that is not as late might be held till its scheduled departure time, or held to fulfill
some headway requirement. The departure time from a depot is usually the scheduled departure time.

Furthermore, it is possible that an incident occurs when the train is at the station and this causes a delay in the train’s departure. In that case, the departure time function needs to account for this additional delay. The structure of the departure time function is discussed below, in the absence and presence of an incident. Section 4.4.3 contains more details on dwell time models for the Piccadilly Line.

### In the absence of an incident

The projected departure time from a through station in the absence of an incident is taken to be:

\[ PDT = AAT + DwT \]  \( (4.1) \)

where \( PDT \) is the projected departure time from that station, \( AAT \) is the actual arrival time at that station, \( DwT \) is a realization from the dwell time distribution for that station for that time of day.

The projected departure time for a reversing station in the absence of an incident is taken to be:

\[ PDT = \max \left\{ AAT + RevT_{\text{min}}, SDT \right\} \]  \( (4.2) \)

where, \( RevT_{\text{min}} \) is the minimum reversing time at that station, and \( SDT \) is the scheduled departure time. This ensures that the train does not depart before its scheduled departure time.

### In the presence of an incident

When an incident occurs at a through station, it is assumed that the projected departure time is pushed back by an amount of time equal to the duration of the incident. Therefore the projected departure time is calculated as:

\[ PDT = AAT + DwT + INCIDENT\_DURATION \]  \( (4.3) \)
In the case of a reversing station, the expression used is:

\[
PDT = \max \left\{ \frac{AAT + RevT_{\text{min}} + INCIDENT\_DURATION}{SDT} \right\}
\]

Therefore, the incident may or may not cause a delay in the projected departure time. If the train is ready to leave before the scheduled departure time, even after accounting for the incident and the time taken to reverse, then the projected departure time is equal to the scheduled departure time. Otherwise, the train is projected to leave as soon as it is ready to leave.

### 4.3.4 Arrival time function

The arrival time function clearly depends on the running time distributions for that segment. It needs to ensure that trains do not overtake each other, and in fact, maintain a minimum separation between each other. The running time distributions that are inputted account for the normal variability in running time. In addition, the arrival time function also accounts for any additional running time due to incidents. The arrival time function is first discussed for undisrupted conditions and then for disrupted conditions.

#### Undisrupted conditions

The simplest form of the arrival function would be to simply draw a value of the running time from the running time distribution for that segment and for that departure time, and to add that running time to the departure time. Expressed mathematically:

\[
AT^n_b = DT^n_a + (RT_{ab}|DT^n_a)
\]

where \(AT^n_b\) is the arrival time of train \(n\) at location \(b\), \(DT^n_a\) is the departure time of train \(n\) at location \(a\), and \((RT_{ab}|DT^n_a)\) is a sample from the running distribution of segment \(ab\), conditional on the departure time \(DT^n_a\) from \(a\). However, there are two important shortcomings of this model: (i) it does not ensure that a minimum separation is maintained between successive trains - in fact, it does not even prevent overtaking, and (ii) it does not capture any correlation between the running times of successive trains. The first point is clearly an issue. The second point may or may not be an issue depending on the context. In the case of a high-frequency rail service such as the Piccadilly Line, there is likely to be significant interaction between successive trains as shown below.
Figure 4-7: Scatter plot of running times of successive trains (ACT-AGR)

Figure 4-8: Scatter plot of running time deviations of successive trains (ACT-AGR)
Figure 4-7 is a scatter plot showing the running times of pairs of successive trains between ACT and AGR. The Y-axis denotes the lead train, and the X-axis denotes the following train, and the red line is the OLS line. It is clear that the two are strongly correlated, and indeed this is borne out by a correlation coefficient of 0.829. This is expected for two reasons: (i) successive trains have the same or similar scheduled running times, and (ii) when a train is delayed/early, the train behind it tends to respond in a similar fashion. Indeed, the second observation is confirmed by Figure 4-8 which is a scatter plot showing the schedule runtime deviations (in seconds) of pairs of successive trains between ACT and AGR. Once again, the Y-axis denotes the leading train, and the X-axis denotes the following train, and the red line is the OLS line. The correlation in this case is 0.735.

Therefore, when we draw from the distribution of running time, we should not make independent draws. Rather, we need to generate the sample in such a way that this correlation is captured. The following procedure is employed to generated correlated random variables:

- Let $RT_{ab}(n-1)$ be the realized running time of the train $n - 1$ that precedes train $n$ on the current segment $ab$. We need to determine $RT_{ab}(n)$ in a way that reflects the correlation between the running times of the two trains.
- Let $(RT_{ab} | DT_a^n)$ be a random draw from the running time distribution of segment $a$, conditional on the departure time $DT_a^n$ of train $n$ from $a$.
- Set $RT_{ab}(n) = \rho_{ab} R_{T_{ab} | DT_a^n} (n-1) + \sqrt{1 - \rho_{ab}^2} (RT_{ab} | DT_a^n)$, where $\rho_{ab}$ is the correlation between the running times of successive trains on segment $ab$.

The above procedure ensures that generated running times of successive trains are correlated as desired. The simplicity of the procedure stems from the assumption that the random variable has a normal distribution. This assumption is not strictly true in the case of running times, because the distribution is asymmetric. However, for the sake of mathematical simplicity we make this assumption.

To ensure that a minimum time separation is maintained between successive trains, we force the arrival time of a train to be greater than or equal to the arrival time of the preceding train, plus the minimum separation $h_{min}$ of sixty seconds. Expressed mathematically:

$$AT_b^n = \max_{AT_b^{n-1}} \left\{ DT_a^n + RT_{ab}(n), AT_b^{n-1} + h_{min} \right\} \quad (4.5)$$

where $RT_{ab}(n)$ is calculated using the procedure described above. This formulation solves both of the drawbacks mentioned at the beginning the section.
Disrupted conditions

This refers to the case when an incident that specifically hinders train movement is introduced into the simulation. This introduces a delay in addition to the variability in running time that has been modelled previously. For the purpose of this simulation, an incident is defined as an event that occurs at a certain location and direction on the line at a certain time, and has the effect of preventing the movement of trains through that location in the disrupted direction for a certain duration of time. To model the effect of disruptions on train movements, we make the following assumptions:

- It is possible to unambiguously identify the train, say $n$, that is directly affected by the incident, i.e. all the trains ahead of this train are unaffected by the incident. This assumption may not always reflect what happens in practice as trains that escaped the incident might still be deliberately held by service control to reduce the gap in from the of the disrupted train.

- The arrival time of this train $AT^n_b$ at the next location $b$ (i.e. the end point of its current segment) is simply taken to be the current estimated arrival time, plus the duration of the incident. Therefore, when inputing the distribution of incidents it is necessary to understand that the duration of the incident is understood to be the duration for which train movement is disallowed. Again this assumption does not perfectly reflect reality because the disrupted train is likely to experience greater than average dwell times because of the gap in service. For the same reason, it is also less likely to be slowed by red or amber signals, and therefore might have shorter than average running times.

Under disrupted conditions, the arrival times of some of the trains right behind the disrupted train will be defined by the second expression in equation 4.5. In other words, as the disruption is cleared and the trains resume movement, the trains keep arriving (for some time) one after another separated by just the minimum headway $h_{\text{min}}$. This makes intuitive sense and is also what is observed.

Figure 4-9 shows the waterfall diagram during an incident that occurred on the central section of the Piccadilly Line at around 6PM on 10 October 2012. The red oval shows the train that was affected by the disruption, i.e. the train either suffered the disruption directly (as in the case of a passenger alarm), or was just behind the disruption (as in the case of a faulty signal). This train remained stationary for nearly 20 minutes. The trains ahead of this train are unaffected by the disruption. The trains behind this train are closely bunched together. Once movement resumes, the trains arrive one after another separated by just the minimum headway, as shows by the black oval. It is possible to model the minimum headway $h_{\text{min}}$ as a random variable with a distribution, say, between 60 and 90 seconds, to model the variability involved. But this variable is difficult to estimate, and this variation is quite small in comparison to the end-to-end running times on the line.
Figure 4.9: Waterfall diagram showing train movement around an incident.
4.3.5 Service control

The operations plan dictates the movements of trains and crew in a system. While the operations plan represents the ideal target state of the system, it is not always possible to stick to the operations plan due to routine variability and unexpected incidents. This can be a very serious issue, especially in the context of a precise high-frequency operations such as the Piccadilly Line, where any blockage of train movement can result in an immediate pile-up of trains. Depending on the severity of the disruption, trains may be running late, running out of order or not running at all. Consequently, the drivers operating those trains would also be unable to adhere to their crew schedule, and might be late for reliefs or changeovers. Therefore there is a need for active service control to ensure that train lateness is kept in check, and drivers are at the right place at the right time.

Service control interventions change the operating plan and these actions can take place at any time. Therefore, every time an event is processed, the service controller function is called. This function examines the current state of the system and makes changes to the operating plan based on the service control strategies that are programmed. The simulation model assumes that the appropriate crews are present at the right time and place to operate the service as per the operations plan. It is the responsibility of the service control logic to maintain the feasibility of the operations plan by making real-time changes to the operations plan as necessary, in accordance with the inputted service control strategies.

Section 2.3 provided a general introduction to service control – what it is, why it is needed, the types of interventions that are possible, and the overall strategy behind service control. Section 4.4.5 describes the specific service control policies that are implemented in the simulation model of the Piccadilly Line.

4.3.6 Simulation output: Performance evaluation

The outcome of one replication of the simulation is a set of observations on train and crew movements from which it is possible to calculate daily performance measures. These measures could be operations focused measures (measuring train or crew performance) or customer focused measures. For instance, it is possible to calculate the distribution of simulated running times for each segment, train lateness at crew change locations, percentage of scheduled kilometers operated, number of spare drivers used, headway variance, headway proxy etc. These performance measures are aggregated over many replications and are used both to validate the simulation, and to evaluate the robustness of the crew schedule.

If we were to simulate the performance of the system given the same incident (and
holding all other inputs also constant), i.e. if we were to run two replications of the simulation with the same incident, the performance measures are almost certainly going to be different. This is because the running times and dwell times are randomly drawn from distributions, and each realization of these random variables is likely to be different. Therefore, even for a given day with a given incident, it is advisable to run the simulation multiple times to increase the sample size of observations from which the performance measures are calculated.

4.4 Simulation model of the Piccadilly Line

One of the advantages of the simulation technique is the ability to use the large amounts of automated data that transit agencies generate to estimate the distribution of parameters empirically rather than trying to fit them to a theoretical distribution. The simulation framework described in section 4.2 requires the following inputs: route structure, timetable, crew schedule, running time distributions, dwell time distributions, service control policies and incident distributions. In this section, we discuss the specifics of the first six inputs in the context of the Piccadilly Line. Incident distributions are discussed in Chapter 5 along with the validation and application of the model.

4.4.1 Route structure

The route structure defines the basic geometry of the line. It is the collection of locations at which train and crew movements are modeled. An obvious choice of route structure for the Piccadilly Line is to use the real network with all 53 stations. However, it is important to bear in mind that the purpose of this simulation is to model train and crew movements so as to evaluate the robustness of crew schedules. The arrival and departure times at intermediate stations are not as important as the arrival and departure times at the crew change locations. Therefore, the route structure that is used in the simulation consists of only the crew change locations and termini. That gives us sufficient detail to model train movements, while allowing us to concentrate on the crew change locations. Figure 4-10 shows the reduced route structure of the Piccadilly Line that is used in the simulation.

Acton Town and Arnos Grove are the crew bases. Cockfosters, Uxbridge, Rayner’s Lane and the two Heathrow terminals are the termini, i.e. the locations where trains reverse. Arnos Grove also doubles as a terminus. This network does not include the depots or any of the intermediate stations. Because all the termini are included, the end-points of the trips in the model are the same as the end-points of trips in reality. Also, the inclusion of crew change locations in the route structure ensures that a crew change does not happen in the middle of a segment. A limitation of this route
structure is that it does not include other reversing locations that are used to turn trains around during emergencies (such as Hammersmith, King's Cross etc.) As we shall see later, this is not a serious limitation given the range of operating conditions that we are being tested here. Moreover, this can easily be incorporated into a future version of the model.

An incidental advantage of this model is that we need running time models for only 6 segments and dwell time models for only 7 stations, as opposed to running time models for 52 segments and dwell time models for 53 stations. The running times are also likely to be more accurate because having more long segments usually results in smaller errors than many short segments, because in the latter case the error is compounded more times. The disadvantage is that interaction between successive trains will not be modeled as accurately as in a micro-simulation.

4.4.2 Running time models

The arrival time function calculates the arrival time of a train at its next location given its departure time from its previous location and the state of the system at the time of departure. The arrival time function uses the running time distribution which is one of the input parameters to the simulation. The arrival time function itself is described in section 4.3.4. In this section, we discuss the running time distributions for the Piccadilly Line.

The running time can be derived in multiple ways: it is possible to construct a detailed micro-simulation model for each segment, it is possible to construct analytical expressions that predict running time based on system state variables (such as regression equations), or it is possible to estimate empirical distributions of segment running time based on AVL data and draw from that distribution. The latter approach is used here because it is simpler than performing a micro-simulation, while
also being data-driven and capturing more variability than an analytical expression.

Clearly, running times vary by time-of-day because dwell times at intermediate stations will be higher during the peak. Furthermore, there are more trains operating at peak times and there is a greater chance that a train’s movement is hindered by the train preceding it. Unlike buses where overtaking is possible, trains generally cannot overtake each other. Therefore, the arrival times of successive trains are not independent. And finally, there is no question that running times will look very different during a disruption. For all these reasons, it is not appropriate to just construct a distribution of running times and draw from it. We need to take a more nuanced view of running times.

As noted in section 4.3.1 the distribution of running times that the simulation needs are the running times in undisrupted conditions. To estimate undisrupted running time distributions, a panel of days with “excellent performance” were selected. These are eight days in 2012 (September 19, 25, 26 and October 2, 4, 18, 24, 29) which have Headway proxy scores greater than 98.5%. The headway proxy score is generated for every LU line at the end of a day’s operations, and is a measure of how many headways were within acceptable limits. The Piccadilly Line has a target headway score of 97%, and any day with a score greater than 98% is considered to have had excellent performance. Furthermore, none of these days had any major incidents, based on the service manager’s comments.

Having selected a panel of days without disruptions, the NetMIS (AVL) dataset for those days and for the stations in the reduced network is examined. The NetMIS dataset for each location (and direction - eastbound or westbound) has an entry for every train arrival/departure event at that station in that direction. Each entry records the train number, trip number, arrival time, and departure time. However, the NetMIS dataset for the Piccadilly Line is far from perfect and there are many entries that are missing or have junk values. For every location, NetMIS entries with incomplete or junk entries for any of the fields (arrival and departure time, train number or trip number) are excluded, which leaves around 80% of the entries.

We start off by looking at the distribution of running times from ACT (Acton Town) to AGR (Arnos Grove) eastbound. Since every NetMIS entry corresponds to an arrival/departure event at a station, it is necessary to associate the departure event of a train at ACT with the corresponding arrival event at AGR. This matching is done using the train and trip numbers. The actual running time of that train is then calculated as the difference between the arrival time at AGR and the departure time at ACT. Performing this calculation on all the NetMIS entries from the selected panel of eight days gives us 2,031 data points for running time between ACT and AGR.
Figure 4-11: Histogram of running times from Acton Town to Arnos Grove

Figure 4-11 shows a histogram of the running times on those eight days from ACT to AGR. The minimum running time is about 48 minutes, and most of them are less than one hour. The mode of the distribution is 53 minutes, and the average is 53.5 minutes. 95% of the runtimes lie between 50 and 58 minutes.

Figure 4-12 shows the variation in running time between ACT and AGR by time of day. The X-axis is departure time from ACT. The left Y-axis shows running time and the right Y-axis shows the standard deviation of the running time. Each red point is one observation from the panel of eight days. The black line shows the average running time for each thirty minute period (i.e. 5:45-6:15, 6:15-6:45, ..., 21:45-22:15), and the green line shows the standard deviation of running time for each thirty minute period. The blue line shows the scheduled running time for every trip.

We observe that:

- There is a significant scatter of observed running time (red points) around the scheduled running time (blue line) which shows the inherent variability in running time that is observed even on “good” days. This scatter is not symmetrical about the scheduled running time (blue line), nor is it consistent through the day.

- The average observed running time (black line) follows the same trends as the scheduled running time (blue line). However, the deviation from the schedule is not the same through the day. For instance, the lines are much closer during the mid-day period (11:00-14:00) than in the evening peak.

- The variability in the running time which is measured by the standard deviation of observed running times (green line) also shows significant variation through the day.
Figure 4.12: Actual and scheduled runtime (ACT-AGR) by time of day.
Given these observations, it might be possible to postulate an analytical expression for the running times. But one of the advantages of simulation and the fact that we have large amounts of AVL data is the ability to use empirical distributions, and that is the approach that is used here. During the simulation, the following procedure is followed to make a random draw from the running time distribution of a particular segment for a particular thirty minute period:

- Using a random number generator, draw from a uniform \([0,1]\) random variable. Let the drawn value be \(x\).
- The value of \(y = F^{-1}(x)\) needs to be calculated, where \(F\) is the CDF (cumulative density function) of the running time distribution under consideration.
- Let the running time distribution have \(N\) observations which are arranged in increasing order. Then the \(n\)th observation \((n = 1, 2, \ldots, N)\) has a percentile of \((n - 1)/(N - 1)\). The smallest observation is at the 0th percentile, and the largest value is at 100 percentile.
- If \(x\) is directly equal to one of the \(N\) percentile values calculated above, then \(y\) is equal to the corresponding value of running time.
- If \(x\) falls between two percentile values, say \(x_1\) and \(x_2\), then let \(y_1\) and \(y_2\) be the running time values corresponding to \(x_1\) and \(x_2\). Then we calculate the value of \(y\) by linearly interpolating between the two values. Mathematically,

\[
y = y_1 + (x - x_1) \frac{y_2 - y_1}{x_2 - x_1}
\]

Therefore, the running time distribution for a segment is composed of a set of ordered observations for every thirty minute period. This process is repeated for every segment, for both directions.

### 4.4.3 Dwell time models

The dwell time model determines how long a train stops at a location to unload and load passengers, and as such it is only relevant for through stations, namely Arnos Grove, Acton Town and Rayner’s Lane. At the termini, Cockfosters, Uxbridge, the Heathrow terminals (and Arnos Grove, for trains that reverse there), the departure time is determined by how long it takes the driver to reverse the train, and not the dwell time. The one exception to this rule is the case of a reversal at Arnos Grove which involves a step-back, as the driver does not have to reverse the train himself. Therefore, the dwell time model is only developed for these three locations.

Dwell time is clearly a function of time-of-day, and we expect higher than average dwell times during peak hours. However, apriori, this effect is not expected to be
pronounced given that none of the three stations being modeled are among the most crowded stations on the Piccadilly Line. Figure 4-13 is a scatter plot of dwell times at Acton Town eastbound. We observe that there is a pretty wide spread of dwell times, from a minimum of 8 seconds to a maximum of 3 minutes and 59 seconds. The average dwell time, however, is only 22 seconds. It is also quite clear that there is no correlation with the time-of-day.

The dwell time at Acton Town and Arnos Grove might also depend on whether a crew change is involved. From talking to the DRMs (Duty Reliability Manager) who regularly ride in the train cabs, I gathered that the dwell time was not impacted by a crew change as long as the entering driver was ready and waiting on the platform. For the purpose of this simulation, it is ensured that the service controller always maintains operational feasibility, including providing sufficient walking time for the drivers to reach their trains. It is also assumed in the simulation that drivers follow the instructions given to them, and therefore the ONA (Operator Not Available) scenario does not arise. In practices, ONAs do in fact occur, and they can be modeled as incidents that occur at a crew change location.

It is also important to note that the dwell time at these intermediate stations is quite small compared to the overall trip journey time. For instance, the average dwell time at Acton Town (for the panel of good days) is only 22 seconds compared to the journey from Cockfosters to Uxbridge, which takes around 100 minutes.
The dwell time model is estimated for each thirty minute period. A random sample is drawn using the same method outlined in the previous section on running times.

### 4.4.4 Station details

In this section, we discuss the geometry of the stations in the simplified route structure, and the type of arrival and departure queues that are used as a result.

**Cockfosters**

Cockfosters is a terminus and has three platforms that are used interchangeably. Therefore the three platforms have a single arrival queue and a single departure queue. The minimum reversing time at Cockfosters is taken as four minutes, which is the time required for an operator to walk to the other end of the train.

**Arnos Grove**

Arnos Grove has trains both going through and reversing. It has three platforms, with the outer platforms usually used for through services, and the centre platform usually reserved for reversing trains. Therefore the three platforms have their own arrival
queues. The eastbound platform has its own departure queue. The center (reversing) platform and the westbound platform share a departure queue. The minimum reversing time at Arnos Grove is also four minutes in the case of trains where there is no crew change. For reversals that involve a crew change (i.e. a step-back operation), the dwell time is drawn from the dwell time distribution because the new driver is already waiting in position at the other end of the platform, and the dwell time is determined by how long the passenger alighting and boarding process takes.

**Acton Town**

Acton Town is not used to reverse trains, but it is unique because the line splits into two branches west of the station. Acton Town has four platforms, with two platforms usually reserved for the District Line and two for the Piccadilly Line. The Piccadilly Line can make use of the District Line platforms, but this usually only happens during severe disruptions. For this model we assume that the Piccadilly Line exclusively uses its two platforms. Both the westbound and eastbound platforms have their own departure and arrival queues. It must be noted that the eastbound arrival queue has trains from both branches present in it. Both in the case of Arnos Grove and Acton Town, we implicitly assume that if a crew change is scheduled to take place, the operator is ready in position to take over the train. ONA (Operator Not Available) scenarios are not considered.

**Rayners Lane**

Rayners Lane has trains going through to Uxbridge, and trains reversing. Platform occupancy may be an issue due to interactions with the Metropolitan line, with which the track is shared. However, this interaction is not explicitly modeled in this model. Delays in arriving at Rayner's Lane due to merging with the Metropolitan Line are partly captured by the running time distribution from Acton Town to Rayner's Lane. We note that in calculating the departure time of a through train using equation 4.2, $DT(n - 1)$ should represent the departure time of the last train departing from Rayner's Lane – be it Metropolitan or Piccadilly. But that is beyond the scope of this simulation. Also, the reversing move at Rayners Lane involves moving from the westbound platform to a central siding, walking to the rear of the train, and then moving back to the eastbound platform. Based on discussions with the Piccadilly Line staff, the minimum reversing time at Rayner's Lane is taken as six minutes.

**Uxbridge and Heathrow Terminals**

Trains only reverse at these stations. The minimum reversing time for Uxbridge is assumed to be four minutes because the reversing move is similar to Cockfosters. The reversing move at Heathrow Terminal 5 is similar to Rayners Lane, and therefore the minimum reversing time is assumed to be six minutes. Trains do not actually reverse
at Heathrow Terminal 4 – it is actually a through move. Therefore, the minimum reversing time is taken as zero minutes.

4.4.5 Service control in the Piccadilly Line

One of the components of the Simulation architecture discussed in Section 4.3 is real-time service control, whose purpose is to keep the operation feasible and running smoothly by monitoring the system and making changes to the operations plan as required. The role of service control was discussed in the general context of simulating the operations of a rapid transit line in Section 2.3. In this section, the specific service control policies that are implemented in the simulation of the Piccadilly Line’s operations are discussed.

Service control is as much an art as a science. The nature of the job is such that there is a need to be able to respond quickly to a wide variety of situations. In the case of the Piccadilly Line, it was the author’s observation, based on four weeks of close interactions with the control center staff, that there is no fixed rulebook which lays out clearly defined policies to deal with every kind of disruption. While there are certainly general guidelines in place, these have been developed in a somewhat informal manner over many years of experience in dealing with disruptions. The exact actions that will be taken in dealing with a problem vary from situation to situation, and from controller to controller.

That being said, there are certainly observable patterns in the way disruptions are handled. This is especially so in the case of most minor disruptions, where train lateness is of the order of ten to twenty minutes. In such cases, it was the author’s observation that service controllers act to reduce train lateness by a combination of short-tripping and renumbering of trains, and employing spare drivers.

On the other hand, when there are more severe disruptions, the range of control actions greatly increases. Controllers often cancel many trips and stable trains in an effort to deal with reduced line capacity. It is also common to see trains short-tripped at many more locations. Furthermore, the large number of factors that enter the decision making process during severe disruptions make every disruption unique, and it is difficult to find generalizable service control policies during severe disruptions.

Furthermore, it is not likely that crew schedules should be expected to be resilient to major disruptions, as doing so would require the schedules to have large amounts of slack resources, making them highly inefficient. For these reasons, the Piccadilly Line crew schedule is only simulated under relatively minor disruptions, which are far more commonplace, and for which it is possible to lay out structured service control
The nature of service control policies make them extremely specific to the geometry and context of a line. While the interventions that are used, such as short-turning or expressing, are common to most rail transit systems (as discussed in Section 2.3.2), the specific combination of interventions that are used in a situation is dependent on the features of that line, and it is impossible to generalize these service control strategies to all rail transit lines.

As discussed in Section 2.3.3, crew reliefs act as fixed points in the schedule and controllers make an effort to ensure that crew reliefs take place as scheduled. In the case of the Piccadilly Line, crew reliefs are scheduled either at Arnos Grove or Acton Town, in the eastbound or the westbound directions. In addition, there is a fifth kind of relief which occurs when a crew relief takes place along with a scheduled reversal at Arnos Grove. We first discuss some general principles that apply in all of these cases.

Firstly, a train that is less than ten minutes late does not call for intervention. This is because nearly every duty in the Piccadilly Line crew schedule has a minimum slack time of ten minutes, and therefore it is not critical from a driver lateness standpoint that the train be put back on time. Furthermore, a train that is currently ten minutes late, might not be ten minutes late when it actually arrives at the crew relief point. Unless further disruptions occur, there is a good chance that the train can make up some of its lateness through recovery time at the terminals. And finally, lateness of less than ten minutes is fairly commonplace even in the absence of any disruption, and service controllers would be fighting a losing battle if they attempted to put every single train back on time. It must be noted that this ten minute threshold is not a strict rule, but more a guideline, which has evolved over many years of operations. A crew schedule that has significantly lower or higher minimum slack would necessitate a change in this threshold.

Secondly, it is important that a driver starts his/her run approximately on time, even though it may not be possible to ensure that a driver finishes his/her run on time. This is important because if a driver starts a run late, some or all of his/her slack time will have been consumed at the start of the run itself. This leaves the driver vulnerable in the event that a further delay is encountered.

Based on these general principles, we discuss the service control rules that have been implemented in this simulator. It must be noted that these rules do not represent the optimal service control policies for those situations. They also do not represent the full spectrum of service control policies that are applied in practice, because it is not possible to generalize every single action that is taken by a controller. Instead,
They represent a simplified version of the service control policies that are used on the Piccadilly Line on a daily basis to deal with everyday lateness and minor disruptions. Simulating the performance of different crew schedules with these service control policies and under normal or slightly disrupted conditions will allow us to test the performance of those crew schedules under the type of operating conditions that are most commonly observed.

As mentioned in Section 4.3.5, the service control function is invoked every time an event is processed because service controllers continuously monitor the system and can make changes at any time. Every time the service control function is invoked, the list of currently running trains is scanned. Service control interventions are only applied to trains that are more than ten minutes late and have not yet been reformed. The exact intervention that is applied depends on where the train is scheduled to have its next crew change. As discussed previously, there are five possible crew change locations, and these cases are now discussed.

**Acton Town EB/WB reliefs**

The most common way of reducing train lateness is to short-trip the train. Trains heading to Cockfosters are often turned at Arnos Grove because of the low passenger flow between Arnos Grove and Cockfosters. This can save a train about 20-25 minutes of time. Trains heading to Uxbridge are often turned at Rayner’s Lane because the Metropolitan line provides service between those two stations. This can save about 25-30 minutes of time. In cases when both interventions are possible, the choice is made based on the degree of train lateness.

It is uncommon to turn trains heading to Heathrow Airport (especially during minor disruptions) because of the potentially large negative passenger impacts of doing so. This action is usually reserved for more serious disruptions, and is not used in this simulation.

However, the preferred strategy for bringing trains back on time is renumbering. To understand how it works, consider the following example: train 240 is scheduled to have a crew change at ACT-WB at 10:00. The train is running 15 minutes late. We then look to see if there is another crew change that happens at ACT-WB around 10:15. Let us assume that train 241 is scheduled for a crew change at 10:14. Therefore, the new driver for 241 would be on the platform by 10:14. When train 240 arrives at 10:15, he/she takes over the train, changes it to 241 and leaves approximately on time. When train 241 arrives, say at 10:30, the new driver for 240 gets into the train, changes it to 240 and leaves. He/she is however, departing 30 minutes late. In other words, the lateness of train 241 has been transferred to train 240, and train 240 needs to be put back on time. If it is bound for Uxbridge, it can be turned at Rayner’s...
Lane. If it is a Heathrow train, it would have to be turned at Northfields, which is avoided. Instead, the train could be turned in the eastbound direction at Arnos Grove, if it is scheduled to go all the way to Cockfosters.

Because renumbering causes the lateness of one train to be transferred to another train, it is used when lateness is in the 10-15 minute range. If the lateness is greater than that, it makes sense to just turn individual trains around. The advantage with renumbering is that two trains are put back on time with a single short-turn operation. The renumbering strategy can be extended to a chain of more than two trains, and this is done occasionally in practice, though it is difficult to find opportunities to do so. However, this operation is not implemented in the simulator.

Arnos Grove EB/WB reliefs

The strategy for handling AGR WB/EB reliefs is very similar to ACT WB/EB reliefs. If a train is less than fifteen minutes late, the possibility of a renumbering operation is explored by searching for a candidate train for renumbering. The candidate train must be scheduled to have a crew change at the same location at the projected time of arrival of this train. Just as it was observed in the ACT case, when a renumbering operation is performed the combined lateness of the two trains is transferred to a single train. This train is put back on time by short-turning it at Arnos Grove instead of traveling all the way to Cockfosters. If renumbering is not possible, or if the train is more than 15 minutes late, the train is put back on time by short-turning it. Again, trains going to Heathrow are not turned around at Northfields.

Arnos Grove reversals

The overall strategy is the same as in the previous two cases. The only difference is that after renumbering, the train cannot be put back on time by short-turning at Arnos Grove, as the train is already scheduled to be turned at Arnos Grove. Therefore, the only possibility for putting the train back on time is to turn it at Rayner's Lane if it is an Uxbridge-bound train. Figure 4-15 shows a flowchart representation of the above service control policy as it is implemented in the simulation model.

Spare operators

In this simulation, spare operators are used in two cases:

- Despite the service control intervention, an operator may be late getting off his/her train. If the operator is scheduled to get onto another train after this (either immediately, or after a meal break), it is possible that the operator is
If it is a CFS-bound train that has not yet crossed AGR-EB

If it is an UXB-bound train that has not yet crossed RLN

Figure 4-15: Service control strategy

107
late for that pickup. In that case, a spare operator is deployed in his/her stead. How often this situation arises would depend on the distribution of slack time in the crew schedule.

- Service controllers usually don’t actively monitor driver parameters. This is partly because this information is not presented to them in an easily accessible manner, and partly because it is difficult to process such large quantities of information while making a decision. During minor disruptions, drivers are not very likely to violate their parameters. But when they are in danger of doing so, the type of relief has a bearing on what happens. It has been observed that drivers who are on their last run for the day and are scheduled to get off their trains at AGR-WB, are sometimes unwilling to go past AGR-EB. This is because they know that they will be late for their relief. In that case, a spare operator is deployed to operate the round trip from Arnos Grove to Cockfosters and back. The same thing is observed in the case of ACT-EB reliefs. This is in fact the reason for the service controllers’ dislike of AGR-EB and ACT-WB reliefs.

A crew schedule that has a higher average slack time, or a higher minimum slack time would be expected to use fewer spare operators. Similarly, a crew schedule that has a better alignment between slack time and typical lateness (other things being equal) would also be expected to use fewer spare operators. Therefore, the number of spare operators used in place of operators who are late for their pickups is an indicator of schedule robustness and is directly influenced by the distribution of slack time in the schedule.

Other things being equal, a schedule that has a better relief mix would use fewer spare operators. Similarly, a schedule that has a lower value of maximum duty length and maximum continuous driving time is also likely to result in fewer drivers hitting their parameters. Therefore, the number of spare operators used to cover for drivers who have hit their parameters is also an indicator of schedule robustness, and this is influenced by the distribution of relief directions, duty length and driving time.

In practice, the deployment of spare drivers is handled by service controllers and the duty managers at the crew bases. It is also possible that there are no spare drivers available when they are needed. For the purpose of this simulation, it is assumed that there is an unlimited supply of spare operators. Therefore, the utilization of spare drivers is not representative of real operations. However, when performing a comparative analysis of crew schedules, the utilization of spare drivers for two different crew schedules is definitely an indicator of robustness. Other things being equal, a crew schedule that has more slack and a better relief mix is likely to use fewer spare operators.
4.5 Conclusion

This chapter presented a framework for a simulation-based evaluation of the robustness of a crew schedule. A simulation model for a general rail transit line was developed, and the details of this model were examined in depth. Finally, we discussed the details of implementing this simulation model for the Piccadilly Line, including the estimation of running time and dwell time models, and service control policies. The following chapter deals with the validation of this model, and its application to performing a comparative analysis of the robustness of two Piccadilly Line crew schedules.
Chapter 5

Simulation-based evaluation of Piccadilly Line crew schedules

The previous chapter described a simulation-based framework that can be used to evaluate the robustness of a crew schedule. Furthermore, a simulation model was developed for the London Underground's Piccadilly Line which can simulate the performance of the line under a range of operating conditions. In this chapter, this model is first validated by performing simple sanity checks under undisrupted and disrupted operating conditions to ensure that the model's output is consistent with what is observed in reality. The model is then applied to two Piccadilly Line crew schedules to evaluate and compare their performance under a range of operating conditions. The comparative evaluation of their simulated performance is viewed in the context of their structural differences (described in Chapter 3) to develop insights on how to improve the robustness of Piccadilly Line crew schedules.

5.1 Model validation

The model is validated by performing simple sanity checks to ensure that the results of the simulation are consistent with what is observed in practice. This is first done for undisrupted conditions, and then for disrupted conditions.

5.1.1 Undisrupted conditions

The 2012 Piccadilly Line crew schedule (schedule 48) is simulated without any incidents. The simulation model is run for a total of 100 replications, and the results are averaged to estimate the performance of schedule 48 under undisrupted conditions. The observed data comes from the panel of eight days with good performance that were used to estimate the running time and dwell time models for the simulation (see Section 4.4). The simulation model is applied to Schedule 48 as the observed data
also comes from Schedule 48.

Figure 5-1 shows the running times of successive trains between Acton Town and Arnos Grove. The X-axis shows the running time of the leading train, and the Y-axis, that of the trailing train. Figure 4-7 showed the same scatter plot but using observed (rather than simulated) running times from undisrupted days. The two figures are very similar which is expected given that the running times used in the simulation are drawn from distributions that are constructed using observed running times.

![Figure 5-1: Running times of successive trains between ACT and AGR (Schedule 48, undisrupted)](image)

The correlation between the simulated running times of successive trains is slightly lower than what is actually observed in Figure 4-7. This can be explained by the fact
that the simulation uses a simplified network structure where many stations are not explicitly represented. For instance, there are no stations between Acton Town and Arnos Grove in the simulation, whereas there are 22 stations in reality. If train movements were explicitly simulated between every one of these stations, there would be a greater degree of interaction between successive trains which would result in greater correlation between the running times of successive trains.

Figure 5-2 shows the average lateness of trains arriving at a crew change location (i.e. ACT-EB, ACT-WB, AGR-EB, AGR-WB or AGR). The average is calculated over all trains arriving in fifteen minute intervals starting at 7AM and ending at 9PM. The figure shows that lateness rarely exceeds three minutes (on average). There is a distinct peak corresponding to the morning peak, and the mid-day period experiences little lateness. Lateness again increases during the evening peak, and then persists late into the night. The observed and simulated lateness trends are very similar, as expected.

Figure 5-3 shows the average headway of trains departing from AGR-EB. The average is calculated over all trains departing in fifteen minute intervals starting at 7AM and ending at 9PM. The peaks have a headway of around 2.5 minutes which corresponds to 24 trains per hour, and the mid-day period has 3-3.5 minute headways which correspond to 18-20 trains per hour. The observed and simulated headways are very similar, as expected.

In all three cases (running times, train lateness and headways), the simulated outputs are very similar to the observed outputs, which is expected given that the running time and dwell time inputs used for the simulation come from the same data. However, this still serves as a simple check to ensure that the simulation is performing as expected in undisrupted conditions.

5.1.2 Disrupted conditions

The 2013 crew schedule (schedule 49) is simulated under two different disrupted conditions: Scenario 1 simulates a disruption in the trunk portion of the Piccadilly Line and Scenario 2 simulates a disruption in the Rayner’s Lane branch.

Scenario 1: Disruption in the trunk

In this scenario, an incident at 10AM, halfway between Arnos Grove and Acton Town (westbound), that lasts forty minutes is simulated. Therefore, the first train to cross this point on the line after 10AM has its running time increased by forty minutes, and the running times of the trains behind it are calculated as described in Section
Figure 5.2: Average lateness of trains arriving at crew change locations (Schedule 48)

The graph illustrates the lateness (in minutes) of trains arriving at crew change locations across different times of the day. The data is compared between observed and simulated lateness for Schedule 48.
Figure 5-3: Average headway between trains departing AGR-EB (Schedule 48, undisrupted)
4.3.4. The delays caused by this incident would normally result in service controllers intervening in the service in an effort to reduce train lateness. To illustrate the effect of service control, this scenario is simulated with and without the presence of service control. In each case, the simulation is repeated 100 times, and the results are then averaged.

While a forty minute incident is quite severe, it was used for the purpose of vividly demonstrating the effect of service control, which might not be as evident with a minor incident. It is not, by any means, a "typical" incident.

Figure 5-4 illustrates the effect of service control on train lateness, and consequently, driver lateness. The graph above depicts the average train arrival lateness across all stations in the network, calculated for every fifteen minute interval between 6AM and 11PM. The graph below depicts the same quantity, but only for trains arriving at Acton Town westbound. In both cases, the blue line marks train lateness when no service control actions are taken, i.e. when trains are not put back on time through renumbering and/or short-tripping. The orange line marks train lateness when the service control policies described in Section 4.4.5 are applied. The Y-axis in both cases is trains lateness in minutes.

The presence of the incident is identified by the sudden upsurge in train lateness shortly after the incident occurred. This increase is first detected at Acton Town westbound, as it is immediately downstream of the incident. The lateness at ACT-WB shoots up to forty minutes as soon as the disrupted train arrives there. But the incident is also reflected, albeit to a lesser extent, in the average lateness across all stations. It is clear from the figure that the service control has a positive effect in reducing train lateness. Both when looking at train lateness at ACT-WB and when looking at train lateness across all stations, the lateness caused by the incident persists for much longer without active service control. This persistent lateness creates problems for crew changes because drivers may be too late to pick up their next train. Drivers may also have to be relieved early if the lateness causes their parameters to be violated. Finally, lateness also causes drivers to end their duties late, essentially working overtime (though this may not necessarily be compensated). Table 5.1 illustrates many of these effects.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Without SC</th>
<th>With SC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of short-turns</td>
<td>0</td>
<td>12.8</td>
</tr>
<tr>
<td>Number of pickups missed</td>
<td>18</td>
<td>10.6</td>
</tr>
<tr>
<td>Number of drivers relieved early</td>
<td>14.5</td>
<td>7.8</td>
</tr>
<tr>
<td>Overtime operator-hours</td>
<td>6.5</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Table 5.1: Performance measures for Scenario 1
Figure 5-4: Average lateness with and without service control for Scenario 1.
In the absence of service control, trains are not short-tripped (short-turned). However, with active service control, an average of 12.8 trains are short-tripped, either at Arnos Grove (for trains going to Cockfosters) or at Rayners Lane (for trains going to Uxbridge). As explained in Section 4.4.5, Heathrow-bound trains are not short-tripped in this simulation. The effect of all this short-tripping is to reduce train lateness as illustrated in Figure 5-4. The reduced lateness in the presence of service control manifests itself in fewer missed pickups, fewer drivers needing to be relieved early, and fewer overtime hours.

Drivers may miss pickups after their meal breaks, or in the case of spells with more than one train run, within a spell. In the case of a meal break, the driver is provided the standard thirty minute break time plus walking time. Depending on how late he/she arrived, and how much slack the duty had, the driver may or may not be able to pick up their next train. In the case of a spell with more than one run, the driver is assumed to be ready to drive the next train immediately after getting off the previous train, and therefore this type of pickup is missed only when trains run out of order.

Drivers are relieved early only when they are scheduled to finish their duties at AGR-WB or ACT-EB, and they have significant lateness at AGR-EB or ACT-WB. If their lateness is more than the End of Duty (EoD) slack, then they are relieved early and a spare driver is used for the remainder of the trip. As described in Section 3.2.5, the EoD slack is defined as the minimum of: (i) buffer time between end of the last run and end of the duty, and (ii) time before hitting a parameter (either duty length or driving time). When a driver is at AGR-EB (or ACT-WB) and knows that it is not possible to get back to AGR-WB (ACT-EB) on time, the driver often asks to be relieved. This is the reason that AGR-EB and ACT-WB reliefs are preferred.

A driver is said to have worked overtime, if the driver’s lateness when finishing a duty is more than the EoD slack, and the difference between the two is the amount of overtime worked. This quantity is summed over all the drivers and averaged over the simulation runs to calculate the average number of operator-hours of overtime.

Though this is not the main objective of this research, the simulation-based framework can also be used as a tool to evaluate the efficacy of service control policies. This service control policy makes a trade-off between short-tripping trains (which hurts customers) and reducing operating costs (by reducing spare driver deployment, overtime etc.) The simulation can be used to test multiple service control policies and understand the tradeoffs between operational costs and customer impacts.
Scenario 2: Disruption on the Rayner’s Lane branch

In this scenario, a fifteen minute incident occurring at 10AM midway between Acton Town and Rayners Lane is simulated with schedules 48 and 49. The simulation is repeated 100 times and the results are averaged as before.

Figure 5-5 shows the average train arrival lateness under the two schedules at Rayners Lane Westbound and at Heathrow Terminal 5. It is interesting to see that even though the disruption occurred on the Rayners Lane branch at 10AM, the effect was also felt at Heathrow Terminal 5. But this happened nearly three hours later, which reflects the amount of time taken for a train to travel from Rayners Lane to Cockfosters and then to Heathrow. Both schedules 48 and 49 perform in line with expectations.

The two scenarios show the simulated performance of the two crew schedules under disrupted conditions. The outputs of the simulation are consistent with our expectations of the line’s performance. While this does not represent a statistically rigorous validation exercise, it serves to ensure that the simulation is performing as intended.

In the case of the Piccadilly Line, service controllers maintain a manual log of the specific interventions (renumbering, short-turns, cancellations) that they carried out, and this makes it very difficult to collect a large sample of incidents and associated interventions with which to validate the simulation model. The service manager creates a daily report which notes the number of trip cancellations, but does not talk about train renumbering and short-turns. Therefore, it is difficult to perform a large scale validation exercise to explicitly check if the simulation outputs (especially service control interventions) are consistent with what happens in practice. Furthermore, the assumptions in the model that there is an unlimited supply of spare drivers and that trains are never cancelled are simplifications of reality. Notwithstanding these shortcomings, the simulation model still provides a reasonable way of understanding the impacts of the crew schedule’s structure on the line’s performance.

5.2 Comparative evaluation of Piccadilly Line schedules 48 and 49

Having validated the simulation model, it is now used to evaluate and compare the performance of Piccadilly Line crew schedules 48 and 49 under disrupted conditions. As explained in Section 4.2, the schedules must be simulated under a representative set of incidents in order to evaluate robustness. It is pointless to simulate the schedules in the absence of incidents, or even under very minor incidents, as these situations will not stress the schedules and bring out the differences between them. At the same time, we do not simulate the schedules under extreme incidents because
Average Train Lateness at Rayners Lane WB

Average Train Lateness at Heathrow Terminal 5

Figure 5.5: Average lateness for Scenario 2
making a schedule resilient to extreme incidents would require large slack resources (slack time and spare operators) making it inefficient. Furthermore, the service control logic that is implemented in this simulation is not applicable to extreme incidents. Based on these considerations, incidents in the range of 15-35 minutes are simulated as explained below.

5.2.1 Distribution of incidents

Freemark[5] characterizes the distribution of incidents on the Piccadilly Line, and these results are used to construct the distribution of incidents against which the crew schedules are tested. The distribution of incidents is characterized by their frequency, duration, time of occurrence and location. A panel of 29 disrupted days is used to characterize the distribution of incidents on the Piccadilly Line.

Freemark notes that there are usually multiple incidents in a day, most often between two and four. Based on this, we simulate three incidents for each day, one each in the AM peak, midday and PM peak. This is not a restrictive assumption as the duration of the incidents are allowed to vary for each of the three periods of the day. Incidents are not simulated early in the morning and late in the evening because the service control actions for these incidents can be very different. For instance, service controllers might respond to an incident late in the night by stabling trains early rather than trying to reduce lateness.

Freemark observes that there are distinct differences in the durations of incidents throughout the day. Incidents that take place at the peaks are usually shorter than those that take place at other times, which might reflect the fact that greater importance is accorded to ensuring smooth operations during the peak hours. Based on this data, Table 5.2 specifies the distribution of the duration of incidents (in minutes) for different times of day that is used in the simulation. The duration of the incident is assumed to be uniformly distributed between the minimum and maximum values specified in the table. The time of occurrence of the incident is also assumed to be uniformly distributed within each interval.

<table>
<thead>
<tr>
<th>Period</th>
<th>Definition</th>
<th>Average</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>AM Peak</td>
<td>7:00 - 10:00</td>
<td>20</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>Midday</td>
<td>10:00 - 16:00</td>
<td>30</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>PM Peak</td>
<td>16:00 - 19:00</td>
<td>17.5</td>
<td>15</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 5.2: Distribution of incident duration by period (in minutes)

Freemark also characterizes the distribution of incident locations. A correction is
made to account for the differences in network structure. In Freemark's characterization, the trunk ends with King's Cross whereas in this simulation it extends to Arnos Grove. Based on the data, Table 5.3 shows the spatial distribution of incidents. It is observed that approximately 40% of the incidents take place at the stations in the network, and the remaining 60% in between the stations in the network. These incidents could also be happening at stations, but they are not explicitly modeled as such because of the simplified network structure used here.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Disruption Mins.</th>
<th>Adjusted Disruption Mins.</th>
<th>Prob. of Incident</th>
</tr>
</thead>
<tbody>
<tr>
<td>UXB-RLN</td>
<td>272</td>
<td>272</td>
<td>9%</td>
</tr>
<tr>
<td>RLN-ACT</td>
<td>203</td>
<td>203</td>
<td>7%</td>
</tr>
<tr>
<td>Heathrow-ACT</td>
<td>231</td>
<td>231</td>
<td>8%</td>
</tr>
<tr>
<td>ACT-AGR</td>
<td>641</td>
<td>741</td>
<td>25%</td>
</tr>
<tr>
<td>AGR-CFS</td>
<td>709</td>
<td>346</td>
<td>12%</td>
</tr>
<tr>
<td>At ACT</td>
<td>234</td>
<td>333</td>
<td>11%</td>
</tr>
<tr>
<td>At AGR</td>
<td>114</td>
<td>263</td>
<td>9%</td>
</tr>
<tr>
<td>At CFS</td>
<td>333</td>
<td>114</td>
<td>4%</td>
</tr>
<tr>
<td>At UXB</td>
<td>263</td>
<td>234</td>
<td>8%</td>
</tr>
<tr>
<td>At RLN</td>
<td>203</td>
<td>203</td>
<td>7%</td>
</tr>
</tbody>
</table>

Table 5.3: Spatial distribution of incidents on the Piccadilly Line

Based on the distributions specified above, three incidents are generated for each simulation run (which represents one day of operation), one incident for each of the three periods in the day. The location, time of occurrence and duration of the three incidents are assumed to be independent. The incident distribution used in this experiment is designed to generate representative incidents that simulate a reasonable range of disrupted operating conditions. Evaluating the performance of the crew schedule under such operating conditions allows us to draw insights into what makes a crew schedule robust.

5.2.2 Simulation results

The simulation experiment described in the previous subsection was repeated 500 times, and the results averaged to produce daily performance measures. Crew schedules 48 and 49 are both tested on exactly the same set of incidents. Table 5.4 shows the performance of the two schedules on a number of metrics.

Given the exact same set of incidents, schedule 49 required slightly fewer short-turns than schedule 48, though this difference is not large enough to be significant. More significantly, schedule 49 had far fewer missed pickups than schedule 48, which might be attributed to the improved matching of slack time to typical train lateness in schedule 49. It should be noted that 17 missed pickups is not equivalent to 17 spare drivers
<table>
<thead>
<tr>
<th>Schedule</th>
<th>48</th>
<th>49</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of short-turns</td>
<td>7.5</td>
<td>6.0</td>
</tr>
<tr>
<td>Number of pickups missed</td>
<td>17.2</td>
<td>10.3</td>
</tr>
<tr>
<td>Number of drivers relieved early</td>
<td>18.5</td>
<td>7.7</td>
</tr>
<tr>
<td>Overtime operator-hours</td>
<td>1.3</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Table 5.4: Performance measures for schedules 48 and 49

being needed. Firstly, a spare driver can perform more than one run. Secondly, when a driver misses a pickup, the driver is often “converted” to a spare, especially if the driver has no other runs for the day.

Schedule 49 also has far fewer drivers who need early reliefs at AGR-EB and ACT-WB, and this can be directly attributed to the improved mix of relief directions in schedule 49. Schedule 49 also has slightly higher EoD slack time which would have also helped to improve this metric. Once again, it must be noted that the one driver being relieved early does not equate to one spare driver.

However, there is a small but significant increase in the overtime hours accumulated by drivers in schedule 49, because of the reduction in early reliefs for drivers. Overtime has negative impacts on operational costs and on crew morale, so this is a trade-off that must be recognized and analysed. But the magnitude of overtime hours here is very small compared to the scheduled number of driver hours, and therefore is not a major cause for concern. It should be noted that different agencies have different ways of dealing with overtime. In some agencies, drivers are monetarily compensated for having to work beyond their scheduled times. In other cases, drivers are compensated informally in non-monetary ways, such as giving them easier duties or by relieving them early on another day. For this reason, the impact of overtime varies from agency to agency.

Figures 5-6 and 5-7 show the relationship between meal break slack and train lateness at Acton Town. The blue points represent meal reliefs and the Y-value of the point is the meal break slack associated with that relief. The orange line shows the 95th percentile of train lateness (from the simulation output) for each fifteen minute interval between 7AM and 10PM. Note that this does not represent typical train lateness, which would be calculated for the full range of operating conditions (from undisrupted to severely disrupted conditions). Rather this represents the 95th percentile of lateness under the disrupted conditions that have been simulated in this experiment, which will be significantly higher than the typical lateness. It is observed that Schedule 49 has, in general, slightly lower lateness than Schedule 48, and fewer duties whose slack is less than the 95th percentile of lateness.
Figure 5.6: Slack versus Lateness at ACT EB
Slack vs Lateness - ACT WB (Schedule 48)

Slack vs Lateness - ACT WB (Schedule 49)
Figure 5-8 shows that the spatial distribution of missed pickups is significantly different for the two schedules. Because the mix of relief directions is so different in the two schedules, this is to be expected. Figure 5-9 shows the temporal distribution of missed pickups. Most of the missed pickups occur in the periods just after the morning and evening peaks, which is when most pickups are scheduled. It is also interesting to note that there are more missed pickups in the evening than the morning. This is because there is typically more lateness in the evening than in the morning, because incidents in the mid-day have an effect through the evening. This reasoning is also borne out by the lateness profiles observed in figures 5-6 and 5-7.

![Figure 5-8: Spatial distribution of Missed pickups](image)

As explained previously, missed pickups can take place after a meal break or during intermediate reliefs (i.e. during a spell with multiple runs). Considering only the former, Table 5.5 characterizes the meal break slack associated with the missed pickups. As discussed in Section 3.2.3, the average meal break slack across all the duties in schedule 48 is 20 minutes, versus 18 minutes in schedule 49. The average meal break slack associated with a driver who missed a pickup (after a meal break) in both cases is 14 minutes. It is logical that the drivers who miss their pickups are those with smaller than average meal break slack times. For each driver, we calculate the number of simulation runs (i.e. frequency) in which that driver missed his/her pickup, and then arrange the drivers in descending order of frequency of missed pickups. The average slack across the duties that comprise the top 50% of missed pickups is 11 minutes for both schedules. As expected, this is lower than the average slack across all missed reliefs.
### Average meal break slack across:

<table>
<thead>
<tr>
<th></th>
<th>Schedule 48</th>
<th>Schedule 49</th>
</tr>
</thead>
<tbody>
<tr>
<td>All duties in schedule</td>
<td>20 minutes</td>
<td>18 minutes</td>
</tr>
<tr>
<td>Duties that had missed pickups</td>
<td>14 minutes</td>
<td>13 minutes</td>
</tr>
<tr>
<td>Duties that make up the top 50% of missed pickups</td>
<td>11 minutes</td>
<td>11 minutes</td>
</tr>
</tbody>
</table>

**Table 5.5:** Meal break slack across missed pickups

It is easy to say that robustness could be improved by tightening the distribution of slack time. While it is not possible to increase the average amount of slack in a crew schedule, given a certain vehicle schedule, without increasing manpower, it is possible to tighten the distribution of slack, by taking away slack from duties that have more slack than needed, and redistributing it to duties that can use more of it. But the degrees of freedom needed to do this have to come from relaxing other constraints.

### Average EoD slack across:

<table>
<thead>
<tr>
<th></th>
<th>Schedule 48</th>
<th>Schedule 49</th>
</tr>
</thead>
<tbody>
<tr>
<td>All duties in schedule</td>
<td>7 minutes</td>
<td>8 minutes</td>
</tr>
<tr>
<td>Duties that had early reliefs</td>
<td>4 minutes</td>
<td>5 minutes</td>
</tr>
<tr>
<td>Duties that make up the top 50% of early reliefs</td>
<td>3 minutes</td>
<td>3 minutes</td>
</tr>
</tbody>
</table>

**Table 5.6:** EoD slack across early reliefs

Similarly, Table 5.6 characterizes the EoD (end of duty) slack associated with early reliefs. We observe the same kind of trend as in meal break slack. It must be noted that improving the distribution of EoD slack does not have as large an effect on robustness as improving the distribution of meal break slack. This is because a driver who is finishing his/her duty does not have any more pickups for the day. Having more EoD slack can reduce the number of early reliefs; but this has already been significantly reduced by the improvement in relief directions. Having more EoD slack also reduces the amount of overtime incurred, but this is not very large to begin with. But to increase the EoD slack, it is necessary to relax some constraints, or increase manpower, or decrease the amount of meal break slack. This ultimately becomes a trade-off between the meal break slack and the EoD slack, which is a trade-off between more robustness and lower overtime costs. The right balance between the two depends on the priorities of the agency and the simulation can be used as a tool to evaluate this trade-off.

### 5.2.3 Simulating hypothetical scenarios

It was observed from Table 5.5 that the duties that had missed pickups most often also had, on average, significantly lower meal break slack. This suggests that increasing the minimum amount of meal break slack, especially at those times of the day when trains tend to run late, would reduce missed pickups and increase robustness.
To check the validity of this hypothesis, we would ideally create a new crew schedule that enforced a higher minimum slack, of say, 15 minutes, by relaxing some other constraints or by increasing manpower, and then simulate its performance under disrupted conditions, as was done previously.

**Scenario 1: Relaxing the meal break length constraint**

In this scenario, we attempt to check this hypothesis by relaxing the constraint on the duration of the meal break. The meal break is normally fixed at thirty minutes. By relaxing this constraint, we can ensure that duties that have less than fifteen minutes of slack time get at least fifteen minutes of slack time. Duties that already have at least fifteen minutes of slack time do not suffer a reduced meal break. Ideally, the minimum slack time would vary with time-of-day and relief location depending on the typical train lateness that is observed at that relief location at that time-of-day. However, for the purpose of this scenario we assume that the minimum slack is fixed at fifteen minutes for all reliefs regardless of when or where they take place.

It is necessary to note that this is merely a workaround that allows us to increase the minimum slack time without creating a new crew schedule. Ordinarily, the length of the meal break would not be the constraint that is relaxed. But doing so allows us to quickly evaluate the impact of having a tighter distribution of slack time. The correct way of testing the impacts of a tighter slack distribution would be to create a crew schedule that relaxes some other constraints (such as fixed-link) and then simulate that schedule. In this scenario, we simulate the performance of schedule 49 under the same distribution of incidents as before, while relaxing the length of the meal break to ensure that every duty has at least fifteen minutes of slack time.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Normal</th>
<th>Relaxed Meal Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of short-turns</td>
<td>6.0</td>
<td>6.1</td>
</tr>
<tr>
<td>Number of pickups missed</td>
<td>10.3</td>
<td>6.1</td>
</tr>
<tr>
<td>Number of drivers relieved early</td>
<td>7.7</td>
<td>7.5</td>
</tr>
<tr>
<td>Overtime operator-hours</td>
<td>3.0</td>
<td>2.8</td>
</tr>
</tbody>
</table>

Table 5.7: Performance measures for schedule 49 with meal break length constraint relaxed

Table 5.7 shows the result of the simulation. As expected the number of missed pickups declines. The number of short-turns does not change because controller intervention depends on train lateness and not driver lateness. The number of early reliefs and overtime hours also decline very slightly, though this might just be noise in the model, and not because of the increased slack.
Scenario 2: Reduced walking time at Arnos Grove

Ever since the completion of the foot bridge at Arnos Grove, there has been talk of reducing the walking time at Arnos Grove from the current $7/8$ minutes to $3/4$ minutes. This has not yet been implemented in the crew schedules because of failure to reach agreement with the labour unions. In this scenario, we simulate the effect of reducing the walking time at Arnos Grove to a flat $4$ minutes. This has the effect of increasing the slack time for all Arnos Grove duties by $6-8$ minutes which should improve robustness. We should note that if the walking time at Arnos Grove was really reduced, the corresponding crew schedule would end up splitting the benefits of increased slack between Arnos Grove and Acton Town. However, for the purposes of this scenario, it is assumed that only the Arnos Grove duties benefit from increased slack time. The scenario is simulated under the same distribution of incidents as before.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Normal</th>
<th>Reduced AGR Walking Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of short-turns</td>
<td>6.0</td>
<td>5.7</td>
</tr>
<tr>
<td>Number of pickups missed</td>
<td>10.3</td>
<td>5.5</td>
</tr>
<tr>
<td>Number of drivers relieved early</td>
<td>7.7</td>
<td>7.4</td>
</tr>
<tr>
<td>Overtime operator-hours</td>
<td>3.0</td>
<td>2.6</td>
</tr>
</tbody>
</table>

Table 5.8: Performance measures for schedule 49 with reduced walking time at Arnos Grove

Table 5.8 shows the result of the simulation. As expected, the performance of the schedule improves, and in fact, the performance in scenario 2 is even better than in scenario 1. This is to be expected given that in scenario 2 all AGR duties get an additional $6-8$ minutes of slack time, whereas in scenario 1, only those duties with less than $15$ minutes of slack time, get additional slack.

From both of these hypothetical scenarios, we can conclude that improving the distribution of slack time by raising the minimum amount of slack, even while keeping the average slack constant can improve the robustness of a schedule. To do so, it is necessary to relax certain constraints, and the simulation can be used to understand the trade-off involved.

These scenarios provide an example of how the simulation model can be used as a tool to inform labour negotiations by providing management with an understanding of the impact of proposed changes. Whenever changes are made to labour agreements or other crew-related matters, it is often assumed that the only impact will be lower or higher manpower costs. But it is often the case that such changes will also impact the performance of the line. For instance, the fixed-link agreement at Arnos Grove resulted in a sub-optimal mix of crew relief directions, and a loss of flexibility to
adjust the distribution of slack time. These unintended consequences had a negative effect on performance because the new crew schedule was found to be less flexible during disruptions. Had management known this before the negotiation process started, they might have negotiated differently. Labour agreements are reviewed periodically, and it is in the management's interest to know beforehand how any proposed changes might impact the line's performance.

5.3 Conclusion

In this chapter, the simulation tool was validated by running it under a variety of operating conditions and looking at many different indicators to ensure that the output of the simulation is consistent with what is observed in practice. Train lateness, running times and headway data from a simulation of undisrupted conditions were compared with observed data from undisrupted days, and the two were found to be consistent. The effect of service control was demonstrated by simulating a disruption in the trunk of the line, with and without service control interventions. In the latter case, train lateness persisted for much longer than when service controllers intervened to put trains and drivers back on time. The simulation was also used to highlight the differing impacts of incidents occur on the branches as compared to incidents on the trunk portion of the line. All of these exercises serve to verify and validate the simulation model of the Piccadilly Line.

The simulation tool was then used to evaluate and compare the performance of Piccadilly Line crew schedules 48 and 49 under a reasonable distribution of incidents. Three incidents were generated per simulation run (i.e. per day), one each in the AM peak, midday and PM peak periods. Based on prior research that characterized the occurrence of incidents on the Piccadilly Line, distributions for the location, duration and time of the incident were obtained. The distributions of simulated incidents are designed to represent a reasonable level of stress on the crew schedules that allows us to compare the robustness of the two crew schedules.

The comparison shows that Schedule 49 outperformed Schedule 48 on a number of key indicators such as number of missed pickups, number of early reliefs and number of short-turns. Some of these results, such as the decrease in the number of early reliefs can be predicted based on the structural comparison of the two schedules which showed that schedule 49 had a vastly improved mix of relief directions compared to schedule 48. However, schedule 49 also had slightly lower average slack, more duties that were close to the driving time parameter, and better matching between lateness and slack. The first two changes would likely hurt robustness, while the last change would help robustness – but overall it is unclear, based on just the structural evaluation, whether schedule 49 would perform better or worse than schedule 48. However,
the simulation shows that schedule 49 has fewer missed reliefs and short turns, and is therefore more robust than schedule 48. Herein, lies the true power of the simulation in that it allows us to evaluate the impact of changes where the effect on robustness is unclear.

Upon closer inspection of the simulation results, we find that the average slack time across the missed pickups is significantly lower than the average slack time in the crew schedule. This suggests the hypothesis that increasing the minimum slack time improves robustness. Given a certain vehicle schedule and manpower level, the average slack time cannot be changed. However, it is possible to change the allocation of slack time between duties, though doing so requires degrees of freedom that can only be obtained by relaxing other constraints. We relax the constraint on the length of the meal break to ensure that every duty has at least fifteen minutes of slack time. Ideally, this minimum slack would vary by time-of-day and relief location, and it would be enforced while creating the crew schedule and not during the simulation by relaxing a constraint. Nevertheless, it provides an approximate way of confirming the hypothesis that improving the minimum slack time improves robustness.

A second hypothetical scenario is tested in which the walking time at Arnos Grove is reduced by 3-4 minutes. This is an issue that has been in negotiation for several years, and the simulation allows us to understand the impacts of the reduction on schedule performance. As expected, the performance of the schedule improves, and this demonstrates how the simulation model can be used to inform labour negotiations.
Chapter 6

Conclusion

The first part of this chapter summarizes the thesis and the key findings from the research. It also presents the main recommendations that are made to the Piccadilly Line as a result of this analysis. Finally, the limitations of this work and the scope for more research that arises from this thesis are discussed.

6.1 Summary

This thesis has taken a broad approach to the topic of robust crew scheduling on a high-frequency rail transit line. Since the focus of the research and applications were focused on the London Underground Piccadilly line, several of the findings are specific to that line. However, the overall approach—the simulation-based framework for evaluating the robustness of a schedule, the development of a simulation model for a high-frequency rail transit service, the measures for evaluating the structure and robustness of a schedule, the interaction between service control and the crew schedule—is not specific to the Piccadilly Line. Therefore, many of the conclusions are also applicable to other metro lines and systems. Though they would have to be adapted to the characteristics of that line and to the available data, the overall approach could largely follow the one described here. Moreover, as other transit agencies may have better train location and crew data available, not all the limitations encountered in the analysis of crew management on the London Underground may apply to the analysis of lines on other metro systems.

The second chapter of this thesis introduced the context in which to understand robustness, by discussing the types of incidents that take place on a line, the service control process that is necessary to deal with incidents, and how service control policies are affected by crew constraints and the crew schedule. We showed that an understanding of the robustness of a crew schedule can only be obtained in the con-
text of the service control policies that make use of those robustness elements. The main drivers of service control are often considerations about crew management, the level of service to passengers, rolling stock management, safety and infrastructure maintenance.

Aside from these considerations, virtually all decisions are influenced by uncertainties regarding the outcome of an intervention and concerns about the manageability of the service. It was seen that the reliability of the system depends on many factors which are endogenous to it and which may previously not have been recognized. In the absence of official policies or effective decision support, the management of these factors is often governed by rules of thumb. As a result service control not only works to manage unreliability caused by exogenous events but can also be the cause of unreliability as controllers work to meet other objectives and constraints.

A review of the literature in this field showed that while the robustness of schedules has been evaluated before, the focus has mostly been on vehicle schedules (or timetables) and not crew schedules. Furthermore, most of the literature does not capture the close relationship between service control and crew schedule robustness, which is particularly critical for high-frequency systems. While there has been some research in the airline industry that considers this aspect, there is a lack of such research in rail and transit systems.

Therefore, any effort to improve service control (and thus, operations in general) on a specific metro line must build on a solid understanding of how that line operates. While this may previously have been a task which was achieved mainly with the help of models, the roles of modeling and data analysis in transit operations are shifting in light of the increased availability and accessibility of automatically collected operations and passenger travel data. This thesis makes use of the availability of vehicle tracking data to build a simulation model that can evaluate the performance of crew schedules.

While crew schedules have traditionally been designed in such a way that the vehicle schedule can be operated at minimum cost, there has been a growing awareness within the London Underground of the potential for improving (or hurting) performance by tweaking the crew schedule. Feedback from the service controllers on what works in the crew schedule, what does not, and what changes would be helpful, reaches the crew scheduling team through the line management, and schedulers attempt to incorporate this feedback, given the labour constraints that need to be met and the available manpower.

This leads to different crew schedules with similar underlying timetables having very different structures as is the case with Piccadilly Line crew schedules 48 and 49. In
Chapter 3, these differences were highlighted by comparing the two crew schedules along a number of different metrics. For instance, schedule 49 has a better mix of relief directions than schedule 48, and a time-varying distribution of slack that is more aligned with typical train lateness. However, schedule 49 also has slightly lower slack time on average and more duties that approach the driving time limit.

These metrics, such as relief mix and slack time, were chosen because they are understood to have direct impacts on how the schedule performs during disruptions. While the impact of some of these metrics (such as relief directions) on the line's performance is clear, the same might not be true of other metrics. Also, a schedule might improve on some metrics and be worse with respect to other metrics. Therefore, a structural evaluation alone is not always sufficient to tell us how all of the different aspects of the schedule come together to influence performance.

Furthermore, it is not always clear how to use additional degrees of freedom that are obtained by relaxing other constraints (perhaps as a result of labour negotiations or changes in physical infrastructure) or by more manpower. The additional freedom could be used in a number of ways to improve the schedule's performance, and a structural evaluation alone does not tell us the best way to deploy these resources. These shortcomings in the structural evaluation of crew schedule performance motivate the creation of a simulation-based framework for evaluating a crew schedule's performance, which is the subject of Chapter 4.

Simulation is a powerful tool as it allows us to capture the interactions between multiple objects such as trains, drivers, incidents, service controllers etc. This allows the system to be represented at a level of detail that is impossible with analytic models. Furthermore, simulation is ideally suited to make use of the large amounts of disaggregate data available from train tracking systems, fare collection systems etc. The model simulates the operations of a high-frequency rail transit line for a range of incidents and this allows the evaluation of the robustness of crew schedules.

In Chapter 4 a general simulation architecture is described that can be applied to any high-frequency transit service, and a model is developed for the Piccadilly Line, which is the key contribution of this research. The simulator is event based and the events correspond to the arrival and departure of trains from stations. The model simplifies the line's network by focusing only on reversing locations and crew change locations, and aggregating all the other stations into segments. This allows for simpler running time and dwell time models. The model ensures that train movements on the simplified network are representative of the actual network by accounting for train blocking, platform capacity constraints, correlation between running times, and temporal variation of running times and dwell times.
Based on the general motivations behind service control and the author’s close observation of the Piccadilly Line control center, service control policies for minor disruptions on the Piccadilly Line are implemented, which are primarily train renumbering and short turning. These service control policies do not cover the entire gamut of possible controller interventions, but are meant to be representative of service control practice on the Piccadilly Line during minor incidents. The inclusion of service control in the simulation model allows for minor incidents to be simulated. It enables us to see how different schedules perform given the same kind of incidents, and this is the subject of Chapter 5.

In this chapter, the simulation tool is first validated by running it under a variety of operating conditions and looking at different indicators (such as train lateness, running times and headway data) to ensure that the output of the simulation is consistent with what is observed in practice. The effect of service control was demonstrated by simulating a disruption in the trunk section of the line, with and without service control interventions. In the latter case, train lateness persisted for much longer than when service controllers intervened to put trains and drivers back on time. The simulation was also used to highlight the differing impacts of incidents that occur on the branches as compared to incidents on the trunk portion of the line. All of these exercises served to verify and validate the simulation model of the Piccadilly Line.

The simulation model is then used to evaluate and compare the performance of Piccadilly Line crew schedules 48 and 49 under a reasonable distribution of incidents. Based on prior research that characterized the occurrence of incidents on the Piccadilly Line, distributions for the location, duration and time of the incident were obtained. The distributions of incidents that are simulated are designed to represent a reasonable level of stress on the crew schedules that allows us the comparison of robustness of the two crew schedules.

The comparison shows that Schedule 49 outperformed Schedule 48 on a number of key indicators such as the number of missed pickups, number of early reliefs and number of short-turns. These results could not have been predicted by merely comparing the structures of the two crew schedules, and therein lies the power of the simulation-based approach.

Upon closer inspection of the simulation results, we find that the average slack time across the missed pickups is significantly lower than the average slack time in the crew schedule. This suggests the hypothesis that increasing the minimum slack time improves robustness. We test a hypothetical scenario by relaxing the constraint on the length of the meal break to ensure that every duty has at least fifteen minutes of slack time. Ideally, this minimum slack would vary by time-of-day and relief location, and it would be enforced while creating the crew schedule and not just during the simulation by relaxing a constraint. A second hypothetical scenario is tested where
the walking time at Arnos Grove is reduced by 3-4 minutes. This is an issue that has been in negotiation for several years, and the simulation allows us to understand the impacts of the reduction on schedule performance. As expected, the performance of the schedule in both cases improves, and this demonstrates how the simulation model can be used to inform labour negotiations.

6.2 Recommendations

At the outset, the complexity of the problem must be recognized. Multiple teams such as planning, scheduling, service control and crew management come together to make the line's operations possible. Understanding this context, the following recommendations are made to the Piccadilly Line and the London Underground based on this research. The recommendations are divided into three topics: crew scheduling, service control and data collection.

6.2.1 Crew scheduling

- The importance of having the right mix of relief directions has been recognized both in practice by controllers over many years, and through the simulation. While it is desirable to have the majority of crew reliefs in the preferred directions, it is necessary to have reliefs of different types for operational flexibility. The simulation can be used as a tool to test and find the right mix of reliefs.

- The Piccadilly Line crew schedules have large amounts of slack time in the form of many duties with more slack time than can be reasonably used, and other duties that could use more slack time. However, to reallocate the slack time between duties it is necessary to provide the schedulers more freedom by relaxing other constraints which can only happen through labour negotiations. Some of the constraints that could be considered are: the fixed-link constraint at Arnos Grove, which is particularly onerous and might well be removed with the proposed splitting of the Arnos Grove crew base; the cap on the number of drivers who are allowed to book on/off remotely at the depots, resulting in other drivers needing to book on at the crew base and waste time travelling to the depot. The Piccadilly Line management might be interested in seeing if any of these labour agreements are worth re-negotiating by creating a crew schedule for a hypothetical scenario where the agreement is relaxed (or changed in some way) and using the simulation model to understand the impacts of that change.

- In reallocating slack time, it is also important to tie slack time to lateness. The average (and minimum) slack times for reliefs at a certain location and certain time of the day should depend on the typical train lateness observed at that
location at that time of day. The typical lateness can be calculated using Net-MIS data and using appropriate percentiles. For instance, management may wish to tie the minimum slack to the 90th percentile of lateness and the average slack to the 95th percentile of lateness. Such analysis assumes the availability of large quantities of reliable train tracking data which is not the case at present.

- When re-negotiating labour agreements it is advisable to understand the impacts that the negotiations will have on the crew schedule structure, and consequently on the schedule's performance. The simulation model can aid in that process. Such negotiations may be likely in the future because of the new crew base that has opened at Northfields, and another that may be opening shortly at Cockfosters/Oakwood. Though these changes would necessitate an update to the service control policies implemented in the model, the effort might be worth it to predict in advance, the impacts of these changes.

### 6.2.2 Service control

Section 2.4 showed the strong links between train service control and crew management, and described strategies which have been developed by the controllers on the Piccadilly Line to tie crew management constraints into their decision-making process. The problems surrounding crew management are deeply related to information availability and management. In addition to the continuous stream of real-time information on the train service which controllers compare to (mostly static) timetable information to make service control decisions, they need crew data to make effective control decisions. Some of this information is present in the crew schedules, but it is difficult to find in a high-stress situation. Controllers on the Piccadilly Line are currently not provided with any live crew data (such as their driving time, slack time etc.). At the same time the provision of information on crew movements and crew schedules in a raw format adds another level of complexity for the controllers which can be very difficult to handle. Moreover, controllers do not have direct authority over drivers, which are primarily the responsibility of the DMT (Duty Manager - Trains). Yet, since every change to a train trajectory is also a change to a driver's schedule, the situation in which controllers have insufficient information about and a lack of authority over crews places an unnecessary constraint on their flexibility when managing the service.

This problem is exacerbated by the fact that management of train service is centralized whereas crew management is decentralized and there is no central tracking of drivers, driver lateness and spare driver availability. The information about these variables is spread over the crew depots, without a continuous feed of information into the control center. It was observed that even at the level of the individual crew depots, DMTs often only track their drivers and driver lateness in the immediate vicinity of their crew depot. Drivers are often the only ones who know whether they
are late with respect to the crew schedule and whether they are at risk of violating driving time or duty length constraints. If controllers need to get a reliable picture of driver lateness and driver availability for unscheduled train movements, they generally have to request the information by telephone from the DMTs or by radio from the drivers. These communications are time consuming. Since many interventions need to be made under significant time pressure, this is a limiting factor which may cause a controller to choose a solution which is robust in terms of crew management or which minimizes the need for communication.

Many of the current service control policies are designed to deal with these shortcomings in the controller's environment, and given better information, it might well be possible to make better control decisions. For instance, the controller's decision on whether to put a late train back on time is triggered by the train's lateness which is assumed to be equal to the driver's lateness, when in fact the primary motivation for putting a train back on time comes from crew management concerns. It would be very useful to provide controllers information on the driver who is aboard each train, and what the driver's parameters are like. Even better than just providing information would be the provision of real-time decision support systems that process the information and provide controllers with useful outputs, such as the predicted impact of certain control actions. A system which anticipates conflicts and alerts controllers ahead of time can help them make decisions at an earlier stage, allowing more time for mitigating the impact of service control interventions.

### 6.2.3 Data collection

- The Piccadilly Line currently has very poor electronic crew management data, and it could gain significant value from more reliable crew data. Having better crew data can improve service control decisions, and can also help with better planning, analysis and scheduling. This could be achieved either by enforcing operator logins more strictly and establishing clear rules on how spare drivers log in or by the deployment of a personal chip card which drivers must insert into a reader in order to enable train controls.

- As preliminary work towards improving the modeling of line behavior after disruptions with TSM, which could ultimately be extended to include a rudimentary set of service control interventions, service controller logs should be digitized. It would be very helpful to understand what service controllers perceived as the incident, and what their interventions were. Currently, the Piccadilly Line service controllers maintain manual logs of the actions that they took while reforming the service. Having this information in a standardized digital format would allow for it to be used in computerized models, such as this simulation model and TSM. The primary reason for not explicitly validating the service control policies implemented in this model is the lack of this data.
Having this data would also allow the management or an analyst to look back in time to see how incidents were handled, and come up with best practices for service control.

- The quality of train tracking (NetMIS) data on the Piccadilly Line is poor. Very often the train number and/or trip number is incorrectly identified. This hampers service control, though the MEL system has been a workaround for this. The poor quality of NetMIS data also hampers planning and analysis, such as this simulation. Given better quality NetMIS data, it would be possible to generate more accurate train lateness distributions which can improve slack time allocation, and also make simulation models (such as this model and TSM) more accurate.

### 6.3 Limitations of this work

As with any model, simplifications of the real world system are made in the interests of model tractability, which impose limitations on the model’s use, and there is always scope for improvement. The following are some of the simplifications that this model makes that present opportunities for improvement:

- **Service control**: This model has embedded fairly simple service control policies. While these policies are fairly representative of controller interventions in response to minor incidents, they are certainly not comprehensive. For instance, this simulation does not cancel trains and stable them, which is a fairly common practice. Control strategies also vary by time-of-day with different strategies being used early in the morning and late at night. Also, there is a certain amount of stochasticity in control interventions because of variations in the thought process and styles of different controllers and because there are few fixed rules in place for service control. And finally, the Piccadilly Line plans to split the crew bases at Acton Town and Arnos Grove, and a new crew base at Northfields has already been opened. This will definitely impact control strategies as there are now more relief locations, and crew management is a major factor affecting control actions. For all these reasons, it is possible to make significant improvements to the service control module of this simulation.

- **Train movements**: This model uses a simplified route network that focuses on train movements at crew change and reversing locations. This was a deliberate design choice because the focus of the model is ultimately on the crew, and the objective was to develop a train movement model that was capable of modeling train lateness at crew change locations. Clearly, the train movement model can be improved by increasing the level of detail, by more carefully modeling the interaction between successive trains, by having more detailed station movement models, or by interfacing with micro-simulation models such as TSM.

- **Effect of incidents**: In this model, an incident is modeled as an event that
increases the running time of the affected train by a certain amount, which is the duration of the incident. In reality, the affected train experiences a further increase in running time because of higher than average dwell times at downstream stations due to the large gap in service preceding the affected train. This phenomenon is not explicitly modeled in this simulation. Instead, it is assumed that the duration of the incident accounts not only for the duration when the train was stopped, but also for this increase in running time. Modeling this phenomenon would improve the representation of the effects of the incident on train movement.

- **Spare drivers:** In this model, it is assumed that there is effectively an unlimited supply of spare drivers so that a train is never held in a station because of a missing driver. In reality, this situation, called an ONA (Operator Not Available) does occur. This is because service controllers do not have a continuous stream of information on (regular and spare) driver movements, and it is left to the Duty Managers at the crew depots to ensure that every train has a driver. Due to miscommunication, lack of information or other reasons, this does not always happen. Furthermore, when a spare driver is used instead of a regular driver (for instance, because the regular driver was late returning from a meal break), the regular driver often joins the pool of spare drivers for the rest of his/her duty. Service controllers tend not to rely heavily on spare drivers because of their uncertain availability. A more realistic representation of spare driver availability and utilization would improve the realism of the model.

- **Explicit modeling of incidents:** In this model, the running times and dwell time that are used are derived from undisrupted days, and incidents are superimposed by the simulation. This is an explicit way of modeling incidents. An implicit way of modeling incidents would be to use running times and dwell times from a large collection of days (say, a year) that includes the full gamut of operating conditions, including undisrupted days and days with varying levels of disruption. In doing so, the simulation would not explicitly generate incidents. The effect of the incidents would automatically be accounted for in the running time (and dwell time) distributions. These distributions would then have longer tails, and when randomly sampling from them, a larger running time would occasionally be picked, representing an incident. There are two major advantages to this approach: (i) there is no need to construct distributions of incidents and then draw from them, and (ii) there is no need to explicitly model the effect of the incidents on train movements. However, the advantage of explicitly modeling incidents is that we have the ability to simulate the effects of specific types of incidents, which we would be infeasible if incidents were modeled implicitly. However, implicitly modeling incidents is a very viable alternative approach to building this kind of model. This alternative was rejected in the case of the Piccadilly Line because of a lack of reliable train movement data that covers the full spectrum of operating conditions.
6.4 Directions for future research

The following are some directions for future research that emerge from this thesis:

- **Evaluating service control strategies**: Service control plays a pivotal role in the operations of a transit line, especially during disruptions. When an incident causes significant deviations from the operations plan, service controllers need to intervene in order to maintain or restore the feasibility of the operations plan, and minimize passenger and operational costs. In doing so, there are always trade-offs to be made. Certain control strategies, say canceling or short-turning many disrupted trains, might help restore service quickly, but at great passenger inconvenience. On the other hand, other strategies that aim to minimize passenger inconvenience due to the incident, might end up incurring large operational costs in the form of driver lateness and overtime costs. These trade-offs are especially stark during more severe disruptions. And while it is clear that a balance needs to be struck between the two objectives based on the management's service and labour policies, it is not clear what the best control strategy to do so is. The simulation can be used to evaluate the impact of different service control strategies in response to incidents. The performance measures in the model can easily be expanded to capture more passenger-centric metrics so that the trade-off between operational and passenger impacts can be evaluated.

- **Real-time decision support**: Service control is a centralized function that monitors the state of the system in real-time and responds to deviations from the operations plan. In the case of the Piccadilly Line, controllers have fairly good knowledge about the state of the trains in the system (i.e. location, destination, lateness etc.), though they have next to no information about the operators in those trains. They do not know, for instance, which driver is on a train, how close the driver is to his/her parameters etc. Some of this information can be inferred from the crew schedule, but it is difficult to do so under pressure. They often need to speak to the operator or a depot manager to get that information. While it may be possible to provide them with that information, it is not clear if a controller would be able to process it to make better control decisions, especially during the high-stress situations during disruptions. A real-time decision support system goes beyond just providing controllers with raw information, and instead synthesizes that information to provide useful tools in aiding decision making. For instance, it could predict the outcome of certain control actions, or even suggest control actions. To do so, it would be necessary to simulate different control scenarios in real-time and evaluate those scenario in terms of relevant metrics, which is exactly what this simulation model does.

- **Improving operations at crew bases**: Currently, the duty manager at each crew base is responsible for the drivers who report to there. Section 6.2.2 described the way that authority is shared and information flows between the crew
bases and the control centre, and the many problems with this structure. It is the author's belief that it is possible to significantly improve the operations at the crew bases, and that this might have a greater positive impact on service reliability than improvements to the crew schedule. Having an electronic system in place that monitors the state of all drivers, including where they are currently, how close they are to their parameters, and providing alerts about possible infeasibility in their trajectories would go a long way in streamlining crew operations at the bases. This information could be provided both to the duty managers at the depots, and to the service controllers and would help both parties make better decisions, and would reduce the amount of time wasted on unnecessary communication. Depending on how authority is split between the control centre and the crew bases, it may be desirable to have a decision support system for the crew bases that helps duty managers adapt to control interventions (such as short-turns and renumbering) and ensure that they are able to staff every train as required. Alternatively, this decision support system could be centralized at the control centre and it could give instructions to the duty managers that they would then have to implement.

Another important aspect of the operations at crew bases is the deployment of spare operators. While this was not a focus of this thesis, it is an area that has not received much attention in the past, and it may hold scope for potentially improving the service. As things stand, the deployment of spare operators is left to the duty managers at the depots. Service controllers tend not to rely on spare drivers if they can because they do not have reliable information on the availability of spare drivers. It would interesting to study how service control strategies can be improved if service controllers are aware of the exact availability of spare drivers at the start of each day. This also automatically leads to the question of whether the scheduling of spare drivers can be improved. The simulation model can potentially be used to test both questions.
Bibliography


