Application of Smart Card Fare Payment Data to Bus Network Planning in London, UK

by

Catherine Whitney Seaborn

B.A. in Economics and Geography
McGill University
Montreal, Canada (2002)

Submitted to the Department of Urban Studies and Planning and the Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degrees of

Master in City Planning

and

Master of Science in Transportation

at the

Massachusetts Institute of Technology

June 2008

Author ....
Department of Urban Studies and Planning
Department of Civil and Environmental Engineering
May 19, 2008

Certified by..
Professor Nigel H.M. Wilson
Department of Civil and Environmental Engineering

Accepted by..
Professor Daniele Veneziano
Chair, Departmental Committee for Graduate Students
Department of Civil and Environmental Engineering

Accepted by..
Professor Langley Keyes
Chair, Master in City Planning Committee
Department of Urban Studies and Planning

Application of Smart Card Fare Payment Data to Bus Network Planning in London, UK

by

Catherine Whitney Seaborn

Submitted to the Department of Urban Studies and Planning and the Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degrees of Master of City Planning and Master of Science in Transportation at the Massachusetts Institute of Technology (June 2008)

Abstract

This research contributes to an emerging body of literature on the application of smart card data to public transportation planning. It also addresses interchange planning, a key component of public transportation network planning that has lately received renewed interest as evidenced by planning literature on integrated and intermodal networks. The research objective of assessing the potential application of smart card data to bus network planning is met, first, by a review of existing data systems and the bus network planning approach at Transport for London (TfL). Then, three potential interchange combinations: bus-to-Underground, Underground-to-bus, and bus-to-bus are examined to gain an understanding of interchange behavior in London and to formulate recommendations for elapsed time thresholds to identify interchanges between journey stages for each passenger on the TfL network. Other TfL data are compared with the results of linking journey stages into complete journeys based on these elapsed time thresholds. Finally, the complete journey data are applied to bus network planning case studies that illustrate the value of new contextual and quantitative information that would be available to network planners using smart card data without necessitating additional data from other sources such as Automated Vehicle Locators or Automated Passenger Counters.

Recommended elapsed time thresholds for identifying interchanges across the London network are: 20 minutes for Underground-to-bus interchanges, 35 minutes for bus-to-Underground interchanges, and 45 minutes for bus-to-bus interchanges, but a range of values that account for variability across the network are provided. Key findings about intermodal travel behavior in London include evidence of complex bus travel patterns during the Midday and PM Peak time periods and of land uses near Underground stations influencing interchange behavior. Moreover, complete journey data result in approximately 2.3 daily public transportation journeys per passenger, 1.3 journey stages per public transportation journey, and 25 percent of Underground journeys including a linked bus journey stage. Finally, examples of new contextual information for bus network planning include connectivity of bus routes, intermodality of bus journeys, duration of bus access journeys, and duration of Underground-to-bus interchanges. New quantitative information about the number of passengers transferring between two bus routes can be used in cost-benefit analyses of service changes to the bus network.

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

Acknowledgements

I would like to thank Nigel Wilson and John Attanucci for their guidance, support and attention to detail from beginning to end. Also, thanks to Karen Polenske and Christopher Zegras for their research support and mentorship during my time at MIT. I am grateful to the staff at Transport for London, especially Rosa McShane, John Barry, and Hélène Bataille for providing the motivation for this work. Michael Frumin, Peter Lau, and Martin Milkovits provided indispensable technical assistance. My classmates from the MCP and MST programs have helped to keep everything in perspective, and I cannot imagine the past three years without my teammates on sMITe. Thank you all. Any remaining errors or omissions are entirely my own.

Table of Contents

List of Acronyms 7
List of Tables 8
List of Figures 9
CHAPTER 1: INTRODUCTION 11
1.1 The Case for Public Transportation Network Planning 11
1.2 Research Objectives and Approach 15
1.2.1 The London Public Transportation Network 17
1.3 Literature Review 19
1.3.1 Smart Card Data Analysis 19
1.3.2 Passenger Interchange 21
1.4 Thesis Organization 24
CHAPTER 2: BUS NETWORK PLANNING IN LONDON 25
2.1 Planning Objectives 25
2.2 Existing Data Systems 28
2.3 Smart Card Data Systems 35
2.3.1 Data Source and Format 37
2.3.2 Data Quality 39
2.4 Summary 43
CHAPTER 3: CHARACTERSITICS OF BUS PASSENGER INTERCHANGE 45
3.1 Definition of Interchange 45
3.2 Journey Stage Patterns 48
3.3 Underground-to-Bus Interchanges 53
3.3.1 Time Periods 54
3.3.2 Fare Zones 56
3.3.3 Underground Stations 57
3.3.4 Summary 60
3.4 Bus-to-Underground Interchanges 61
3.4.1 Time Periods 62
3.4.2 Fare Zones 63
3.4.3 Underground Stations 65
3.4.4 Summary 67
3.5 Bus-to-Bus Interchanges 67
3.5.1 Time Periods 69
3.5.2 Pay-As-You-Go versus Pass Holders 70
3.5.3 Summary 72
3.6 Summary 73
CHAPTER 4: MULTI-STAGE JOURNEYS BY BUS AND UNDERGROUND 75
4.1 Assumptions and Methodology 75
4.2 Results of Linking Journey Stages to Form Complete Journeys 79
4.2.1 Total Daily Journeys 79
4.2.2 Public Transportation Journeys per Passenger 80
4.2.3 Stages per Public Transportation Journey 83
4.2.4 Modal Patterns 84
4.3 Summary 86

CHAPTER 5: APPLICATION OF COMPLETE JOURNEY INFORMATION TO BUS NETWORK PLANNING 87
5.1 Expected Findings 87
5.2 Case Studies of Bus Passenger Interchange 89
5.2.1 Comparison of Bus Route Case Studies 92
5.2.2 Route 293: Suburban Underground Station Access 93
5.2.3 Route 221: Suburban Connector 98
5.2.4 Route 69: Complex Urban Link 104
5.2.5 Comparison of Underground Station Case Studies 109
5.2.6 Vauxhall Station 111
5.2.7 North Greenwich Station 114
5.3 Summary 119
CHAPTER 6: IMPLEMENTATION AND CONCLUSIONS 121
6.1 Research Contributions 121
6.1.1 Interchange Time Thresholds 122
6.1.2 Intermodal Travel Behavior in London 123
6.1.3 New Information for Bus Network Planning 124
6.2 Implementation of Complete Journeys Information Using Smart Card Data 125
6.3 Future Research 129
REFERENCES 131
Appendix 1 SQL Code for Identifying Complete Journeys 135
Appendix 2 List of Routes Missing from Analysis 141

List of Acronyms

APC – Automated Passenger Counter

AVL – Automated Vehicle Locator

BCMS – Bus Contract Management System

BODS – Bus Origin-Destination Survey

BREMS – Bus Revenue and Mileage System

ETM – Electronic Ticketing Machine

GIS – Geographic Information System

GLBPS – Greater London Bus Passenger Survey

LATS – London Area Travel Survey

LTDS – London Travel Demand Survey

PAYG – Pay-As-You-Go

TfL – Transport for London

List of Tables

Table 2-1 Data Systems and Needs for Bus Network Planning 29
Table 2-2 Comparison of Oyster Bus Journey Stages with Average Weekday from GLBPS .. 40
Table 2-3 Comparison of Weekly Oyster Bus Journey Stages with BREMS 41
Table 3-1 Top 10 Journey Stage Patterns per Passenger 51
Table 3-2 Recommended Elapsed Time Thresholds for Identifying Interchanges 73
Table 4-1 Ranges of Elapsed Time Thresholds for Identifying Interchanges 78
Table 4-2 Total Daily Journeys Relative to Expected Value 80
Table 4-3 Average Daily Public Transportation Journeys per Passenger 81
Table 4-4 Average Stages per Complete Journey by Elapsed Time Thresholds 84
Table 4-5 Share of Underground Journeys with Bus Stages by Elapsed Time Thresholds 85
Table 4-6 Summary of Metrics from Oyster-based Complete Journeys and LTDS 86
Table 5-1 Daily Passenger Volumes on Case Study Routes 90
Table 5-2 Modal Combinations of Journeys on Case Study Routes 93
Table 5-3 Probable Return Journeys on Case Study Routes 93
Table 5-4 Top 10 Routes in Journeys on Route 293 96
Table 5-5 Top 10 Routes in Journeys on Route 221 100
Table 5-6 Modal Patterns of Journeys on Route 221 with Underground Interchange 102
Table 5-7 Daily Interchange Volumes between Route 221 and Top 3 Underground Stations 103
Table 5-8 Top 10 Routes in Journeys on Route 69 106
Table 5-9 Top 5 Stations in Journeys on Route 69 107
Table 5-10 Individual Journey Pattern on Route 69 109
Table 5-11 Modal Patterns of Journeys through Vauxhall and North Greenwich Underground Stations 110
Table 5-12 Return Journeys through Vauxhall and North Greenwich Underground Stations 110
Table 5-13 Top 7 Routes in Journeys through Vauxhall Underground Station 111
Table 5-14 Top 7 Routes in Journeys through North Greenwich Station 116

List of Figures

Figure 1-1 Two Approaches to Interchange in Public Transportation Networks 13
Figure 3-1 Conceptual Diagram of Elapsed Time Thresholds versus Interchange Type 47
Figure 3-2 Bus and Underground Journey Stages per Day by 15-Minute Interval.................. 49
Figure 3-3 Journey Stages per Passenger 50
Figure 3-4 Journey Stage Start Times, Passengers with Two Underground Journey Stages.... 52
Figure 3-5 Journey Stage Start Times, Passengers with Two Bus Journey Stages.................. 52
Figure 3-6 All Potential Underground-to-Bus Interchanges 54
Figure 3-7 Potential Underground-to-Bus Interchanges Across Time Periods........................ 55
Figure 3-8 Potential Underground-to-Bus Interchanges by Fare Zone 57
Figure 3-9 Potential Underground-to-Bus Interchanges at Highest Exit Volume Stations 58
Figure 3-10 Potential Underground-to-Bus Interchanges at Highest Interchange Volume Stations 60
Figure 3-11 All Potential Bus-to-Underground Interchanges 62
Figure 3-12 Potential Bus-to-Underground Interchanges by Time Period 63
Figure 3-13 Potential Bus-to-Underground Interchanges by Fare Zone 64
Figure 3-14 Potential Bus-to-Underground Interchanges, Highest Entry Volume Stations 66
Figure 3-15 Potential Bus-to-Underground Interchanges, Highest Interchange Volume Stations 66
Figure 3-16 All Potential Bus-to-Bus Interchanges 68
Figure 3-17 Potential Bus-to-Bus Interchanges Across Time Periods.................................... 70
Figure 3-18 Potential Interchanges, Passengers with Two Bus Boardings on Different Routes 71
Figure 3-19 Potential Interchanges, Passengers with Two Bus Boardings on Same Route 71
Figure 4-1 Daily Public Transportation Journeys per Passenger, LTDS vs. Oyster Complete Journeys 82
Figure 4-2 Daily Public Transportation Journeys per Passenger for Time Thresholds within and Outside Suggested Ranges 83
Figure 4-3 Stages per Complete Journey for Sets of Time Thresholds vs. LTDS 85
Figure 5-1 Images of Vauxhall and North Greenwich Stations 91
Figure 5-2 Map of Route 293 in South West London 94
Figure 5-3 Elapsed Time Distributions for Interchanges between Routes 293 and 93 95
Figure 5-4 Elapsed Time Distribution for Interchanges from Route 293 to Morden Station ... 97

Figure 5-5 Elapsed Time Distribution for Interchanges from Morden Station to Route 293....98
Figure 5-6 Map of Route 221 in North West London 99
Figure 5-7 Elapsed Time Distributions for Interchanges between Routes 221 and 263 101
Figure 5-8 Elapsed Time Distribution for Interchanges from Route 221 to All Underground Stations..103
Figure 5-9 Map of Route 69 in North East London...105
Figure 5-10 Elapsed Time Distributions for Interchanges between Routes 69 and 55 108
Figure 5-11 Elapsed Time Distributions for Interchanges between Routes 69 and 34 108
Figure 5-12 Map of Bus Routes Near Vauxhall Station..112
Figure 5-13 Elapsed Time Distributions for Interchanges from Routes 36 and 344 to Vauxhall Underground Station...113
Figure 5-14 Elapsed Time Distributions for Interchanges from Vauxhall Station to Routes 36 and 344...114
Figure 5-15 Map of Bus Routes Near North Greenwich Station.......................................115
Figure 5-16 Elapsed Time Distributions for Interchanges from Routes 472 and 486 to North Greenwich Station...117
Figure 5-17 Elapsed Time between Bus Boardings for All Routes to Route 472 and Reverse in North Greenwich Access or Egress Journeys 118

CHAPTER 1: INTRODUCTION

This research explores the application of smart card data to transport planning with a focus on bus interchanges using London as an example. To assess the potential application of smart card data to bus network planning in London, existing data systems and the approach to bus network planning taken at London Buses are reviewed. Then, three potential interchange combinations: bus-to-Underground, Underground-to-bus, and bus-to-bus are examined to gain an understanding of interchange behavior in London and to formulate recommendations for elapsed time thresholds to identify interchanges between journey stages for each passenger. Then, the outcomes of linking journey stages into complete journeys based on different elapsed time thresholds within these ranges are compared. Finally, complete journey data is applied to network planning case studies that illustrate how this new information that will be of value to bus network planners in London.

To set this work in context, this chapter begins with a discussion of the case for public transportation network planning (Section 1.1), further elaborates on the research objectives (Section 1.2), and finally presents a literature review dealing with the application of smart card data to network planning as well as the role of interchanges in public transportation planning.

1.1 The Case for Public Transportation Network Planning

In cities around the world, public transportation systems provide people with communal transportation services, most commonly by bus or train. Whether the service is managed and delivered by public or private sector agents, or some combination thereof, it is usually organized into a local network serving an urbanized area defined by political boundaries. Of course, public transportation networks also interact with and are influenced by other daily transportation options such as driving, walking, and cycling.

We are interested in studying and improving public transportation networks for three reasons: economics, the environment, and quality of life. First, it is difficult to imagine cities such as New York, London, or Tokyo functioning as global economic centers without underground rail lines to bring commuters into financial districts that are at the heart of their economic strength. If instead these commuters were to drive to work and therefore require space to park, the agglomeration economies associated with tightly packed employment centers served by public transportation would be greatly diminished. Conversely, greater dependence on non-motorized options such as walking or cycling would be difficult because of the resulting need for housing in central business districts. For most cities of regional economic importance, investment in public transportation is an essential means of connecting people with their jobs in high-density urban areas and thus promoting economic growth (Graham 2007; Banister and Berechman 2001).

Second, climate change and urban air pollution both predicate a need to find viable alternatives to private automobiles as the dominant form of travel in developed nations, as well as solutions to the traffic congestion in most major cities around the world. The provision of high quality, ubiquitous public transportation services, in combination with other policies such as congestion charging, high-density land development, and gasoline taxes to encourage substantial changes in travel behavior, is an important part of the solution to these pressing environmental problems (Hickman & Banister 2007, Monbiot 2006).

Third, public transportation can improve the quality of life in our cities. It provides an alternative means of accessing a wide range of employment and social activities for many people who cannot or choose not to drive. Public transportation also allows for and is made viable by denser land use patterns that often entail quality of life benefits such as more shopping and housing options within a neighborhood (Newman and Kenworthy 1996). Given the general benefits of public transportation and the positive feedback effect of providing high quality service, we might ask what makes a good public transportation network.

One challenge in definitively characterizing a good public transportation network is that the appropriate design depends largely on passenger demand, which in turn depends on factors such as service quality[1], land use density, and demographics. Moreover, the network layout is most often constrained by the existing spatial structure of the region it serves. Certainly, we observe some cities with a grid network structure, for example Toronto and Mexico City, and others with a predominant radial structure, for example Chicago's rail and Munich's S-Bahn regional rail system. Regardless of overall structure, a primary objective in public transportation planning should be to provide direct service between trip origins and destinations for which there is sufficient demand, for example from high-density suburbs to the central business district (The Institute of Transport and Logistics 2000). What happens

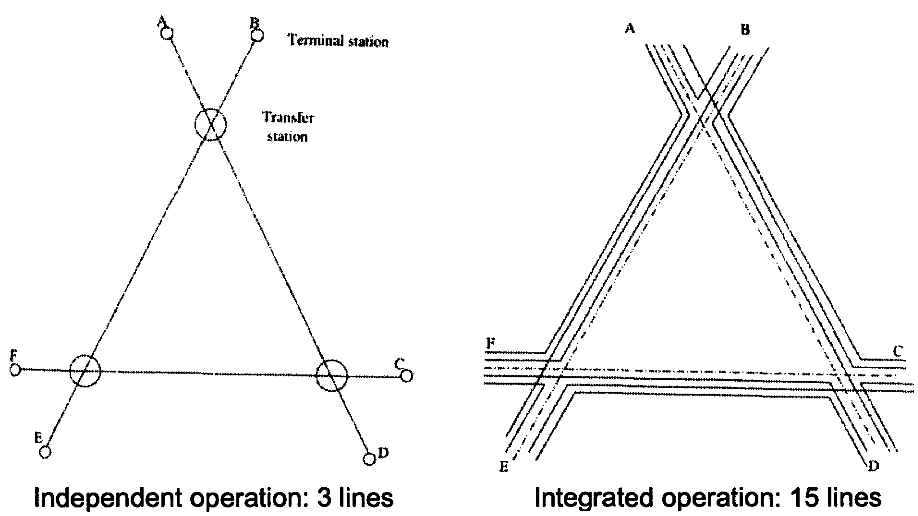
---

[1] Service quality is generally measured in terms of travel time, frequency of service, reliability, and comfort.

when there is insufficient demand to justify direct service according to the transport authority's network planning criteria? Then, people who choose to travel by public transportation may have to transfer, or interchange[2], between modes or between different services of the same mode (Transport for London 2001). (Note that in this case, 'interchange' refers to the act of transferring, but it may also refer to the physical location where interchange occurs, which may or may not be a purpose-built facility. The distinction should be evident from the context.)

Vuchic (2005) characterizes two planning approaches to the design of public transportation networks, which address interchange differently. The integrated (or overlapping) transit line approach avoids the need for passengers to interchange by creating as many direct bus or rail services as possible. This approach is illustrated conceptually in Figure 1-1 (reproduced from Vuchic (2005)) which shows that 15 lines are needed to connect all 30 origin-destination pairs for the 6 destinations, but that passengers using such a system would never have to interchange. On the other hand, the independent transit line approach, also illustrated in Figure 1-1, connects all 6 destinations using only 3 transit lines if passengers are willing to interchange once. The inconvenience of interchanging may be mitigated by more resources being available to invest in higher frequency service on the 3 lines as opposed to running the 15 lines in the integrated operation.

Figure 1-1 Two Approaches to Interchange in Public Transportation Networks



Source: Vuchic, V. R. 2005, *Urban Transit: operations, planning, and economics*, John Wiley & Sons, Hoboken.

[2] 'Interchange' is used because it is the preferred term at Transport for London, which is the focus of this thesis.

Of course, most public transportation networks have some elements of both independent and integrated lines. For example, Vuchic (2005) characterizes the London underground rail network as "moderately overlapping" (pp. 196). This, he argues, is good because it provides the benefits of integrated lines in terms of more diversified service and more direct services, thereby reducing the number of passenger interchanges, but there is enough distinction between independent lines so as to avoid confusion by passengers. In either case there are trade-offs between attracting passengers and maximizing operating efficiency, and finding a good balance between these two objectives is a key element of public transportation network planning.

Interchange planning, in terms of network structure, scheduling and the quality of the interchange facility, is just one aspect of public transportation service delivery. Nonetheless, it is an extremely important consideration as the need to interchange is believed to be a strong deterrent to people choosing public transportation over the door-to-door convenience of traveling by car (Guo and Wilson 2006, Wardman and Hine 2000). Analyzing where and when people chose to interchange in their daily travels provides insight into their behavioral response to this inconvenience of public transportation. On the other hand, interchange facilities may also be viewed as a positive element of public transportation networks because they allow for a greater range of travel paths. Some argue that good interchange facilities, characterized by elements such as frequent and reliable service, ease of access, shops and other amenities, and passenger safety, actually enhance the public transportation experience for passengers so that networks should be planned around interchanges rather than seeking to avoid them (Vuchic 2005, The Institute of Transport and Logistics 2000). If you consider this view, then it would be helpful to analyze how passengers behave in existing networks and to evaluate the quality of their interchange experiences. To summarize, public transportation networks may be improved both by reducing the need to interchange and, where necessary, enhancing the quality of the interchange experience for passengers, thereby encouraging increased demand which then generates a positive feedback effect on the economic, environmental, and quality of life benefits described above.

A comprehensive analysis of interchanges on a network-wide basis in a city such as London is beyond the scope of this research. However, new data available from smart card transactions may be useful in the analysis of key elements of interchange behavior for selected routes and stations. As outlined in the following section, this thesis attempts to develop new methods for analyzing if and how it may be appropriate to reduce the need for passengers to interchange and, where the network is planned around interchange, how to improve the interchange experience for public transportation passengers.

1.2 Research Objectives and Approach

The goal of this research is to test the hypothesis that smart card fare payment data can be a valuable input in improving bus network planning by providing new contextual and quantitative information on passenger demand between routes and with Underground stations. Thus, the research is focused on bus passenger interchange behaviour and how bus services interact with the rest of the network at interchange points. The recent expansion of bus services and associated 52 percent growth in ridership between 2001 and 2007 are testimony to the high quality of the network in London; however this growth has been achieved with only modest advances in the methods and data systems used for network planning (see Chapter 2) (Transport for London 2008).

Smart cards, such as the Oyster card in London, are owned by individuals and generally record the time and place of every transaction the card holder makes on the public transportation system (e.g., bus boarding, Underground station exit, etc.). There are several types of analyses that can be done with smart card data, including estimating origin-destination matrices, measuring passengers' behavioural reactions to service changes, and evaluating service quality through travel patterns (see 1.3 Literature Review). The key contribution of this research, however, is to develop a methodology for describing aggregate passenger interchange behaviour to, from and within the bus network in London using smart card data and to offer a framework for implementation in the context of London Buses.

Interchange is of particular interest in London because of the multitude of modes, complex travel patterns and an organic urban spatial structure that has evolved from historic town centres. The staff at London Buses is interested in how the bus network serves both Underground and National Rail stations, however due to data scarcity the focus of this research is on the interaction between bus and Underground services in London. Nevertheless, the methodology developed is expected to be applicable to National Rail in the near future as Oyster card usage increases on surface rail services.

The analysis of smart card data for public transportation planning is a new and emerging area of study (see 1.3 Literature Review). In London, there are specific questions relating to bus network planning that may be answered with the help of smart card data, for example:

- What would the impact of free bus-to-bus transfers be on current pay-as-you-go riders?
- What is the impact of bus service improvements on the number of people interchanging at an Underground station?
- What is the typical duration of bus journeys to access Underground or rail stations?
- What is the average number of interchanges passengers are willing to make as a function of journey duration?
- What can variations in interchange time tell us about the quality of interchange facilities?

Due in part to the contribution of academic research, transport agencies are slowly integrating smart card data analysis into their daily operations and planning, for example at the Seoul Metro Company and the London Underground at Transport for London (TfL) (Park and Kim 2008, Transport for London 2007a). By beginning to address some of the questions above, this research will contribute to that effort as well as provide new information about bus passenger interchange in London. To the best of my knowledge, smart card data has not been used to study how passengers travel across multiple modes in London, although some similar analysis has been done in Chicago (Cui 2006).

In order to test the hypothesis that smart card data can be used to improve bus network planning and monitoring at TfL, two data samples are used: (a) a 5% sample of all Oyster cards for one month, and (b) a 100% sample of all Oyster cards for one day during that month (see Chapter 2). The general approach involves querying and summarizing millions of records, each record representing one journey stage, of Oyster card transaction data in a systematic way, and then comparing these results with existing data sources and contextual knowledge. To begin with, journey stage volumes for bus and Underground over the course of a day are examined to understand how the network is used by passengers at an aggregate level. Then, time differences between journey stages for each passenger are analyzed in order to determine which stages to link together into complete journeys, which varies depending on the modal combination. Next, common features of complete journeys across the network are described, for example average number of public transportation stages per journey. Finally, in order to address interchange in a network planning context, case studies of interchange behaviour at two purpose-built bus-Underground stations, Vauxhall and Stratford, are presented. In addition, case studies of three bus routes are provided in order to illustrate the application of new information on complete journeys (i.e., including interchange) to bus network planning. Finally, this thesis concludes with recommendations for implementation in the context of bus network planning at TfL.

1.2.1 The London Public Transportation Network

This thesis uses the London public transportation network, defined by the geographic area controlled by the Greater London Authority (GLA), as a basis for analyzing the application of smart card fare payment data to bus network planning with a focus on interchange behavior. In this section, the basic characteristics of the London network are described and then the general applicability of the research results to public transportation planning in other cities is discussed. The results will, at a minimum, be of interest to planners and policymakers in the London context but similar analyses could be of value in many other large multi-modal public transportation networks having modern automated fare collection systems.

In 2006, 28 million journey stages[3] were made per day in Greater London, which has an area of 1,584 square kilometers and a population of about 7.5 million residents (Transport for London 2008, Greater London Authority 2005). Remarkably, slightly more than 37 percent of these journey stages were made using public transportation, including bus or tram (19%), Underground (10%), and rail (8%). In addition, public transportation represents about two-thirds of all weekday, work-related journey stages, demonstrating its critical role in the economy of the city. Finally, journey stages by public transportation grew 18 percent between

[3] "'Journey stages' refers to the component parts of a complete trip between transport interchanges. Thus, a journey stage is made by a single mode of transport [and vehicle…] within a trip that may comprise several journey stages by different modes [and vehicles]." (Transport for London 2008, pp. i) For example, every time a passenger boards a different bus, it is counted as a separate journey stage. In 2006, 28 million journey stages represented over 24 million trips or journeys.

2000 and 2005 while travel by private vehicle peaked in 2002 and has since declined (Transport for London 2007b).

The London public transportation network is best described as complex; a result of historic development patterns, a multitude of modes and the dense built environment of the largest city in Europe. In fact, London's network geometry is so complex that Vuchic (2005) is unable to categorize it and simply describes it as "irregular" (pp.245). Few cities can boast a public transportation network that includes: 275 subway stations on 12 lines, 8,200 buses serving over 700 routes, light rail, tram lines and ferry services, 10 fare zones, and 8 major central heavy rail termini (Mayor of London 2007).

Despite the ongoing challenge of keeping the existing infrastructure in a state-of-good-repair, major network expansions are being realized by TfL in response to projected population and economic growth as well as the 2012 Olympics. The Greater London public transportation network is directed by TfL, delivered by public-private partnerships for the Underground, and operated by private companies under contract for buses. Most overground rail services are regulated by the national Office of Rail Regulation and delivered by private rail companies. One recent success in London has been the widespread adoption of the Oyster smart card fare payment system, which was used for 73 percent of journeys made on London's public transportation network in 2007 (Mayor of London 2007).

With reference to the previous discussion of independent versus integrated transit lines, the London Underground network was described as moderately integrated. The bus network might be described as extremely integrated (or overlapping), which may be due in part to the limited road space available for bus services that results in many routes running on shared corridors. Nonetheless, the London network certainly includes a plethora of interchange opportunities. In a recent report, TfL categorized and evaluated over 600 interchange facilities (Transport for London 2002) and in addition to these facilities there are innumerable intersecting bus routes that provide opportunities for on-street transfers. In short, like public transportation networks in many large cities, it is extremely difficult to characterize the London network according to a simple model, thus providing further impetus for using interchange as a focal point to gain insight into how passengers exploit the network.

Many of the challenges faced by TfL, including maintaining the existing network in a state-of-good-repair, taking advantage of changing technology, and encouraging a positive shift in residents' attitudes toward public transportation are common to most, if not all, transport authorities. So, although the statistics and results about bus passenger behavior presented in Chapters 3 through 5 are specific to London, Chapter 6 draws broader lessons from the application of smart card data to public transportation planning in complex urban environments. The widespread adoption of the Oyster smart card fare payment system in London and the potential application of resultant data to bus network planning would likely be of interest to other transport agencies implementing similar fare payment systems around the world.

1.3 Literature Review

This literature review is divided into two parts because this thesis is informed by previous research on both the use of smart card data in public transportation planning and various approaches to interchange planning and quality assessment. Smart card data analysis is a new and emerging theme in the transportation literature, especially in North America where it has been combined with Automated Vehicle Locator (AVL) and Automated Passenger Counter (APC) data to estimate origin-destination matrices. However, smart cards have been adopted by approximately 22 public transportation agencies in Europe and it may be that they are moving forward with implementing systems to take advantage of the new data without publicizing their experiences. More than 30 cities in Asia are using smart card fare payment systems for public transportation but studies of the application of the resultant data to network planning are only beginning to emerge (Park and Kim 2008). On the other hand, interchange planning has long been recognized as a key component of public transportation network planning as described in Section 1.1, but new literature on how to approach interchange planning as well as passenger attitudes toward interchanges (i.e., the "transfer penalty") have recently emerged in the United Kingdom and to some extent in North America.

1.3.1 Smart Card Data Analysis

Using the Chicago Transit Authority (CTA) as an example, Utsunomiya, Attanucci and Wilson (2006) discuss the potential usage of and barriers to increased data availability after smart card implementation in public transportation agencies, concluding that agencies need to tailor their smart card implementation plan to make the most of the increased data availability it offers and that smart card penetration as a fare payment method is the key to its effective use for the analysis of passenger behavior. Bagchi and White (2004) examine three cases of smart card implementation in bus networks in the United Kingdom. They find that the advantages of smart card data include larger samples than existing data sources and the ability to analyze travel behavior over longer periods, but there are also limitations, particularly in the case of bus travel in which cards are only validated upon entry to the system (i.e., bus boarding). Certain types of information such as journey purpose are absent from smart card data and would have to be collected through other methods. Therefore, they conclude that smart card data cannot replace existing survey methods for data collection but may complement them. In addition, the authors estimate smart card turnover rates and trip rates per card, and infer the proportion of all bus boardings to linked trips (i.e., with bus-bus transfers) in each network. For the small areas covered by the cards in their study, Bagchi and White (2004) link together two bus journey stages that begin within 30 minutes of each other as recorded by smart card transactions, but assert that in larger cities a wider time window would be needed to identify complete trips. They refer to a similar study for a larger city by Hoffman and O'Mahony (2005) in which 90 minutes is used to link bus journey stages as recorded by magnetic stripe electronic ticketing technology. However, the highest rate of interchange in the Hoffman and O'Mahony (2005) study occurred between 18 and 28 minutes after boarding the first bus. Bagchi & White (2004) find a typical implied ratio of boardings to complete (linked) trips of 1.25 and Hoffman and O'Mahony (2005) find a similar ratio of 1.21. As a final example of interchange identification based on time windows, Okamura, Zhang and Akimasa (2004, cited in Trepannier and Chapleau 2006) define an interchange as two journey stages that are provided by different operators and occur within 60 minutes at the same location. They use this definition to analyze interchange wait time at major transit hubs.

Trépanier, Tranchant and Chapleau (2007) develop a model to estimate the destination of individual bus passengers using smart cards based on two basic assumptions: (1) a passenger's journey stage destination is the first stop of their following journey stage, and (2) at the end of the day, passengers return to the stop where they first boarded. The authors' trip destination estimation experiment has a success rate of 66 percent in the first application but reaches 80 percent in peak hours when there is more trip regularity. However, they find that the destination estimation is not straightforward and that significant data pre-processing is required. In this application, the transit agency estimates that 80 percent of journeys are captured by smart card transactions. Further to this work, Chu and Chapleau (2007) develop methods for enriching smart card data for transit demand modeling including inferring the arrival time of bus runs at the stop level using schedule constraints and linking journey stages based on both location and time constraints. Thus, they avoid the need to make arbitrary interchange time assumptions, but the methodology is complex and computationally intensive.

Perhaps the most fully developed application of smart card data to bus network planning is provided by Cui (2006) who demonstrates a methodology for estimating a bus origin-destination matrix for Chicago using supplementary data from APC and AVL systems. These additional data sources are necessary to determine precise bus boarding locations and to infer passenger alightings. Cui concludes that bus origin-destination estimation using smart card data provides benefits in terms of reduced cost, larger sample size, and a more automated system but that it should be combined with targeted surveys to validate the estimation and obtain additional socio-economic and trip purpose information.

On the rail side, Zhao (2007) uses bus journey stages with associated AVL data to infer rail stage destinations in the development of an algorithm for rail origin-destination estimation using smart card data in Chicago. The basic inference is that the destination of one journey stage is also likely to be the origin of the next stage. In London, Chan (2007) estimates an origin-destination matrix for the Underground using station entry and exit information from smart card data, and concludes that this method can reduce survey costs by shifting to targeted surveys for the calibration of Oyster-based estimation at specific locations. Finally, Park and Kim (2008) analyze the reliability of smart card data as a basis for describing characteristics of public transportation users in Seoul and conclude that the results are not statistically different from survey data on travel volumes.

1.3.2 Passenger Interchange

The literature review on passenger interchange continues the earlier discussion of the role of interchange within public transportation networks. It includes academic studies, educational references and policy documents from the United Kingdom.

Interchange is an important component of most transit networks that may be considered from many perspectives, including passenger experience, operator priorities, network design, scheduling and the layout of transfer stations. As stated by Vuchic (2005), "All transit systems offering all-day area-wide services rely heavily on passenger transfers among lines and modes [...] Inadequate transfer arrangements usually create major obstacles to present passengers and discourage potential passengers from using transit." (pp. 223) Indeed, the need to interchange has been shown to be a significant barrier to people choosing to travel by public transportation instead of the door-to-door convenience of traveling by car (Hine and Scott 2000). Conversely, in their review of the literature on the costs of interchange, Wardman and Hine (2000) point to studies that argue that quality interchange facilities create an opportunity to provide a greater number of travel paths, or "journey opportunities", to public transportation users. Vuchic (2005) confirms this view and adds that "[i]n recent decades, there have been many innovations in facilitating transfers, such as new designs of intermodal stations, increased use of escalators, better information, and joint fares." (pp. 223) However, Wardman and Hine (2000) recommend that more emphasis be put on the bus market in analyzing interchange values and the behavioral response to interchange because the vast majority of the research on travel behavior to date has been conducted in the rail industry.

Crockett (2002) argues that three main elements need to be considered in assessing and improving transit service connectivity: system elements, facility elements, and service elements, and that service elements have the greatest potential for cost effective improvements to transfer (or interchange) time. Crockett also suggest that system elements may be assessed by volumes of transfer types for each area of the system and that service elements may be assessed by examining real connection times as compared to scheduled times. Crockett cites a lack of reliable quantitative data on travel patterns at interchange locations as a limitation to her work.

The Institute of Logistics and Transport (ILT) in the UK published a report in 2000 about passenger interchanges in response to the Government's White Paper, *A New Deal for Transport* that brought attention to this key element of integrated transport. The report discusses interchange penalties and behavioral attitudes toward interchange from a policy perspective, and then details network factors, physical factors, and finance issues related to quality interchange provision. The network and physical factors closely mirror Crockett's elements of improved transit service connectivity, with service elements included in the network factors. Accompanying this report, ILT has produced a manual, *Joining up the Journey* (2000) that provides practical guidance to local authorities on network assessment and improving interchange facilities. In the United States, recent research by Smart (2007) asks what makes a good stop, station or transfer facility from the perspective of transit service operators, finding that safety, security and the absence of conflicts between pedestrians and vehicles are the most important elements followed by ease of transferring and cost-effectiveness.

Given the UK national policy context, it is not surprising that improving public transportation interchange in London is a key element of the *Mayor's Transport Strategy* (GLA 2001). In order to attract more people to public transportation, policymakers in London emphasize better system integration, including planning for improved interchange services and facilities. The TfL *Interchange Plan* (Transport for London 2002) follows on another report (Transport for London 2001) about best practice guidelines for intermodal transport interchange in London that was published by a consortium of public sector agencies. The *Interchange Plan* categorizes 614 interchange facilities into five groups ranging from major central London termini to interchanges of local importance and then prioritizes them for infrastructure improvement. The prioritization is based almost entirely on a qualitative analysis of the discrepancy, or 'Quality Gap', between (a) the strategic importance of the facility based on policy objectives and (b) the physical quality of the facility based on a Mystery Shopper Survey (MSS). The report states that reliable data on such basic metrics as the number of people interchanging at each facility was not available.

This literature review raises four key points of particular relevance to the objectives of this research. First, smart card penetration is critical to its effective use for the analysis of passenger behavior. This point supports Oyster-based research in London where 73 percent of all journeys on the TfL network were made by Oyster card in 2007 (Mayor of London 2007). Moreover, the value of analyzing passenger behavior using smart card data should be considered relative to existing survey-based data systems. Second, a range of interchange time assumptions between 30-90 minutes has been used in previous studies for linking bus journey stages to form complete journeys depending on the location and purpose of the study. Third, previous studies provide examples of the development of bus origin-destination matrices using a combination of data from smart cards and other sources, especially APC and AVL systems. Given the ongoing deployment of Tfl's AVL system, iBus, these complementary data sources need to be considered in the next stage of smart card research and analysis for bus network planning in London. Finally, interchange planning has long been recognized as a key component of public transportation planning but has recently enjoyed renewed interest under the policy framework of developing more integrated and intermodal networks. Oyster smart card data is of particular value in capturing individual travel on multiple modes so, in combination with a strong interest in interchange planning in the United Kingdom and at TfL, supports the emphasis of this research on bus passenger interchange behavior.

1.4 Thesis Organization

This thesis is organized into six chapters, including this first one that sets the context for studying interchange behavior in London using smart card data, defines the research question, and reviews related literature. Chapter 2 follows with a discussion of the current network planning policy context at London Buses as well as available data systems, and concludes with a description of the smart card data used for the analysis presented in the subsequent chapters. Chapters 3 and 4 present a methodology for linking journey stages as recorded by smart card transactions to create complete journeys. In Chapter 3, the emphasis is on inferring interchange time windows for different modal combinations based on observed interchange behavior whereas in Chapter 4 the outcome of applying these time windows to the smart card data for London is assessed. The case studies presented in Chapter 5 move toward putting the results of creating a complete journey dataset into practice using metrics that would be helpful in assessing interchange behavior at the station or route level. Finally, Chapter 6 considers the general implications of the London-specific results presented in Chapters 3 through 5, and then presents a framework for implementation that may be of interest to other agencies that wish to apply smart card data to public transportation network planning.

CHAPTER 2: BUS NETWORK PLANNING IN LONDON

This chapter sets the context for the potential use of smart card fare payment data for bus network planning at London Buses, a division of TfL. It begins by describing bus network planning objectives in London (Section 2.1) and the informational requirements for meeting these objectives (Section 2.2). Then, the availability, format, and quality of new data from the Oyster smart card fare payment system that may complement, or eventually replace, existing data systems are described (Section 2.3). The final section also serves as an introduction to the data used for this research.

2.1 Planning Objectives

The Network Development team at London Buses is responsible for designing and planning bus services in Greater London. Possible changes are tested in a Cost/Benefit framework and against funding availability using the following guidelines:

- Comprehensive, "providing service to all areas and recognizing the needs of local people from all sections of the community";
- Frequent, "with adequate capacity for the peaks";
- Simple, "easy for passengers to understand and remember, and well-integrated with other public transport"; and,
- Reliable, "providing even service intervals when frequencies are high and running to time when they are low". (Transport for London 2004, pp.1)

The first objective is to provide a comprehensive network, serving residential and employment centers and giving access to local amenities. The key guideline to achieve this is bringing all households within a 5 minute walk of the bus network, which is about 400 meters at the average walking speed. The 400 meter guideline is used alongside other indicators of accessibility. Currently, over 90 percent of households in Greater London meet this criterion (Transport for London 2003). The guideline to achieve the second key objective of a frequent service is that the majority of passengers should be able to use buses on a "turn-up-and-go" basis, which is defined as running service every 12 minutes or better. Services are also planned so that the average load per bus, in the peak and at the busiest point, is between 75 percent to 80 percent of the total capacity of buses on the route, and where this is not possible to ensure that passengers should not have to wait more than 10 minutes for a bus that has room for them to board. The "simple network" guideline is a reaction to the diverse demand for bus travel and resulting complexity of the overall network in London. The idea is to keep individual services as simple and consistent as possible, for example having the last bus leave at the same time every day when there is no alternative night service. The reliability of the network needs to be built in at the service design stage by, for example, allowing additional recovery time for longer routes and basing schedules on current traffic conditions and passenger demand (Transport for London 2004).

Although the planning guidelines provide a framework for bus network development, decisions about route changes are based on an estimate of social benefit relative to cost. The social benefits of proposed route-level changes such as frequency changes, restructuring, re-routeing and extensions to provide new links are measured in terms of time savings to existing passengers and, where appropriate, account for the benefits of new "induced" demand or relief from crowding. Existing passenger counts are estimated both for the route under review and routes running between the same origin-destination pairs as the route under review. Wait time and/or travel time savings for existing passengers are converted to consumer surplus using an elasticity of demand based on service frequency and distance traveled. Costs from proposed service improvements are more straightforward to calculate and include labor, fuel, vehicles and overheads, which vary with frequency of service and length of route. A social benefit to cost ratio of 2 to 1 is usually required for proposed route-level changes to be implemented. In the case of new routes or new direct links network planners tend to search for similar contexts as proxies for what to expect in terms of passenger demand at the location where new services are being considered.

The Network Development team builds a business case for all proposed service changes by evaluating costs and benefits and taking account of contributions from other teams such as stakeholder comments, operational issues and vehicle requirements. This review provides an important link between passengers, planning, performance and operations. Once a new or modified bus service is approved, the Specifications Team compiles a detailed specification from Network Development's plan, which is then distributed to the private bus companies who tender and run bus services in London. Most contracts last for five years with the possibility of a two year extension under the current incentive scheme. Routes are reviewed at least every five years and in some cases more frequently where issues arise. Public participation is critical to the planning process and this is facilitated by members of the Stakeholder Engagement Team who maintain relationships with the London boroughs and carry out public consultation. Finally, the Performance team uses data collected primarily from Electronic Ticketing Machines (ETM) and manual on-street "traffic checks" or surveys to assess service reliability, which highlights poor performance and leads to recommendations for improvements. This information is also useful in the review process. Although all these groups benefit from the existing data systems at London Buses, Network Development is an intensive user of the systems described in Section 2.2 because of their service planning role.

Using this planning framework, London Buses has delivered major improvements to bus services since it was made a priority by the *Mayor's Transport Strategy*[4] in 2001. The 25.5 percent increase in bus kilometers operated, from 373 million to 468 million between 2001 and 2007, and the replacement of the entire fleet of over 8,000 buses with low-floor models are among the most notable improvements, many of which support congestion charging in Central London. Fifty two percent growth in the number of bus passengers over the same period is testimony to the impact of these improvements on travel behavior in London (Transport for London 2008, Transport for London 2003).

London's strong economic performance and associated population growth are expected to continue in the foreseeable future, and expansion of the bus network is the only short-term solution to address the city's growing transportation needs due its flexibility, relatively low cost and low infrastructure requirements (Transport for London 2003). There is a strong need to continue to ensure network efficiency from the operational perspective while continuing to provide a high-quality, comprehensive and passenger-oriented network, which is well integrated with other services. At present, Network Development uses survey and ETM data (usage record at point of entry on the bus), as well as experiential knowledge, to evaluate service changes. These tools have been well used to develop the network to deliver the Mayor's growth target. With the ongoing challenge of developing the bus network to meet the needs of Londoners, smart card data could be used to expand and/or enhance information on travel behavior which might include the following:

[4] The *Mayor's Transport Strategy* proposed 40 percent growth in passenger journeys from 2001 to 2011, about twice the rate of growth for the previous 20 years (Transport for London 2003).

- Passenger flows between two or more routes to provide support for direct links that reduce the need for passenger interchange;
- Interchange volumes between bus routes at an Underground station to show which routes are the most important means of accessing the station and then adjust station design and/or bus routing accordingly;
- A comparison of interchange times for Underground-to-bus with scheduled bus frequency at an Underground station could highlight reliability or crowding problems;
- A comparison of bus journey stage durations to access an Underground station with route length would show the area served by a station and demonstrate whether longer or shorter routes may be worthwhile;
- Evidence of multi-stage journeys (e.g., bus-Underground-bus) might support route redesign to create a direct bus link that reduces the need to interchange and relieves congestion on the Underground; and
- Identification of regular passenger travel on a route or sequence of journey stages could demonstrate strong reliance on a bus service under review and influence decision-making about that route.

2.2 Existing Data Systems[5]

In this section, the major existing information systems used by Network Development are briefly reviewed in order to further explain the context for assessing the potential application of smart card data to bus network planning, with a focus on interchange, in London. Table 2-1 summarizes general data needs for bus network planning and how each of the existing data systems at Network Development serves those needs. A synopsis of the contribution of each data system is provided in the remainder of this section, but it begins with a discussion of the need for information on passenger demand because it is the key metric in providing comprehensive service passenger demand and can be measured in various ways. Moreover, passenger demand is most directly related to the exploration of the value of Oyster smart card data to bus network planning that is presented in this thesis.

---

[5] The reference material for this section is an internal TfL report titled: "*Estimation of Daily Passenger Journeys – Brief, Version 00d, 02/06/2006 (Transport for London, Surface Transport)*".

Table 2-1 Data Systems and Needs for Bus Network Planning

Data System			Data Needs for Bus Network Planning					
Name	***Description***	***Recurrence***	***Passenger demand (ridership per unit time)***	***Max load point & peak load (capacity)***	***Run time & reliability***	***Existing schedule & route design***	***Operating costs & revenue***	***Value of time***
BODS	Holds detailed information about passenger travel patterns for each route, including the number of people boarding and alighting at each stop, the purpose of travel, the location of the beginning and end of each journey, and how passengers get to the bus boarding point and from the bus alighting point to their final destination. Informs the detailed model of how passengers behave in response to a service change.	Every 5-7 years for each route.	Number of passengers boarding and alighting at each bus stop for each route and parallel routes for day and night services.	Max load point and peak load.	Can indicate reliability issues.	Surveyed bus trips compared with scheduled bus trips in BODS. Expansion factors added to account for non-return of survey cards and non-surveyed bus trips.		
Keypoints	Survey of boarding, alighting and load figures for routes at 400 key bus stops, which is used in conjunction with other data to determine the busiest point at the busiest hour.	Every 2 years for each location.	Boardings, alightings and load at a fixed key location.	Keypoints updated with BODS travel pattern.	Can indicate reliability issues.			
BREMS	Holds passenger journey stages, scheduled mileage, operated mileage and lost mileage per route per day for all day types, which is used to assess usage trends at route and network level. Data downloaded from BCMS.	On-going with 6 week time lag.	Current daily passenger numbers by route used to update BODS results to the present.	Provides trends in usage over time.	The level of non-operated mileage can indicate reliability issues.			

Data System			Data Needs for Bus Network Planning					
Name	***Description***	***Recurrence***	***Passenger demand (ridership per unit time)***	***Max load point & peak load (capacity)***	***Run time & reliability***	***Existing schedule & route design***	***Operating costs & revenue***	***Value of time***
Route Results	An extension of BREMS that holds passenger journey stages and revenue per route per day for all day types.	On-going with 6 week time lag.					Provides revenue by route from all passenger transactions.	
BCMS	Holds all bus contract details including cost and all variations to contracts following award from tender. Also holds ETM data recorded by drivers on each bus on each route, which is used to determine the number of passengers over a 24 hour period for any time band.	On-going.	Daily passenger volumes per timeband throughout the day for all day types. Used to inform BODS to determine weekend usage.	Peak vehicle requirement and type of bus along with frequency of service determine capacity.	The level of non-operated mileage can indicate reliability issues.		Provides all contract costs and daily on-bus revenue.	
Caesar	Holds all specifications and schedules for each route with point to point mileages and run times.	On-going.			Informs run time for each bus trip in the schedule.	Informs mileage for each bus trip in the schedule.		
BusNet / GIS	Historic record of the specific point to point routing for each bus service, including all operational details and BODS stops with GIS map outputs. Linked to Caesar.	On-going.				Exact routing recorded in both directions.		
STABS	Quality of Service survey of all routes passing specific timing points, for different timebands throughout the day during each quarter.	Every 12 weeks.			Used to calculate Excess Wait Time (EWT) which measures reliability from the passenger's point of view.			

Data System			Data Needs for Bus Network Planning					
Name	***Description***	***Recurrence***	***Passenger demand (ridership per unit time)***	***Max load point & peak load (capacity)***	***Run time & reliability***	***Existing schedule & route design***	***Operating costs & revenue***	***Value of time***
GLBPS	Survey of bus journeys by ticket type, zones, day type and time of day, including night services.	Approx. 400 driver duties quarterly.	Aggregated by fare zone, time of day and ticket type. Emphasis on non-electronic tickets.	Provides trends in usage over time for different categories of passenger.				
LATS / LTDS	TfL's household travel diary survey.	Continuous as of 2005, approx. 8,000 households surveyed per year.	Calibration of TfL's strategic transport models.					Yes

Depending on the planning application, information on bus passenger demand (or ridership per unit time) may be required:

- between stops or stop zones on a route,
- between stops or stop zones on parallel routes,
- between stops or stop zones on a route and intersecting routes or stations,
- between ultimate origins and destinations, or
- for an entire route or group of routes.

Network Development currently has well-developed systems for assessing passenger demand between stops or stop zones for a route and *parallel* routes as well as for an entire route, but that information on passenger demand between stops or stop zones on a route and *intersecting* routes or stations, and between ultimate origins and destinations is less well provided for. The source of information on passenger demand between stops or stop zones on a route and on parallel routes is the Bus Origin-Destination Survey (BODS) and for an entire route or group of routes it is the Bus Revenue and Mileage System (BREMS).

A BODS survey is conducted on average every six years for each route and is used to develop an origin-destination matrix for the route. Automated reports from BODS include boardings, alightings, and loads at each stop (or stop zone) along a route, as well as parallel routes. BODS could be considered the primary data system used by Network Development because it provides the detailed, disaggregated passenger demand information that is necessary to calculate social cost/benefits in terms of time loss/savings to passengers. A major limitation of this type of survey is that it records passenger travel for one day per route which, combined with substantial network growth and a changing network, necessitates supplementary data from other sources. Moreover, although surveyed passengers are asked for their ultimate origin and destination in addition to their travel on the route itself, this information is rarely transferred from the paper surveys into the BODS database and is therefore not readily available to network planners. Another challenge in measuring passenger demand along a route is that it is dynamic; new demand for bus transportation may be generated by the development of commercial or residential buildings, or by increasing transportation services. New demand is difficult to estimate so network planners tend to search for similar contexts as proxies for what to expect at the location where new demand is foreseen. In many cases, BODS data is complemented by Keypoints, a survey that records boardings, alightings and loads at 400 key bus stops in the bus network on a two year cycle.

Timely route-level passenger demand data is currently gathered from Electronic Ticketing Machine (ETM) transactions, which are downloaded from bus garages to the Bus Contracts Management System (BCMS). BCMS is used to determine the number of journey stages over an entire day (24 hours) for any timeband (e.g., peak usage from 7 a.m. to 9:30 a.m.) as recorded by the bus driver. Summarized data from BCMS forms the basis of BREMS, which provides week by week passenger volumes for all ticket types (including Oyster cards) for all day types on all routes. This provides trends in usage over time and is used to update BODS results to current journey stage levels. Moreover, ETM data is used to inform BODS to determine weekend passenger volumes. Route Results is a further development of BREMS that incorporates the allocation of off-bus revenue to each route and combines the results with the on-bus revenue already recorded and copied from BCMS. Revenue is a key element in reviewing a route as this is used along with the cost (from BCMS) to calculate the cost recovery for each service.

BODS, Keypoints and BREMS are currently used together to inform demand patterns on each route including determining the location and timing of the busiest point and the origin-destination matrix on the route, which is the basis of evaluation when a route is reviewed. Peak load and the maximum load point are important metrics for ensuring that there is adequate capacity for peak demand as specified in the service planning guidelines discussed in Section 2.1.

Another data system, the Greater London Bus Passenger Survey (GLBPS) is used primarily to estimate revenue allocation for trips made on cash and paper tickets, and to estimate aggregate patterns of bus travel across Greater London (i.e., between fare zones). It is not used explicitly for bus route planning because the resulting information about passenger demand is too aggregate, but it does provide an important input to adjusting BREMS passenger volumes for non-ETM tickets. GLBPS data is made available to network development planners in spreadsheet format on a quarterly basis and basic demographic characteristics of bus passengers are also available through the survey.

A final data system that provides information on passenger demand is TfL's household travel diary survey, the London Travel Demand Survey (LTDS). Network-wide origin-destination information for all travel in Greater London used to be estimated using the London Area Travel Survey (LATS), which was conducted every 10 years until 2001. However, with the rapid rate of change of travel behaviour in London TfL decided to replace LATS with LTDS. This new survey reaches approximately 8,000 households, or 12,000 individuals, across London annually. The most recent LTDS survey was completed in April 2006, and preliminary results were released in March 2007. They show, for example, that bus is the most frequently used form of transport in London, with 57 percent of Londoners using buses at least once a week (Transport for London 2007c).

LATS and now LTDS do not provide enough detail on passenger behaviour to be useful for route-level network planning, but rather are employed primarily by other divisions of London Buses such as Strategic Planning to inform higher-level policies and by TfL Corporate to calibrate its strategic transportation models. The LATS and LTDS data are stored in Microsoft Access databases. Data from surveys such as LATS and LTDS are also used to develop estimates of the elasticity of demand with respect to frequency, distance, congestion and interchange – all key components of bus network planning.

Finally, other data systems that support network development directly are Caesar, STABS, BusNet and the MapInfo-based Geographic Information System (GIS). Caesar is the schedules system, which holds the specification for each route and all subsequent implemented schedules. The schedule is key to understanding the current level of service and can be viewed in full, along with calculated run times and mileage for each trip. STABS provides Quality of Service (QSI) information which enables planners to review the reliability of a route from the passenger's point of view. BusNet includes a history of all changes on each route and is used regularly by network planners to compile and visualize the operational details of existing routes, which is the starting point for any review. The GIS is a compilation of route maps, stop location maps, and demographic information by ward. The GIS is used on an ad-hoc basis to perform more complex spatial analyses related to proposed network changes.

2.3 Smart Card Data Systems

Having reviewed the planning guidelines and data systems used for bus network planning in London, I turn now to a potential new source of travel information, Oyster smart card data. Every time an Oyster card is used to pay a transit fare, the smart card system records the time, location and other information about the transaction. This recording occurs when the cardholder taps their card on a reader as they go through a ticket gate or board a bus. In the case of London, ticket gates are located at the entry and exit to Underground stations so information about where and when the cardholder entered and exited the Underground is captured. For buses, the card records information about the time of boarding and the bus route number that has been associated with the reader on that bus.

Because these entries and exits to the system are associated with each card, the resulting data provides a wealth of information about each cardholder's travel patterns on the TfL network. This data yields valuable information about actual travel behaviour over time and potentially real-time information about network usage. On the other hand, as with any system, smart card data has its limitations. Contrary to survey data, it does not provide socioeconomic information about the cardholder nor details of their journey purpose. Information about people's choice between public transportation and other modes is also lacking. Finally, the volume of data available poses some data management challenges and barriers to implementation as part of a day-to-day planning system.

Despite these limitations, the breadth of smart card data would be prohibitively expensive to replicate with any travel survey. Journeys paid for using Oyster cards currently account for 73 percent of all journeys on TfL's network and because they are classified by card number they can be monitored at the individual level over time[6]. So, like other sections of TfL, London Buses has begun to explore the extent to which smart card data can provide new planning information over and above ETM and survey data, or perhaps provide the same information at lower cost or in a timelier manner. As discussed in the previous section, Network Development currently has little, if any information about passenger demand from intersecting bus routes or stations. Therefore, smart card data could be used to fill this gap and increase understanding of how bus passengers interact with other components of the London public transportation network through interchanges.

In the remainder of this section, the source, format and quality of the smart card data used for this research and potentially by London Buses for network planning is explained.

[6] Individual monitoring is subject to privacy regulations that are addressed by randomly generating encrypted ID numbers for each card.

2.3.1 *Data Source and Format*

The smart card data that is gathered from the millions of Oyster fare payment transactions that occur on the TfL network each day is held primarily for monitoring purposes. However, the raw data collected from the card readers on each bus and at each ticket gate has been compiled into 'sequenced journeys tables' for this research. In this format, each record represents one journey stage by bus, Underground or another mode where Oyster cards are valid. A list of the attributes associated with each journey stage is provided below, but the main point is that the journey stages for all modes (primarily bus and Underground) are indexed by day, card ID, and a unique sequence number for each journey stage on the card.

The sequenced journey tables used for this research were supplied to researchers at the Massachusetts Institute of Technology (MIT) by Prestige[7] staff. Normally, the text format data files are transferred to MIT via a File Transfer Protocol (FTP) and researchers at MIT use an Oracle server to manage the monthly and ad-hoc Oyster sequenced journeys tables. The monthly sequenced journey tables include a continuous random sample of 5 percent of all Oyster cards in the TfL system so, for each encrypted Oyster card ID, the table includes an ordered series of journey stages by day over the course of a four week period. Unfortunately, a 5 percent sample of cards was insufficient for the route level of detail used in the case studies presented in Chapter 5 so a sequenced journey table for all Oyster cards in the system was supplied to MIT for two weeks during November 2007. This '100 percent sample' amounts to over 8 million records, or journey stages, per weekday and approximately 5 million records per weekend day. November is generally believed to a representative month in terms of normal travel behavior in London.

The attributes included in the sequenced journey tables are:

- Day
- Oyster Card ID (encrypted)
- Journey Stage Sequence Number
- Mode, primarily bus or Underground but also includes tram, Docklands Light Rail (DLR) and some National Rail

---

7 Prestige (Procurement of Revenue Services Ticketing Information Gates and Electronics) is the project name for the integrated ticketing and revenue system (i.e., Oyster) delivered by a partnership between TfL and TranSys, a private consortium selected by TfL for the project.

- Start Location, station code for Underground, DLR and National Rail; farestage code representing a route number for buses and trams
- End Location, station code for Underground, DLR and National Rail; direction (1 or 2) for buses
- Inner Fare Zone, innermost fare zone for bus routes (missing data)
- Outer Fare Zone, outermost fare zone for bus routes (missing data)
- Route Distance, length of bus route
- Start Time, journey stage start time, recorded in minutes past midnight
- End Time, journey stage end time, recorded in minutes past midnight and always '0' for buses
- Zonal Validity, zones in which the ticket is valid
- Fare Type, pay-as-you-go (PAYG) or period ticket fare payment category
- Daily Capping Flag, indicates whether PAYG reached daily price cap
- Capping Scheme, scheme journey was capped under
- Full Fare, fare that would have been charged at full price
- Discounted Fare, fare charged with daily capping
- Ticket Product Code
- Ticket Time Validity, length of time the ticket has been kept valid

The attributes of each journey stage that were used for this research are:

- Day
- Card ID (encrypted)
- Journey Stage Sequence Number
- Mode
- Start Location
- End Location
- Start Time
- End Time
- Fare Type

Unless otherwise noted, this research uses the sequenced journeys table for Wednesday, November 14, 2007. This day was chosen to represent a typical weekday and was compared to other weekdays in November 2007 from both the 5 percent and 100 percent samples to ensure data consistency. The sequenced journeys table for this day includes 8,134,887 records, each record representing one journey stage on the TfL network. A journey stage on the Underground includes "behind-the-gate" interchanges, which means that when passengers change Underground lines without exiting and re-entering the system it is counted as a single journey stage. This is consistent with the TfL definition: "Underground journey stages are counted by station entries and interchanges within stations are ignored." (Transport for London 2008, pp.2). However, "bus journey stages are counted as starting a new journey stage each time a new bus is boarded." (Transport for London 2008, pp.2) The vast majority of the journey stages recorded by Oyster are by bus or Underground, but approximately 4 percent are on other modes including National Rail, trams, and Docklands Light Rail (DLR). These records were excluded because of data scarcity, leaving 7,963,425 records for bus and Underground only. Data on National Rail usage in London is expected to increase substantially in the near future as Oyster cards are accepted by more Train Operating Companies (TOCs) through revenue agreements with TfL. At that point, the methodology presented herein for analyzing passenger interchange Underground stations could easily be applied to National Rail stations as well.

An example of the Structured Query Language (SQL) code that was used to manipulate the sequenced journeys tables is provided in Appendix 1 because it may help with the development of a customized database for Network Development that is supported technically by information management groups at TfL. It should be noted that although this research employs Oyster smart card data, the general data structure and quantitative findings should apply to any smart card data system implemented at TfL in the future.

2.3.2 Data Quality

This section describes the extent to which bus and Underground network usage as captured by Oyster smart cards is representative of actual TfL network ridership as well as general issues pertaining to data quality. Both modes are discussed in terms of overall passenger volumes and bus journey stages are disaggregated to the route level.

In terms of bus journey stage volumes, estimates from GLBPS show that 6.54 million bus journey stages took place on an average weekday in London during 2006/07 (Transport for London 2008). Moreover, according to recent TfL statistics, Oyster smart cards were used for

76 percent of bus journey stages in November 2007 (Transport for London 2007a) so we would expect the number of bus journey stages in the sequenced journeys table to be approximately 76 percent of 6.54 million, which is 4.97 million. As summarized in Table 2-2, the sequenced journeys table for November 14, 2007 includes 5.10 million bus journey stages, or 78 percent of average weekday journey stages from GLBPS. This figure holds for the other weekdays during the week of November 12 to 16, 2007. The slight difference in actual versus expected bus journey stage volumes may be accounted for by on-going growth in bus travel (Transport for London 2008, Transport for London 2007a) and by seasonal variations in travel behavior. Nonetheless, the volume of Oyster bus journey stages in the sequenced journey table relative to the GLBPS estimates are close enough to the expected value of 76 percent to be used for this research.

Table 2-2 Comparison of Oyster Bus Journey Stages with Average Weekday from GLBPS

Period	**Journey Stages** (million)	**Ratio Oyster / GLBPS**
Average weekday from GLBPS	6.54	-
Single weekday from Oyster*	5.10	78%
Average weekday from Oyster**	5.13	78%

*Wednesday, November 14, 2007
**November 12 to 16, 2007

In terms of Underground journey stage volumes, TfL estimates that Oyster smart cards were used for 70 percent of all Underground journeys in November 2007 (Transport for London 2007a). Moreover, they estimate that 3.28 million Underground journey stages took place on the average weekday in 2006/07 based on Underground entry counts so we would expect the sequenced journeys table to include approximately 70 percent of 3.28 million, or 2.3 million Underground journey stages. The sequenced journeys table for Wednesday, November 14, 2007 includes a total of 2.51 million Underground journey stages with valid start locations[8]. This is 77 percent of TfL's average weekday figure, which is higher than expected but can also be explained by seasonal variation and on-going growth in Underground journeys (Transport for London 2008, Transport for London 2007a). For example, year-on-year growth in Underground journey stages from November 2006 to November 2007 was 6.7

[8] Oyster smart cards with more than 10 journey stages were also excluded because the data is unreliable.

percent (Transport for London 2007a). Thus, it can be concluded that it is acceptable to proceed using this sequenced journeys table as a representative day.

At the route level, a comparison of weekly bus journey stage volumes from the Oyster sequenced journeys tables and BREMS[9] reveals that Oyster volumes for all bus routes with valid route codes represent 82 percent of total BREMS passenger volumes on the corresponding routes. This high share of Oyster-based journey stages is good because it means that travel behavior as revealed by the sequenced journey table is likely to be representative of all travel on the bus network. However, higher journey stage volumes on Oyster relative to BREMS (82 percent) than on Oyster relative to GLBPS (78 percent) may be cause for concern. There are two possible explanations for the discrepancy: first, journey stage volumes as recorded by BREMS may simply be lower than the GLBPS estimates and, second, that routes with invalid route codes in the sequenced journey table may have a lower Oyster card penetration rate than the average route. In either case, and as would be expected due to varying penetration of Oyster smart card use across routes, there is significant variation in bus journey stage volumes between BREMS and the sequenced journeys file at the route level. For example, 73 percent of routes have Oyster journey stage volumes between 75-100 percent of the corresponding volumes from BREMS, but, on the other hand, a handful of Oyster route volumes are higher than the corresponding volumes from BREMS. These issues merit further investigation so the three main discrepancies between Oyster and BREMS route-level passenger volumes as revealed by the sequenced journeys tables for the week of November 11-17, 2007 are now discussed.

Table 2-3 Comparison of Weekly Oyster Bus Journey Stages with BREMS

Weekly Journey Stage Data Source*	**Journey Stages** (million)
Oyster with valid route code	30.36
BREMS for routes with valid route code in Oyster	36.94
Ratio Oyster / BREMS	*82%*

*November 11 to 17, 2007

[9] BREMS provides route-level passenger volumes based on Oyster and non-Oyster fare payment transactions recorded by the Electronic Ticketing Machine (ETM) installed in all buses operated under TfL contract. It is discussed in detail earlier in this chapter.

First, about 40 currently operational routes, or about 6 percent of over 700 routes in London, are missing from the sequenced journeys table used for this research. Although journey stages on these routes are included, their bus boarding location is invalid. This is because the farestage codes that identify the route number for each journey stage in the Oyster sequence journeys table have not been updated since 2003. As a result, the routes that have been added to the network since 2003 are not captured in this research. That being said, in general bus journey stages on routes that existed prior to 2003 are correctly assigned to the route on which they occurred. This problem of missing routes has been addressed for future work with Oyster data by adding an update-to-date Route ID attribute to the sequenced journey tables that should be used instead of the farestage code for identifying the route number.

Bus journey stages for each passenger are assigned to the route on which they occurred based on a farestage location code that, unfortunately, is not generally accurate below the route level. In the past, TfL bus fares were based on distance traveled so a system of farestages (or fare zones) was used to determine the fare owed by each passenger. In theory, this farestage code would then provide information on the zone of the route in which the passenger boarded but in actual fact bus fares are now flat rate so operators tend to leave the ETM set to the first farestage on the route. Thus, bus boarding location is accurate at the route level but cannot provide reliable information about where the passenger boarded along the route. It is expected that, in the near future, TfL's new AVL system, iBus, could be linked to the time record for Oyster-based bus boardings to provide location information at the stop level.

Second, 12 routes exhibit journey stage volumes that are in the order of 150 percent higher in Oyster than BREMS. This should not be possible because BREMS is supposed to include Oyster transactions *plus* magnetic tickets and cash fares. The notable similarity between the routes in question is that they are all articulated bus routes that have Oyster card readers installed inside the rear doors as well as at the front of the bus. It may be that Oyster transactions recorded at rear doors have not been included in BREMS reports and as a result BREMS underestimates ridership on articulated buses. This issue is being investigated by TfL, but in the meantime these routes have been included in the analysis as there is no evidence that the Oyster data is incorrect.

Third, total journey stages on night bus routes are significantly lower in the Oyster data than in BREMS on Sunday only. Total journey stages on night bus routes for Sunday, November 10 generally represent about 25 percent of BREMS volumes on the same day. This may be because days are defined to begin at 4 a.m. in Oyster data so journeys on night buses from Saturday night at midnight to Sunday at 4 a.m. would not be included in total journey stages for Oyster on Sunday whereas they would be included in BREMS (assuming the day starts at midnight). This research relies on data for Wednesday, November 14 so the discrepancy between night bus journey stage volumes on Sunday can be ignored, but it should be investigated further if weekend data for night bus routes is to be analyzed in the future.

The routes with invalid route codes are listed in Appendix 2 and route-level discrepancies in journey stage volumes are taken into consideration in the case studies presented in Chapter 5. Despite these discrepancies, journey stage volumes are in the range of 80 percent for the majority of routes so it appears to be reasonable to proceed with the research using the sequenced journeys file for Wednesday, November 14, 2007.

In addition to verifying aggregate and route-level journey stage volumes and removing cards with journey stages on modes other than bus or Underground, some error correction of the sequenced journeys file is required as follows:

- delete duplicate records and records with non-unique *ID-Sequence Number* combinations (approximately 1,100 records, or less than 0.01 percent, of records),
- delete cards with more than 10 journey stages per day because they represent outliers (approximately 170,000, or 2 percent, of records representing 1 percent of cards), and
- generate new sequence numbers that increase by a margin of one for each card to simplify subsequent coding.

2.4 Summary

This chapter began with a description of the general planning guidelines and procedures of bus network planning in London to provide a context within which to evaluate whether smart card data will be of value to this effort. It emphasized the tremendous growth in bus services and patronage delivered by London Buses in recent years due in no small part to the efforts of the Network Development team. Their existing survey- and ETM-based data systems provide detailed information about passenger demand at the route and stop zone level. However, there appears to be a gap in information in terms of passenger demand between intersecting routes and between buses and other modes at specific locations, including the Underground.

At present, Oyster card data has had limited use by Network Development due to lack of access, uncertainty about its value, and the substantial up-front investment required to develop a database system that meets their needs. However, the description of smart card data, including its strengths and limitations, illustrates that it could be of great value in terms of filling in gaps in knowledge about passenger demand.

Finally, the analysis of aggregate and route-level journey stage volumes showed that Oyster smart card data is representative enough of travel on the bus network in London to inform a methodology for joining journey stages into complete journeys and to then apply the results to network planning case studies. The review of route-level bus network passenger volumes exposed some outstanding data issues but also revealed the possibility that Oyster data may actually be more reliable than ETM data in some instances, for example for articulated bus routes.

CHAPTER 3: CHARACTERSITICS OF BUS PASSENGER INTERCHANGE

To the best of my knowledge, this research represents the first comprehensive attempt to combine bus and Underground journey stage data derived from smart card fare payment transactions into complete journeys using informed elapsed time assumptions to identify interchanges. The availability of complete journey information, albeit approximate, would be an advance in knowledge for network planners in evaluating the costs and benefits of changes to the bus network. Thus, the goal of this chapter is to develop ranges of elapsed time thresholds that could be used to link bus and Underground journey stages into complete journeys using smart card data.

The chapter begins by defining interchanges (Section 3.1) and presenting aggregate statistics on journey stages in London (Section 3.2). Then it turns to developing recommended ranges of elapsed time thresholds for identifying interchanges through the analysis of times between journey stages at the passenger level for three modal combinations: Underground-to-bus, bus-to-Underground, and bus-to-bus (Sections 3.2 – 3.4). The first two modal combinations are analyzed in terms of time periods, fare zones, and stations, and bus-to-bus combinations are examined across time periods and by ticket type. This analysis not only informs recommended ranges of elapsed time thresholds but also provides insights into bus passenger interchange behavior in London.

Unless otherwise noted, the statistics presented in this chapter are drawn from a dataset of all journey stages made on the TfL network using Oyster smart cards on Wednesday, November 14, 2008. A description of this sequenced journeys table is provided in Chapter 2.

3.1 Definition of Interchange

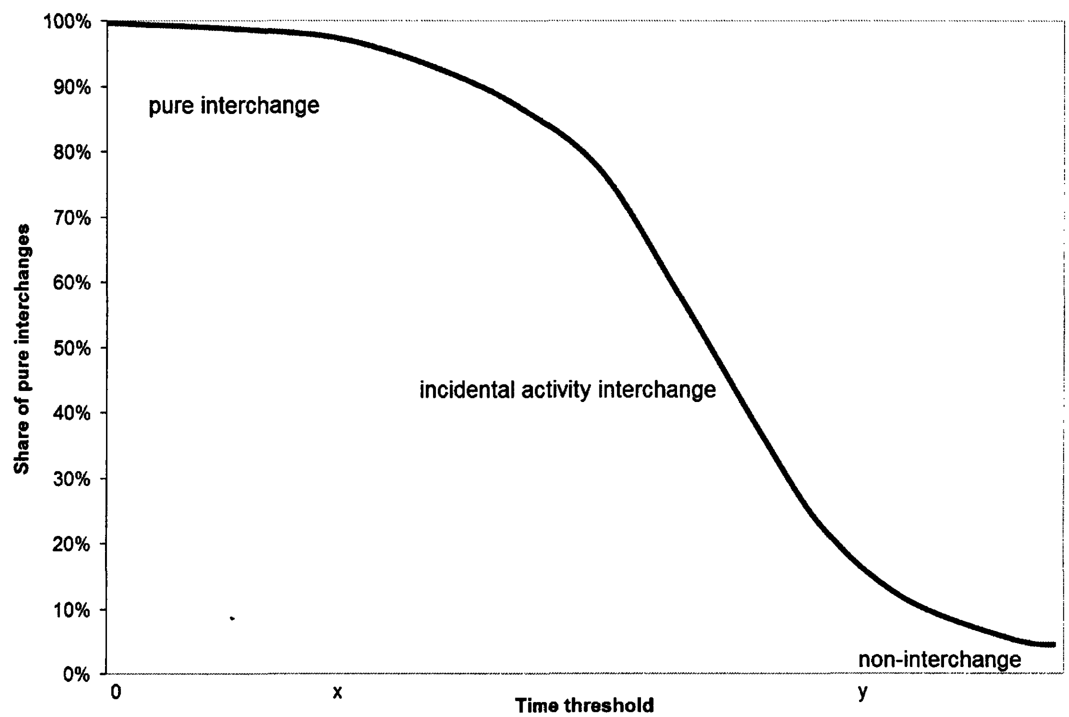
As defined in Chapter 1, to interchange is the act of transferring between modes or between different services (i.e., vehicles) of the same mode (Transport for London 2001). For an example of a transfer between bus services, when a passenger alights from Route 221 and then waits for and boards Route 263 in order to continue his or her journey to the ultimate destination it is considered to be an interchange. However, in the context of this research, "behind-the-gate" interchanges in the Underground network, i.e. transfers between trains without exiting the system, are ignored so that all the components of an Underground journey stage from when the passenger enters through a ticket gate until they exit through a ticket gate are considered part of a single Underground journey stage with no interchanges. But, it is not quite as simple as this. If the passenger who is transferring between bus routes takes time to buy a magazine at the newspaper stand on the corner, is this still considered an interchange? What if they instead walk around the corner to their favorite bakery to buy a snack? Or, they take 15 minutes to pick up their child from school? Or, what if they meet a friend for lunch? The point is that there is a spectrum of activities that people may engage in between journey stages and often that activity is actually the purpose of the journey, for example meeting a friend for lunch, rather than some non-travel activity incidental to the interchange, such as buying a newspaper. In the case of incidental activities, the passenger would consider the two journey stages to be part of the same complete journey so they should be linked together, however in the case of an activity that was the main purpose of travel we do not want to link the journey stages together into a complete journey even if the activity duration is very short.

One limitation of smart card data is that there is no way of determining what activities people are engaged in between journey stages. All that is known is where and when they traveled on the TfL network. So, the objective of this chapter is to determine time thresholds between journey stages within which most people are likely to be interchanging, allowing time for incidental activities only. The elapsed time threshold would be the maximum allowable interchange time (or bus in-vehicle plus wait time) for two sequential stages to be considered part of same journey.

This concept is illustrated in Figure 3-1, in which the horizontal axis represents elapsed time thresholds and the vertical axis represents the share of interchanges that are pure interchanges as the time threshold increases. In other words, if the time threshold is set between 0 and x then all potential interchanges are pure interchanges whereas if the time threshold is set between x and y then an increasing share of interchanges will include incidental activities. Elapsed time thresholds above y would mislabel a large share of non-interchanges as interchanges.

In other words, with the formation of complete journeys based only on time thresholds between journey stages, some inaccuracy may result in certain sequential stages being linked when in reality no interchange took place. However, interchanges that include incidental activities such as buying a newspaper between journey stages should be classified as an interchange because the application of complete journey data to planning issues should reflect passengers' perceptions of their travel experience rather than a minimum time threshold to transfer. Since the difference between incidental and purposeful activities between journey stages cannot be identified with certainty using smart card data it is necessary to rely on an analysis of typical interchange times to determine which journey stages to link together. The assumption is that so long as the elapsed time thresholds are fairly representative for the system as a whole, then when a particular route is reviewed the actual journey patterns for passengers on that route relative to network norms as well as the physical and scheduling context can be considered. Thus, the general approach to determining appropriate elapsed time thresholds is to include all pure interchanges and incidental activity interchanges whilst minimizing the number of non-interchanges to the extent possible. This could be thought of in terms of tolerating Type 1 error (i.e., including a potential interchange that actually involves non-travel activities) and minimizing Type II error (i.e., excluding a potential interchange when in fact it is a pure interchange).

Figure 3-1 Conceptual Diagram of Elapsed Time Thresholds versus Interchange Type

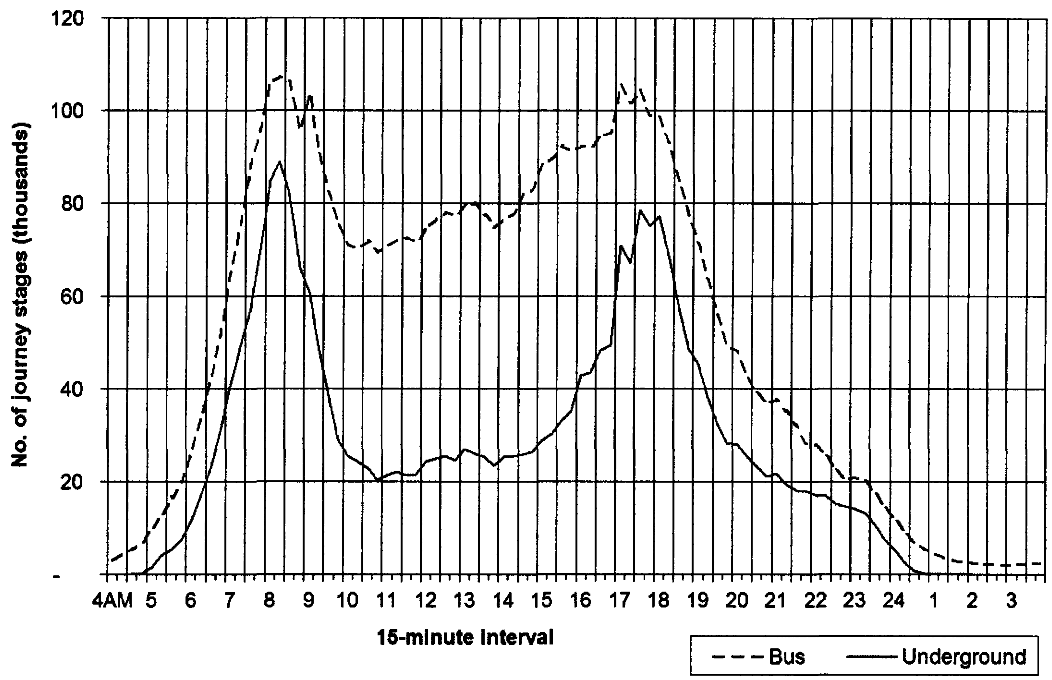


In order to determine which journey stages to link together into complete journeys, three specific interchange scenarios are examined: Underground-to-bus, bus-to-Underground, and bus-to-bus. Interchanges between bus and other modes such as National Rail are not considered due to data scarcity but a similar approach could be taken to include them in the future. Potential Underground-to-bus interchange is characterized by the time in minutes between Underground station exit and bus boarding on a route that serves the vicinity of that station. Both the station exit and bus boarding must be recorded by a transaction, or "tap", with a unique smart card. On the other hand, for a potential bus-to-Underground interchange, the "interchange" time threshold includes bus travel time in addition to the time it takes to walk to the Underground station ticket gates after alighting from the bus, which we expect to be no more than a few minutes. Thus, for most passengers the time difference observed between bus boarding and Underground station entry is almost entirely bus travel time if the person does not engage in any other activities between alighting from the bus and entering the station. It does not include the initial wait time for the bus or for the Underground train after entering the ticket gate. Therefore, the potential bus-to-Underground access journey time will give us an indication of how long people are willing to travel by bus to access the Underground and will provide an indication of an appropriate elapsed time threshold for joining bus journey stages with subsequent Underground journey stages for each passenger. Finally, for potential bus-to-bus interchanges, the time threshold includes not only the in-vehicle travel time on the first bus stage, but also the wait time for the second bus. Therefore, it could also be called a bus journey access time for another bus route.

3.2 Journey Stage Patterns

Before looking at potential interchanges, this section presents summary statistics of journey stage patterns in London over the course of the day and across passengers. These aggregate statistics help to set the context for the interpretation of potential interchanges between bus and Underground journey stages presented later in this chapter. First, Figure 3-2 shows that using journey stage start times both bus and Underground travel is highest in the morning around 8:30 a.m. and in the afternoon around 5:30 p.m., however the afternoon peak is less sharp, particularly for bus. Moreover, there is a significant difference in midday travel between bus and Underground with far greater bus activity during the middle of the day. Bus journey stage volumes are consistently higher but this is probably an artifact of Underground journey stages including "behind-the-gate" interchanges by definition. Certainly, some of the bus journey stages are actually part of multi-stage journeys and should be linked to either Underground or bus journey stages to form complete journeys. This is done in Chapter 4, but for now I would add only that the trends in this graph are consistent with published TfL data (Transport for London 2008).

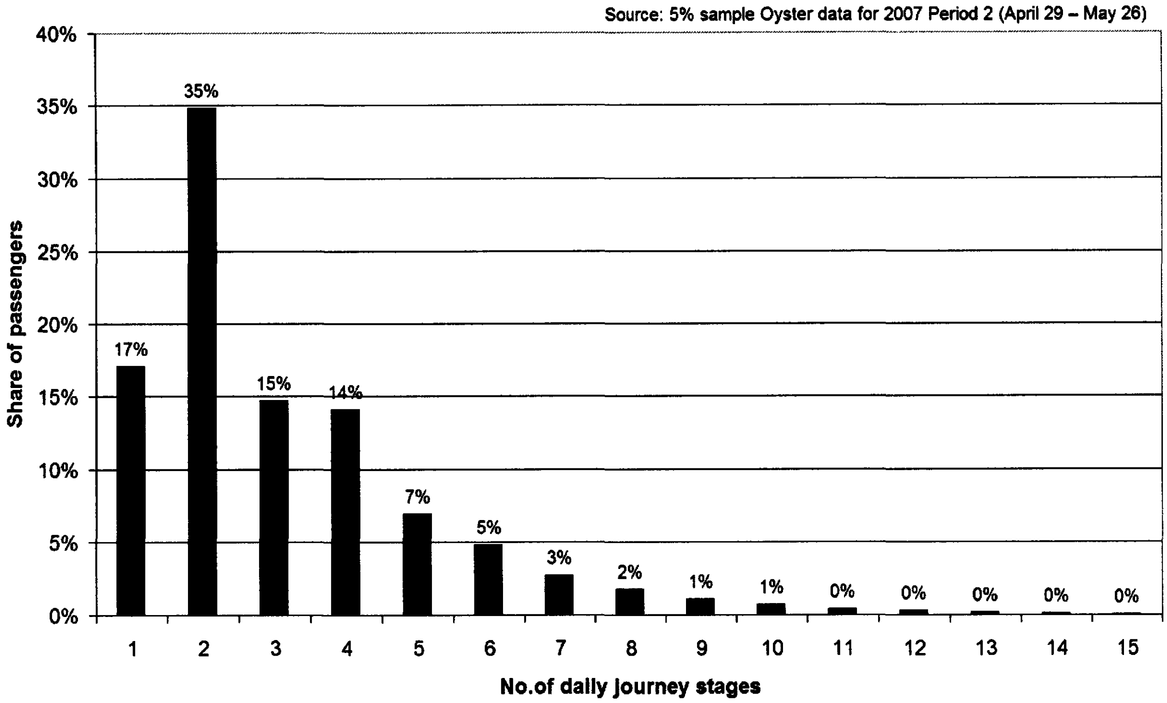
Figure 3-2 Bus and Underground Journey Stages per Day by 15-Minute Interval



Next, Figure 3-3 shows the number of daily journey stages per passenger as a percentage of all passengers, for example 35 percent of passengers make two journey stages per day using their Oyster smart cards[10]. This summary is useful in determining whether to include cards with a large number of journey stages in subsequent analysis. The graph illustrates that any card with more than 10 journey stages per day could be excluded from the analysis without loss of generality. It is also worth noting that a surprisingly large share of passengers (17 percent) only make one daily journey stage on the TfL network. This indicates that people have complex travel patterns that do not necessarily include a single return journey per day.

[10] The graph shows average daily results for a random 5 percent sample of Oyster smart cards for four weeks. A similar summary of all smart cards for a single weekday results in the same pattern of use.

Figure 3-3 Journey Stages per Passenger



To provide more detail on journey stages per passenger, the most common modal patterns of daily bus and Underground journey stages across all passengers can be examined. Table 3-1 shows that the most common daily pattern is two Underground stages, followed closely by two bus stages. It is somewhat surprising that 10 percent of passengers make only one daily bus journey. This may be explained by short journeys for which the return journey is made by walking, or by bus being used to access a National Rail journey stage (not recorded by Oyster) but not for the egress or vice versa. These scenarios may be better identified in the future as Oyster card penetration on National Rail in London continues to increase. It is also important to note that only a small share of passengers with both bus and Underground journeys appear in the top ten journey patterns. This suggests that relatively few complete journeys include interchanges between bus and Underground (see Chapter 4 for further discussion), but the modal patterns also indicate that bus-to-bus interchanges are likely to be prevalent across the TfL network.

Table 3-1 Top 10 Journey Stage Patterns per Passenger

Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6	Passengers	Share	Cumulative Share
U	U	-	-	-	-	416,082	16%	16%
B	B	-	-	-	-	401,356	16%	32%
B	-	-	-	-	-	266,561	10%	42%
B	B	B	-	-	-	150,781	6%	48%
B	B	B	B	-	-	144,275	6%	54%
U	-	-	-	-	-	125,528	5%	59%
B	U	U	B	-	-	77,353	3%	62%
B	B	B	B	B	-	72,943	3%	65%
U	U	U	-	-	-	65,190	3%	67%
B	B	B	B	B	B	50,485	2%	69%

U = Underground, B = bus

Finally, the top two daily journey stage patterns are presented in terms of the start time of each stage. The contrast between two Underground journey stages (see Figure 3-4) and two bus journey stages (see Figure 3-5) is remarkable. The two Underground journey stages show a clear commuting pattern with the start of the first stage occurring in a sharp AM peak and the start of the second journey stage taking place in a slightly wider PM peak. Conversely, the peaks are less sharp in the case of two bus journeys, in part because the second journey stage tends to start much earlier, likely representing onward journeys rather than return journeys as for the Underground. Of course, these differences are in part an artifact of Underground-to-Underground interchanges taking place "behind-the-gate" and therefore not being recorded in the smart card data for the TfL network, but they also illustrate the added value of determining which bus journey stages to link together to form complete journeys that allow a distinction to be made between return journeys and one-way, two-stage journeys.

Overall, this section has shown that a simple examination of bus and Underground journey stages by time period reveals some interesting travel patterns but that it is difficult to compare bus and Underground travel behavior, as well as journeys involving both modes, without linking journey stages into complete journeys based on identified interchanges.

Figure 3-4 Journey Stage Start Times, Passengers with Two Underground Journey Stages

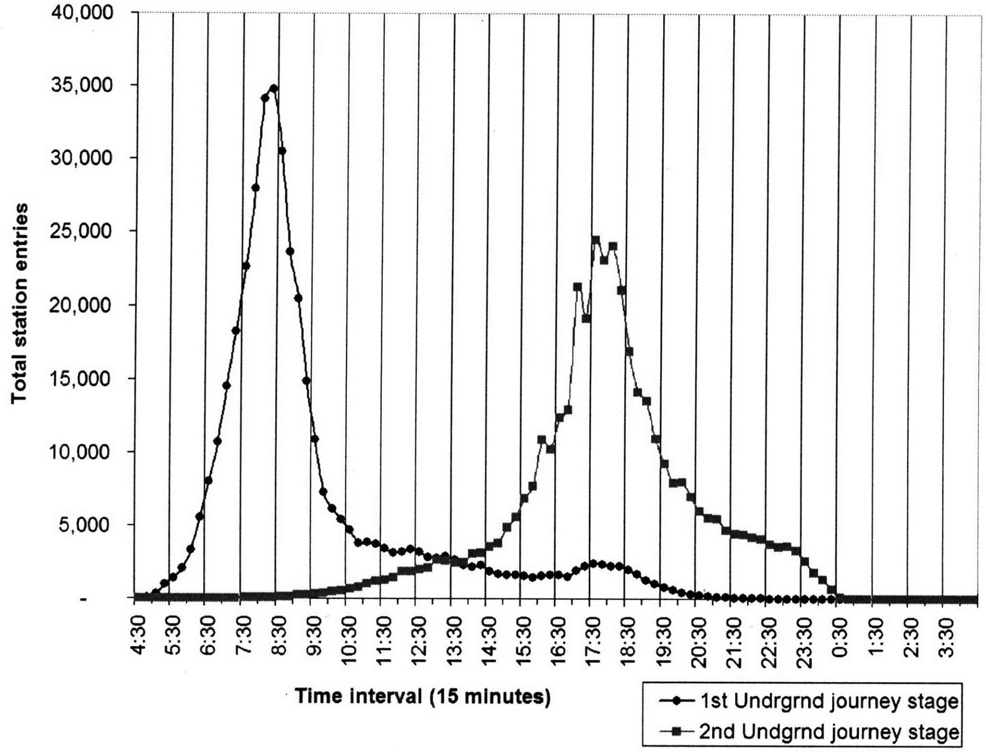
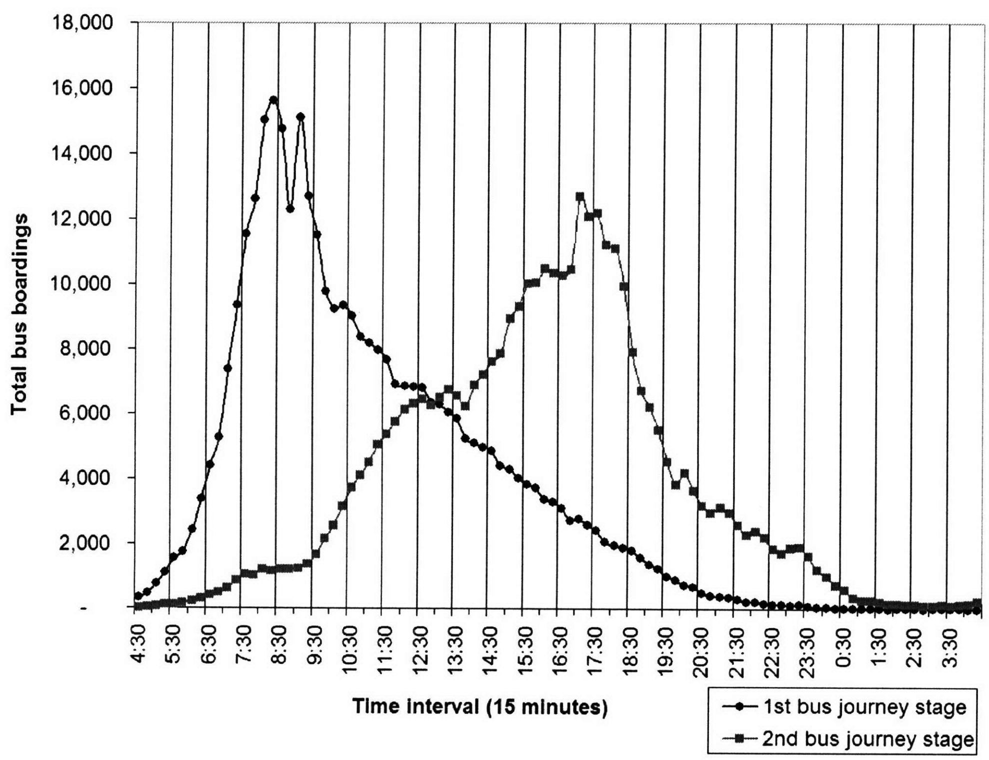


Figure 3-5 Journey Stage Start Times, Passengers with Two Bus Journey Stages



3.3 Underground-to-Bus Interchanges

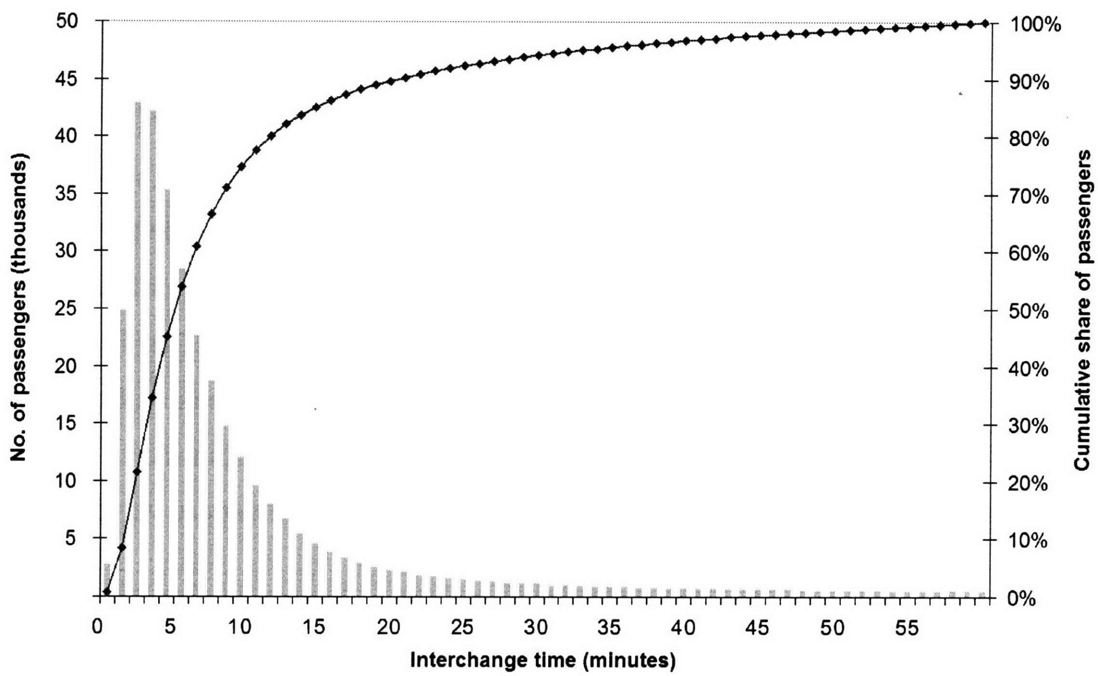
The purpose of this section is, first, to determine an appropriate network-wide interchange time assumption for Underground to bus in London, and, second, to examine variability in interchange times by location and across time periods. This will inform the identification of complete journeys on the TfL network by linking Underground and bus journey stages for each smart card user. It also provides a baseline of observed interchange behavior for future studies of selected Underground stations or interchange facilities.

Recall that a potential Underground-to-bus interchange is characterized by the time in minutes between Underground station exit and bus boarding onto a route that serves the vicinity of that station. Both the station exit and bus boarding must be recorded by a transaction, or "tap", with a unique smart card. Routes that serve the vicinity of each station were identified through a manual review of published TfL "spider" maps. Despite the high frequency of bus services in London and the low likelihood of people being willing to wait more than an half an hour for a bus, it nonetheless seemed prudent to examine all potential Underground-to-bus interchanges of less than 60 minutes in order to expose variability by location and across time periods.

Based on TfL's network planning guidelines that specify a turn-up-and-go bus service with maximum 12-minute headways whenever feasible, we would predict an expected wait time across the network of about 6 minutes assuming random passenger arrivals. However, the multitude of routes serving many Underground stations means that the observed headway for many passengers would actually be less than 6 minutes. Indeed, TfL reports an actual average wait time on high frequency services of 5.5 minutes, which is 1.1 minutes in excess of their expected average wait time of 4.4 minutes (Transport for London 2008).

Figure 3-6 shows that 90 percent of potential Underground-to-bus interchanges shorter than 60 minutes occur within 20 minutes of station exit. The cumulative share of potential interchanges increases quite sharply for the first 10 minutes and the highest volume is observed at just 1 to 3 minutes. However, after 10 minutes the cumulative share of potential interchanges starts to increase at a lower rate until it reaches a low, stable rate at about 30 minutes. This trend suggests that potential interchanges shorter than 10 minutes are pure Underground-to-bus interchanges as people board a bus almost immediately after exiting the Underground. Conversely, those potential interchanges over 30 minutes are unlikely to be pure interchanges as people board buses randomly after completing some activity that was the purpose of the first journey stage. In between these two bounds, the observed trend likely represents a mix of pure interchanges and incidental activity interchanges.

Figure 3-6 All Potential Underground-to-Bus Interchanges



Thus, for the purposes of joining journey stages to form complete journeys, an appropriate interchange time threshold for Underground-to-bus in London lies somewhere between 10 and 30 minutes. In order to further inform the Underground-to-bus interchange time assumption, observed behavior across time periods and by location, which we might expect to vary as a result of station physical design, bus level of service (i.e., frequency and reliability), and passenger behavior is examined.

3.3.1 Time Periods

Figure 3-7 shows the cumulative share of potential Underground-to-bus interchanges across time periods. For time period analyses throughout this chapter, the time periods used in the *London Travel Report 2007* (Transport for London 2008) were adopted:

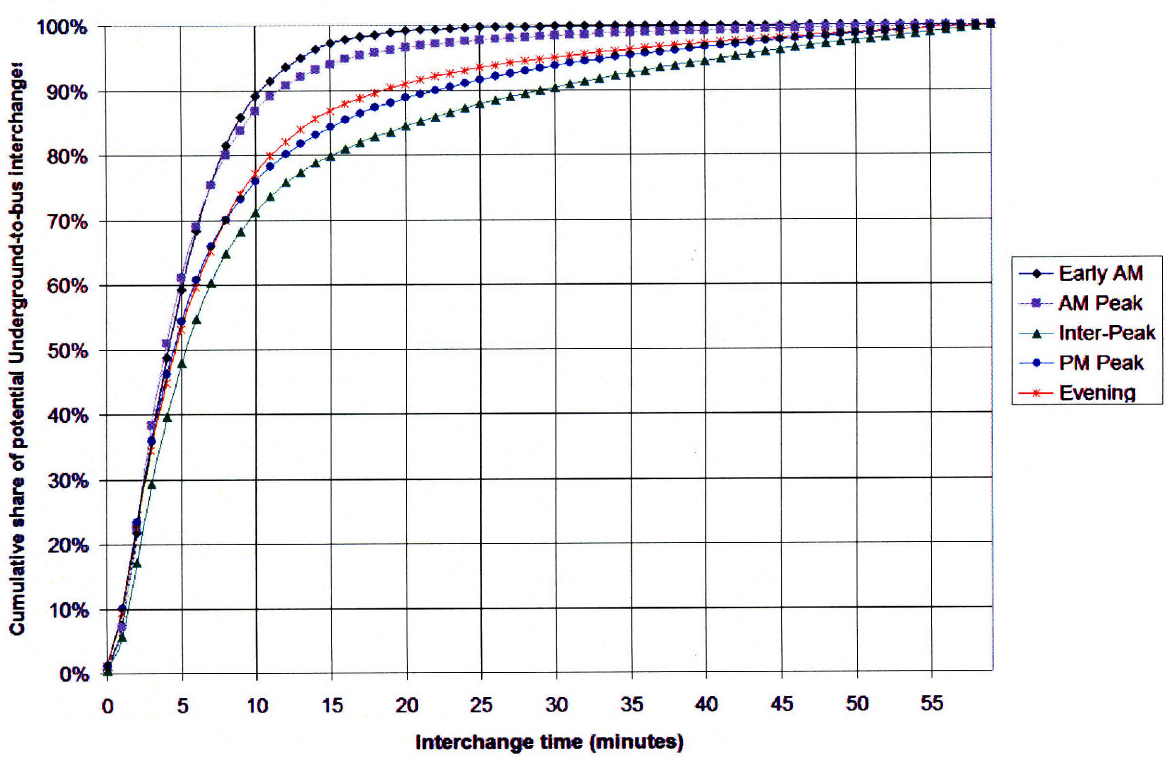
- Early AM: 0400 – 0700,

- AM Peak: 0700 – 1000,
- Inter-peak: 1000 – 1600,
- PM Peak: 1600 – 1900, and
- Evening: 1900 – 2200.

Nighttime (2200 – 0400) was excluded from time period analysis due to data scarcity.

Given that we expect the level of bus service to be similar in the AM and PM Peaks, it is interesting that the aggregate observed behavior in Figure 3-3 is quite different across these two time periods. Although it is possible that less reliable service or congested buses in the PM Peak is causing more people to wait longer for the bus, this is likely to be manifested in the shift of the curve to the right relative to the AM Peak rather than the earlier flattening of the curve, which suggests that during the PM Peak more people are engaging in other activities before starting their bus journey stage.

Figure 3-7 Potential Underground-to-Bus Interchanges Across Time Periods



The sharp rise in the cumulative share of potential transfers in the Early AM and AM Peak is not surprising if we expect that people are less likely to engage in activities other than interchanging between journey stages during these time periods. Moreover, comparing this observation to the Inter-peak trend lends support to the intuition that the cumulative share of potential transfers begins to level off due to people engaging in other activities before boarding a bus for their onward journey. Finally, the Evening trend suggests a mixture of behaviors, both interchange as soon as possible after exiting the Underground and engagement in other activities before the second journey stage. Interestingly, the aggregate distribution shown in Figure 3-6 (previous) is almost identical to the Evening.

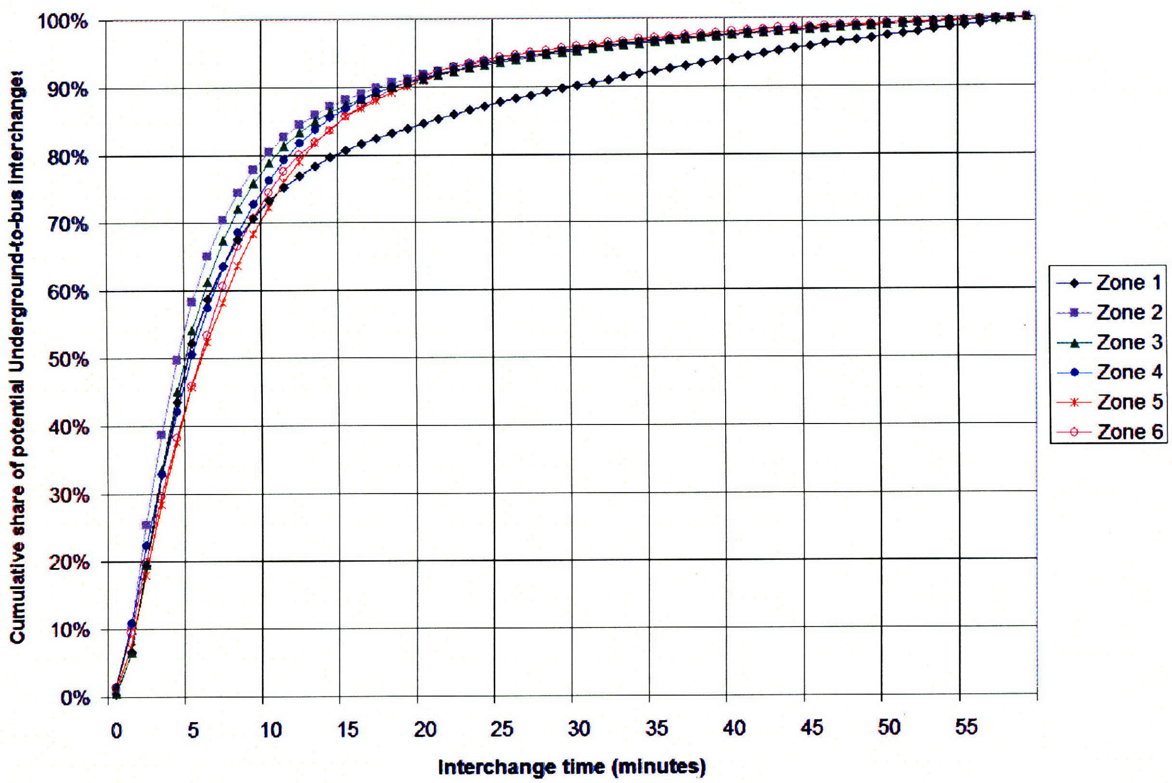
These results imply that an appropriate interchange time threshold lies within a narrower band than 10 to 30 minutes: using 10 minutes would miss pure interchanges in the Early AM and AM Peak but using 30 minutes would include many potential interchanges during which people are engaging in non-incidental activities. Therefore, 15 to 25 minutes is probably a better range for an Underground-to-bus interchange time threshold.

3.3.2 Fare Zones

TfL fare zones serve as a rough proxy for location, with Zone 1 representing Central London and Zone 6 representing the suburban areas. Throughout this chapter, TfL Zones 1 through 6 are used to represent urban location with Zones 7, 8, and 9 excluded due to data scarcity. In general, we might expect interchange times from Underground-to-bus to be lower in Central London because of higher network density that reduces the expected waiting time for passengers who can choose among several buses running along shared corridors to reach their destination.

Figure 3-8 shows the cumulative share of potential Underground-to-bus interchanges for Zones 1 through 6. Zone 1 exhibits a different distribution than the other zones, not reaching 90 percent until 30 minutes whereas the other zones are similar to the aggregate trend shown in Figure 3-2. The distribution for Zones 2 to 6 converge at about 20 minutes, covering 90 percent of potential interchanges. Clearly, something is different in Zone 1. Perhaps people are more likely to engage in other activities between Underground station exit and bus boarding in Central London, inducing a greater mix of behaviors after 12 minutes than in the rest of London. The Zone 1 distribution suggests that most pure Underground-to-bus interchanges occur within 12 minutes of Underground station exit, but across the other zones this bound is at least 15 minutes. In addition, the cumulative share of potential interchanges increases most quickly for Zone 2, then 3, 4 and so on, which may be due to a slight influence of lower frequencies and less route overlap in bus service further from Central London.

Figure 3-8 Potential Underground-to-Bus Interchanges by Fare Zone

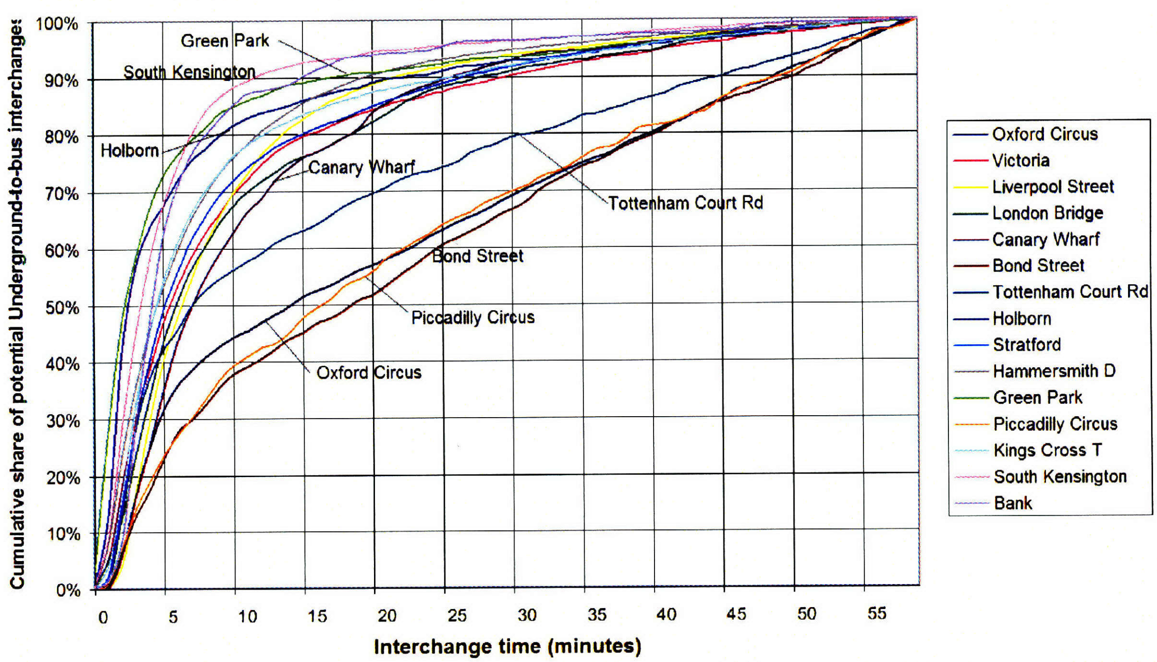


To summarize, Figure 3-8 suggests once again that 30 minutes is a generous time window if we want to be sure to exclude potential interchanges that are not pure interchanges, but even 15 minutes may include many incidental activities for travelers in Zone 1. On the other hand, beyond Zone 1 the minimum interchange time assumption that is needed to be confident of including most pure interchanges is 15-20 minutes.

3.3.3 Underground Stations

So far, level of service does not appear to be a significant factor in explaining observed variations in Underground-to-bus interchange behavior by fare zone or across time periods. However, each category is comprised of a mix of station types so the aggregate trends may mask important differences in passenger behavior across stations. For example, Figure 3-9 shows the cumulative share of potential interchanges for the fifteen largest Underground stations by exit volume based on Oyster smart card data. These stations are important to consider because they have the potential to be included in a large number of multi-stage journeys. It is somewhat surprising that these stations exhibit quite different distributions in the cumulative share of potential Underground-to-bus interchanges. In particular, potential interchanges at stations such as Oxford Circus, Piccadilly Circus, and Tottenham Court Road appear to include many passengers engaging in non-interchange activities in less than 5 to 10 minutes from station exit whereas other stations such as South Kensington, Green Park, and Bank do not exhibit this tendency. In fact, the cumulative share of potential interchanges in the later group increases more quickly than the aggregate trend across all stations.

Figure 3-9 Potential Underground-to-Bus Interchanges at Highest Exit Volume Stations

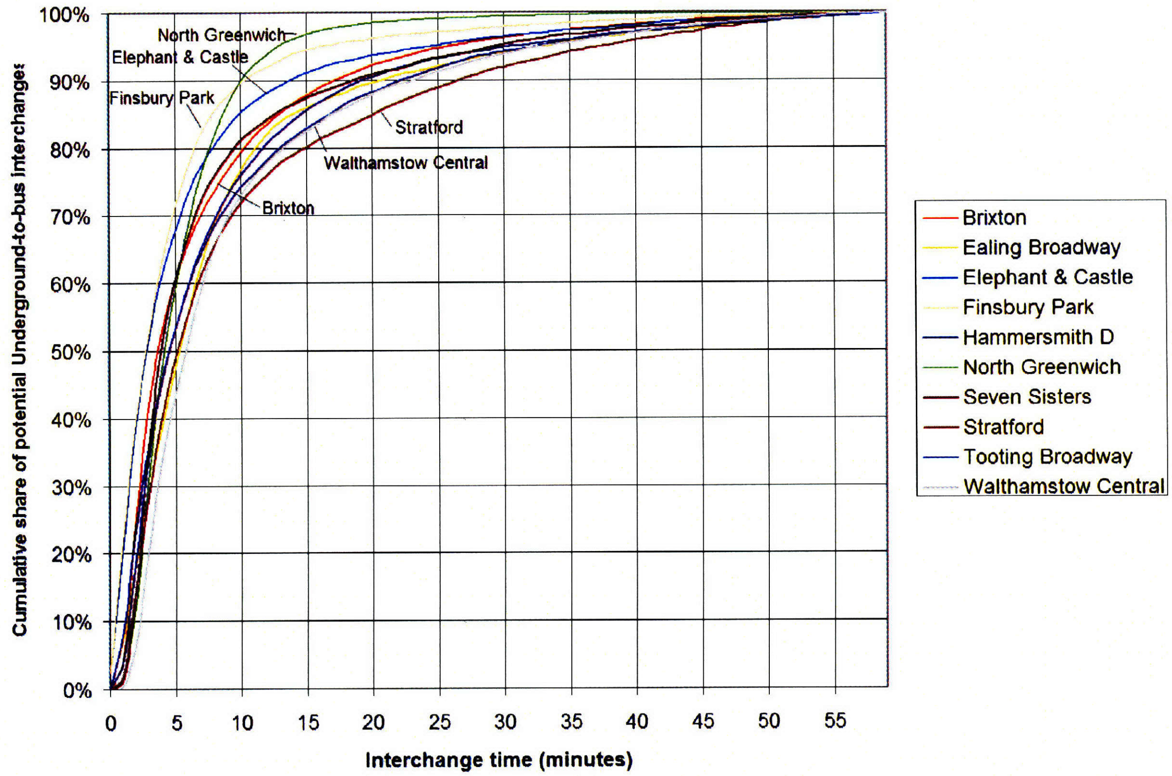


Clearly, people are behaving differently at these large, Central London Underground stations. One possible explanation is that the large, varying mix of potential activities near these stations, especially the availability of shops, is influencing people's behavior. Another possible explanation is the complex nature of the interchange facilities at the first group of stations, including stops that are spread out on surrounding streets and many routes that serve the vicinity but are not feeder routes for those particular stations. As a further example, the stations in the mid range, including Victoria, Hammersmith (District & Circle), Kings Cross, and Stratford are all large interchange facilities that include several modes (i.e., DLR or National Rail), a mix of surrounding land uses including some shops, and dedicated bus station areas. This mix of factors makes it more difficult to discern an appropriate time threshold for interchange at these locations.

For this group of large Underground stations, we can be fairly confident that a 10 minute interchange time threshold would capture pure interchanges, but an examination of the stations with the lowest exit volumes (not illustrated) indicates that a 20 minute assumption would be more appropriate in order to include all pure interchanges and avoid Type I error. A 20 minute time threshold assumption would mean that incidental activity interchanges at large Underground stations would be included.

A final way to compare Underground stations is to look at those with the largest volume of potential transfers. Figure 3-10 shows the Underground stations with highest number of potential Underground-to-bus interchanges under 60 minutes. The first thing to note is that the stations included here are mostly different than the largest exit volume stations and, second, that the cumulative distributions are more similar to each other than in the previous case. Nonetheless, the difference between Stratford (bottom right) and North Greenwich (top left), for example, is notable. One possible explanation is that North Greenwich has very few nearby activity generators whereas at Stratford people may engage in other activities such as shopping before continuing their journey – or starting a new journey if their activity at Stratford was the purpose of the first journey stage. At a station such as North Greenwich it is very easy to determine an appropriate interchange time assumption whereas at a more complex station such as Stratford it is more challenging.

Figure 3-10 Potential Underground-to-Bus Interchanges at Highest Interchange Volume Stations



3.3.4 Summary

The analysis in this section leads to three preliminary conclusions about Underground-to-bus interchange in the TfL network. First, at the aggregate level, TfL's maximum 12 minute turn-up-and-go bus network planning guideline is evident in the cumulative distribution of potential Underground-to-bus interchanges. The data shows the highest volume of interchanges at 1 to 3 minutes and that most potential interchanges occur in less than 15 minutes. Nonetheless, the second insight is that interchange behavior differs across Underground stations and time periods due to factors other than level of service, probably including station design and surrounding land uses. Acknowledging this variability, the final insight is that a reasonable Underground-to-bus interchange time threshold would be between 15 and 25 minutes for the network as a whole but may range from 10 to 30 minutes for any given station.

3.4 Bus-to-Underground Interchanges

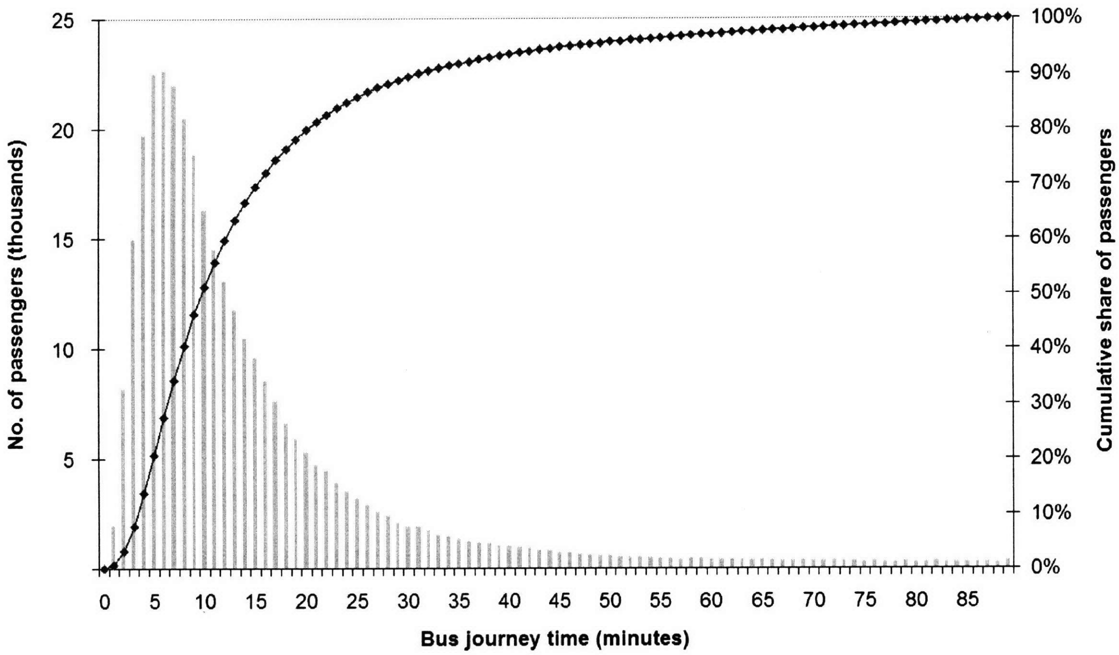
In order to join journey stages recorded by unique Oyster smart cards into complete journeys, it is also necessary to make a time threshold assumption about the maximum duration of bus travel to access the Underground station. Since Oyster smart card transactions do not provide any information about where the cardholder gets off the bus, I limit the geographic scope for the start location of the second journey stage to Underground stations in the vicinity of the bus route used for the first journey stage.

Figure 3-11 shows the cumulative share of potential bus access journey times less than 90 minutes aggregated over all stations. Single bus journey stages longer than 60 minutes are rare in London where the average journey stage length is 3.7 kilometers and the average traffic speed in Central London in the morning peak is 16.4 km per hour, resulting in an approximate journey stage duration of about 13.5 minutes (Transport for London 2008). Nevertheless, in order to account for differences across stations, time periods, and fare zones, a much wider window of potential bus access journey times is initially examined. Indeed, Figure 3-11 indicates that the volume of potential bus access journeys longer than 60 minutes is low and is more likely to be evidence of people taking a bus journey then later entering a station served by that route after engaging in some other activity.

Although the highest volume of potential bus access journeys is just 5 to 7 minutes in duration (including interchange time at the station), the volume increases rapidly from 2 to 20 minutes. This suggests that any bus boarding followed by an Underground station entry within 20 minutes is very likely to be a pure bus-to-Underground interchange. However, in the range between 20 minutes and 60 minutes, the cumulative share of potential bus journey access times increases more slowly and suggests a mix of behaviors: people interchanging directly from bus to Underground and others engaging in incidental activities before starting a new journey at the Underground station. The question of where to draw the line between activities that are incidental to the interchange and those that are the purpose on the journey is somewhat arbitrary and is further complicated by a wider range of bus access journey times than expected wait times for Underground-to-bus.

Based on Figure 3-11, I propose that an appropriate bus access journey time threshold lies between 20 and 60 minutes, but this is a fairly wide range so next differences in potential bus access times across time periods and by location are discussed with the purpose of determining a narrower range of appropriate elapsed time thresholds and of illustrating variability in passenger behavior.

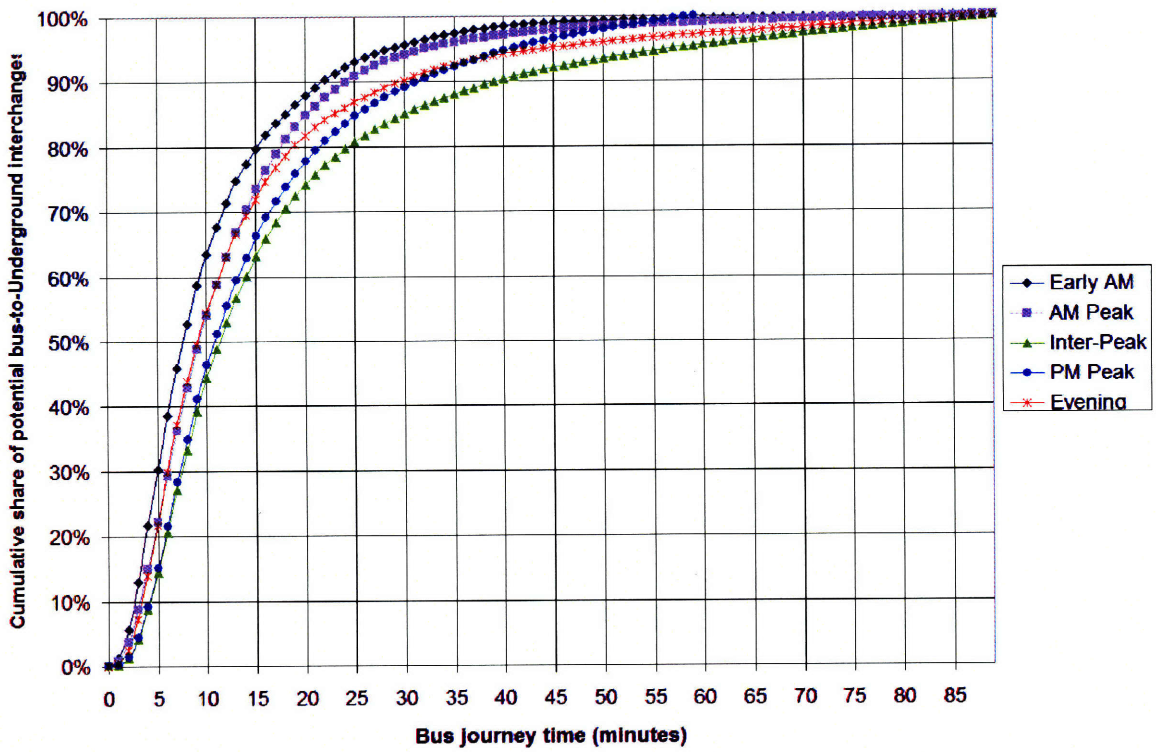
Figure 3-11 All Potential Bus-to-Underground Interchanges



3.4.1 *Time Periods*

Similar to the Underground-to-bus time period analysis, Figure 3-12 makes intuitive sense in terms of the Early AM and AM Peak trends relative to the other time periods. People commuting to work in the morning are less likely to engage in other activities between their first bus journey stage and entering the Underground so the distribution of potential access times is more tightly distributed and therefore the transition from a sharply rising curve to a near-flat trend line is quite abrupt compared with the other periods. Conversely, it appears that people are most likely to engage in non-travel activities between their bus journey stage and entering the Underground during the Inter-peak period. The Inter-peak and PM Peak distributions are shifted further to the right than other time periods which suggests that, on average, bus journey stages may be longer during these periods. Finally, it is worth noting that differences in observed travel behavior are probably not entirely due to different levels of service or congestion delays because the AM Peak distribution differs from the PM Peak even though we would expect them to be similar based on these factors.

Figure 3-12 Potential Bus-to-Underground Interchanges by Time Period

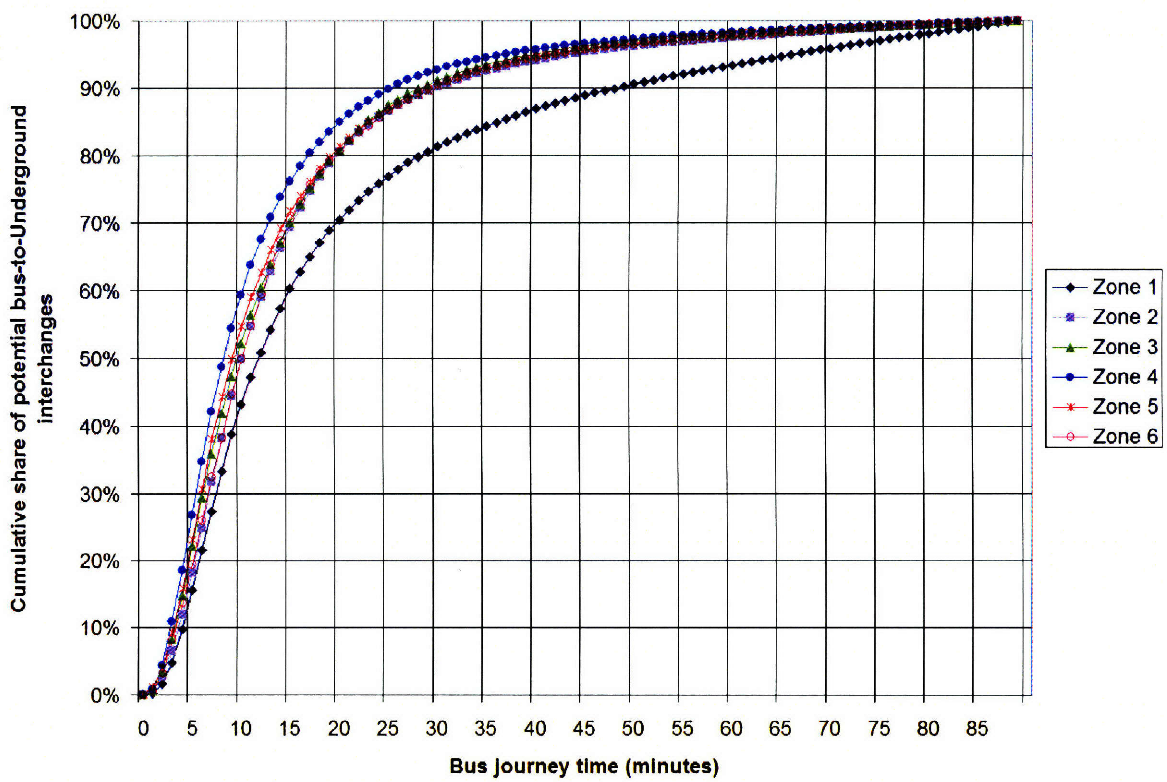


In terms of an appropriate "interchange" time threshold, Figure 3-12 suggests that a minimum of 25 minutes is needed to capture the pure bus-to-Underground access journeys in the Early AM and AM Peak. On the other hand, the upper threshold should be less than 60 minutes because otherwise it is likely to include too many non-travel activities in the Inter-Peak and Evening time periods. Therefore, I contend that an appropriate bus access journey time threshold lies between 25 and 45 minutes, by which point the distributions for all the time periods have passed the 90th percentile so we can be confident that all pure interchanges and most incidental activity interchanges have been included.

3.4.2 Fare Zones

Another way to group potential bus-to-Underground access journeys is by location. Figure 3-13 shows the cumulative distribution of potential bus-to-Underground access journeys by TfL fare zone. It is again immediately apparent that the observed behavior in Zone 1 differs considerably from the other zones. The fact that Zone 1 includes the central areas of London where many stations are surrounded by a high concentration of activity generators (i.e., business, shops, entertainment) lends support to the intuition that people are more likely to engage in other activities between their bus journey and entering the Underground in Zone 1. Nonetheless, it appears that average journey stage duration is fairly similar to the other zones as the distribution initially increases at a similar rate and also begins to flatten at approximately the same time threshold of 30 minutes. Conversely, the distribution for Zone 4 exhibits a slightly sharper transition from a rising cumulative share of potential bus access journeys to a nearly flat distribution. This suggests that fewer people engagie in non-travel activities between bus and Underground journey stages in Zone 4.

Figure 3-13 Potential Bus-to-Underground Interchanges by Fare Zone



Overall, Figure 3-13 lends support to the assertion that an appropriate time threshold for assuming a bus-to-Underground access journey lies between 30 and 50 minutes, where the cumulative distribution function for Zone 1 reaches the 90^{th} percentile so we can be confident that all pure interchanges and most incidental activity interchanges have been included.

3.4.3 Underground Stations

The similarity in observed bus-to-Underground journey access times across Zones 2 to 6 in the previous section may well mask important differences across individual stations. For example, Figure 3-14 shows the cumulative distribution of potential bus access times for the ten largest Underground stations by entry volume. The wide range of observed passenger behavior across stations is immediately apparent. For example, it is clear that most bus access journeys to Canary Wharf Station are less than 20 minutes long and that there are probably few people engaging in non-travel activities between journey stages. However, a threshold of 35 minutes would be necessary at Hammersmith Station (District and Circle Lines) in order to have the same level of confidence of including all pure bus access journeys. This comparison suggests that, on average, passengers travel longer by bus to access Hammersmith Station than Canary Wharf Station. Moreover, observed passenger behavior at several stations make it impossible to discern a threshold assumption for pure bus access journeys because the cumulative share of potential access journeys rises at a near constant rate for the entire 90 minutes under consideration. This occurs at stations such as Oxford Circus, Leicester Square, and Tottenham Court Road that are urban destinations in their own right. The observed trend is likely a result of people engaging in non-travel activities between journey stages. In either case, the difficulty in determining an appropriate access journey time threshold for these stations implies that in order to avoid Type 2 error almost any elapsed time assumption will allow incidental activity interchanges to be included in bus-to-Underground journey stage sequences.

Another way to classify individual stations is by volume of potential interchanges. Figure 3-15 shows the cumulative share of potential bus-to-Underground access journey times for the ten stations with the most potential interchanges of that type. North Greenwich Station stands out immediately as a location with both longer bus access journeys and very few people engaging in non-travel activities between journey stages.

Figure 3-14 Potential Bus-to-Underground Interchanges, Highest Entry Volume Stations

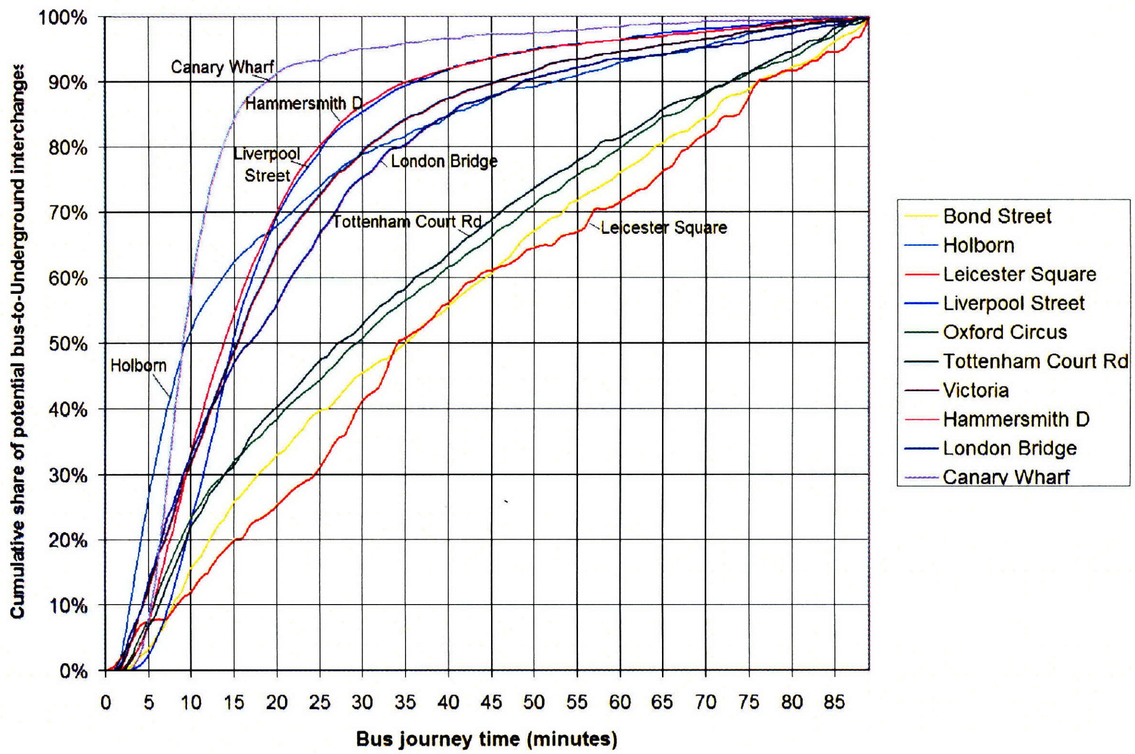
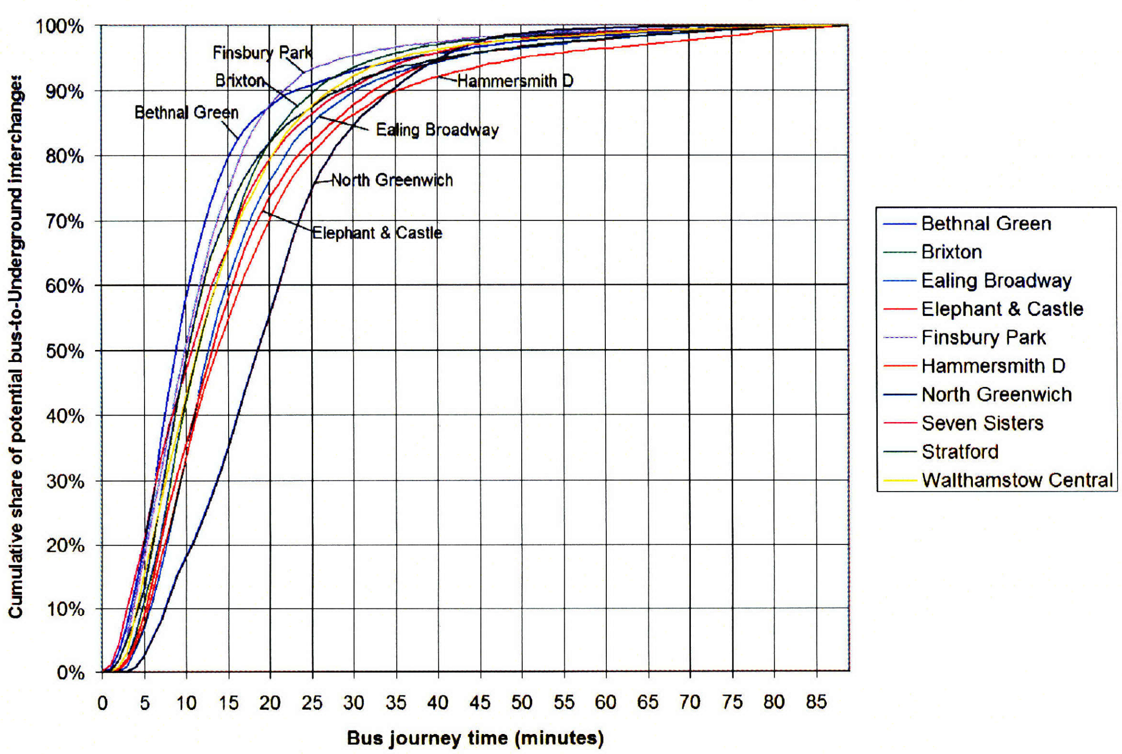


Figure 3-15 Potential Bus-to-Underground Interchanges, Highest Interchange Volume Stations



For North Greenwich, 45 minutes would be an appropriate bus access journey time threshold, but using this threshold would likely capture many non-travel activities at the other stations where a more appropriate time threshold lies in the 25 to 40 minute range. These variations in bus access journey times across stations are not unique to large stations, but these examples illustrate the variability in behavior behind the aggregate distribution shown previously. Unlike potential Underground-to-bus interchanges where the variability is due largely to differences in the amount of non-travel activities occurring between journey stages, this variability is the result of both differences in typical bus in-vehicle travel times and in the relative prevalence of non-travel activities during potential interchanges.

3.4.4 Summary

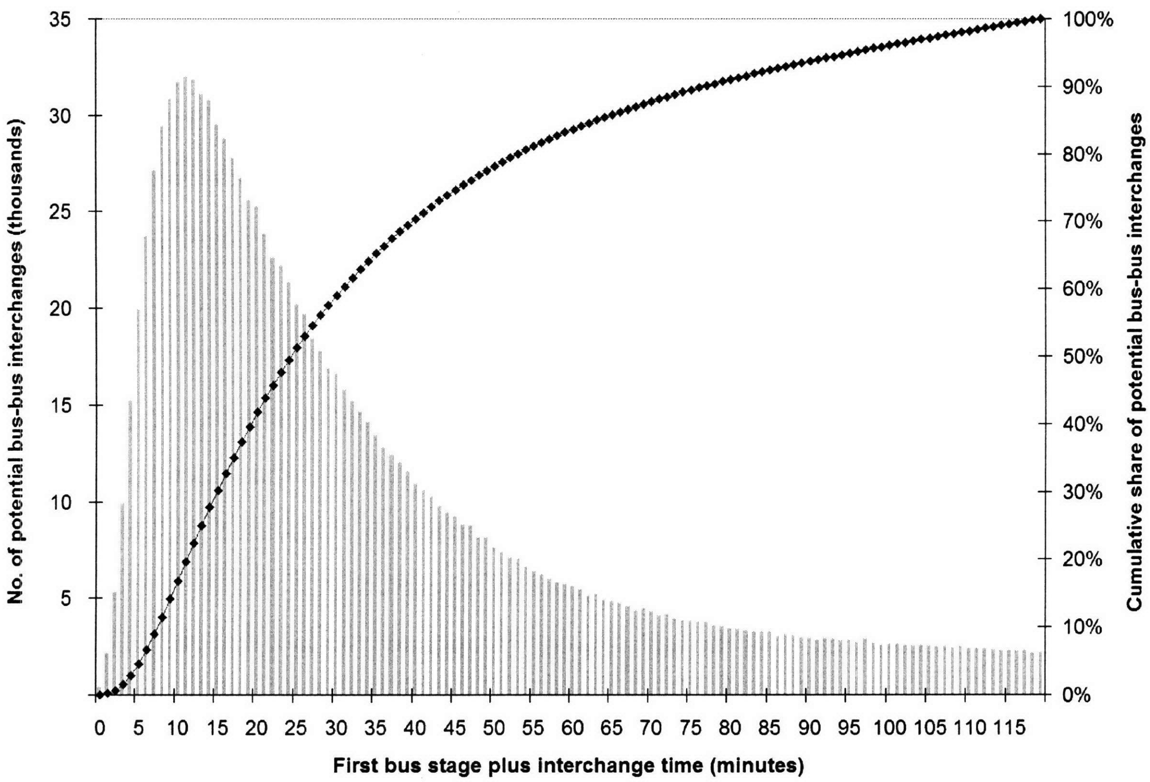
Based on this review of potential bus-to-Underground access journey times across locations and time periods, I conclude that an appropriate threshold for maximum bus journey access time over the entire network lies in the range of 30 to 50 minutes. Using a threshold of 40 minutes would under-represent pure bus-to-Underground interchanges at a few stations such as North Greenwich, but could nevertheless include a significant number of non-interchanges at other stations. In Chapter 4, thresholds of 30 to 50 minutes are tested against extreme scenarios of 10 and 70 minutes for bus-to-Underground access journey time.

3.5 Bus-to-Bus Interchanges

A third combination of journey stages that could be linked to create complete journeys on the TfL network is potential bus-to-bus interchanges. Pure bus-to-bus interchanges are difficult to identify in the Oyster sequence journey table for two reasons: first, the complexity of the bus network and, second, the lack of spatial detail currently available for bus boardings. In other words, the complexity of the bus network compounds the fact that only the route number is known for each bus smart card transaction. In a simple network, it would be possible to manually identify plausible bus-to-bus interchange points where routes intersect across the network. However, in the case of over 700 curvilinear bus routes in London this task of identifying plausible interchange points would need to be either automated or constrained to a few routes. Nonetheless, in this section, aggregate patterns of time differences between sequential bus boardings per passenger are examined in order to determine an appropriate time threshold for linking a second bus journey stage to its preceding bus journey stage.

To this end, Figure 3-16 shows the cumulative share of potential bus-to-bus interchanges for which the start of the second journey stage occurs less than 120 minutes after the start of the first journey stage. It also shows the total volume of potential bus-to-bus interchanges at each time threshold. The counts of potential bus-to-bus interchanges exclude two sequential journey stages recorded on the same route by the same smart card because this clearly constitutes a return journey. That being said, in London many buses run along shared corridors for part of their route, so there are often opportunities for people to make a return journey on a different route than they used for their outward journey.

Figure 3-16 All Potential Bus-to-Bus Interchanges



This inability to exclude all return journeys from potential bus-to-bus interchanges may provide a partial explanation for the lack of a clear transition from a rapidly increasing share of potential bus-to-bus interchanges to very few potential interchanges as compared to the observed behavior in the bus-to-Underground and Underground-to-bus scenarios discussed previously. Another possible explanation for the shape of the curve is that there is a lot of variation in bus-to-bus interchange behavior, including a significant number of people continuing journeys on another bus route after engaging in non-travel activities between journey stages.

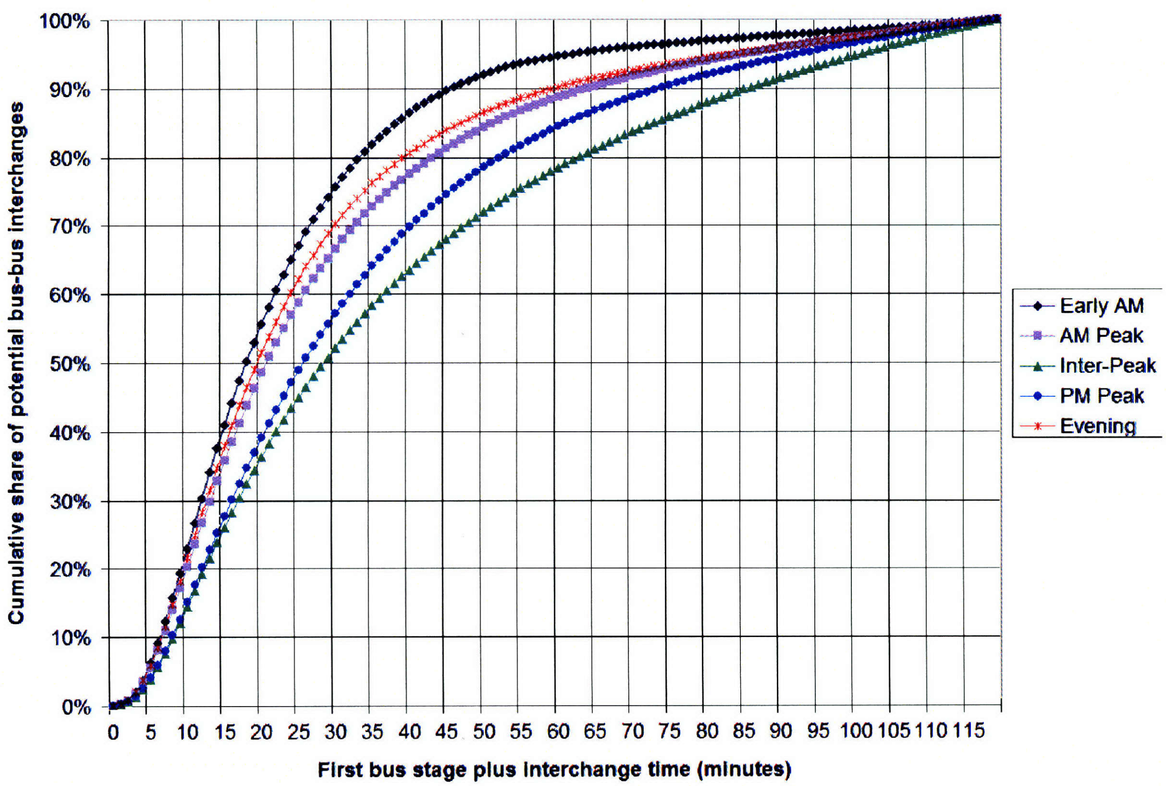
Overall, it can be inferred from Figure 3-16 that a reasonable maximum time threshold for assuming bus-to-bus interchange based on the time between first and second bus boardings is in the range of 40 to 80 minutes. However, this wide range could allow for non-interchanges to be identified as interchanges, or, conversely, miss out on a significant number of pure interchanges. Next, differences across time periods and types of users are examined in order to narrow the range of reasonable time thresholds for assuming bus-to-bus interchange.

3.5.1 Time Periods

By dividing potential bus-to-bus interchanges into time periods based on the start time of the first bus journey stage (see Figure 3-17), it becomes evident that the aggregate trend shown in Figure 3-16 is influenced significantly by observed travel behavior in the Inter-Peak period. This is not surprising because potential bus-to-bus interchanges in the Inter-Peak period represent 44 percent of all potential bus-to-bus interchanges less than 120 minutes. Travel patterns in the Inter-Peak period are clearly complex as there is a wide distribution of time differences between the start times of sequential bus journey stages. Nonetheless, in the Early AM time period potential bus-to-bus interchanges, which include travel time for the first journey stage, exhibit a relatively clear transition in the 45 to 55 minute range from mostly pure interchanges to the random effects of people engaging in non-travel activities between bus journey stages. Likewise, the Evening trend exhibits a transition zone lying in the 40 to 60 minute range and is fairly similar to the AM Peak.

Is it likely that people tend to take longer initial bus journeys in the PM Peak and Inter-Peak periods? Or, could it be that interchange wait times are longer during these two periods? Both situations are plausible, but given the symmetry one would expect in commuting behavior (i.e., AM Peak vs. PM Peak) and a priori knowledge of the complexity of travel behavior in the Inter-Peak period, at least some of the observed variation in distributions is likely due to more people engaging in short non-travel activities between sequential bus journey stages during these time periods. Regardless, the combination of non-travel activities, travel time variability, and wait time variability makes it difficult to infer a single appropriate time threshold for assuming bus-to-bus interchange. More information is provided by comparing different types of passengers on return and non-return journeys below.

Figure 3-17 Potential Bus-to-Bus Interchanges Across Time Periods



3.5.2 Pay-As-You-Go versus Pass Holders

Because it is too difficult to categorize bus routes by spatial location for the purposes of this analysis, potential bus-to-bus interchange behavior for different types of ticket holders is examined as another means of understanding variations in interchange behavior across the system. Figure 3-18 shows the raw and cumulative volume of elapsed times between the start times of the first and second journey stages for people who take only two bus journey stages a day on *different* routes. The passengers are grouped into two fare payment categories, Pay-As-You-Go (PAYG) and Pass Holders.

Figure 3-18 Potential Interchanges, Passengers with Two Bus Boardings on Different Routes

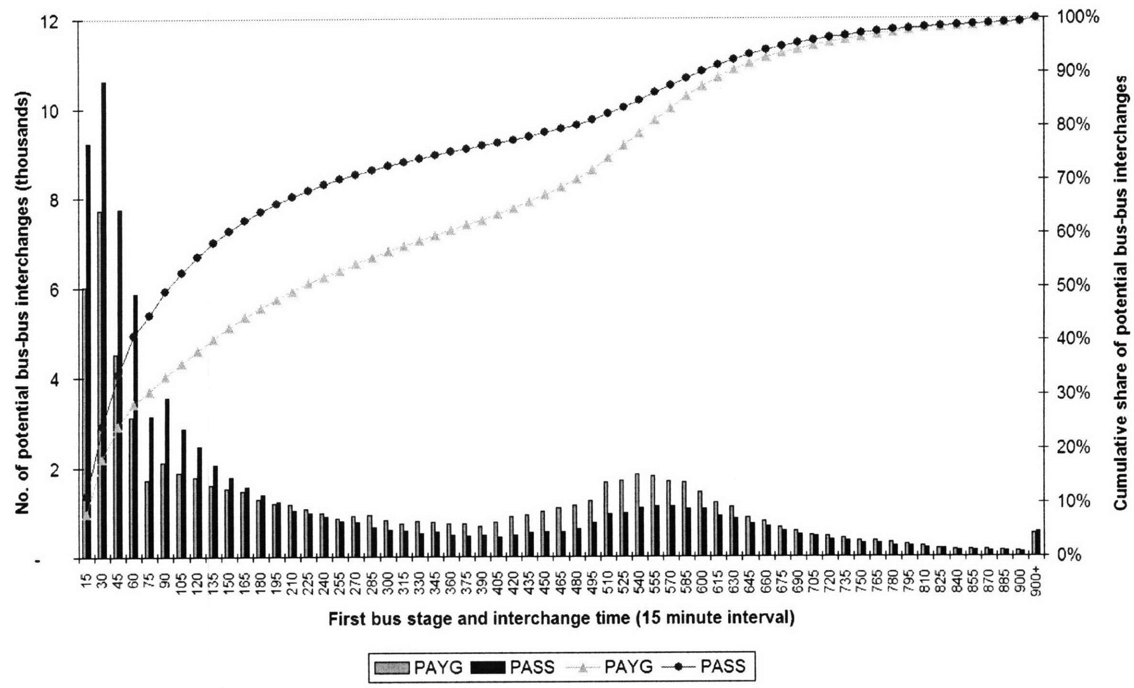
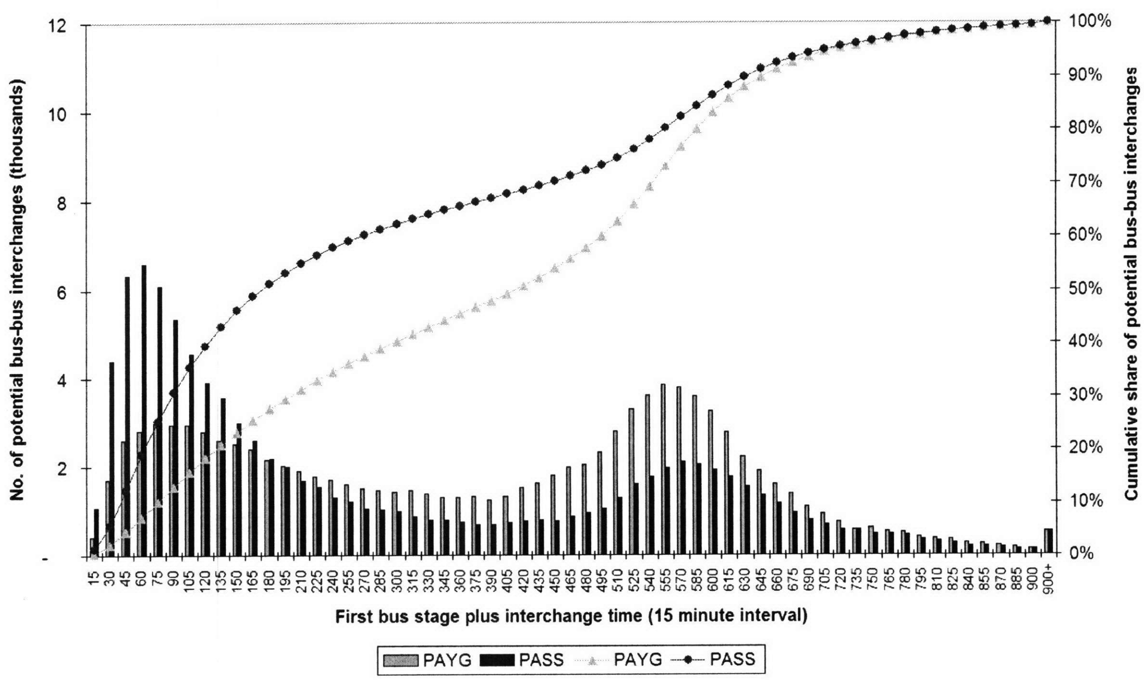


Figure 3-19 Potential Interchanges, Passengers with Two Bus Boardings on Same Route



The PAYG fare category means that the cardholder is paying separately for each bus journey stage whereas the Pass category means that a season ticket, for example Monthly Bus Pass or Weekly Travelcard, has been loaded onto the Oyster card thereby allowing the passenger to travel on unlimited bus journey stages during the days covered by the season ticket. Figure 3-19 shows this information at the same scale for people who take two bus journey stages a day on the *same* route.

Together, Figures 3-18 and 3-19 illustrate the difficulty in determining an appropriate bus-to-bus interchange time threshold. The graph of journeys on the same route shows that for Pass Holder passengers the highest volume of return journeys occurs between 45 and 60 minutes from the time of the first bus boarding whereas for PAYG passengers the highest volume occurs between 60 and 75 minutes. This suggests that Pass Holders are more likely to make short return journeys than PAYG passengers. It is reasonable to assume that the highest volume of return journeys on different routes would likewise occur within these same time thresholds. (The prevalence of return journeys on different routes is supported by the second rise in time difference volumes in Figure 3-18 that occurs about 9 hours after the start time of the first journey stage.) Thus, it is likely that a share of the potential bus-to-bus interchanges on different routes actually represent return journeys within elapsed time thresholds of 45 to 75 minutes. In short, it is most likely that two sequential bus journey stages on different routes with 30 minutes or less between bus boardings represent pure bus-to-bus interchanges (i.e., onward journeys), but above that threshold a significant proportion of potential interchanges would actually include return journeys and/or non-travel activities between journey stages.

3.5.3 Summary

Overall, bus-to-bus interchanges are the most difficult potential interchange sequence to identify as true interchanges because of (a) the lack of spatial detail and (b) the complexity of bus travel behavior. Nevertheless, by examining differences in behavior across types of users and time periods, it can be inferred that the most appropriate maximum time threshold between sequential bus boardings for assuming a bus-to-bus interchange is between 40 and 60 minutes. This range is explored in the following chapter on linking journey stages to create complete journeys.

3.6 Summary

This chapter provided a general definition of interchange as the act of transferring between modes or between different services of the same mode, but also emphasized that although the definition is clear, the ability to accurately identify interchanges using smart card data is limited by a lack of information on journey purpose. Thus, an exploration of elapsed times between journey stages for passengers using Oyster smart cards to travel by bus or Underground on the TfL network was used to develop recommended elapsed time thresholds for identifying interchanges across the network.

Table 3-2 Recommended Elapsed Time Thresholds for Identifying Interchanges

Interchange Type	Elapsed Time Threshold
Underground-Bus*	15-25 minutes
Bus-Underground*	30-50 minutes
Bus-Bus**	40-60 minutes

*Not restricted by physical proximity of route and station.
**Limited to journey stages on different routes.

As shown in Table 3-2, this analysis led to the conclusion that a reasonable Underground-to-bus interchange time threshold would lie between 15 and 25 minutes for the network as a whole but ranges from 10 to 30 minutes for any given station. Additionally, an appropriate threshold for maximum bus journey access time over the entire network lies in the range of 30 to 50 minutes. And, finally, elapsed time thresholds between 40 and 60 minutes are most appropriate for bus-to-bus interchanges. These recommended ranges are explored in the next chapter to determine unique elapsed time thresholds for identifying interchanges that are then applied to the bus network planning case studies presented in Chapter 5.

CHAPTER 4: MULTI-STAGE JOURNEYS BY BUS AND UNDERGROUND

The purpose of this chapter is to describe the methodology for, and results of, linking journey stages to form complete journeys on the TfL network using smart card data. The sequenced journeys table used for application throughout this chapter is Wednesday, November 14, 2007.

The chapter begins by summarizing suggested time threshold ranges for linking journey stages into complete journeys that are informed by the analysis of potential interchanges in the preceding chapter. The simple methodology for identifying complete journeys is explained and then I propose to examine the results of applying this methodology for three sets of time thresholds within the suggested ranges and, for purposes of comparison, with two sets outside them (Section 4.1). Next, aggregate travel patterns revealed by complete journeys formed using assumptions within the ranges are compared to those outside the ranges in terms of total journeys generated, average stages per journey, average number of journeys per passenger and modal patterns. These summary statistics are also compared to expectations based on the LTDS from 2005/06 (Section 4.2). Based on this analysis, I conclude that a set of complete journeys formed using maximum elapsed time thresholds of 20 minutes for Underground exit to bus boarding, 40 minutes for bus boarding to Underground entry, and 50 minutes for bus boarding to bus boarding should be used for the application of complete journey data to selected network planning case studies to be presented in Chapter 5 (Section 4.3).

4.1 Assumptions and Methodology

In order to create a network-wide dataset of complete multi-stage journeys for a single weekday using the smart card data currently available in London, it is necessary to make assumptions about acceptable elapsed time thresholds between sequential journey stages[11] for each passenger. This is difficult because the spatial accuracy of bus boarding locations is limited to the route level and there is no information on bus alighting time or location. The

[11] Recall that a 'journey stage' is defined as travel by a single mode and vehicle of transport within a complete journey that may be composed of several journey stages by different modes and vehicles. For example, every time a passenger boards a different bus, it is counted as a separate journey stage.

lack of a link between vehicle location at the time of a transaction, or at any other time, makes it difficult to link journey stages. Nonetheless, there is great value in forming complete journeys based on elapsed time thresholds for the purposes of identifying and better understanding passenger demand for bus interchanges as discussed in Chapter 2. Once a network-wide dataset of complete journeys is generated, a bus network planner could then quickly extract information for any subset of routes or stations without having to re-identify complete journeys, which is computationally demanding. Examples of the application of these types of extracts to bus network planning will be presented in Chapter 5.

As discussed in the definition of interchanges in Chapter 3, the formation of complete journeys based only on time thresholds between journey stages may result in certain sequential stages being linked when in reality the passenger was engaged in non-travel activities between stages. It these activities are incidental to travel (e.g., buying a newspaper) then there is no problem because they should be categorized as interchanges, however if the non-travel activity is the purpose of the trip then we have committed a Type 1 error of falsely identifying an interchange. However, this is preferred to failing to identify pure interchanges because the network planning applications that will be presented in Chapter 5 are more relevant if they include all pure interchanges and most incidental activity interchanges. Smart card data does not record journey purpose information so the difference between incidental and purposeful activities between journey stages cannot be identified with certainty and instead we rely on an analysis of typical interchange times to determine which journey stages to link together. So long as the elapsed time thresholds are fairly representative for the system as a whole, then when a particular route is reviewed the actual journey patterns for passengers on that route may be considered relative to network norms.

In Chapter 3, the following ranges of elapsed time thresholds for identifying interchanges in the London network were proposed:

- Underground-to-bus interchanges from 15 to 25 minutes between Underground station exit and bus boarding;
- bus-to-Underground interchanges, including bus in-vehicle time, from 30 to 50 minutes between bus boarding and Underground station entry; and
- bus-to-bus interchanges, including bus in-vehicle time and waiting time for the second bus stage, from 40 to 60 minutes from first bus boarding to second bus boarding.

To be clear, an elapsed time threshold of 15 minutes for Underground-to-bus interchanges means that any passenger who boards a bus 15 minutes or less after tapping their smart card at an Underground exit gate would then have their Underground journey stage linked to their bus journey stage to form (part of) a complete journey. With the 15 minute maximum time threshold, the two journey stages are considered separate journeys if it takes the passenger 16 minutes or more to move from the Underground exit gate to the bus stop, to wait for the bus to arrive and then to board it. In the case of bus-to-Underground interchanges, under an elapsed time threshold of 30 minutes the two journey stages would be linked if the passenger were to tap their smart card at an Underground entry gate 30 minutes or less after they tapped their smart card on a bus, rode the bus, and walked to the Underground entry gate. Finally, for bus-to-bus interchanges the elapsed time threshold includes the time spent waiting for the second bus as well as the in-vehicle time on the first bus. Therefore, assuming that bus journeys tend to be the same duration when used to access either another bus or the Underground, it makes sense for the bus-to-bus range to be higher than the bus-to-Underground range where the passenger generally does not have to wait to enter the Underground station. Note that more than two stages may be linked together under elapsed time thresholds, resulting in complete journeys with three or more stages.

Underground-to-Underground interchanges are not included explicitly because the vast majority take place behind the ticket gate and are therefore not recorded in the smart card data. That being said, there are a few locations in London where a pure Underground-to-Underground interchange can occur outside the ticket gates, for example Hammersmith, District and Piccadilly Lines Station to the nearby Hammersmith & City Line Station. This research focuses on bus network planning so these infrequent cases of Underground-to-Underground interchanges are ignored.

As presented in Table 4-1, I create three complete journey datasets based on the middle and extremes of each range to see whether they yield similar results in terms of journeys per passenger, stages per journey and modal patterns. Assuming the ranges of elapsed time thresholds are a reasonable reflection of reality, the results from the three datasets are expected to be similar. Moreover, the three complete journey datasets based on reasonable elapsed time thresholds are compared to metrics from LTDS to verify whether they are similar to survey-based results from TfL. I also create two complete journey datasets based on arbitrary assumptions 75 percent higher and 75 percent lower than the ends of the ranges to confirm that the suggested ranges, based on the analysis in previous chapter, actually yield different results than an arbitrary set of interchange time assumptions. The purpose of these comparisons is to decide on a single set of network-wide elapsed time thresholds for application to the case studies presented in Chapter 5.

Table 4-1 Ranges of Elapsed Time Thresholds for Identifying Interchanges

	Within Reasonable Range			**Outside Reasonable Range**	
Interchange Type	*Low*	*Middle*	*High*	*Very Low*	*Very High*
Underground-Bus**	15*	20	25	5	35
Bus-Underground**	30	40	50	10	70
Bus-Bus***	40	50	60	12.5	87.5

* Numbers represent maximum elapsed time threshold in minutes.
**Not restricted by physical proximity of route and station.
***Limited to journey stages on different routes.

An iterative method was used to link bus and Underground journey stages for each of the 2,366,693 passengers in the sequenced journeys table for Wednesday, November 14, 2007. First, the query loops through each smart card ID and flags the journey stages, ordered sequentially by time, that should be linked to their successor based on the specified modal combination and elapsed time threshold. Sequential bus journey stages that meet the elapsed time threshold are not linked if they occur on the same route because this most likely represents a return journey[12]. The query then iterates a second time to flag and number each complete journey for each passenger. Note that complete journeys may be comprised of a single journey stage. This process takes about an hour to execute for the sequenced journeys table of approximately 7.2 million daily journey stage records[13]. Sample SQL code used to implement this process is included in Appendix 1.

[12] It would also possible to restrict linked Underground and bus journey stages to routes in the vicinity of Underground stations as was done for the analysis in Chapter 3, but this is computationally intensive and should not be necessary if an appropriate interchange time thresholds are applied.
[13] Smart cards with more than 10 journey stages as well as those with journey stages on modes other than bus or Underground were excluded from the analysis.

4.2 Results of Linking Journey Stages to Form Complete Journeys

In this section, the aggregate results of the three reasonable time threshold approaches (Low, Middle, and High in Table 4-1) are compared to expected results based on the LTDS. They are also compared to the unreasonable sets of assumptions for elapsed time thresholds (i.e., Very Low and Very High in Table 4-1). The metrics used for comparison are total journeys, journeys per passenger, stages per journey, and modal patterns. These metrics will help to determine which approach to use going forward in the route- and station-level examples to be presented in Chapter 5.

4.2.1 Total Daily Journeys

The most recent figure from TfL for average weekday journeys on the Underground is 4 million (Mayor of London 2008). Average daily bus journeys (as opposed to journey stages) in 2006/07 stood at 3.2 million[14] (Transport for London 2008). Therefore, we expect total daily weekday journeys in November 2007 on the bus and Underground to be approximately 7.2 million. However, travel by Oyster only represents 73 percent of daily journeys, so we expect the complete journeys datasets created using reasonable elapsed time thresholds to include 73 percent of 7.2 million, or approximately 5.3 million complete journeys. Note that it is necessary to rely on published TfL figures for this estimate because LTDS only includes London residents so the total number of bus and Underground journeys derived from the survey does not represent all the non-residents who enter London and use the public transportation network on any given day.

As shown in Table 4-2, the sets of elapsed time thresholds within the suggested ranges yield a total of 5.3 to 5.5 million daily complete journeys and these values are within 4 percent of the expected value of 5.3 million complete journeys based on published TfL statistics. Conversely, the total number of daily weekday journeys formed by the unreasonable, very low set of elapsed time thresholds is 23 percent higher than the expected value. This suggests that, as expected, the unreasonably low set of thresholds misses a significant share of actual complete journeys. However, the total of 5.1 million daily journeys formed by the unreasonably high set of thresholds is not very different from the expected value, probably

[14] Comparable statistics for average *weekday* bus journeys were not available. Average weekday bus journey stages are 38 percent higher than average weekend day bus journey stages (Transport for London 2008).

because all true complete journeys are captured by the set of elapsed time thresholds at the high end of the reasonable range so by increasing the thresholds still further the methodology simply links a small number of additional journey stages onto the already formed journeys.

Table 4-2 Total Daily Journeys Relative to Expected Value

Elapsed Time Thresholds	**Total Daily Journeys** (million)	**Difference from Expected Value**
Very Low – outside range	6.5	23%
Low – in range	5.5	4%
Middle – in range	5.4	2%
High – in range	5.3	0%
Very High – outside range	5.1	-4%
Expected Value	*5.3*	

In general, the total number of complete journeys formed by the low, middle, and high sets of elapsed time thresholds are close to the expected value of 5.3 million, although at the low end they may overestimate complete journeys slightly (i.e., not linking enough journey stages together). This comparison provides some confidence that the proposed ranges of elapsed time thresholds are indeed appropriate in the London context, but several other metrics are examined to determine whether the low, middle or high end of the ranges is preferred.

4.2.2 Public Transportation Journeys per Passenger

Another way to evaluate the outcome of linking Oyster-based journey stages to form complete journeys is in terms of public transportation journeys per passenger. According to my calculations using LTDS data from 2005/06, residents of Greater London who travel by bus or Underground make an average of 2.05 complete journeys on these modes per weekday. For the almost 2.4 million passengers in the Oysters sequenced journeys table, the elapsed time thresholds within the proposed ranges yield 2.23 to 2.33 daily journeys per passenger. As shown in Table 4-3, these results are slightly higher than the expected value based on LTDS[15]. This difference could be explained by London residents making slightly fewer daily journeys

[15] The LTDS data includes approximately 5,000 households or 12,000 individuals in 2005/06. Results are weighted in accordance with the sampling methodology.

by public transportation per day than the average TfL passenger, or by the difference in methodologies used to identify complete journeys.

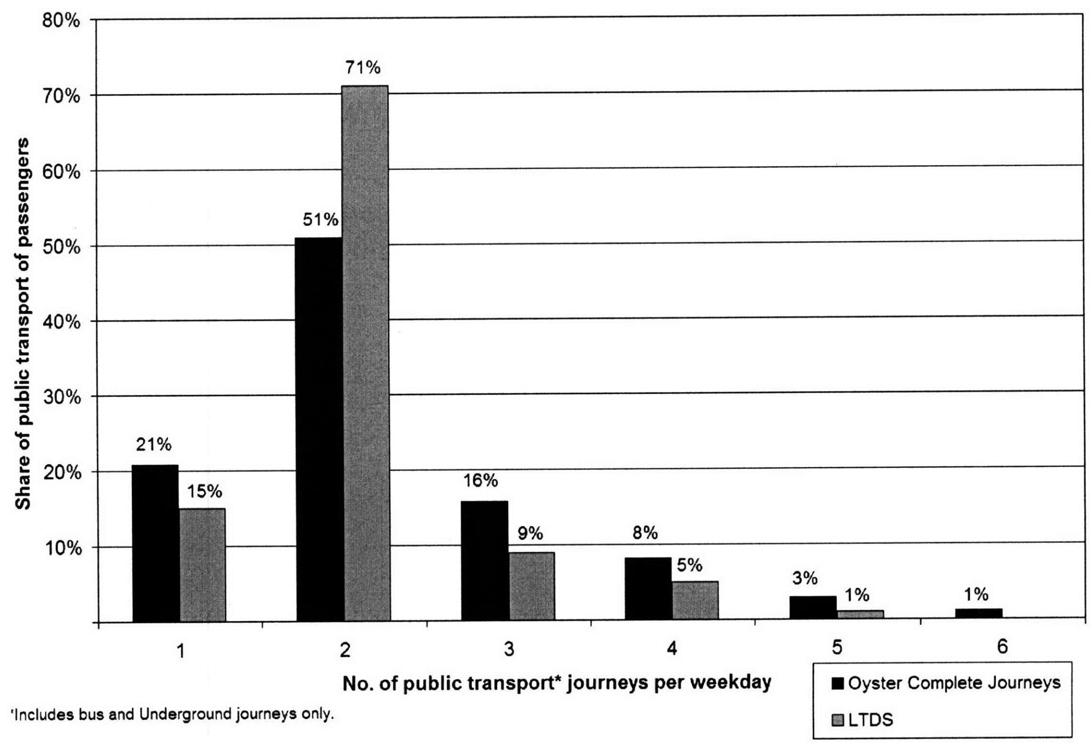
Table 4-3 also shows that, once again, the unreasonably low elapsed time thresholds yield a value that is very different from the expected value – that is far too many daily public transportation journeys per passenger – but that the unreasonably high elapsed time thresholds yield a value that is fairly close to the sets of thresholds within the reasonable range. Similar explanations apply in this case as for the total daily journeys metric discussed above.

Table 4-3 Average Daily Public Transportation Journeys per Passenger

Elapsed Time Thresholds	Daily Public Transport Journeys per Passenger	Difference from Expected Value
Very Low – outside range	2.73	33%
Low – in range	2.33	14%
Middle – in range	2.27	11%
High – in range	2.23	9%
Very High – outside range	2.16	5%
Expected Value from LTDS	*2.05*	

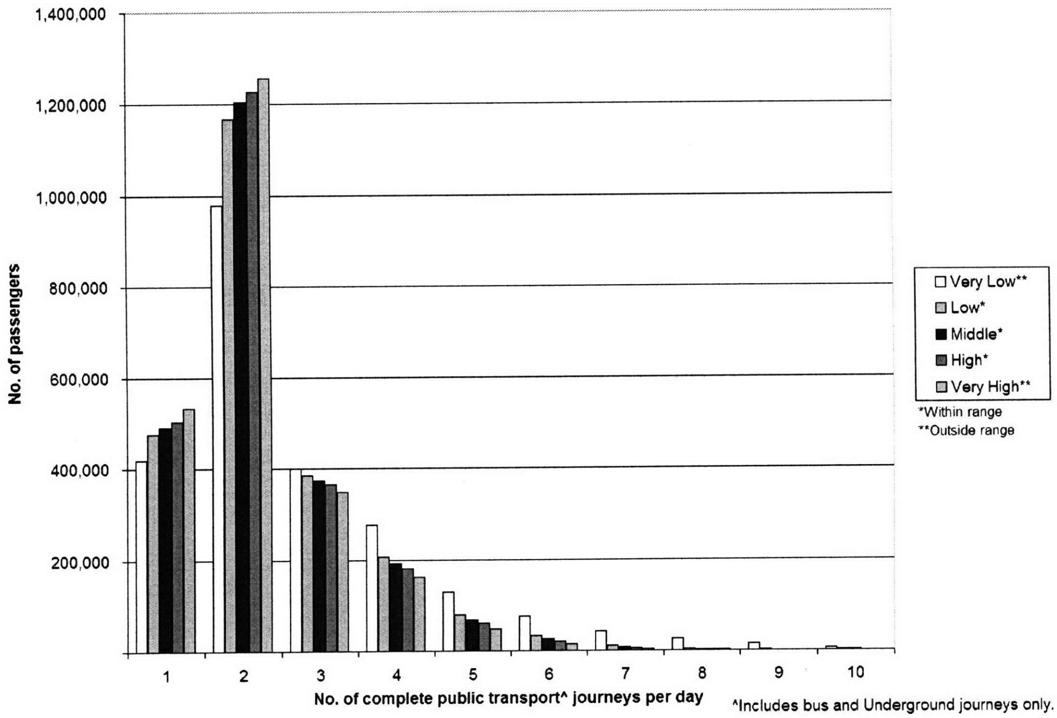
To better understand the differences between the expected value and the Oyster-based value of daily public transportation journeys per passenger, the share of passengers who make one to six journeys per day are examined for both cases. Figure 4-1 shows that for elapsed time thresholds in the middle of the suggested ranges (i.e., Underground-bus = 20 minutes, bus-Underground = 40 minutes, and bus-bus = 50 minutes), 51 percent of passengers make two journeys per day and a further 37 percent make one or three journeys per day. The graph also shows that according to LTDS, 71 percent of passengers make two public transportation journeys per day whereas only 26 percent make one or three journeys per day. The differences between the Oyster-based and LTDS results are not negligible, but they are difficult to address. If the Oyster-based approach resulted only in too many people making one journey per day, then the elapsed time thresholds could be decreased to address the discrepancy with LTDS. Conversely, if the Oyster-based approach resulted only in too many people making three or more journeys per day then the elapsed time thresholds could be increased to form fewer journeys and thereby address the discrepancy with LTDS. However, since both the number of passengers making one journey and the number of passengers making three or more journeys per day are too high, there is no obvious way to change the elapsed time thresholds to make the results more consistent with LTDS.

Figure 4-1 Daily Public Transportation Journeys per Passenger, LTDS vs. Oyster Complete Journeys



The linking of journey stages into complete journeys using Oyster smart card data is based on simple elapsed time assumptions and does not explicitly take into account passenger's perceptions of what constitutes a complete journey as would be the case in LTDS. Future versions of the methodology could use different elapsed time thresholds per station or apply more spatial restrictions on which journey stages become linked. Both methods would help to more accurately identify true complete journeys but some discrepancy with LTDS is still expected due to the differences in methodology and LTDS being a survey of London residents only whereas Oyster smart card data is representative of travel on TfL services across a larger population. For the moment, Figure 4-2 shows that the number of daily journeys per passenger is fairly consistent across sets of time thresholds within the proposed ranges. It also shows that the very low time thresholds outside the proposed ranges are even less consistent with LTDS, and also that the very high time thresholds do little to improve the results.

Figure 4-2 Daily Public Transportation Journeys per Passenger for Time Thresholds within and Outside Suggested Ranges



Overall, there is some cause for concern that the complete journeys methodology does not accurately combine journey stages into two journeys per day for a sufficient number of passengers. On the other hand, Oyster card data constitutes a very large sample of public transportation users and may simply reflect a greater variety of travel behavior than is captured by the small sample from LTDS. Moreover, this metric should be considered in combination with the other metrics presented in this chapter in evaluating the elapsed time threshold approach.

4.2.3 Stages per Public Transportation Journey

A third way to assess the results of forming complete journeys based on elapsed time thresholds is in terms of stages per journey. According to LTDS data for 2005/06, the average number of stages per public transportation journey is 1.25. This is similar to empirical results from other smart card data analyses that find a typical implied ratio of bus boardings to complete journeys on bus-only networks of 1.21 to 1.25 (Bagchi and White 2004, Hoffman and O'Mahony 2005). Table 4-4 shows that the average number of stages per journey for complete journeys formed from the Oyster sequenced journeys table by the reasonable sets of elapsed time thresholds is 1.30 to 1.36. These values are slightly higher than the expected value from LTDS and previous smart card analyses in other cities, which suggests that overall the elapsed time thresholds may be a bit high and therefore the best option is the lowest set of thresholds (i.e., Underground-bus = 15 minutes, bus-Underground = 30 minutes, and bus-bus = 40 minutes).

Table 4-4 Average Stages per Complete Journey by Elapsed Time Thresholds

Elapsed Time Thresholds	Average Stages per Journey	Difference from Expected Value
Very Low – outside range	1.11	-11%
Low – in range	1.30	4%
Middle – in range	1.34	7%
High – in range	1.36	9%
Very High – outside range	1.40	12%
Expected Value from LTDS	*1.25*	

Additionally, Figure 4-3 shows that the distribution of one-, two-, and three-stage journeys is almost identical between LTDS and the complete journeys formed using the low ends of the elapsed time threshold ranges. Therefore, the data presented in this graph further support using the low end of the elapsed time thresholds for creating complete journeys.

4.2.4 Modal Patterns

Finally, the share of Underground journeys that include at least one bus journey stage as a measure of modal patterns formed by Oyster-based complete journeys versus LTDS is examined. According to LTDS data for 2005/06, 23 percent of Underground journeys include at least one bus journey stage. As shown in Table 4-5, this is the same share of Underground journeys that include bus stages in the complete journeys formed by the low end of the elapsed time thresholds. It is also very close to the 25 percent of Underground journeys that include bus stages for complete journeys formed using the middle and high end of the elapsed time threshold ranges. Therefore, we can conclude that based on this metric it is preferable to use the low end of the proposed ranges but that the results from the middle, or even the high end of the ranges would not be drastically different from expected values.

Figure 4-3 Stages per Complete Journey for Sets of Time Thresholds vs. LTDS

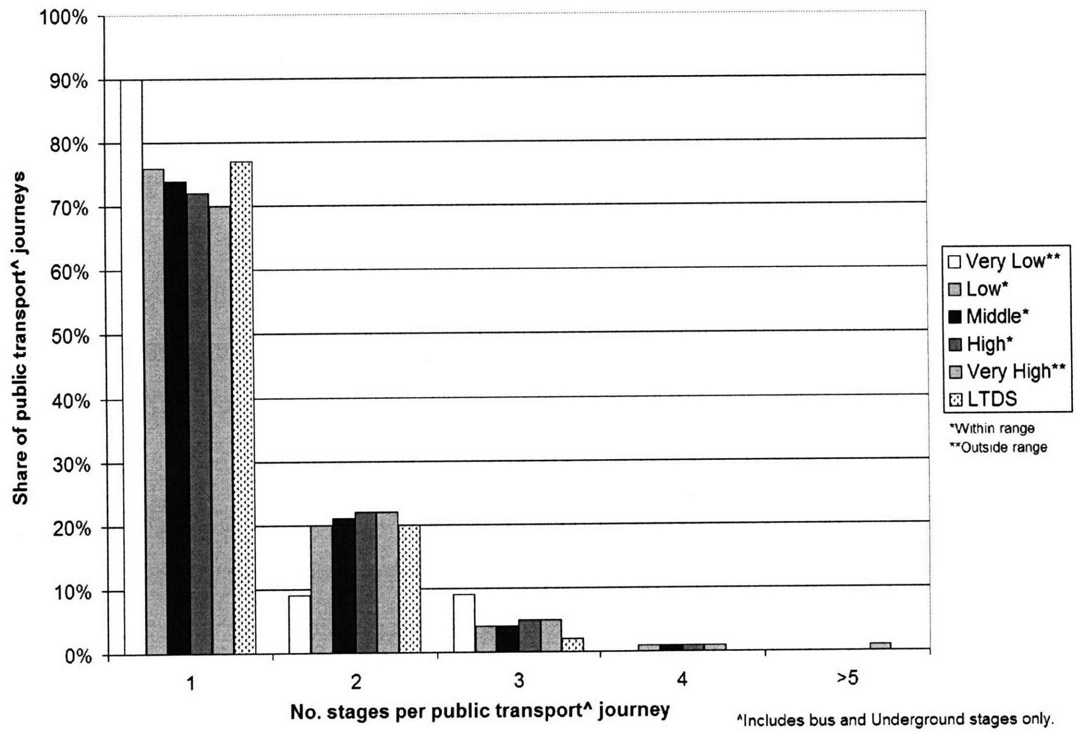


Table 4-5 Share of Underground Journeys with Bus Stages by Elapsed Time Thresholds

Elapsed Time Thresholds	Share of Underground Journeys with Bus Stages
Very Low – outside range	14%
Low – in range	23%
Middle – in range	25%
High – in range	25%
Very High – outside range	26%
Expected Value	*23%*

4.3 Summary

Table 4-6 summarizes the metrics examined in this chapter for the low and middle values of the elapsed time thresholds ranges. Both are compared to the expected values from LTDS, although it should be noted that there are two reasons why there is likely to be some discrepancy between the LTDS metrics and those from Oyster smart card data. First, LTDS is a survey of London residents only whereas Oyster data captures anyone who travels on the TfL network, and, second, the LTDS travel diary methodology for a small, statistically-constructed sample reveals how London residents perceive and recall their travel by public transportation as opposed to the network-wide representation of actual travel by bus and Underground obtained from the Oyster data.

Returning to Table 4-6, it shows that the stages per journey and number of Underground journeys with bus stages metrics indicate that elapsed time thresholds at the low end of the ranges yield results closest to the expected values, but the daily public transportation journeys per passenger metric indicates that the middle of the ranges (or even higher) yields results closer to the expected value. Thus, two metrics suggest that the best set of elapsed time thresholds are at the low end of the range, but the third suggests that the elapsed time thresholds are too low and should be set higher. Nonetheless, in all cases, the results within the ranges are fairly similar. Since there is no clear winner and the middle of range is very close to the expected values (which are approximate themselves) I propose to use a complete journeys dataset generated from those values for application to bus network planning case studies in the next chapter. However, I believe that in reality either the low end or middle of the range would provide similar results and so could be used for implementation at London Buses.

Table 4-6 Summary of Metrics from Oyster-based Complete Journeys and LTDS

Metric	Oyster-based Complete Journeys		LTDS
	Low thresholds*	Mid-range thresholds**	
Daily public transportation journeys per passenger	2.33	2.27	2.05
Stages per journey	1.30	1.34	1.25
Underground journeys with bus stage	23%	25%	23%

*Elapsed time thresholds: Underground-bus=15 min., bus-Underground=30 min., bus-bus=40 min.
**Elapsed time thresholds: Underground-bus=20 min., bus-Underground=40 min., bus-bus=50 min.

CHAPTER 5: APPLICATION OF COMPLETE JOURNEY INFORMATION TO BUS NETWORK PLANNING

The purpose of this chapter is to use several case studies to illustrate how complete journey data derived from applying elapsed time thresholds between journey stages for individual passengers is useful for bus network planning in London. The case studies of three bus routes and two bus-Underground interchange stations are intended to be illustrative rather than comprehensive, providing examples of the range of new information that could be available to bus network planners should an information system using smart card data be developed and implemented for day-to-day use.

The chapter begins with a discussion of the types of new information we expect to gain from the complete journey data (Section 5.1). Next, the five case studies are presented with the key insights from each (Section 5.2). The final section summarizes the findings from the case studies and their implications for bus network planning (Section 5.3).

5.1 Expected Findings

New information that may be gained from smart card data needs to be considered relative to the existing data systems used by Network Development at London Buses that were discussed in Chapter 2. Smart card data is beneficial if it provides either new information not previously available to planners on a day-to-day basis, or information in a more timely and efficient manner than that which is currently available. This chapter is focused on information not available from current data systems that may be especially relevant to bus network planning. This new information comes from forming complete journeys using smart card data and can be divided into two categories: (1) contextual knowledge about a route or station that may be quantitative or qualitative, and (2) quantitative inputs to the cost-benefit models used to evaluate frequency, capacity, or restructuring changes to bus routes.

Contextual knowledge about a route or group of routes in the vicinity of an Underground station that may be gained from examining all complete journeys that pass through a selected route or station can be roughly categorized into information about:

- connectivity of the route or station with the public transportation network,
- intermodalism of journeys on the route or through the station,

- bus access journey duration, and
- Underground-to-bus interchange time.

Examples of contextual information about the **connectivity** of a selected route or station are:

- the ratio of single stage journeys to multi-stage journeys,
- the number of passengers who transfer to or from any other route or station in the network,
- the proportion of passengers that make probable return journeys (i.e., two journeys) per day, and
- individual public transportation travel patterns.

Examples of contextual information about the **intermodalism** of a selected route or station are:

- the ratio of multi-stage journeys that include Underground stages to those that include bus only,
- the most common multi-stage journey patterns at either the modal or individual route/station level of aggregation,
- the ratio of journeys through a station that begin with a bus-Underground modal sequence to those that begin with an Underground journey stage in order to determine the relative importance of bus as an access mode to that station, and
- the ratio of journey stages ending at a station that are followed by a bus stage to those that terminate at the station in order to determine the relative importance of bus as an egress mode at that station.

Examples of contextual information about **bus access journey stage duration** for a selected route or station are:

- to deduce how far passengers are traveling by bus to access another route or an Underground station by comparing elapsed bus travel plus interchange time with scheduled frequency and run time for that route, and
- the average duration of bus journey stages to access a given station across different routes.

Finally, an example of contextual information about **interchange time** for a selected route or station is:

- the distribution of interchange times from an Underground station to selected routes that could be used to calculate wait time savings or increases due to changes in bus service frequency.

Quantitative inputs to the **cost-benefit models** used by network planners to evaluate proposed changes in capacity or frequency of a route may also be augmented by information from the complete journeys data. With this data, it is possible to count the number of people interchanging from one route to another over the course of a day or during any time period of interest. If frequency is increased on a low-frequency route, there may be greater benefit for interchanging passengers than those who access the route by walking because they have less control over their arrival time due to it being constrained by the schedule of the first service. Therefore, their wait time savings may be weighted more heavily in the cost-benefit model than for non-interchanging passengers. (Remember that passenger benefits are measured in terms of time converted to pounds based a given value of time.) Conversely, a decrease in frequency on the second route could have a greater disbenefit for interchanging passengers and should be assessed accordingly.

Similarly, in an evaluation model for a route restructuring, if two routes currently have a unique intersection then the complete journeys data can be used to count how many passengers would no longer need to interchange if a new direct link were added. Additional passenger benefits in terms of wait time savings and no interchange penalty could then be included in the model.

In the following section, five case studies are used to illustrate the new information available from complete journey data as compiled from smart card data using the elapsed time threshold approach described in Chapter 4.

5.2 Case Studies of Bus Passenger Interchange

The case studies selected for this thesis are intended to be illustrative rather than to provide a comprehensive assessment of bus passenger interchange in London. This section begins with a brief description of the three bus routes and two Underground stations that are used to draw examples of new information available from complete journey data as discussed above.

Routes 293, 221, and 69 are located in different areas of Greater London and each has a different level of potential interchange patterns from simple (Route 293) to complex (Route 69). They have different passenger volumes, but each route has daily volumes as recorded by Oyster that represent a large share of those reported in BREMS that includes both Oyster and non-Oyster passengers[16] (see Table 5-1). Moreover, the farestage route codes currently used to identify the route number in Oyster transaction data are accurate for all three routes[17]. For simplicity, interchange and passenger volumes are reported based on Oyster transactions rather than adjusting for expected values from BREMS and other sources.

Table 5-1 Daily Passenger Volumes on Case Study Routes

Route No.	**Oyster Daily Passenger Volumes*** (1)	**BREMS Daily Passenger Volumes**** (2)	**(1) / (2)**
293	1,989	2,915	68%
221	13,837	16,568	84%
69	21,830	28,763	76%

*Wednesday, November 14, 2007
**Weekday average for November 11-16, 2007

Route 293 serves only Morden Underground Station, which could be considered a typical suburban station. It is a radial route located in Southeast London and links with only one other station – Epsom National Rail Station – and it has several intersecting bus routes. By contrast, Route 221 is a circumferential suburban route with numerous intersecting routes but few running on the same corridor. It links two branches of the Northern Line and the Piccadilly Line in Northwest London. Finally, Route 69 was suggested as a potentially interesting case by network planners because it serves the East London regeneration area. It is a long-established route and serves a relatively high-density urban area, which means that it has numerous intersecting routes, routes running on shared corridors and serves five large Underground stations. A map of each route is provided in the relevant section below.

One of the key advantages of using smart card data for bus network planning is the ability to include information about interchange between Underground and bus. So, Vauxhall

[16] Oyster-based journeys represent 73 percent of all journeys on the TfL network (Mayor of London 2007).
[17] As discussed in Chapter 2, approximately 20 percent of current bus routes are either missing or partially incorrectly identified using the farestage route code that appears in the sequenced journey table used for this research. This issue has been addressed for future versions of the sequenced journey table.

and North Greenwich Stations were selected to provide examples of Underground station information that could be of interest to bus network planners responsible for routes serving these stations.

Vauxhall Station was suggested by network planners who mentioned that the purpose-built bus station there is busy, but it is not clear why because ultimate origin-destination information for passengers using routes in the vicinity of the station is not available. Overall bus-passenger flows to and from Vauxhall Underground Station are examined, but National Rail services are also provided at the station which makes it difficult to gain a full picture of what is happening there from Oyster card data alone. On the other hand, North Greenwich Station is a bus and Underground interchange station on the Jubilee Line. Most routes that serve the station also terminate there, so both station access time by bus and Underground to bus interchange time can easily be compared across routes. Also, North Greenwich Station appears in many of the 3-stage journeys on Route 69, which makes it interesting to explore further. As illustrated in Figure 5-1, both bus stations have attractive, modern designs and are laid out with designated bus bays on an exclusive roadway loop. A map of the routes serving each station is provided in the relevant section below.

Figure 5-1 Images of Vauxhall and North Greenwich Stations

Vauxhall Station
Source: unknown

North Greenwich Station
Source: http://www.stacey.peak-media.co.uk (with permission)

The details of each case are loosely organized according to the categories described in the previous section (i.e., connectivity, intermodalism, etc.), but the information is not exhaustive for each case. Instead, some interesting observations that could be explored further by a planner doing a full review of the route or bus interchange station in question are highlighted.

For each case, all journeys that include the selected route or station from the complete journeys data for the entire network are extracted. This makes the subsequent queries and summaries of journeys involving each case study route or station more efficient. The complete journeys data used for the cases is based on elapsed time thresholds of 20 minutes for Underground-to-bus, 40 minutes for bus-to-Underground, and 50 minutes for bus-to-bus. In other words, the median interchange time assumptions from Chapter 4 were selected because they fall in the middle of the recommended range and neither the low- or high-end of the ranges were clearly preferable. The results of the cases illustrate that these times are adequate for a network-wide approach to creating complete journeys, but final recommendations about appropriate elapsed time thresholds will be presented in Chapter 6.

5.2.1 Comparison of Bus Route Case Studies

Before presenting detailed examples for each route, aggregate modal and connectivity metrics for the three routes are compared. (A similar comparison of the two case study Underground stations is presented later.) Table 5-2 shows that the most suburban route, Route 293, has the highest share of single-stage bus journeys whereas the most urban route, Route 69, has the lowest. This is probably explained by the higher number of interchange opportunities, both with the Underground and other bus routes, for Route 69 as compared to Route 293. However, it is interesting that while Routes 293 and 221 have similar shares of single-stage bus journeys, the relative shares of multi-stage journeys are quite different: Route 221 journeys tend to involve more bus-bus interchanges whereas Route 293 journeys tend to involve more interchanges between bus and Underground. This is interesting because Route 293 connects only with one Underground station, Morden, but perhaps not surprising given the large number of bus routes that intersect with Route 221.

Table 5-2 Modal Combinations of Journeys on Case Study Routes

Route No.	Single-Stage Bus Journeys	Multi-Stage Bus Journeys	Journeys with Underground Stage(s)
293	44%	34%	21%
221	41%	43%	15%
69	33%	30%	35%

Total % per route may not sum to 100 due to rounding.

Route 221 serves large areas with no Underground service so is more likely to be used for home-work trips than Route 69 which mostly runs parallel with Underground lines and could therefore be viewed as a local connector. Possible explanations for the seemingly low number of passengers with only one daily journey on each route include: a short bus journey with return journey by foot, a return journey on a different bus route that serves the same corridor, and a single stage journey that that forms part of a complex trip chain (e.g., go to a fitness centre after work on Route 69 then take another bus route or the Underground to get home). Once again, these results illustrate the complexity of travel behavior in London.

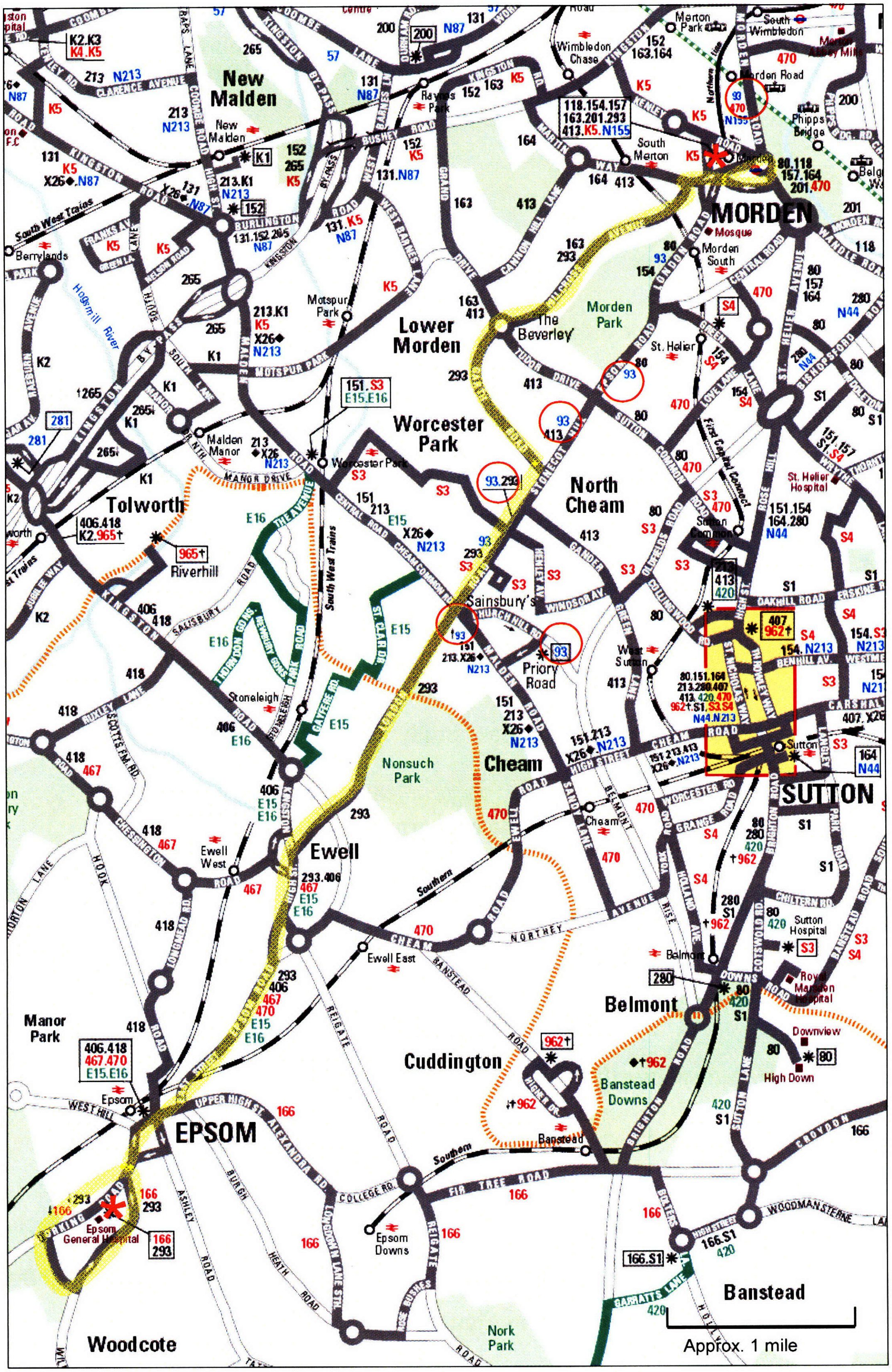
Table 5-3 Probable Return Journeys on Case Study Routes

Route No.	Passengers with 2+ Daily Journeys on Route	Share of All Journeys on Route
293	34%	52%
221	40%	60%
69	28%	47%

5.2.2 Route 293: Suburban Underground Station Access

Route 293 has a fixed headway of 20 minutes between 8 a.m. and 7 p.m. and an off-peak runtime of 41 minutes between Morden Underground Station and Epsom General Hospital. It runs every 15 minutes from 7 to 8 a.m., and approximately every half hour from 7 p.m. until midnight. A map of Route 293 that shows its relations with other routes and stations discussed in the case study is provided in Figure 5-2.

Figure 5-2 Map of Route 293 in South West London

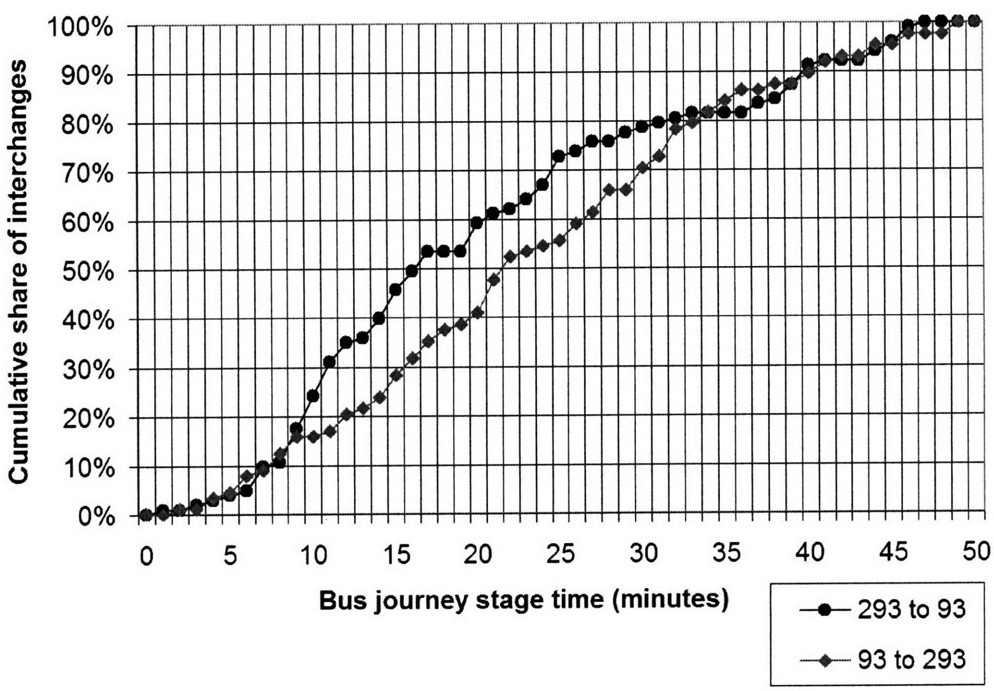


Source: Transport for London, South West London Bus Map, www.tfl.gov.uk

Route 93, which also serves Morden Station and continues northward, has the highest volume of bus-to-bus interchanges with Route 293. As shown in Figure 5-3, the elapsed time trend between bus boardings for the 103 daily interchanges[18] from Route 293 to Route 93 is different from the 88 daily interchanges of the reverse pattern, that is Route 93 to 293. This is probably because Route 93 is a higher-frequency route that runs every 5-7 minutes on weekdays between 7 a.m. and 7 p.m. so passengers can generally board a bus shortly after alighting from Route 293.

On the other hand, passengers interchanging from Route 93 to the lower-frequency Route 293 have a wider distribution of wait times that result in the nearly linear trend line shown in Figure 5-3. Another possible explanation for the difference in trend lines is more variation in the duration of journey stages on Route 93, but this would have to be investigated through other means. This type of comparison is useful for deducing the distribution of typical journey durations on Route 293 (approximately 6 to 30 minutes including wait time for the Route 93 bus) for passengers who continue their journey on Route 93.

Figure 5-3 Elapsed Time Distributions for Interchanges between Routes 293 and 93



[18] Daily interchanges are only counted for two- and three-stage journeys.

Recall from Table 5-2 that 34 percent of journeys on Route 293 involve interchange with another bus route, as opposed to 21 percent that involve interchange with the Underground. The other routes that appear most often in the subset of journeys that include stages on Route 293 are consistent with routes available for interchange. As shown in Table 5-4, the majority of bus-to-bus interchanges involving Route 293 take place at Morden Underground Station, but numerous street-corner interchanges with Routes 213 and 151, which intersect Route 293 near the middle of the route, also occur. This summary, which represents approximately 700 daily bus-to-bus interchanges, illustrates that Morden Underground Station is also an important bus-to-bus interchange location and might support further improvements to bus interchange facilities there.

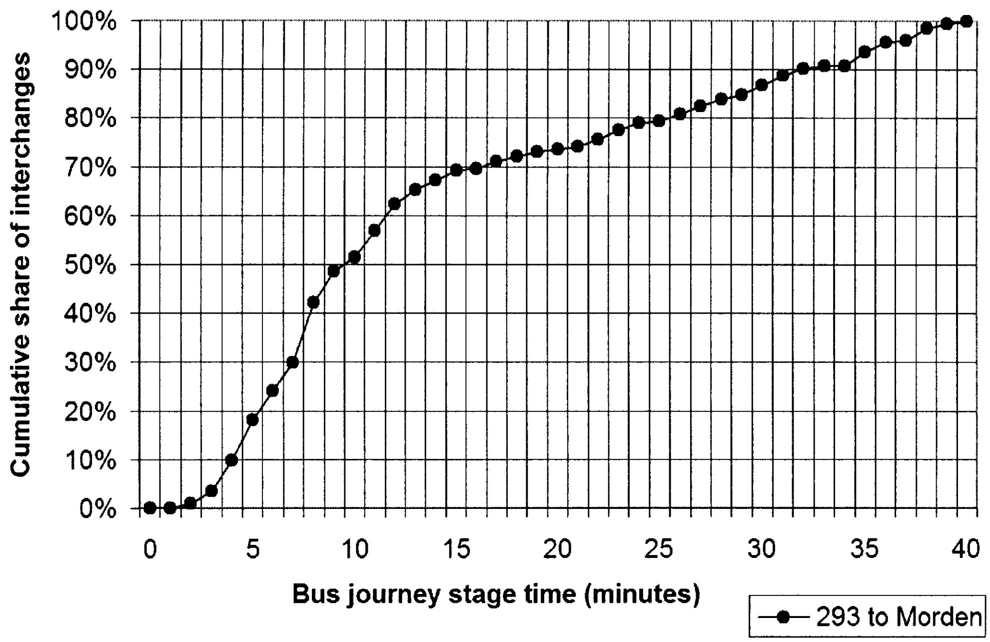
Table 5-4 Top 10 Routes in Journeys on Route 293

Route No.	Share of Top 10 Routes in Journeys on Route 293 *	Station Intersection with Route 293	Road Intersection with Route 293
93	29%	Morden	Stonecot Hill**
213	16%	-	Malden Road/London Road
118	11%	Morden	-
163	10%	Morden	Hillcross Avenue**
151	8%	-	Malden Road/London Road
80	5%	Morden	-
164	5%	Morden	-
201	5%	Morden	-
406	5%	Epsom Rail	Epsom Road**
157	5%	Morden	-

*Excluding Route 293.
**Indicates shared corridor with Route 293.

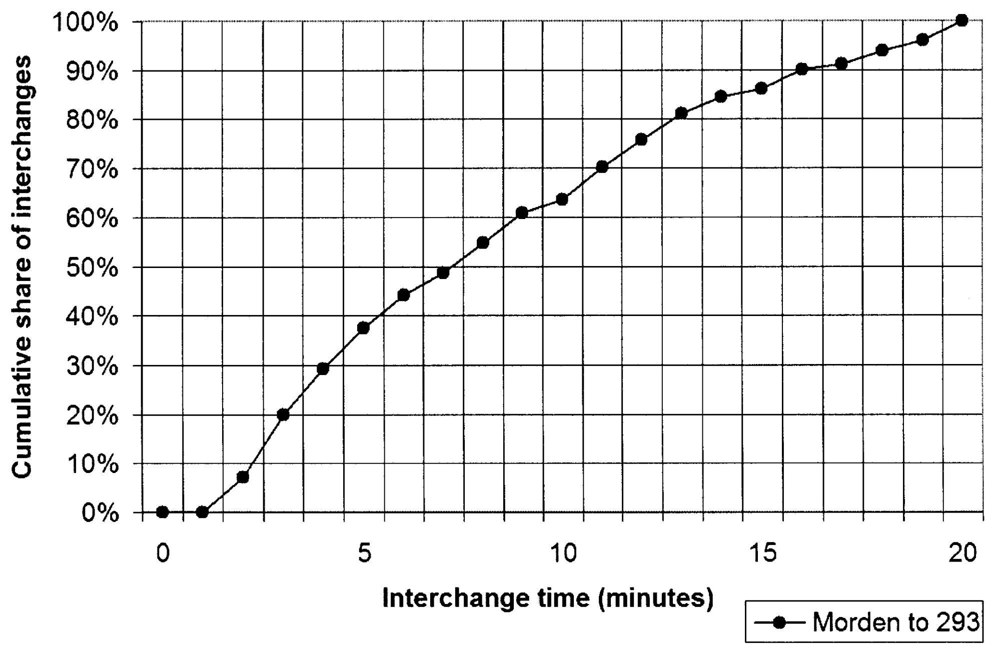
In terms of bus access journeys, Figure 5-4 indicates that of the 204 daily journeys on Route 293 to access Morden Station, 70 percent take no more than 15 minutes, however, many of the rest take at least 22 minutes. This is interesting because Route 293 runs on the same road as the high-frequency Route 93 in the section of its route from which it takes 15 to 20 minutes to reach Morden Station. Moreover, as illustrated in Figure 5-2, Route 93, which has a daytime headway of 5-7 minutes, follows a more direct route that only takes 9 to 10 minutes to reach Morden Station from the shared section with Route 293. This suggests that passengers boarding a bus to access Morden Station on the shared section will tend to board Route 93, but that some people do travel longer than 22 minutes on Route 293 to access the station. This analysis provides insight into typical travel behavior on Route 293, but because of the small sample size, it would have to be confirmed by comparing the results across multiple days.

Figure 5-4 Elapsed Time Distribution for Interchanges from Route 293 to Morden Station



Finally, Figure 5-5 shows that the elapsed times for the 184 daily interchanges from Morden Underground Station to Route 293 are fairly evenly distributed from 2 to 20 minutes, which is reasonable given the route's 20 minute headway and assuming random arrivals by Underground. From this graph, we can conclude that for Morden Station the network-wide elapsed time assumption of 20 minutes for Underground-to-bus interchanges is appropriate. In addition, the graph suggests that in general passengers are not timing their arrivals at the station to coincide with Route 293 bus departure times, which could be helpful in assessing potential schedule changes.

Figure 5-5 Elapsed Time Distribution for Interchanges from Morden Station to Route 293

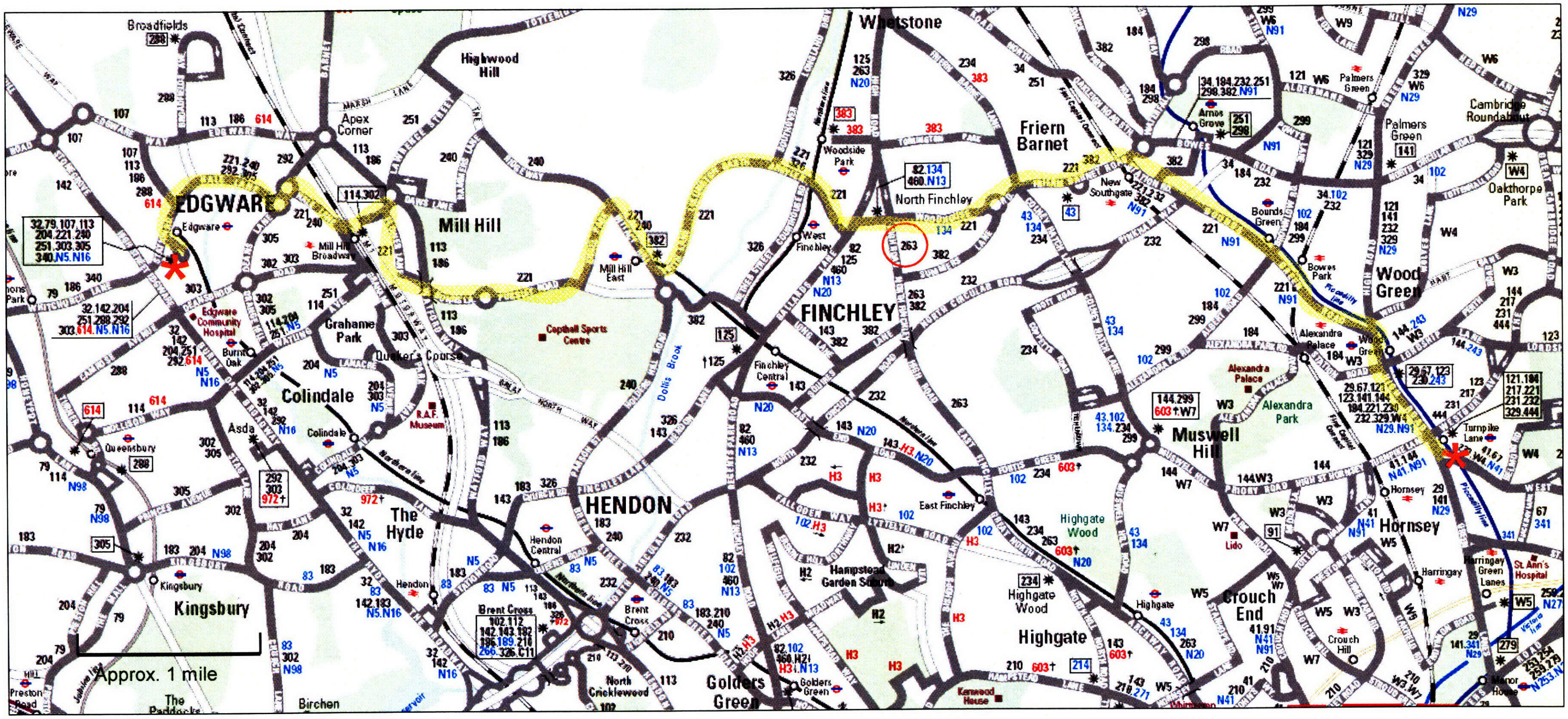


5.2.3 Route 221: Suburban Connector

Route 221 runs between Turnpike Lane and North Finchley every 5-8 minutes until 3 p.m., then every 2-6 minutes from 3 to 4 p.m., and then every 6-12 minutes until midnight. However, the route extends westward from North Finchley to Edgware with a daytime frequency of 9-13 minutes. The off-peak runtime for the entire route from Turnpike Lane to Edgware is 59 minutes (see Figure 5-6).

Recall from Table 5-2 that 43 percent of journeys on Route 221 involve interchange with other bus routes as opposed to only 15 percent that involve interchange with the Underground. Similar to the previous case, the other bus routes that appear most often in journeys on Route 221 are consistent with routes available for interchange. In fact, the top three routes, as presented in Table 5-5, are routes that intersect Route 221 in the middle of its route (near North Finchley) where no Underground interchange is available. The top eight routes, comprising a total of 2,782 journey stages, are simply intersecting routes meaning they do not run on the same corridor as Route 221. As such, we can be very confident that journey stages on these routes represent true interchanges with Route 221.

Figure 5-6 Map of Route 221 in North West London



Source: Transport for London, North West London Bus Map, www.tfl.gov.uk

However, for routes that run on the same corridor as Route 221, for example Routes 240 and 144, passengers may have made return journeys on these routes that have been erroneously linked into one complete journey using the elapsed time threshold explained in Chapter 3[19]. This limitation could be addressed in the future by lowering the elapsed time threshold between bus boardings for routes that use the same corridor in accordance with the length of overlap and/or by only linking journey stages that are in the same direction into complete journeys. That being said, most of the complete journeys on Route 221 that include journey stages on Routes 240 or 144 are likely to include pure interchanges because the shared corridor with each of these routes is very short. In a comprehensive network planning application, this assertion could be substantiated by more detailed analysis of travel patterns for Route 221 journeys that include stages on Routes 240 and 144. In the longer term, the new GPS-based AVL system that is being implemented at TfL could be used to more accurately determine the boarding location of passengers, which would further inform the identification of return versus onward journeys.

Table 5-5 Top 10 Routes in Journeys on Route 221

Route No.	Share of Top 10 Routes in Journeys on Route 221*	Station Intersection with Route 221	Road Intersection with Route 221
134	12%	-	Colney Hatch Lane
263	12%	-	High Road
125	12%	-	Ballards Lane
W3	11%	Wood Green	-
382	10%	Mill Hill East	-
82	10%	-	Ballards Lane
102	9%	Bounds Green	-
243	9%	Wood Green	-
240	8%	Edgeware	Hale Lane**
144	7%	Wood Green; Turnpike Lane	High Road**

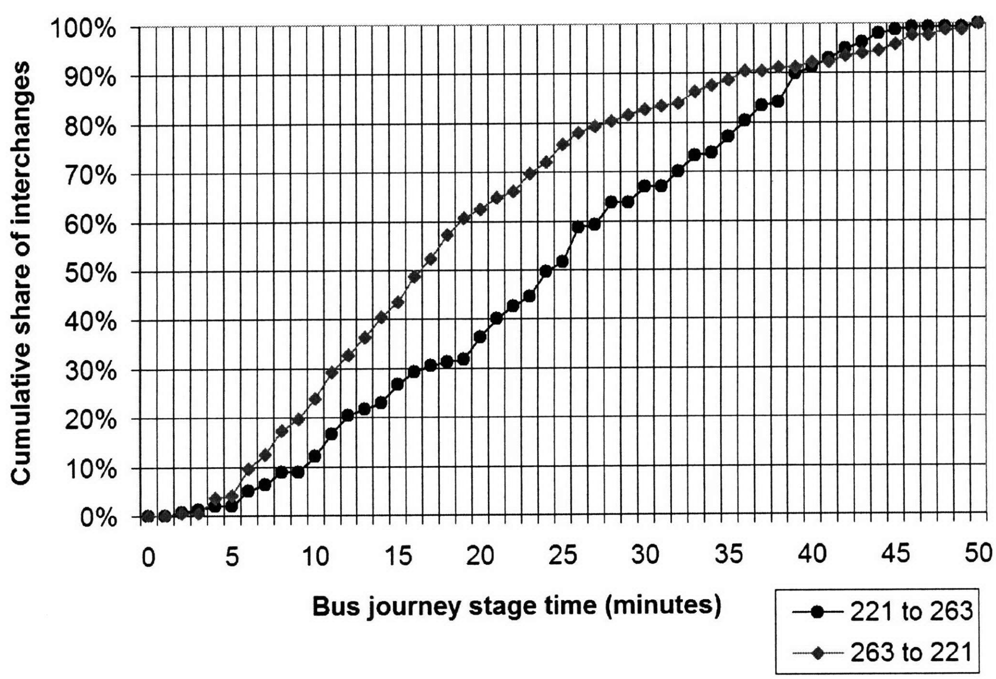
*Excluding Route 221.
**Indicates shared corridor with Route 221.

Route 221 has a higher frequency than most of the routes that intersect it, for example Routes 263, which has a daytime headway of about 10 minutes as opposed to 2-8 minutes for

[19] As discussed in Chapter 3, two sequential journey stages on the *same* route cannot be included in a complete journey.

Route 221. As shown in Figure 5-7, the observed elapsed time trends are similar to those of interchanges between low-frequency Route 293 and high-frequency Route 93 discussed in the preceding case. Similar to that case, Figure 5-7 suggests that passengers are able to board a Route 221 bus shortly after alighting from Route 263, but that in the reverse scenario there is a wider range of wait times for the second bus. This example is repeated to demonstrate that the first case study is not unique, and because the trends are more obvious here. This starkness may be the result of more data points, specifically 160 interchanges in each direction as opposed to 100 in each direction for the previous case. Once again, an examination of bus journey access times is useful for deducing the distribution of typical journey durations on Route 263 (approximately 6 to 35 minutes) for passengers who continue their journey on Route 221. When compared to journey durations from BODS, these results may show that interchanging passengers travel less time on the route than non-interchanging passengers which would help in evaluating the differential impacts of potential service changes.

Figure 5-7 Elapsed Time Distributions for Interchanges between Routes 221 and 263



Turning now to the connectivity between Route 221 and the Underground, recall from Table 5-2 that 15 percent of journeys on Route 221 include journey stages on the Underground. This figure is low compared to the other two case study routes, but, as presented in Table 5-6, it can be further disaggregated to show the modal patterns of two- and three-stage journeys on Route 221 that include the Underground. The similar shares of bus-Underground and Underground-bus journeys lends credibility to the elapsed time thresholds of 40 minutes and 20 minutes respectively that were used to identify these journeys. Moreover, it can be concluded that passengers rarely take two buses to access any of the Underground stations that are served by Route 221.

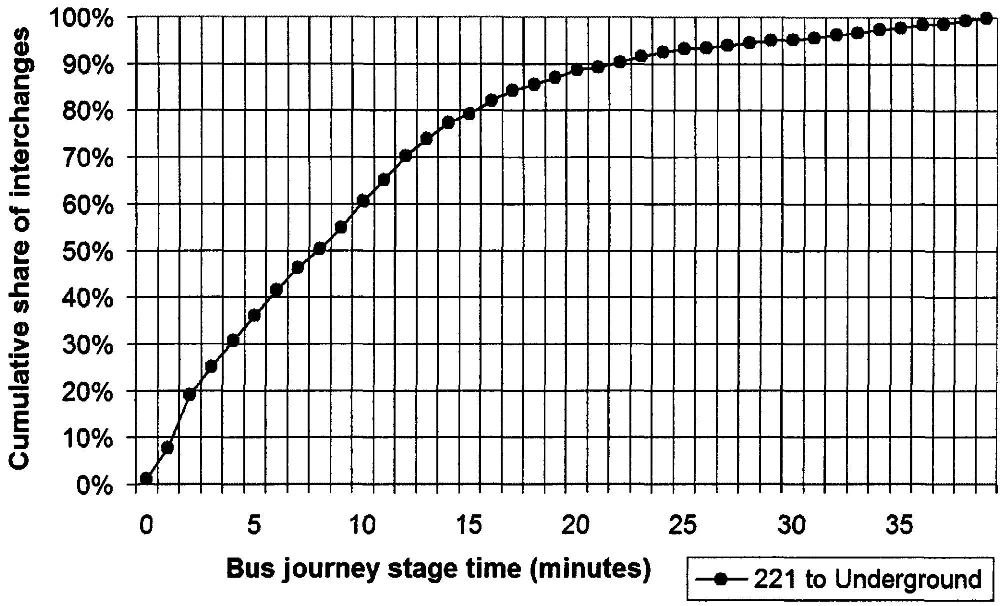
Table 5-6 Modal Patterns of Journeys on Route 221 with Underground Interchange

Mode 1	Mode 2	Mode 3	Share of All Journeys on Route 221
U	B	-	6%
B	U	-	5%
B	U	B	3%
U	B	B	<1%
B	B	U	<1%

U = Underground; B = Bus

In terms of bus journey access time, the elapsed time trend illustrated in Figure 5-8 shows that most journeys on Route 221 to access the Underground are no more than 15 to 20 minutes long. This corresponds to an off-peak run time for Route 221 in which all points are at most 15 minutes from the nearest Underground station. Assuming a short interchange time (<5 minutes) from bus alighting to Underground entry, this shows that for this route at least bus passengers transfer to the Underground at the closest possible station. Although this makes intuitive sense, existing data systems at TfL do not provide readily available information to substantiate this point. This result is also interesting in comparison to bus-to-bus interchanges between Routes 263 and 221 in which passengers appeared to travel up to 35 minutes, albeit including wait time, on Route 263 to access Route 221.

Figure 5-8 Elapsed Time Distribution for Interchanges from Route 221 to All Underground Stations



By disaggregating bus and Underground interchanges by station, it becomes obvious that Bounds Green has the highest interchange volume with Route 221 and, at 1,057 daily interchanges, the volume is four times larger than the next station, Wood Green (see Table 5-7). Turnpike Lane and Wood Green have low interchange volumes relative to Bounds Green, which is the first interchange point between Route 221 and the Piccadilly Line. Surprisingly, Manor House Station, which is not on Route 221 but rather is the next station on the Piccadilly Line after Turnpike Lane, where Route 221 terminates, appears more frequently in journeys involving Route 221 than Edgware Station, which is the other route terminus. This is because Bounds Green to Manor House is the most common Underground origin-destination pair for bus-to-Underground journeys on Route 221.

Table 5-7 Daily Interchange Volumes between Route 221 and Top 3 Underground Stations

Underground Station	Daily Interchanges with Route 221	
Bounds Green	1,057	74%
Wood Green	255	18%
Turnpike Lane	110	8%

This information and an examination of the route map in Figure 5-6 raises the question of whether it is beneficial for Route 221 to run between Wood Green and Turnpike Lane.

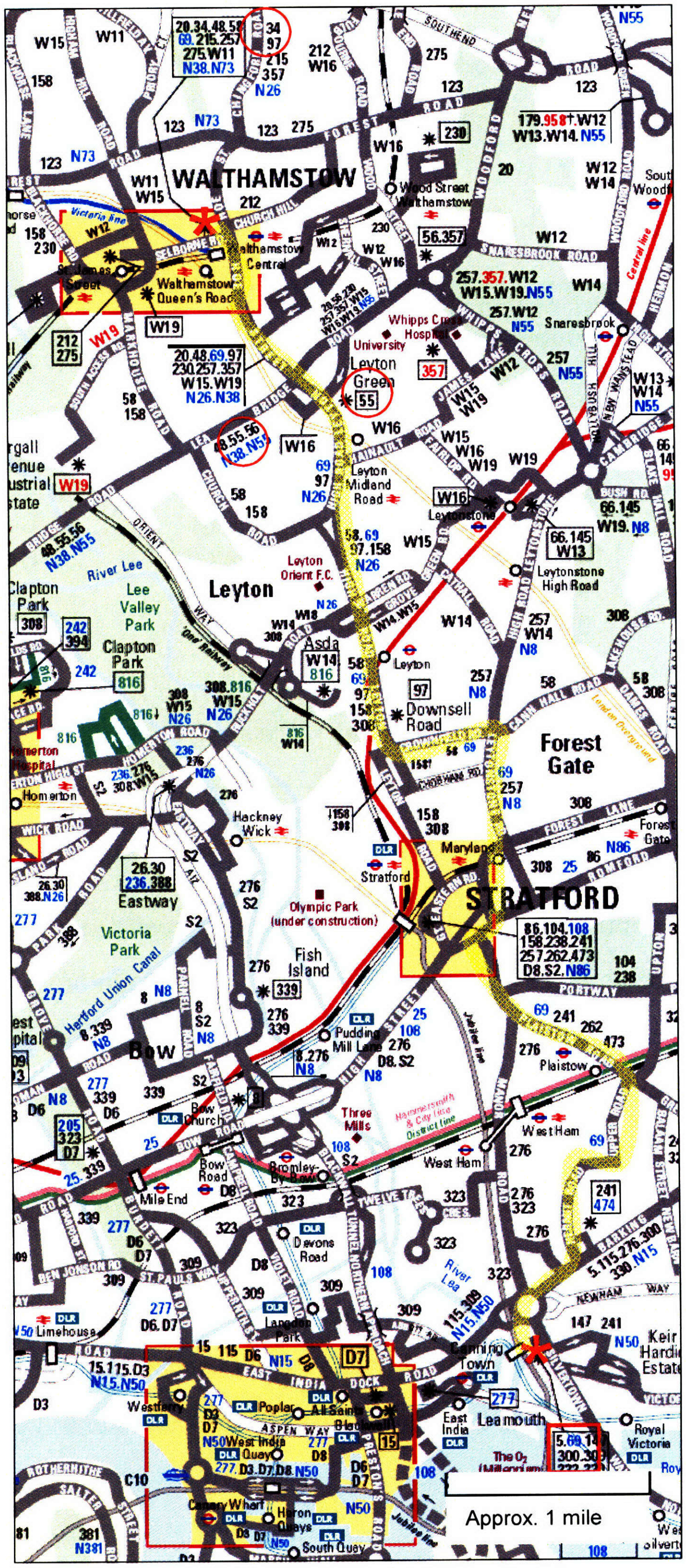
Route 221 replicates the Underground line after Bounds Green and there are numerous other routes on the corridor between Wood Green and Turnpike Lane. According to published schedules, buses at 3:30 p.m. are already short-turned at Wood Green. There are a good number of interchanges with other bus routes that serve Turnpike Lane, for example Route 144 is one of the top routes in terms of interchange volumes with Route 221 (see Table 5-5) but it, and most of the routes that intersect Route 221 at Turnpike Lane, also serves Wood Green. Therefore, the interchange could occur at either station or in the intervening shared corridor. In short, it appears that passengers take Route 221 to Bounds Green or transfer from Route 221 to another bus that is available at Wood Green Station, so perhaps Route 221 should terminate at Wood Green. There may be other factors, such as space for layovers, that may make terminating Route 221 at the purpose-built bus interchange at Turnpike Lane the preferred option, but this analysis nevertheless demonstrates the value of gathering information on current travel patterns to inform a potential rerouting.

5.2.4 Route 69: Complex Urban Link

Route 69 has a daytime headway of 6-10 minutes from 7 a.m. to 9 p.m. and an off-peak runtime of 47 minutes. It serves five Underground stations and other buses run on the same corridor for most of the route as shown in Figure 5-9.

Recall from Table 5-2 that only 28 percent of passengers make two or more journeys on Route 69 per day, representing 47 percent of all journeys on the route. This low share may be because Route 69 is used more heavily for short, one-way journeys than for commuting trips, but it may also be partially explained by the multitude of journey path options in the area so that passengers are likely to make their return journey on another route in the same corridor. On a related note, as shown in Table 5-8, the top three routes with continuing journeys on Route 69 are routes that share a corridor with it. Although they nonetheless represent plausible interchanges (e.g., at the end of the shared corridor) there is reason to be concerned that bus-bus journeys involving these routes actually constitute short return journeys on different routes that were erroneously linked together by the elapsed time thresholds.

Figure 5-9 Map of Route 69 in North East London



Source: Transport for London, North East London Bus Map, www.tfl.gov.uk

This observation is particularly true of Route 97, which has a long shared corridor with Route 69. Journeys involving Routes 257 and 58 are more likely to represent real interchanges because they have a shorter shared corridor with Route 69. Some approaches to addresses these concerns were suggested in the Route 221 case study.

Table 5-8 Top 10 Routes in Journeys on Route 69

Route No.	Share of Top 10 Routes in Journeys on Route 69*	Station Intersections with Route 69	Road Intersection with Route 69
97	16%	Walthamstow Central, Leyton	Long shared corridor*
257	14%	Walthamstow Central	Lea Bridge Road*
58	11%	Walthamstow Central	Church Road*
34	10%	Walthamstow Central	Hoe St.
262	9%	Stratford, Plaistow	Greengate Street*
158	9%	Stratford-T, Leyton	Church Road*
86	8%	Stratford-T	Romford Road
25	8%	Stratford	Romford Road
473	8%	Stratford-T, Plaistow	Greengate Street
56	7%	-	Lea Bridge Road

*Excluding Route 69.
**Indicates shared corridor with Route 69.

In terms of connectivity with the Underground, recall from Table 5-2 that 35 percent of journeys on Route 69 involve interchange with the Underground (as opposed to only 30 percent that involve interchange with other bus routes). The stations that appear most often in journeys on Route 69 are the five stations on the route: Walthamstow Central, Leyton, Stratford, Plaistow, and Canning Town. As shown in Table 5-9, Leyton Underground Station, located near the center of the route, appears most often and twice as often as the next highest station, Stratford. This provides an indication of the most beneficial sections of the route from the perspective of passengers' ability to access the Underground. Unfortunately, Oyster smart cards have limited use on National Rail so the extent to which Route 69 is used to access National Rail at stations such as Stratford and Walthamstow Central cannot currently be determined from Oyster card analysis.

Table 5-9 Top 5 Stations in Journeys on Route 69

Underground Station	Share of Top 5 Stations
Leyton	43%
Stratford	21%
Walthamstow Central	12%
Plaistow	12%
Canning Town	11%

It is interesting that interchanges between Routes 34 and 69 follow a different pattern than interchanges between Routes 55 and 69. This may be because Route 55 meets Route 69 at a crossroad intersection whereas Route 34 meets Route 69 where both routes terminate, at Walthamstow Central Station. Figure 5-10 suggests that passengers tend to travel slightly longer on Route 55 to access Route 69 than they do on Route 69 to access Route 55, but in both cases, the elapsed time between bus boardings is less than 30 minutes for about 80 percent of the complete journeys. In contrast, Figure 5-11 shows interchanges between Routes 34 and 69 for which the elapsed time between bus boardings is at least 5 minutes in all cases and 34-37 minutes for about 80 percent of the complete journeys. This analysis, based on 440 daily interchanges between Routes 69 and 34, and 290 daily interchanges between Routes 69 and 55, demonstrates that a significant number of interchanging passengers would be affected by a change in bus service on any of these routes. Moreover, the cumulative distributions illustrate that passengers tend to spend more time traveling before an interchange at a dedicated station than at a street-corner. Of course, this is only one example but it illustrates that there are differences in travel behavior across route combinations and that the complete journeys data could help to quantify time savings for interchanging passengers, if, for example, a direct link were created to replace two intersecting routes.

Finally, another approach to understanding bus route connectivity is to look at individual journey patterns. Route 472, which terminates at North Greenwich Station and does not intersect with Route 69, nevertheless appears 92 times in journeys involving Route 69. An examination of journeys that include Underground entry or exit at North Greenwich Station yields 103 bus-Underground-bus journeys through North Greenwich in which one bus stage is on Route 69.

Figure 5-10 Elapsed Time Distributions for Interchanges between Routes 69 and 55

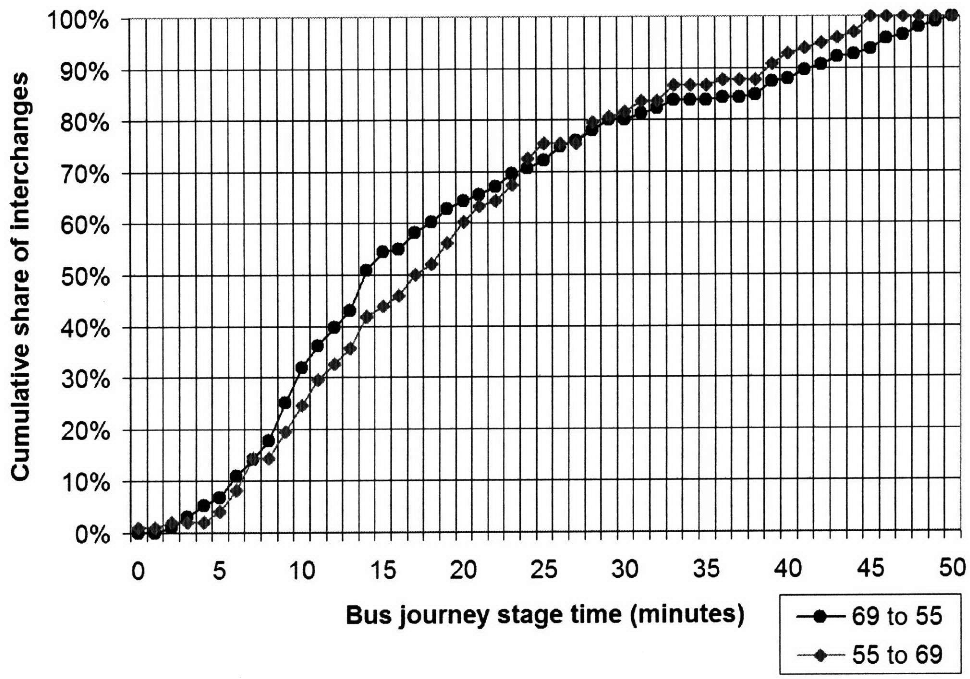
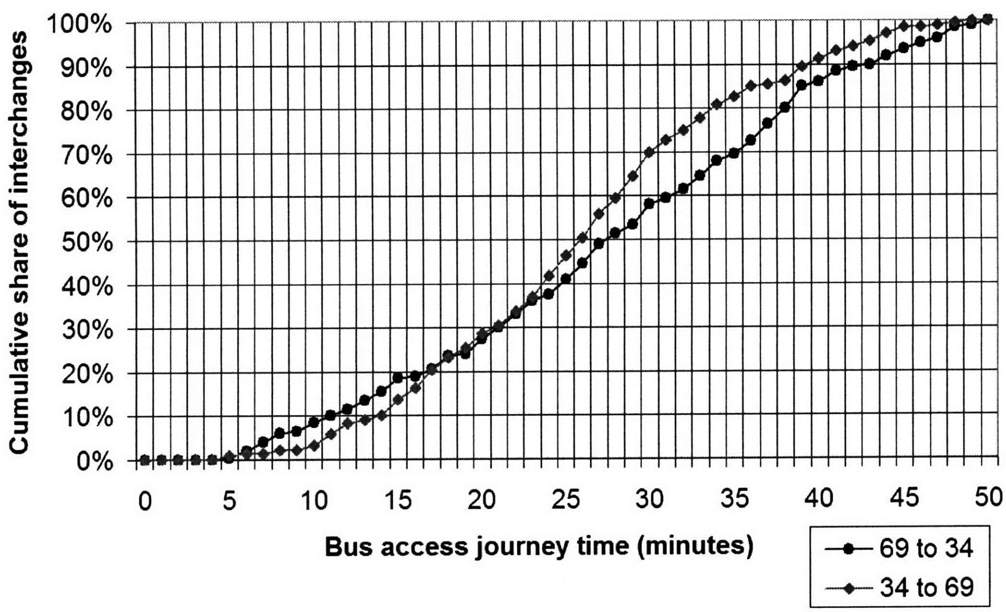


Figure 5-11 Elapsed Time Distributions for Interchanges between Routes 69 and 34



For 26 of these journeys, the Underground stage is between North Greenwich and Canning Town, adjacent stops on the Jubilee Line. Canning Town is also the terminus of Route 69. In addition, several 4-stage journeys involving Route 472 and a journey stage between North

Greenwich and Canning Town were observed, for example the journeys shown in Table 5-10. Referring to the route map in Figure 5-10, if Route 472 were extended across the river to Canning Town Station or Route 69 were extended across the river to North Greenwich Station, this would save passengers the inconvenience of interchanging for a one-stop Underground journey stage and would reduce the number of passengers on the Underground, which could have congestion benefits. In this particular case, the number of people following this travel path is probably too low to justify a change in service, but it does provide an example of using the smart card data to identify potential routing improvements. North Greenwich Station is examined in more detail in the final case study in this chapter.

Table 5-10 Individual Journey Pattern on Route 69

User ID	Journey Number	Stage Number	Mode	Start Location	End Location / Bus Direction	Start Time	End Time
5359	1	1	B	69	2	10:37	0
5359	1	2	U	Canning Town	North Greenwich	10:46	10:53
5359	1	3	B	472	2	11:04	0
5359	1	4	B	53	1	11:34	0
5359	2	5	B	53	2	15:26	0
5359	2	6	B	472	1	15:40	0
5359	2	7	U	North Greenwich	Canning Town	16:02	16:09
5359	2	8	B	69	1	16:17	0

5.2.5 Comparison of Underground Station Case Studies

Next, two case studies of interchange behavior at Underground stations are presented. This is important because network planners currently have limited quantitative information about how bus services are used in conjunction with the Underground network. Before presenting examples specific to each station, overall modal patterns and connectivity in terms of return journeys at the two stations, Vauxhall and North Greenwich, are compared. Table 5-11 shows that Vauxhall has a far higher share of single-stage Underground journeys than North Greenwich. This observation may be because Vauxhall journeys are likely to include interchange with National Rail services that are not captured by Oyster, but what is certain is that bus is a far more important access/egress mode at North Greenwich Station than Vauxhall Station. Not only do almost 50 percent of journeys through North Greenwich include bus stages, but there is a surprisingly high share of 3-stage (e.g., bus-bus-Underground) journeys as well. This may be explained by the lack of National Rail services to North Greenwich Station as well as its isolation from housing and office buildings, so that for the most part bus or car are the only viable means of accessing the station.

Table 5-11 Modal Patterns of Journeys through Vauxhall and North Greenwich Underground Stations

Mode 1	Mode 2	Mode 3	Mode 4	Vauxhall	North Greenwich
U	-	-	-	66%	30%
U	B	-	-	12%	24%
B	U	-	-	14%	22%
B	U	B	-	4%	9%
U	B	B	-	1%	5%
B	B	U	-	1%	5%
B	U	B	B	<1%	2%
B	B	U	B	<1%	2%

U = Underground; B = Bus

Another way to compare the two stations is in terms of probable return journeys. Table 5-12 shows that 67 percent of passengers who enter or exit the Underground at North Greenwich station do so at least twice a day as compared to only 45 percent of passengers at Vauxhall Station. This suggests that North Greenwich is used primarily for return journeys whereas journeys that involve Underground entry or exit at Vauxhall Station are part of more complex daily travel patterns or involve interchanges with National Rail. This knowledge would be of value in determining what information about bus services that should be made available at each station.

Table 5-12 Return Journeys through Vauxhall and North Greenwich Underground Stations

Underground Station	Passengers with 2+ Daily Journeys through Station	Share of All Journeys through Station
Vauxhall	45%	63%
North Greenwich	67%	81%

5.2.6 Vauxhall Station

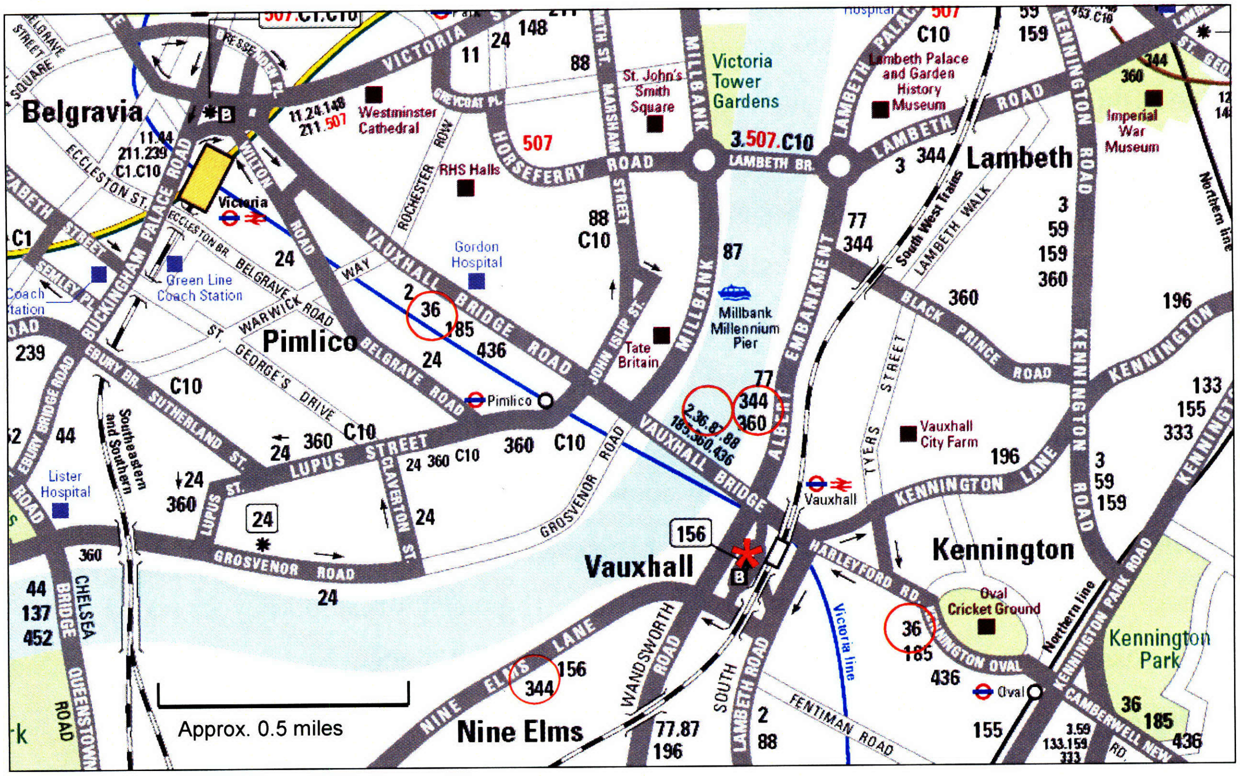
Vauxhall Underground Station is integrated with a purpose-built bus station that has stands for Routes 2, 36, 77, 87 (no data), 88, 156, 185, 196 (missing data), 344, 360, and 436 (missing data). Night buses N2, N87, and N136 also stop at Vauxhall Bus Station but no data was available for these routes. Missing data for certain routes is the result of changes in bus routes since 2003 for which the Oyster database has not been updated (this issue and steps taken to address it are discussed in Chapter 2). A map of current bus routes serving Vauxhall Station is provided in Figure 5-12.

As shown in Table 5-13, the top three routes to appear in Vauxhall Station journeys, namely Routes 344, 77, and 156, run only south of the River Thames on Nine Elms Lane and Albert Embankment. However, the next four routes cross the River Thames on Vauxhall Bridge toward Victoria Station. Approximately 5 to 15 percent of all journey stages on routes that cross the river are part of journeys with interchange at Vauxhall Station. This information suggests that although passengers are using bus feeder routes to access Vauxhall Station primarily from South of the river, some passengers may choose to enter or exit the Victoria Line at Vauxhall Station and use bus services to access their origin or destination across the river in Pimlico or Belgravia. Further examination of individual travel patterns for Routes 36, 2, 88, and 185 across days could be used to substantiate this suggestion and provide insight into how passengers trade off between bus and Underground journey stages.

Table 5-13 Top 7 Routes in Journeys through Vauxhall Underground Station

Route No.	Stage Count	Road In/Out of Vauxhall Station	Share
344	1,056	Nine Elms Lane; Albert Embankment	23%
77	901	Wandsworth Road; Albert Embankment	20%
156	643	Nine Elms Lane	14%
36	579	Vauxhall Bridge; Harleyford Rd	13%
2	542	Vauxhall Bridge; Lambeth Road	12%
88	435	Vauxhall Bridge; Lambeth Road	10%
185	346	Vauxhall Bridge; Harleyford Rd	8%
Total	*4,502*		*100%*

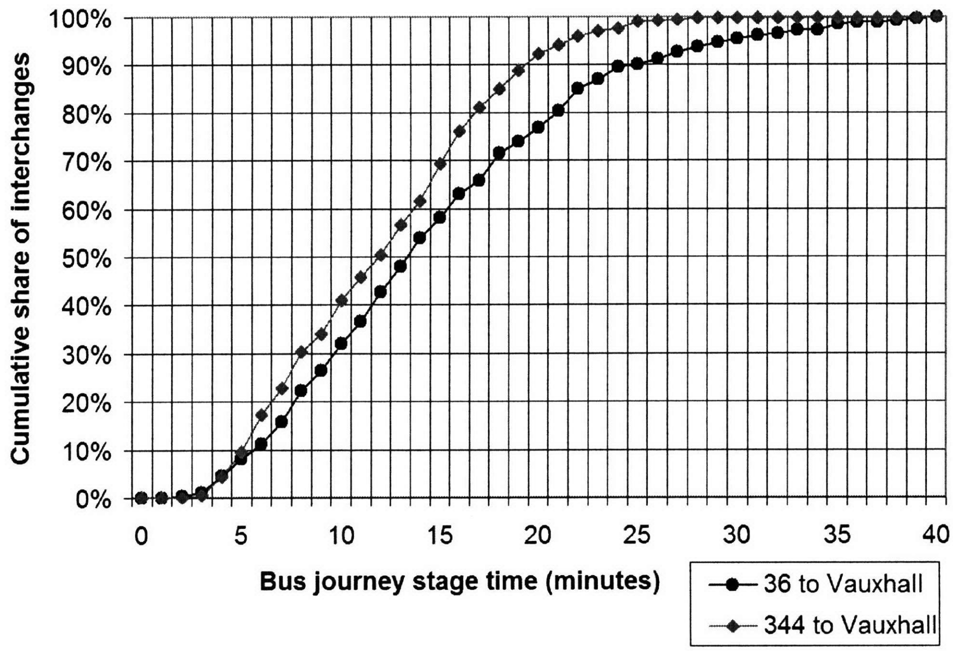
Figure 5-12 Map of Bus Routes Near Vauxhall Station



Source: Transport for London, Central London Bus Map, www.tfl.gov.uk

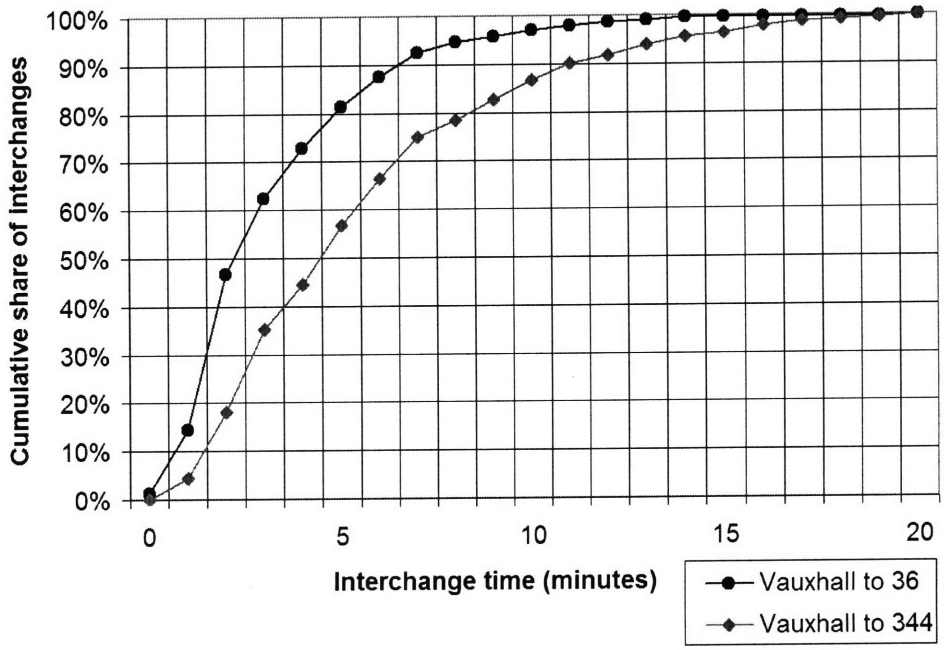
A comparison of elapsed times for journey stages to access Vauxhall Station on two bus routes, Route 344 on the South Bank and Route 36 which comes from the South then crosses the river, reveals bus journey stage times of 3 to 28 minutes and 2 to 40 minutes respectively (see Figure 5-13). In the case of Route 344, 95 percent of interchanges occur within 22 minutes of initial bus boarding whereas in the case of Route 36 this level is only reached within 29 minutes of initial bus boarding. Once again, this observed behavior shows different patterns of demand across routes for Underground access journey stages. Knowledge of this type might inform a decision about whether to shorten or lengthen a route that is designed to serve a particular Underground station.

Figure 5-13 Elapsed Time Distributions for Interchanges from Routes 36 and 344 to Vauxhall Underground Station



Finally, Underground-to-bus interchange times for Routes 36 and 344 could inform the evaluation of passengers' interchange experience at the station. Figure 5-14 shows that 95 percent of bus boardings occur within 8 minutes of station exit for Route 36 versus 14 minutes for Route 344. Route 36 has a daytime headway of every 3-6 minutes at Vauxhall Station between 7 a.m. and 7 p.m. and between 6-12 minutes outside of those times whereas Route 344 has a headway of 4-8 minutes between 7 a.m. and 7 p.m., and between 6-14 minutes outside of that time. Moreover, the stops for Route 344 are slightly further from the Underground station exit than the stops for Route 36, which could increase the elapsed time for interchanges from the Underground to Route 344 by a few minutes. For these two routes, observed interchange behavior reflects scheduled frequencies and is therefore consistent with our expectations about interchange time. However, one could image situations in which severe delays on a particular route increase wait times or crowding prevents some passengers from boarding the first bus. In these instances, the interchange time distribution would be different than our expectations based on planned frequency and station design, and steps could be taken to remediate the problem. Moreover, the elapsed time distributions show the cumulative transfer time for both routes, which would be of value in quantifying passenger time loss/savings in a cost/benefit analyses for possible interchange or network enhancements.

Figure 5-14 Elapsed Time Distributions for Interchanges from Vauxhall Station to Routes 36 and 344



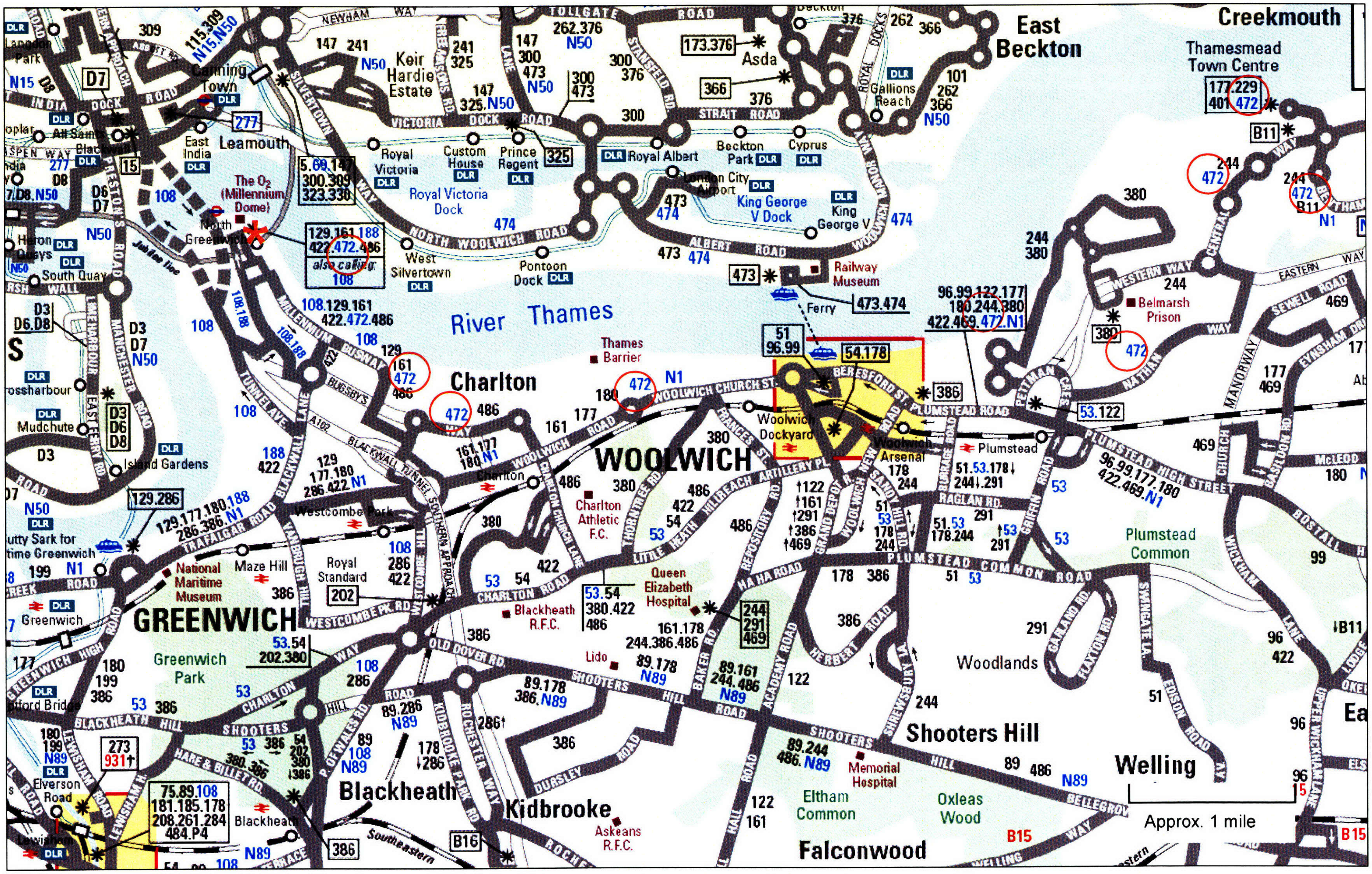
5.2.7 North Greenwich Station

Routes 108, 129, 161, 188, 422, 472, and 486 provide service to North Greenwich Underground Station on the Jubilee Line with all but Route 108 terminating there. Routes in the vicinity of North Greenwich Station, the subject of this last case study, are illustrated in Figure 5-15.

Recall from the earlier discussion of Table 5-11 that North Greenwich Station is an important location for interchanges between bus and Underground. A full 19 percent of journeys with an entry or exit at North Greenwich Station include two bus journey stages and in total 68 percent of Underground journeys through North Greenwich Station have an associated bus journey stage. This figure is quite when compared to the estimate that only 25 percent of all Underground journeys in London have an associated bus journey stage[20]. It is also much higher than Vauxhall Station, where only 32 percent of Underground journeys through the Underground station have associated bus journey stages.

[20] Based on the complete journeys file, 25 percent of Underground journeys have an associated bus journey stage and based on the author's calculations from the 2005 LTDS this figure is 23 percent.

Figure 5-15 Map of Bus Routes Near North Greenwich Station



Source: Transport for London, Central London Bus Map, www.tfl.gov.uk

Given the importance of bus as an access/egress mode at North Greenwich Station, it is valuable to further quantify how the bus network interacts with this Underground station.

Routes 472 and 486 are the most important bus routes for access to North Greenwich Station. Approximately 43 percent of daily journeys on Route 472 and approximately 40 percent of daily journeys on Route 486 involve interchanges with the Underground at North Greenwich. These and the other most common routes in journeys involving interchange at North Greenwich are summarized in Table 5-14. This type of information could help with decisions about bus service changes. For example, Route 188 provides 6 percent of interchanges at North Greenwich Station, but this value only represents 6 percent of journey stages on the route and thus is a relatively small source of revenue from passengers. On the other hand, passengers on Route 129 only represents 4 percent of interchanges at North Greenwich, but these interchange journeys comprise more than half the journey stages on Route 129 so from the operator perspective it is important to maintain the service to North Greenwich Station.

Table 5-14 Top 7 Routes in Journeys through North Greenwich Station

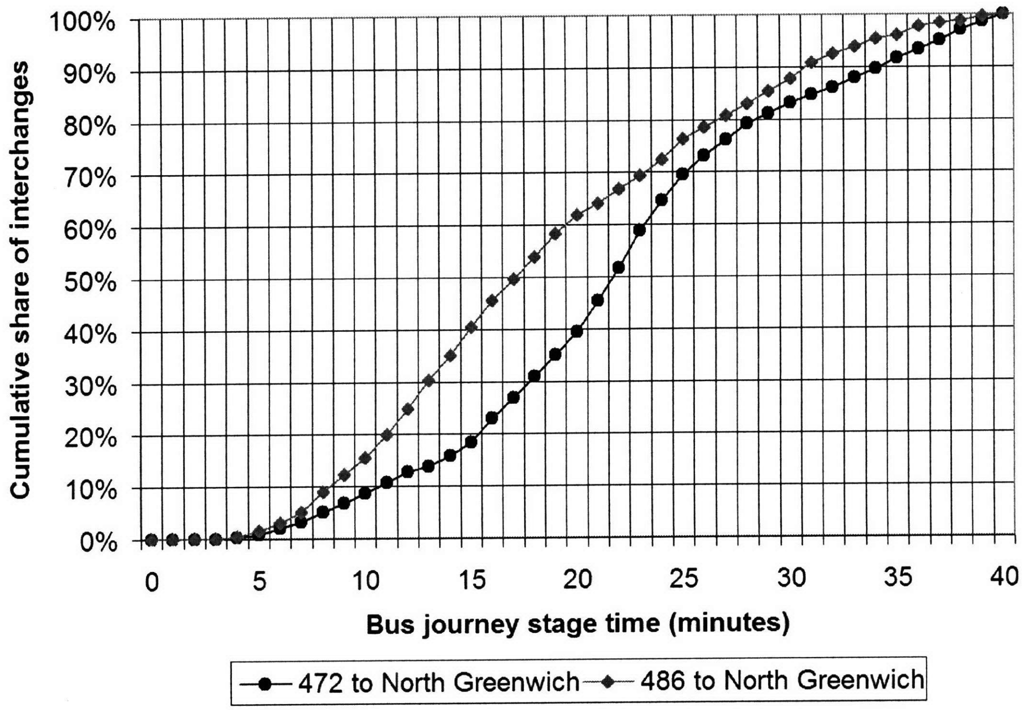
Route Number	**Total Stages in North Greenwich Journeys** (1)	**Share of (1)**	**Total Daily Stages on Route*** (2)	**(1) / (2)**
472	9,004	41%	20,722	43%
486	3,985	18%	10,025	40%
161	3,135	14%	12,224	26%
422	2,316	10%	10,559	22%
108	1,560	7%	7,453	21%
188	1,239	6%	19,387	6%
129	937	4%	1,757	53%
Total	*22,176*	*100%*	*82,127*	*27%*

*Using Oyster data.

Continuing with the top two routes, the average bus journey durations to access North Greenwich Station are 23 minutes on Route 472 and 18 minutes on Route 486. However, the elapsed time distributions for each of these routes, shown in Figure 5-16, illustrate that even near the maximum durations of 40 minutes these are pure bus access journeys because the trend line does not flatten out. This makes intuitive sense because the runtime for Route 472 from Thamesmead Town Center to North Greenwich is 41 minutes and the runtime for Route

486 from Bexleyheath is 42 minutes. A total of 1,950 daily access journeys were recorded on Route 486 and about 4,150 daily access journeys on Route 472.

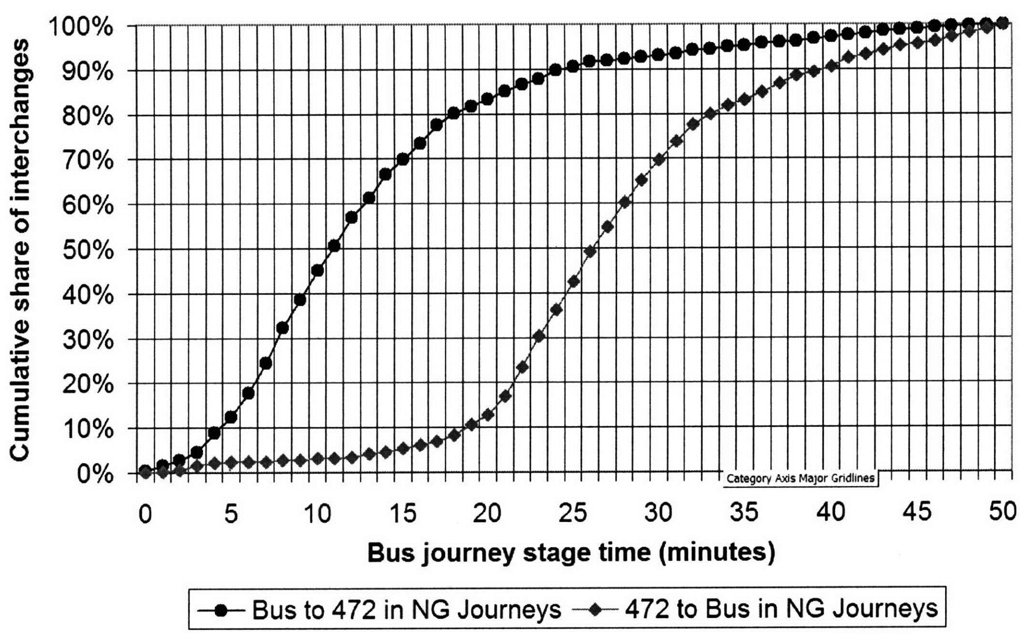
Figure 5-16 Elapsed Time Distributions for Interchanges from Routes 472 and 486 to North Greenwich Station



This shows that people are willing to make longer bus access journeys than have been presented so far and, consequently, that the elapsed time assumption of 40 minutes for bus access journeys is certainly appropriate in this case. Once again, this suggests that ideally the elapsed time threshold should be route specific although this would be difficult to implement given the 700 routes in the TfL network.

The prevalence of bus-bus-Underground and Underground-bus-bus modal patterns in North Greenwich journeys argues for a closer examination of bus routes used in these journeys. The most frequent bus-bus-Underground pattern involves interchange from another bus route to Route 472 and then interchange to the Underground at North Greenwich Station. Therefore, elapsed times between bus boardings for all other routes and Route 472 are examined. Figure 5-17 shows that after exiting North Greenwich Station, passengers generally travel about 20 to 35 minutes (including interchange wait time) on Route 472 before boarding another bus to finish their journey to their final destination. In the reverse case of interchange from other routes to Route 472, passengers travel only 3 to 20 minutes on the other route before boarding Route 472 to access North Greenwich Station. These results make sense because the most common interchanges between Route 472 and another bus are with Routes 177, 53 and 180 which are all at least 11 minutes from North Greenwich Station. In addition, the average bus access journey duration for Routes 177 and 472 together (i.e., interchange from Route 177 to 472) is 35 minutes with a maximum of 70 minutes, and the average bus access journey duration for Routes 35 and 472 together is 34 minutes with a maximum of 70 minutes. This suggests that people are willing to travel well over half an hour on multiple bus routes in order to access an Underground station if that is the best available public transportation option. One way that bus service to North Greenwich Station might be improved for the many passengers that depend on it is through zonal express routes. For example, Route 472A could serve all stops from Thamesmead Town Center to Plumstead Road where it intersects with Route 53 and from there run express to North Greenwich Station while Route 472B could run express from Thamesmead Town Centre to Woolwhich where it would serve all stops and then run express to North Greenwich Station.

Figure 5-17 Elapsed Time between Bus Boardings for All Routes to Route 472 and Reverse in North Greenwich Access or Egress Journeys



5.3 Summary

This chapter provided examples of new information of value to bus network planning that can be gained from smart card data. The new information can be categorized into four areas of contextual knowledge: connectivity, intermodalism, bus access journeys, and underground to bus interchange times, and one key quantitative measure, that is counts of interchanging passengers between routes that can be used as an input to cost/benefit evaluations.

The three bus route case studies demonstrated that expected variation in connectivity and intermodalism across routes is borne out in the complete journeys for these routes, and that the expected intersecting routes and stations also emerge from the data. However, unexpected results such as the disproportionately high number of interchanges between Route 221 and Bounds Green Station or the appearance of Route 472 in journeys on Route 69 draw attention to potential areas for network improvement. In addition, the two station case studies provide examples of the value of comparing bus access times and interchange times across routes, as well as identifying unexpected journey patterns such as the high number of two-stage bus journeys made to access North Greenwich Station. In all cases, the simple metric of volume of interchanges between intersecting routes is a fundamental result.

After completing the network planning applications, I propose to lower the recommended bus-to-Underground elapsed time threshold to 35 minutes and the bus-to-bus threshold to 45 minutes because there was no evidence that pure interchanges would be excluded at this level. The recommended Underground-to-bus elapsed time threshold remains 20 minutes.

CHAPTER 6: IMPLEMENTATION AND CONCLUSIONS

This chapter summarizes the main findings and research contributions of this thesis (Section 6.1). Recommendations are also provided for the implementation of a complete journeys data system at London Buses (Section 6.2) and suggestions for future research in the area of smart card data for bus network planning (Section 6.3).

6.1 Research Contributions

This research contributes to an emerging body of literature on the application of smart card data to transportation planning that has until recently been more heavily focused on rail than buses. It also addresses the issue of planning for interchanges, which has long been established as a key component of network planning but has lately enjoyed a renewed interest in transportation planning practice as evidenced by recent literature on integrated and intermodal networks. Finally, this research contributes to the practical challenge of applying smart card data to bus network planning, with a focus on interchange, in London.

To meet the research objective of assessing the potential application of smart card data to bus network planning in London, I first reviewed existing data systems and the network planning approach at London Buses. Then, I examined three potential interchange combinations: bus-to-Underground, Underground-to-bus, and bus-to-bus to gain an understanding of interchange behavior in London and to formulate some recommendations for elapsed time thresholds to identify interchanges between journey stages for each passenger. Then, I compared other TfL data with the results of linking journey stages into complete journeys based on these elapsed time thresholds. Finally, I applied this complete journeys data to network planning case studies that demonstrated new contextual and quantitative information that would be of value to bus network planners without necessitating additional data from other sources such as AVL or APC systems.

Key findings and contributions in terms of interchange time thresholds, intermodal travel behavior, and new information for bus network planning are discussed below.

6.1.1 Interchange Time Thresholds

Based on an analysis of elapsed times between journey stages for individual passengers across locations and time periods in Chapter 3, I recommend elapsed time thresholds for linking journey stages to form complete journeys for the entire London network as follows: Underground-to-bus interchanges from 15 to 25 minutes between Underground station exit and bus boarding; bus-to-Underground interchanges, including bus in-vehicle time, from 30 to 50 minutes between bus boarding and Underground station entry; and, bus-to-bus interchanges, including bus in-vehicle time, from 40 to 60 minutes from first bus boarding to second bus boarding. I also note that the most appropriate times vary across stations and for certain stations an Underground-to-bus interchange time as low as 10 minutes is recommended.

After comparing the results of linking journey stages into complete journeys using elapsed time thresholds from the middle of the ranges as well as either end with existing travel survey data, I concluded that either the low end or middle of each range was most appropriate for creating complete journeys at the network-wide scale. Nonetheless, as expected, any elapsed time threshold within the recommended ranges gave similar results in terms of complete journeys formed. Again, the goal is to determine elapsed time thresholds that include all pure and incidental activity interchanges in the network, but exclude all non-interchanges. That said, given the network planning applications presented in Chapter 5, the preference is to err on the side of falsely including non-interchanges in the elapsed time thresholds as opposed to falsely excluding pure interchanges. As such, I decided to use the middle of the ranges for the network planning applications presented in Chapter 5 (discussed further below). After completing the network planning applications, I proposed to lower the recommended bus-to-Underground elapsed time threshold to 35 minutes and the bus-to-bus threshold to 45 minutes because there was no evidence that pure interchanges would be excluded at this level but it would reduce the "noise" of linked journeys on shared bus corridors with complex travel patterns. The recommended Underground-to-bus elapsed time threshold remains 20 minutes.

Ideally, we would implement different elapsed time thresholds for each station, or even for each route-station or route-route combination but this was not possible within the scope of this research. Nor is it likely to be possible in the short-term application of Oyster-card data to bus network planning at TfL. However, a means of accomplishing this next step is discussed in the section on implementation below.

Although the time thresholds for identifying interchanges between bus and Underground journey stages at the passenger level were applied to the formation of complete journeys and then to the analysis of bus passenger demand at interchanges in this research, there are other potential applications of this interchange information. One example is to inform fare policy on free or discounted transfers, which are widely recognized to be an important part of creating a truly integrated network. Transfer discounts at TfL currently take the form of period tickets or daily capping, but with daily capping passengers are still required to pay for each journey stage until they reach the cap regardless of how close in time the two journey stages occur. It is my understanding that there was some discussion of discounted transfers at TfL in the past, with proposed elapsed times for discounts to be 60 minutes for bus-to-bus and bus-to-Underground interchanges, and 30 minutes for Underground-to-bus interchanges. These wide time thresholds may be necessary from a public relations perspective (i.e., to accommodate extreme cases), but they do not reflect the general interchange behavior observed through the smart card analysis in this research.

This research on elapsed time thresholds for identifying complete journeys would also be beneficial to network and interchange planners in TfL groups other than London Buses. TfL's current organizational structure was created in part to deliver a more integrated public transportation network in London and yet there appears to be limited intermodal information available across groups for planning purposes. For example, TfL's *Interchange Plan* states that reliable data on such basic metrics as the number of people interchanging at each facility was not available at the time of publication (Transport for London 2002).

6.1.2 Intermodal Travel Behavior in London

Next, a brief summary of three additional findings about interchange behavior for bus and Underground in London is provided. First, variation in elapsed times between bus and Underground, and between two bus journey stages across time periods reflects complex travel patterns in the Midday and PM Peak, particularly for bus journeys. Second, variation in travel behavior in terms of elapsed times between journey stages at different Underground stations appears to depend, in part, on land uses around stations in addition to expected factors such as bus service frequency. Third, variation in interchange behavior between fare zones may also be influenced by land use because a large number of Underground stations in Zones 1 and 2 have high-density, mixed land uses surrounding them. This physical context appears to disperse observed interchange times. Finally, bus passenger interchange behavior is influenced by ticket type, for example PAYG versus pass holders. The implication of these and similar findings is that although travel and interchange patterns are complex, there may be means of influencing people's interchange behavior, for example, by locating shops and restaurants around a station or by implementing fare-related incentives (or disincentives) for interchange.

Some other findings about interchange behavior based on forming complete journeys using elapsed time thresholds between journey stages are that approximately 25 percent of Underground journeys involve bus segments and 75 percent of bus or Underground journeys are single-segment journeys (although they may be linked with National Rail). In addition, 20 percent of complete journeys are two-segment journeys and about 5 percent include three or more segments (mostly three). These types of metrics are currently not readily available at TfL but with smart card data they could be monitored over time to assess the impacts of policy or network planning measures intended to influence interchange behavior in London.

6.1.3 New Information for Bus Network Planning

The final area where this research makes a key contribution is in a simple methodology for providing new, valuable information from smart card data for application to bus network planning in London. Route- and station-level examples demonstrated that with smart card data we can now quantify interchange volumes, or passenger demand, between two intersecting routes and that this information might support rerouting decisions such as direct links that reduce the need for passenger interchange. It could also be used as a cost/benefit input if the value of time for interchanging passengers is considered to be different than for passengers who access a route by walking. Also, the ability to quantify passenger demand for interchange between a bus route and an Underground station is of value in bus station design and other aspects of planning service to and from a station. For example, elapsed times between bus boarding and station entry provide new contextual knowledge about the area served by a station and how long people are willing to travel by bus to access it. Finally, new information about: the percentage of intermodal journeys on a route, the most important interchange locations for a route, the most important routes serving a station, individual travel patterns, and so on can only prove beneficial in terms of informing decisions about bus route planning.

In short, although the spatial accuracy of bus boardings is currently limited to the route level in the London smart card data, the identification of complete passenger journeys using elapsed time thresholds certainly provides new, relevant information about passenger demand for interchanges that supports a more integrated evaluation of bus routes by network planners. However, the challenges related to implementing a new information system based on smart card data are not insignificant and are therefore discussed further in the following section.

6.2 Implementation of Complete Journeys Information Using Smart Card Data

The intent of this research was to provide examples of a methodology for creating value out of smart card data that could actually be applied at London Buses in the short term. In this section, four short-term and two long-term considerations in implementing a new information system using smart card data that follows on the planning examples and complete journeys approach presented herein are discussed. The short-term considerations are user interface and automated reports, data clean-up, data sampling and time frame, and institutional structure.

TfL staff from Network Development and Prestige have already initiated a collaborative process for developing a smart card data system that will be used by bus network planners. One area that requires particular attention is the user interface, including the choice of automated reports. Assuming that the system is based on a sequenced journeys-type file that is used to create a complete journeys dataset, at the most basic level the system should allow users to select a route or station of interest and then to generate a report of summary statistics that includes:

- number of daily journeys on the route,
- number of daily passengers on the route,
- average stages per journey,
- total and share of journeys by number of stages,
- average journeys per passenger on the route,
- total and share of journeys per passenger, and
- a summary of journeys by modal pattern.

The automatic reports should also include cumulative distribution functions of elapsed times for Underground-to-bus interchange, bus-to-Underground access times, and bus-to-bus access times. These automated graphs would be aggregated to the level of the route or station initially selected. It is also crucial that the user be able to automatically generate lists that provide a count of all routes and stations that are part of complete journeys on the selected route (or station). After an initial route or station has been selected, a subset of journeys involving that route or station and any other one could be summarized in terms of interchange volumes, cumulative elapsed time distributions and most common journey patterns detailed to the level of route number and station name. SQL code has been developed in the course of this research to generate all of the reports mentioned above, with the exception of graphing which was done in Microsoft Excel. For all but the largest stations, a raw data file listing all complete journeys for a selected route or station should be short enough to fit into an Excel spreadsheet for further examination by network planners as needed.

Although the raw Oyster data undergoes processing and error correction to create the sequenced journeys tables, some additional data clean-up is needed before creating complete journeys based on elapsed time thresholds. The most important actions are to:

- replace out-of-date farestage codes currently used to identify bus route numbers with an up-to-date Route ID attribute for identifying bus boarding location (automated as of Spring 2008),
- remove cards that have more than 10 daily journey stages because these represent outliers or staff cards that have been used to grant passengers entry at malfunctioning ticket gates,
- remove cards with journey stages by National Rail because they currently represent a small share of Oyster card users and the majority of National Rail journey stages are unstarted or unfinished thus complicating the formation of complete journeys disproportionately relative to the potential gains from including them in the analysis[21],
- ensure that all card IDs are encrypted,
- ensure that all Underground station codes are up-to-date and unique, and

[21] Oyster card usage on National Rail is expected to increase substantially in the near future, at which time it would be preferable to include National Rail stages in the formation of complete journeys with a similar treatment as is given to Underground journey stages.

- assign new sequence numbers for each card per day that start at 1 and increase by increments of one to facilitate coding.

When creating complete journeys, Underground journey stages with missing entry or exit information should be accounted for by adding 55 minutes to the elapsed time thresholds where necessary to account for Underground travel time. Tests show that this adjustment has a negligible affect on aggregate journey patterns but it would affect results for stations with open gates. A similar method could also be used for National Rail journey stages included in the future because entries and exits outside Central London are often missing. Finally, once complete journeys are formed, journeys with five or more stages should be excluded because they represent outliers and complicate subsequent analyses unnecessarily.

Another aspect of implementation that is best considered in the short term is the availability of data across days. Any given route, station or interchange combination in the system is likely to vary with time so the implementation of an Oyster database system for network planning would need to allow for queries across multiple days, and, ideally, across time periods. Due to the low passenger volumes on some routes, it is recommended that all Oyster journey stages be used for any given day but that complete journeys datasets be created for one week out of every month or some similar method for limiting the amount of processing and storage required. Journey patterns are known to be very different on weekends than weekdays but data on weekends is scarce so this is an area where Oyster can provide substantial added value. Therefore, it is important to ensure that weekend days are included in the new data system.

The final short-term consideration relates to institutional structure and information flow. There are already numerous data systems in use at Network Development that require substantial resources to maintain, so a rationalization of information systems and procedures is recommended as part of the process of developing a new smart card data system. One obvious consideration is the eventual integration of the smart card system for analyzing interchange with BREMS, which already relies substantially on smart card data for reports on daily passenger volumes by bus route. Although an integrated system may not be implemented immediately, it should be considered in the system design from the outset to minimize required changes in the future. By rationalizing and comparing existing systems in the process of implementing a new smart card system for interchange analysis, there will be opportunities to discover new information, for example the large discrepancy between passenger volumes on articulated bus routes as reported in BREMS compared to the Oyster sequenced journeys tables. Overall, it is recommended to start with a stand-alone system for speed and ease of implementation but to simultaneously plan for longer term opportunities for integration with or replacement of existing systems, especially BREMS.

Having discussed short-term considerations for the implementation of a new smart card data system for analyzing bus passenger interchanges, two longer term considerations are presented: first, applying elapsed time thresholds by station, and, second, iBus, which is TfL's new GPS-based AVL system. Both considerations relate to improving the accuracy of linking journey stages into complete journeys for the purposes of analyzing more complex travel patterns on the bus network than is possible using data on single journey stages by route.

In the longer term, it would be ideal to apply different elapsed time thresholds to identify interchanges at each Underground station. This would reduce the error rate of linking journey stages that actually form separate journeys or, conversely, of not linking journey stages that are actually part of complete journeys. Information on distributions of elapsed time thresholds for each station was extracted for this research but it proved too complex to implement station-specific thresholds at the network-wide level. That said, it would be possible to first extract all cards that include journey stages for a selected route or station, and then apply station-specific interchange times only to the stations that appear in that subset of cards. This approach would lower the computational requirements relative to applying different elapsed time thresholds at the network-wide level and thus it might be possible to identify complete journeys just for those cards on the fly as the route or station selection is being made by the system user. The additional input, aside from the sequenced journeys table, to this method is a table specifying the elapsed time thresholds for each station for both bus-to-Underground and Underground-to-bus interchanges.

Station-specific elapsed time thresholds do not solve the limitation of route-level spatial accuracy for bus journeys, but this could be addressed through iBus, which is in the process of being rolled out at TfL. At its most basic level, a synchronization of the time stamp clocks for iBus and Oyster card readers would enable the location of the bus to be identified at the time a passenger boards the bus and taps their smart card on the reader. This type of stop-specific spatial information would add incredible richness to the smart card data to a level of detail that would not only make information on passenger interchanges, especially bus-to-bus, more precise by allowing for spatial thresholds as well as time thresholds, but also might eventually modify the sampling methodology for BODS.

6.3 Future Research

Building on the short- and long-term considerations for implementation as well as interesting findings more generally, some potential areas for future research to continue to explore the application of smart card data to bus network planning include:

- Improving the spatial accuracy of bus boarding locations with postcode or iBus (preferred) data and then possibly using this enriched data to build an origin-destination matrix for the bus network;
- Use postcode for registered Oyster card users to better understand tradeoffs between walking and extra bus journey stages in accessing the Underground network or between bus and Underground journey stages across days;
- Study subsets of passengers based on travel patterns, for example two bus journeys per day;
- Model interchange behavior, for example where number of interchanges = *f*(network connectivity, land use, socio-economic variables, service quality, etc.);
- Develop appropriate sampling methodology to minimize the computational requirements of using smart card data for bus network planning while recognizing the low passenger volumes on many routes;
- Compare interchange times and intermodal travel patterns results to other urban areas; and
- Evaluate the impacts of increased data availability from smart card fare payment systems across TfL and other large transport organizations.

REFERENCES

Banister, D. & Berechman, Y. 2001, 'Transport investment and the promotion of economic growth', *Journal of Transport Geography*, vol. 9, no. 3, pp. 209-218.

Bagchi, M. & White, P.R. 2004, 'The Potential of Public Transport Smart Card Data', *Transport Policy*, vol. 12, pp. 464-474.

Chan, J. 2007, *Rail transit OD matrix estimation and journey time reliability metrics using automated fare data*, Master's Thesis, Massachusetts Institute of Technology, Boston.

Chu, K. K. A. & Chapleau, R. 2008, 'Enriching Archived Smart Card Transaction Data for Transit Demand Modeling', *Transportation Research Board 2008 Annual Meeting CD-ROM*, Washington, DC.

Crockett, C. 2001, *Strategies for Improving Service Connectivity*, Master's Thesis, Massachusetts Institute of Technology, Boston.

Cui, A. 2006, *Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems*, Master's Thesis, Massachusetts Institute of Technology.

Graham, D. J. 2007, 'Agglomeration, Productivity and Transport Investment', *Journal of Transport Economics and Policy*, vol. 41, no. 3, pp. 317-343(27).

Greater London Authority 2005, *London's changing population: Diversity of a world city in the 21st century*, Data Management and Analysis Group, Greater London Authority, London.

Greater London Authority 2001. *The Mayor's Transport Strategy*, Greater London Authority, London.

Guo, Z. & Wilson, N. H. M. 2006, 'Modeling Effects of Transit System Transfers on Travel Behavior: Case of Commuter Rail and Subway in Downtown Boston, Massachusetts', *Transportation Research Record*, vol. 2006, pp. 11-20.

Hine, J. & Scott, J. 2000, *'Seamless, accessible travel: users' views of the public transport journey and interchange'*, Transport Policy, vol. 7, no. 3, pp.217-226.

Hickman R. & Banister, D. 2007, 'Looking over the horizon: Transport and reduced CO2 emissions in the UK by 2030', *Transport Policy*, vol. 14, no. 5, pp. 377-387.

Hofmann, M., & O'Mahony, M. 2005, *Transfer Journey Identification and Analyses from Electronic Fare Collection Data*, IEEE Intelligent Transportation Systems 2005 Proceedings, Vienna.

Mayor of London 2008, *Transport – facts and figures*, Available at: http://www.london.gov.uk/mayor/transport/facts-and-figures.jsp.

Mayor of London 2007, *TfL Fact Sheet*, Available at: www.tfl.gov.uk.

Monbiot, G. 2006, *Heat: how to stop the planet from burning*, Anchor Canada, Canada.

Newman, P. W. G. & Kenworthy, J. R. 1996, 'The land use-transport connection: An overview', *Land Use Policy*, vol. 13, no. 1, pp. 1-22.

Okamura, T., Zhang, J. & Akimasa, F. 2004, 'Finding Behavioral Rules of Urban Public Transport Passengers by Using Boarding Records of Integrated Stored Fare Card System', *Proceedings of the 10th World Conference on Transportation Research*, Istanbul.

Park, J. Y. & Kim, D. J. 2008, 'The Potential Use of Smart Card Data to Define the Use of Public Transit in Seoul', *Transportation Research Board 2008 Annual Meeting CD-ROM*, Washington, DC.

Smart, M., Miller, M.A. & Taylor, B. 2007. 'Transit Stops, Stations, and Transfer Facilities: Evaluating Performance from the Perspective of Transit System Managers', *Transportation Research Board 2008 Annual Meeting CD-ROM*, Washington, DC.

The Institute of Logistics and Transport 2000, *Passenger Interchanges: A practical way of achieving passenger transport integration*, The Institute of Logistics and Transport, United Kingdom.

The Institute of Logistics and Transport 2000, *Joining up the Journey: Guidance on improving passenger interchange for those preparing Local Transport Plans and similar documents*, The Institute of Logistics and Transport, United Kingdom.

Transport for London 2008, *London Travel Report 2007*. Transport for London, London.

Transport for London 2007a, *Traffic Report – Period 9 (11 November 2007 – 8 December 2007)*, Fares & Ticketing, Transport for London, London.

Transport for London 2007b, *London Travel Report 2006*. Transport for London, London.

Transport for London 2007c, *Research Briefing: London Travel Demand Survey supplement*, Transport for London, London.

Transport for London 2004, *Guidelines for Planning Bus Services*, London Buses, Transport for London, London.

Transport for London 2003, *The case for investing in London's buses: Presenting the results of the London Buses Strategic Review*, London Buses, Transport for London, London.

Transport for London 2002, *Interchange Plan – improving interchange in London*, Transport for London, London.

Transport for London 2001, *Intermodal transport interchange for London: Best practice guidelines*, Transport for London, London.

Trépanier, M., Tranchant, N. and Chapleau, R. 2007, 'Individual Trip Destination Estimation in a Transit Smart Card Automated Fare Collection System', *Journal of Intelligent Transportation Systems: Technology, Planning and Operations*, vol. 11, no. 1, pp. 1-14.

Utsunomiya, M., Attanucci, J. and Wilson, N. 2006, 'Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning', *Transportation Research Record*, Transportation Research Board, vol. 2006.

Vuchic, V. R. 2005, *Urban Transit: operations, planning, and economics*, John Wiley & Sons, Hoboken.

Wardman, M. & Hine, J. 2000, *Costs of Interchange: A Review of the Literature*, Working Paper 546, Institute for Transport Studies, University of Leeds.

Zhao, J., Rahbee, A., Wilson, N. H. M. 2007, 'Estimating a Rail Passenger Trip Origin-Destination Matrix Using Automatic Data Collection Systems', *Computer-Aided Civil and Infrastructure Engineering*, vol. 22, no. 5, pp. 376-387.

Appendix 1 SQL Code for Identifying Complete Journeys

```
/***Oracle SQL Developer code to create complete journey data from Oyster single day sequenced journeys text file.
    Written by Catherine Seaborn, Massachusetts Institute of Technology. Compiled 14 May 2008***/

/*STEP 1: Import sequenced journeys text file*/

/*a. Create table for sequenced journeys text file*/

CREATE TABLE Journeys
 (DAYKEY INT NOT NULL,
 PID_ENCRYPT INT NOT NULL,
 SEQUENCENO  INT NOT NULL,
 SUBSYSTEMID INT NOT NULL,
 STARTLOC  INT NOT NULL,
 ENDLOC  INT NOT NULL,
 ROUTE_INNERZONE INT,
 ROUTE_OUTERZONE INT,
 ROUTE_DISTANCE INT NOT NULL,
 TB_CTEN INT NOT NULL,
 TB_CTEX INT NOT NULL,
 ZVPPT VARCHAR2 (4000 BYTE) NOT NULL,
 JNYTYP VARCHAR2 (4000 BYTE) NOT NULL,
 DAILYCAPPINGFLAG VARCHAR2 (4000 BYTE) NOT NULL,
 CAPPINGSCHEME VARCHAR2 (4000 BYTE) NOT NULL,
 FULLFARE INT NOT NULL,
 DISCOUNTEDFARE INT NOT NULL,
 PPTPRODUCTCODEKEY INT NOT NULL,
 PPTTIMEVALIDITYKEY INT NOT NULL,
 PPT1VALIDITYPAGECRC INT NOT NULL,
 PPT2VALIDITYPAGECRC INT NOT NULL,
 PPT3VALIDITYPAGECRC INT NOT NULL);

/*b. Use sqlldr or other tool to import sequenced journeys text file to Oracle table*/

/*STEP 2: Create new table to identify next journey stage for each record*/

/*a. Identify duplicate records*/

CREATE VIEW IdSeqno AS
 SELECT pid_encrypt, sequenceno, COUNT(*) AS "duplicates"
 FROM Journeys
 GROUP BY pid_encrypt, sequenceno
 HAVING COUNT(*) = 1;

/*b. Create table to identify next journey stage for each record (no duplicates)*/

CREATE TABLE JrnyNext
 ( pid_encrypt INT NOT NULL,
   sequenceno INT NOT NULL,
   nextseqno INT,
   startseqno INT NOT NULL,
   stage INT NOT NULL );

INSERT INTO JrnyNext
 SELECT j.pid_encrypt, j.sequenceno, MIN( n.sequenceno ) AS "nextseqno", j.sequenceno AS "startseqno", 1
```

```
  FROM IdSeqno j LEFT JOIN IdSeqno n ON ( j.sequenceno < n.sequenceno ) AND ( j.pid_encrypt = n.pid_encrypt )
  GROUP BY j.pid_encrypt, j.sequenceno;

CREATE INDEX JrnyNext_idx ON JrnyNext( pid_encrypt, sequenceno );

/*loop 1*/
DROP TABLE Stages_tmp;

CREATE TABLE Stages_tmp
  ( pid_encrypt INT NOT NULL,
    sequenceno INT NOT NULL,
    nextseqno INT,
    startseqno INT,
    stage INT );

INSERT INTO Stages_tmp
  SELECT j.pid_encrypt, j.sequenceno, j.nextseqno, n.startseqno, ( n.stage + 1 ) AS "stage"
  FROM JrnyNext j LEFT JOIN JrnyNext n
    ON ( j.sequenceno = n.nextseqno ) AND ( j.pid_encrypt = n.pid_encrypt ) AND ( n.stage = 1 );

CREATE INDEX Stages_tmp_idx ON Stages_tmp ( pid_encrypt, sequenceno );

UPDATE JrnyNext j
  SET startseqno = ( SELECT s.startseqno
                FROM Stages_tmp s
                WHERE ( s.pid_encrypt = j.pid_encrypt ) AND ( s.sequenceno = j.sequenceno )
                      AND ( s.startseqno IS NOT NULL ) )
  WHERE EXISTS
    ( SELECT s.startseqno
      FROM Stages_tmp s
      WHERE ( s.pid_encrypt = j.pid_encrypt ) AND ( s.sequenceno = j.sequenceno )
           AND ( s.startseqno IS NOT NULL) );

UPDATE JrnyNext j
  SET stage = ( SELECT s.stage
                FROM Stages_tmp s
                WHERE ( s.pid_encrypt = j.pid_encrypt ) AND ( s.sequenceno = j.sequenceno )
                      AND ( s.stage IS NOT NULL ) )
  WHERE EXISTS
    ( SELECT s.stage
      FROM Stages_tmp s
      WHERE ( s.pid_encrypt = j.pid_encrypt ) AND ( s.sequenceno = j.sequenceno ) AND ( s.stage IS NOT NULL ) );
.
.
.
/*loop 10*/
DROP TABLE Stages_tmp;

CREATE TABLE Stages_tmp
  ( pid_encrypt INT NOT NULL,
    sequenceno INT NOT NULL,
    nextseqno INT,
    startseqno INT,
    stage INT );

INSERT INTO Stages_tmp
```

```
 SELECT j.pid_encrypt, j.sequenceno, j.nextseqno, n.startseqno, ( n.stage + 1 ) AS "stage"
 FROM JrnyNext j LEFT JOIN JrnyNext n
  ON ( j.sequenceno = n.nextseqno ) AND ( j.pid_encrypt = n.pid_encrypt ) AND ( n.stage = 10 );

CREATE INDEX Stages_tmp_idx ON Stages_tmp ( pid_encrypt, sequenceno );

UPDATE JrnyNext j
 SET startseqno = ( SELECT s.startseqno
           FROM Stages_tmp s
           WHERE ( s.pid_encrypt = j.pid_encrypt ) AND ( s.sequenceno = j.sequenceno )
                AND ( s.startseqno IS NOT NULL ) )
 WHERE EXISTS
  ( SELECT s.startseqno
   FROM Stages_tmp s
   WHERE ( s.pid_encrypt = j.pid_encrypt ) AND ( s.sequenceno = j.sequenceno )
        AND ( s.startseqno IS NOT NULL ) );

UPDATE JrnyNext j
 SET stage = ( SELECT s.stage
           FROM Stages_tmp s
           WHERE ( s.pid_encrypt = j.pid_encrypt ) AND ( s.sequenceno = j.sequenceno )
                AND ( s.stage IS NOT NULL ) )
 WHERE EXISTS
  ( SELECT s.stage
   FROM Stages_tmp s
   WHERE ( s.pid_encrypt = j.pid_encrypt ) AND ( s.sequenceno = j.sequenceno ) AND ( s.stage IS NOT NULL ) );

/*STEP 3: Flag journey stages that should be linked for each card based on elapsed time thresholds*/

/*a. Create complete journey data*/

CREATE TABLE jnTrip
 ( pid_encrypt INT NOT NULL,
  sequenceno INT NOT NULL,
  nextseqno INT,
  startseqno INT,
  stage INT );

/*b. Remove stages greater than 10*/

INSERT INTO jnTrip
SELECT * FROM JrnyNext
WHERE stage < 11;

CREATE INDEX jnTrip_idx ON jnTrip (pid_encrypt, sequenceno, nextseqno );

/*c. Remove cards with journeys on modes other than bus or Underground*/

CREATE TABLE bad_ids
 ( pid_encrypt INT NOT NULL);

INSERT INTO bad_ids
 ( SELECT DISTINCT pid_encrypt
  FROM Journeys
  WHERE subsystemid = 3 OR subsystemid = 4 OR subsystemid = 5 OR subsystemid = 256 OR subsystemid = 257
     OR subsystemid = 258 OR subsystemid = 260 OR subsystemid = 261 );
```

```
CREATE INDEX bad_ids_idx ON bad_ids (pid_encrypt);

DELETE FROM jnTrip j
WHERE EXISTS
 (SELECT b.pid_encrypt FROM bad_ids b WHERE b.pid_encrypt = j.pid_encrypt);

/*d. Flag underground-to-bus links*/

ALTER TABLE jnTrip
ADD ( links INT );

UPDATE jnTrip f
SET links = (SELECT ub.links FROM
 (SELECT x.pid_encrypt, x.sequenceno, decode (trunc( (jn."enter2" - x.tb_ctex)/21 ), 0, 1 )links
 FROM Journeys x,
   ( SELECT n.pid_encrypt, n.sequenceno, j.subsystemid AS "mode2", j.tb_cten AS "enter2"
     FROM jnTrip n LEFT JOIN Journeys j
     ON j.pid_encrypt = n.pid_encrypt AND j.sequenceno = n.nextseqno) jn
 WHERE x.pid_encrypt = jn.pid_encrypt AND x.sequenceno = jn.sequenceno AND x.subsystemid = 0
        AND jn."mode2" = 1 ) ub
WHERE f.pid_encrypt = ub.pid_encrypt AND f.sequenceno = ub.sequenceno AND ub.links IS NOT NULL )

WHERE EXISTS
(SELECT ub.links FROM
 (SELECT x.pid_encrypt, x.sequenceno, decode (trunc( (jn."enter2" - x.tb_ctex)/21 ), 0, 1 )links
 FROM Journeys x,
   ( SELECT n.pid_encrypt, n.sequenceno, j.subsystemid AS "mode2", j.tb_cten AS "enter2"
     FROM jnTrip n LEFT JOIN Journeys j
     ON j.pid_encrypt = n.pid_encrypt AND j.sequenceno = n.nextseqno) jn
 WHERE x.pid_encrypt = jn.pid_encrypt AND x.sequenceno = jn.sequenceno AND x.subsystemid = 0
        AND jn."mode2" = 1 ) ub
WHERE f.pid_encrypt = ub.pid_encrypt AND f.sequenceno = ub.sequenceno AND ub.links IS NOT NULL );

/*e. Flag bus-to-underground links*/

UPDATE jnTrip f
SET links = (SELECT bu.links FROM
 (SELECT x.pid_encrypt, x.sequenceno, decode (trunc( (jn."enter2" - x.tb_cten)/41 ), 0, 1 )links
  FROM Journeys x,
   ( SELECT n.pid_encrypt, n.sequenceno, j.subsystemid AS "mode2", j.tb_cten AS "enter2"
     FROM jnTrip n LEFT JOIN Journeys j
     ON j.pid_encrypt = n.pid_encrypt AND j.sequenceno = n.nextseqno) jn
 WHERE x.pid_encrypt = jn.pid_encrypt AND x.sequenceno = jn.sequenceno AND x.subsystemid = 1
        AND jn."mode2" = 0 ) bu
WHERE f.pid_encrypt = bu.pid_encrypt AND f.sequenceno = bu.sequenceno AND bu.links IS NOT NULL )

WHERE EXISTS
(SELECT bu.links FROM
 (SELECT x.pid_encrypt, x.sequenceno, decode (trunc( (jn."enter2" - x.tb_cten)/41 ), 0, 1 )links
 FROM Journeys x,
   ( SELECT n.pid_encrypt, n.sequenceno, j.subsystemid AS "mode2", j.tb_cten AS "enter2"
     FROM jnTrip n LEFT JOIN Journeys j
     ON j.pid_encrypt = n.pid_encrypt AND j.sequenceno = n.nextseqno) jn
 WHERE x.pid_encrypt = jn.pid_encrypt AND x.sequenceno = jn.sequenceno AND x.subsystemid = 1
        AND jn."mode2" = 0 ) bu
```

```
WHERE f.pid_encrypt = bu.pid_encrypt AND f.sequenceno = bu.sequenceno AND bu.links IS NOT NULL );

/*f. Flag bus-to-bus links with spatial constraints (i.e., no return journeys on same route)*/

UPDATE jnTrip f
SET links = (SELECT bb.links FROM
 (SELECT xz.pid_encrypt, xz.sequenceno, decode (trunc( (jn."enter2" - xz.tb_cten)/51 ), 0, 1 )links
  FROM
   (SELECT x.pid_encrypt, x.sequenceno, x.subsystemid, x.tb_cten, z.routeid
     FROM Journeys x, Farestages z
     WHERE x.startloc = z.farestagekey ) xz,
   ( SELECT n.pid_encrypt, n.sequenceno, js.subsystemid AS "mode2", js.tb_cten AS "enter2", js.routeid AS "route2"
     FROM jnTrip n LEFT JOIN
      ( SELECT j.pid_encrypt, j.sequenceno, j.subsystemid, j.tb_cten, s.routeid
        FROM Journeys j, Farestages s
        WHERE j.startloc = s.farestagekey ) js
     ON js.pid_encrypt = n.pid_encrypt AND js.sequenceno = n.nextseqno) jn
 WHERE xz.pid_encrypt = jn.pid_encrypt AND xz.sequenceno = jn.sequenceno AND xz.subsystemid = 1
        AND jn."mode2" = 1 AND xz.routeid NOT IN jn."route2" ) bb
WHERE f.pid_encrypt = bb.pid_encrypt AND f.sequenceno = bb.sequenceno AND bb.links IS NOT NULL )

WHERE EXISTS
( SELECT bb.links FROM
 (SELECT xz.pid_encrypt, xz.sequenceno, decode (trunc( (jn."enter2" - xz.tb_cten)/51 ), 0, 1 )links
  FROM
   (SELECT x.pid_encrypt, x.sequenceno, x.subsystemid, x.tb_cten, z.routeid
     FROM Journeys x, Farestages z
     WHERE x.startloc = z.farestagekey ) xz,
   ( SELECT n.pid_encrypt, n.sequenceno, js.subsystemid AS "mode2", js.tb_cten AS "enter2", js.routeid AS "route2"
     FROM jnTrip n LEFT JOIN
      ( SELECT j.pid_encrypt, j.sequenceno, j.subsystemid, j.tb_cten, s.routeid
        FROM Journeys j, Farestages s
        WHERE j.startloc = s.farestagekey ) js
     ON js.pid_encrypt = n.pid_encrypt AND js.sequenceno = n.nextseqno) jn
 WHERE xz.pid_encrypt = jn.pid_encrypt AND xz.sequenceno = jn.sequenceno AND xz.subsystemid = 1
        AND jn."mode2" = 1 AND xz.routeid NOT IN jn."route2" ) bb
WHERE f.pid_encrypt = bb.pid_encrypt AND f.sequenceno = bb.sequenceno AND bb.links IS NOT NULL );

/*STEP 4: Form complete journeys*/

/*a. Create field to label journeys*/

ALTER TABLE jnTrip
ADD ( tripno INT );

/*b. Use the following rules to form complete journeys:
        i. IF stage = 1 THEN trip = 1
        ii. IF stage = 1 AND links = 1 THEN tripnext = trip ELSE tripnext = (MAX(trip) + 1)
        iii. IF stage = 2 AND links = 1 THEN tripnext = trip ELSE tripnext = (MAX(trip) + 1)
        etc.
*/

UPDATE jnTrip
SET tripno = 1
WHERE stage = 1;
```

```
/*loop 1*/
UPDATE jnTrip t
SET t.tripno = (SELECT l.tripno FROM
( SELECT n.pid_encrypt, n.sequenceno, j.tripno FROM jnTrip j, jnTrip n
WHERE j.stage = 1 AND j.links = 1 AND j.pid_encrypt = n.pid_encrypt AND j.nextseqno = n.sequenceno ) l
WHERE t.pid_encrypt = l.pid_encrypt AND t.sequenceno = l.sequenceno AND t.stage = 2)
        WHERE EXISTS
        (SELECT l.tripno FROM
        ( SELECT n.pid_encrypt, n.sequenceno, j.tripno FROM jnTrip j, jnTrip n
        WHERE j.stage = 1 AND j.links = 1 AND j.pid_encrypt = n.pid_encrypt AND j.nextseqno = n.sequenceno ) l
        WHERE t.pid_encrypt = l.pid_encrypt AND t.sequenceno = l.sequenceno AND t.stage = 2);

UPDATE jnTrip t
SET t.tripno = (SELECT (l.tripno + 1) FROM
( SELECT n.pid_encrypt, n.sequenceno, j.tripno FROM jnTrip j, jnTrip n
WHERE j.stage = 1 AND j.links IS NULL AND j.pid_encrypt = n.pid_encrypt AND j.nextseqno = n.sequenceno ) l
WHERE t.pid_encrypt = l.pid_encrypt AND t.sequenceno = l.sequenceno AND t.stage = 2)
        WHERE EXISTS
        (SELECT l.tripno FROM
        ( SELECT n.pid_encrypt, n.sequenceno, j.tripno FROM jnTrip j, jnTrip n
        WHERE j.stage = 1 AND j.links IS NULL AND j.pid_encrypt = n.pid_encrypt
                AND j.nextseqno = n.sequenceno ) l
        WHERE t.pid_encrypt = l.pid_encrypt AND t.sequenceno = l.sequenceno AND t.stage = 2);
.
.
.
/*loop 9*/
UPDATE jnTrip t
SET t.tripno = (SELECT l.tripno FROM
( SELECT n.pid_encrypt, n.sequenceno, j.tripno FROM jnTrip j, jnTrip n
WHERE j.stage = 9 AND j.links = 1 AND j.pid_encrypt = n.pid_encrypt AND j.nextseqno = n.sequenceno ) l
WHERE t.pid_encrypt = l.pid_encrypt AND t.sequenceno = l.sequenceno AND t.stage = 10)
        WHERE EXISTS
        (SELECT l.tripno FROM
        ( SELECT n.pid_encrypt, n.sequenceno, j.tripno FROM jnTrip j, jnTrip n
        WHERE j.stage = 9 AND j.links = 1 AND j.pid_encrypt = n.pid_encrypt AND j.nextseqno = n.sequenceno ) l
        WHERE t.pid_encrypt = l.pid_encrypt AND t.sequenceno = l.sequenceno AND t.stage = 10);

UPDATE jnTrip t
SET t.tripno = (SELECT (l.tripno + 1) FROM
( SELECT n.pid_encrypt, n.sequenceno, j.tripno FROM jnTrip j, jnTrip n
WHERE j.stage = 9 AND j.links IS NULL AND j.pid_encrypt = n.pid_encrypt AND j.nextseqno = n.sequenceno ) l
WHERE t.pid_encrypt = l.pid_encrypt AND t.sequenceno = l.sequenceno AND t.stage = 10)
        WHERE EXISTS
        (SELECT l.tripno FROM
        ( SELECT n.pid_encrypt, n.sequenceno, j.tripno FROM jnTrip j, jnTrip n
        WHERE j.stage = 9 AND j.links IS NULL AND j.pid_encrypt = n.pid_encrypt
                AND j.nextseqno = n.sequenceno ) l
        WHERE t.pid_encrypt = l.pid_encrypt AND t.sequenceno = l.sequenceno AND t.stage = 10);

/*END*/
```

Appendix 2 List of Routes Missing from Analysis

Route No.	Start Date
332	10/1/2007
347	11/1/2004
385	7/5/2007
427	4/5/2006
435	6/25/2005
452	12/2/2006
481	5/6/2006
498	12/26/2005
605	9/1/2007
608	9/4/2006
628	4/16/2007
636	4/18/2006
637	9/1/2007
639	9/1/2007
641	10/1/2005
650	12/31/2005
658	9/6/2006
687	1/27/2007
692	4/14/2007
696	1/20/2007
15H	new
212D	3/5/2005
606D	
9H	new
K50	
N102	9/1/2007
N128	10/6/2007
N188	7/1/2007
N205	4/4/2007
N220	10/21/2005
N344	new
N472	9/29/2007
N474	11/3/2007
N57	new
N87	new
NC2	C2 night service
PR2	
T130	130 tram fare
T314	314 tram fare
X26	new