Bus Passenger Origin-Destination Matrix Estimation
Using Automated Data Collection Systems

by

Alex Cui

B.S. in Civil Engineering
University of California at Berkeley, 1996

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

at the
Massachusetts Institute of Technology

September, 2006

© 2006 Massachusetts Institute of Technology.
All rights reserved.

Signature of Author

Department of Civil and Environmental Engineering
August 11, 2006

Certified by

Nigel H.M. Wilson
Professor of Civil and Environmental Engineering
Chair, Master of Science in Transportation Program
Thesis Supervisor

Certified by

John Attanucci
Research Associate
Center for Transportation and Logistics
Thesis Supervisor

Accepted by

Andrew Whittle
Chair, Departmental Committee on Graduate Students
Abstract

Automatic Data Collection (ADC) systems can enhance the ability of transit agencies to obtain useful planning information that was previously too expensive to obtain. This thesis documents the development of an algorithm to estimate a Bus Passenger Trip Origin-Destination Matrix (OD Matrix) based on ADC system archived data including Automated Fare Collection (AFC) data, Automatic Passenger Count (APC) data, and Automatic Vehicle Location (AVL) data. This algorithm consists of three steps: data preparation, estimation of a Single Route OD Matrices for all routes, and estimation of a Network Level OD Matrix using transfer flow information. The single route OD matrix estimation requires the development of a "seed" matrix derived primarily from AFC data and the "marginal" control totals, i.e. the boarding and alighting counts, derived primarily from APC data. Both Iterative Proportional Fitting (IPF) and Maximum Likelihood Estimation (MLE) techniques are used to estimate the single route OD matrices based on seed matrices and marginal values, and the results from these two techniques are compared. For the network level OD matrix estimation, we estimate the transfer flows by considering the consecutive transactions from AFC data. The resulting network level OD matrix is provided at a route segment level (3-5 combined stops) of detail. This OD estimation algorithm is illustrated by its application to a selected corridor of the Chicago Transit Authority (CTA) bus network.

This OD estimation algorithm can be practically applied to a full-size bus network. It is recommended that the MLE method be used to estimate the single route OD matrices and the proportional distribution method be used to estimate the transfer flow OD matrix.

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

Thesis Supervisor: John P. Attanucci
Title: Research Associate, Department of Civil and Environmental Engineering
Acknowledgment

This thesis could not have been completed without the help of many people, only a few of whom are listed below.

I am especially grateful to my advisor and mentor Professor Nigel Wilson who has always inspired me to do my best throughout my two years at MIT. I learned a great deal from his wisdom and experiences, and his advices are always candid. He has provided me with some of the most stimulating educational experiences at MIT.

I would like to thank John Attanucci, my advisor, whose experience and intelligence provided crucial guidance to my research. He has always pushed me to think very hard and not to forget common sense. The weekly meetings with Nigel and John were the place where many of the ideas in this research were formed.

Thanks to Professor Haris Koutsopoulos for examining the theoretical foundation of my research and for the comments on my thesis draft. Without your expertise and our numerous meetings and discussions this thesis would not be truly complete.

Thanks to Professor Peter Furth for showing me the application of MLE in OD estimation. Our numerous meetings and discussions and your comments have proved invaluable for my research.

Also thanks to Mikel Murga and Professor Moshe Ben-Akiva for spending time to discuss my research.

Ginny Siggia is so kind and helpful, and I really appreciate her efforts in making everything as easy for us as possible.

And of course I am truly grateful to have Adam Rahbee as my supervisor during my internship at CTA. My experiences during the internship and his advices on my research provided the critical foundation of this thesis. I am also grateful to Jinhua Zhao whose help and advices got me started on the OD estimation project. He also encouraged me to choose this research topic.

Kevin O'Malley and Mike Haynes of CTA provided me with the data without which this research would be impossible. Jeff Sriver and Jeff Busby of CTA have provided encouragements and helps to me for which I am grateful. I also appreciate the efforts by Paul Gross and Kevin O'Malley to get me hired by CTA, and I look forward to working with all of you people at CTA.

This research is generously support, in part, by the Chicago Transit Authority, Transport for London, and Massachusetts Bay Transportation Authority.

Thanks for the support from my fellow students, Gunwoo Lee, Chris Pangilinan, Saumya Gupta, Guo Zhan, Tzu Ching Horng, Kenichi Hodota, Yayoi Uchiyama, Hong Yi, Grace Fattouche, Nobuhiro Iwamura and many others.

I am grateful to my family without whose support I would not be able to get this far. I would like to thank my parents Tian-yi Cui and Li Zhang, my brother Kevin, and my parents-in-law Li Zhenyun and Huang Chongfang. Finally, I am grateful to the most important person in my life, my beautiful wife Qiqi whose support and effort made the completion of this thesis possible.
Table of Content

Abstract ........................................................................................................... 3
Acknowledgement ............................................................................................ 5
Table of Content ............................................................................................... 7
List of Tables ..................................................................................................... 11
List of Figures ................................................................................................... 13
Chapter 1: Introduction ...................................................................................... 15
  1.1 Potential Benefits of an OD Matrix Using ADC ................................. 15
    1.1.1 Reduced cost .................................................................................. 16
    1.1.2 Larger and less biased sample ......................................................... 17
    1.1.3 More frequent estimation .............................................................. 17
    1.1.4 Allow more targeted surveys ......................................................... 18
  1.2 Automatic Data Collection Systems ....................................................... 18
    1.2.1 Automatic Fare Collection (AFC) system ..................................... 19
    1.2.2 Automatic Vehicle Location (AVL) system .................................. 20
    1.2.3 Automatic Passenger Count (APQ) system .................................. 20
  1.3 Research Objectives .................................................................................. 21
  1.4 Thesis Organization ................................................................................... 22
Chapter 2: Literature Review ............................................................................ 23
  2.1 OD Estimation Techniques Using Manual Data ..................................... 23
    2.1.1 Iterative proportional fitting (IPF) .................................................. 24
    2.1.2 Maximum likelihood estimation (MLE) .......................................... 25
  2.2 OD Estimation Using ADC Data Sources ............................................. 27
    2.2.1 Entry-only rail system destination inference .................................. 28
    2.2.2 Bus OD estimation without AVL data ........................................... 29
Chapter 3: General Approach to Estimating Bus OD Matrix Using ADC ........ 31
  3.1 Data Preparation ....................................................................................... 32
  3.2 Single Route OD Matrix Estimation ....................................................... 33
    3.2.1 Iterative proportional fitting (IPF) .................................................. 34
    3.2.2 Maximum likelihood estimation (MLE) .......................................... 34
  3.3 Network Level Bus Passenger Linked Trip OD Matrix Estimation ....... 36
    3.3.1 Total transfer flows ....................................................................... 38
    3.3.2 Obtain the origin or destination distribution of the transfer flow ...... 39
      A. First leg of the transfer .................................................................. 39
      B. Second leg of the transfer ............................................................ 42
    3.3.3 Non-transfer flow OD matrices ..................................................... 43
    3.3.4 Transfer flow OD matrices .......................................................... 43
Chapter 4: Introduction to CTA Case Study .................................................... 45
  4.1 CTA System ............................................................................................. 45
  4.2 CTA ADC Systems ................................................................................... 47
    4.2.1 Automatic Fare Collection (AFC) system ..................................... 47
      (a) Data fields .................................................................................... 48
      (b) Location information .................................................................... 49
      (c) Transfer card data ........................................................................ 50
    4.2.2 Automatic Vehicle Location (AVL) system .................................. 50
    4.2.3 Automatic Passenger Count (APQ) system .................................. 51
      (a) Data description ........................................................................... 51
4.3 Differences of Available Data between CTA Rail and Bus Systems

Chapter 5: CTA Case Study Data Preparation

5.1 Segment Definition

5.2 Boarding Stop Identification for AFC Trips

5.2.1 Look up bus number

5.2.2 Relationship between AFC and AVL records

5.2.3 Boarding stop identification

5.2.4 Identifying boarding stop based on transfer information

5.3 Alighting Stop Inference

5.3.1 Trip-chaining method

5.3.2 Destination reassignment

5.3.3 Seed matrix formation

5.4 APC Data Cleaning

5.5 APC Data Scaling

Chapter 6: CTA Single Bus Route OD Matrix Estimation

6.1 Iterative Proportional Fitting (IPF)

6.2 Maximum Likelihood Estimation (MLE)

6.3 Comparison of the Estimation Results

Chapter 7: Network Level OD Matrix

7.1 Demonstration Corridor

7.2 Transfer Flow Totals

7.2.1 Transfer trips in farecard data

7.2.2 Transfer trips in transfer card data

7.2.3 Transfer flow total

7.3 Proportional Distribution

7.4 Modified IPF

7.4.1 Modifying seed matrix

7.4.2 Modifying boarding and alighting totals

7.4.3 Modified IPF results

7.4.4 Comparison with the results from proportional distribution

7.5 Transfer Flow OD Matrices

Chapter 8: Summary and Recommendations

8.1 Further Discussion of Key Results

8.1.1 Implication of network level OD estimation

8.1.2 Discussion of Potential Bias

8.2 Recommendations

8.2.1 Advantages of MLE

8.2.2 When to use modified IPF for transfer flows

8.3 Potential Future research

8.3.1 Bus/Rail network OD estimation

8.3.2 Validation design

8.3.3 Integrating smart card data to enhance the OD matrix estimation and interpretation

8.3.4 Using MLE technique at network level

8.3.5 Applying methodology to other transit agencies
References ................................................................. 93
Appendix
   A. Single route estimation results for Routes # 20, # 49, # X49, # 126 of CTA …… 95
   B. Transfer flow OD matrices between Routes # 20, # 49, # X49, # 126 of CTA … 103
   C. SQL code used in data preparation .................................................. 107
   D. Main Software Tools Used ........................................................... 134
List of Tables

Table 4-1 Sample AFC Data ................................................................. 49
Table 4-2 Sample Transfer Card Data .................................................. 50
Table 4-3 Sample APC Data ................................................................. 52
Table 4-4 Data Differences between CTA Rail and Bus .......................... 53
Table 5-1 Sample Equipment ID to Bus Number Look Up Table .................. 57
Table 5-2 Seed Matrix for Route 20 Eastbound .................................... 63
Table 5-3 Seed Matrix and Marginal Values for Route 20 Eastbound, 6-9AM ................................................................. 66
Table 6-1 IPF OD Matrix for Route 20 Eastbound 6-9AM .......................... 68
Table 6-2 Unscaled APC Boarding and Alighting Counts .......................... 69
Table 6-3 Estimated Passenger Flows from MLE .................................... 70
Table 6-4 Estimated Seed Matrix Sampling Rates .................................... 70
Table 6-5 Comparison of MLE and IPF Results ...................................... 71
Table 6-6 Route # 20 OD Matrix Comparison with Prior Estimates ............ 73
Table 7-1 Transfer Flow Totals by Route-dir ......................................... 79
Table 7-2 Single Route OD Matrix for Route # 49 Northbound .................... 80
Table 7-3 Total OD Flows Ending at the Transfer Point ............................ 81
Table 7-4 Total OD Flows Originating from the Transfer Point ................... 81
Table 7-5 Transfer-only Flows with Origins Distributed ............................ 81
Table 7-6 Transfer-only Flows with Destinations Distributed .................... 82
Table 7-7 Modified Route # 20 Seed Matrix with Pseudo-columns and Pseudo-rows for Transfers to Route # 49 .................................................. 83
Table 7-8 Boarding and Alighting Counts with Pseudo-entries ................... 84
Table 7-9 Modified IPF Method with Pseudo Columns and Rows .................. 84
Table 7-10 Transfer Flow Results Comparison ........................................ 85
Table 7-11 Transfer Flow OD Matrix ....................................................... 86
List of Figures

Figure 2-1 Trip-chaining Using CTA Example for a Single Passenger and Day ...... 29
Figure 3-1 Network Level OD Matrix ......................................................... 38
Figure 3-2 Example Schematic Network ....................................................... 39
Figure 3-3 Example Partial Single Route Matrices ......................................... 40
Figure 3-4 Illustration of Pseudo-stop and Modified IPF ................................. 41
Figure 3-5 Example Network ........................................................................ 42
Figure 3-6 Transfer Flow Only OD Matrix ..................................................... 44
Figure 4-1 CTA System Map ......................................................................... 46
Figure 5-1 Segment Definition ....................................................................... 55
Figure 5-2 Distribution of Time between AFC Transaction and Matched AVL Records ................................................................. 58
Figure 5-4 Destination Reassignment Examples ............................................. 61
Figure 5-5 Relationship between Seed Matrix and Original Data ...................... 64
Figure 7-1 Demonstration Corridor ................................................................ 75
Figure 7-2 Transfer Flows from Route #20 East to Routes #49/X49 North ........... 88
Chapter 1

Introduction

Over the past few decades, there have been rapid technological advances in automatic data acquisition and data processing resulting in large archives of transactional data. Many industries have undergone fundamental transformations in how they do business after careful examination of the impact of technological changes and implementation of processes that attempt to make the most of the new opportunities. However, the public transit industry has been lagging behind many other industries in embracing the full potential of the recent information technology revolution. This thesis proposes a methodology to estimate a bus passenger Origin-Destination matrix ("OD matrix") from the data obtained from various on-vehicle Automated Data Collection (ADC) systems. This process is scalable and suitable to implement as an automatic process in a production environment. The OD matrix thus derived can provide a firm basis for decision making by various departments within a transit agency. With the availability of an OD matrix generated from ADC systems, the traditional bus passenger on-board survey can be transformed to become a tool to validate and calibrate the automated OD matrix estimation model. It should be focused more on obtaining intensive individual travel data rather than extensive system-wide data considering its relatively high cost but its ability to capture socio-economic, trip purpose and customer perception information.

1.1 Potential Benefits of an OD Matrix Using ADC

A transit passenger flow OD matrix has wide application in a transit agency. It can be used in service planning, in operations analysis, before-and-after impact analysis, and in service management. For example, when considering potential express bus service, we need to know the number of passengers traveling between various stops to estimate total travel time and the potential passenger travel time savings.
The traditional way a transit agency would obtain an OD matrix is as a by-product of occasional on-board passenger surveys and using various techniques to expand the survey results based on manual boarding and alighting counts at the stops. Such passenger surveys and counts are expensive to conduct and thus are extremely infrequent. In practice this means that OD matrices are produced only when a system-level passenger survey is being conducted, typically for some other primary purpose than origin-destination estimation itself. (Ben-Akiva and Morikawa 1989)

A reasonably current system-wide or even single route OD matrix is usually not available in most transit agencies. The primary advantage of using ADC data is to make these OD matrices practically available to a transit agency. The following advantages over using manual data are not merely differences in degree but also differences in kind including:

1. The cost of obtaining an OD matrix is significantly reduced.
2. The resulting matrix is based on a significantly larger sample size.
3. The process is more suitable for automation which will make the process much faster and therefore able to be updated more frequently.
4. This process can be combined with more targeted surveys to obtain a more cost effective and comprehensive picture of passenger travel behavior.

These benefits of using ADC to estimate an OD matrix are discussed further below.

1.1. Reduced cost

The use of data from ADC systems to estimate an OD matrix can give similar information at significantly lower costs compared to the more conventional passenger survey method of estimating an OD matrix. The cost savings are mainly because most of the equipment is purchased primarily for other purposes (with the exception of Automatic Passenger Counting or APC equipment) and these on-board systems are generally designed to collect and archive passenger transactions and vehicle events on all trips by equipped vehicles.
There may also be an comparatively small increase in the cost of data processing because:

1. The data from ADC systems are often not primarily designed to be used to generate an OD matrix, and additional processing is often necessary to make the data usable.

2. The data from ADC system are usually so extensive that only a full scale relational database management system (RDBMS) can properly handle the processing tasks. Therefore, special software licenses and more specialized technical skills are required to take full advantage of ADC system data.

Passenger surveys conducted for other planning or market research purposes can still be used in this process but OD estimation no longer depends on having new survey data.

1.1.2 Larger and less biased samples
The sample size of on-board passenger survey is always limited because of high survey costs. Estimating OD matrix using ADC data can take advantage of a much larger sample size as virtually all passenger fare transactions are recorded. The resulting dramatic increase in sample size should theoretically be free of the sampling error and have minimal bias if all systems are working perfectly together at all times. In any case, estimating an OD matrix using ADC data provides a much more representative raw data source which becomes the basis for further analysis.

1.1.3 More frequent estimation
Using ADC data to estimate an OD matrix will allow much more frequent estimation of the OD matrix for the whole system because of its low marginal cost. It could be applied several times a year or certainly before and after major system changes. Relying on passenger on-board surveys means that OD matrices are estimated very infrequently, typically every 5 to 10 years. (Barry 2002) Any changes in the interval between two surveys can only be interpolated, seasonal factors in the changes of OD patterns can not be captured, and analysis on possible OD changes due to service improvements requires the expense of an additional special survey.
On the other hand, the data from ADC are readily available in electronic form, and their collection is usually a continuous process without the need for human intervention. Therefore, the raw data is available at any time, and the OD matrix estimation process can be applied at any time of the year. The only limit on the length of period for OD matrix estimation is having a large enough set of data to obtain statistically robust results. Specifically, the quantity of data should be large enough such that there is at least one observed trip for any OD pair that should have passenger flows between them. Practically, there is unlikely to be any planning need for an updated OD matrix more than monthly or even quarterly during any given year.

1.1.4 Allow more targeted surveys

The advantage of using ADC data is its automatic generation and availability, the expansive coverage, and the low marginal cost to extend the time period of data collection while surveys have the advantage of being more flexible and can be more detailed. When the OD matrix results using ADC data identify certain areas of special interests, a targeted survey can then be conducted on a particular route, or part of a route, or just at a major transfer stop to get more detailed information that is unavailable from ADC systems.

Data from ADC sources and surveys can complement each other in another area: to maintain the credibility of the OD matrix estimation process, and to verify its underlying assumptions, some on-going monitoring program should be in place besides the initial validation.

1.2 Automatic Data Collection Systems

Over the past several decades, transit agencies have been gradually implementing technological innovations such as Automated Fare Collection (AFC), Automatic Vehicle Location (AVL) and Automatic Passenger Count (APC) systems. Sometimes data acquisition innovations such as APC are primarily designed to provide automatically collected data to support decision making at various levels within a transit agency, but often a large amount of automatically collected data is produced as a byproduct of another primary application (for example fare
payment). However, regardless of the original intention for which the data are collected, these data sources can provide information that the transit agency has been lacking, can provide information that is more accurate and complete, and can provide information much more quickly. Because these data sources cover different aspects of transit operations, they will be able to give more extensive and more accurate information if they are used together.

However, it is not always straightforward to use the acquired data for transit planning and operations analysis. The reliability of the data needs to be examined, bad data needs to be identified and removed, and reliable algorithms must be developed and tested to process the data into useful information. (Zhao 2004)

ADC systems tend to be purchased one at a time without adequate consideration to provide integrated and consistent data jointly with other ADC systems. This is partly because insufficient attention is paid to the data gathering functions of the systems and equipment, and often results in valuable data sitting in separate and isolated databases which is difficult to use and may not be used at all. (Zhao 2004)

The fact that different data sources are gathered independently provides the benefit that different data sources can be used to check and adjust the other sources. Data inconsistency will be more difficult to resolve if the data quality of different data sources is low.

1.2.1 Automatic Fare Collection (AFC) system

Automatic Fare Collection systems are gaining gradual acceptance in the US transit industry after becoming more widely used in other parts of the developed world. The primary purposes tend to be the reduction of costs by eliminating ticket and token clerks and the reduction of fare evasion by reducing the human element in fare collection. However, the installation of AFC systems can also yield valuable information after careful analysis. (Wile 2003) This thesis relies extensively on AFC data.

A transit agency can have a flat, distance based, or zone based fare. This generally governs whether the AFC system relies only on entry control or both entry and exit control. If the AFC system controls both entries and exits, passengers will create electronic transaction
records at both the entry and exit; whereas, the entry-only AFC systems will have records only of passengers entering the system. Therefore, although the destinations of passengers in the entry-exit control system is readily available from the AFC data, the destinations of passengers in the entry-only AFC system can only be inferred based on other information and assumptions. This thesis will focus on bus systems with only entry control, and will rely on destination inference.

Fare boxes onboard buses generally do not know the location of the bus, and thus bus AFC transaction records do not contain detailed stop location information for the transactions. This is in contrast to rail systems in which the stations where the fare transactions occur are fixed. If the bus is AVL equipped, however, the location of this bus can be determined for any AFC transaction even though this will entail additional data processing.

1.2.2 Automatic Vehicle Location (AVL) system

Automatic Vehicle Location (AVL) systems provide vehicle location information through one of the following methods:

1. Global Positioning System (GPS),
2. Odometer and gyroscope combined with predetermined route information,
3. Use radio between signposts at key locations, and
4. Sometimes combinations of these different methods to provide redundancy and ensure internal consistency. (Wilson 2006)

Knowing the location of vehicles is critical in estimating bus OD matrices.

AVL systems can be either connected via wireless communications to a central computer, allowing the transit agency to access real-time vehicle location data, or simply store the data in an onboard computer which are then uploaded to a central computer overnight at the garage. While real-time AVL can have more valuable operational applications, estimation of OD matrices is fundamentally an offline data mining application. In this respect, an offline AVL system can be just as effective. (Wilson 2006)
1.2.3 Automatic Passenger Count (APC) system

APC systems are primarily designed as data collection systems to provide data obtained previously through ride-checks or "on-off" counts. APC systems record the number of passengers boarding and alighting at each serviced stop together with the time and stop location.

There are different methods to detect passenger boardings or alightings. The most common method uses infrared light beams which when broken in sequence indicate a passenger boarding or alighting. This method requires that the sensors be properly aligned. Another method is to use pressure sensitive mats to detect passengers stepping into or out of the bus. This method will not work with low floor buses. There is also the method to use passive infrared sensors to detect passengers based on the difference of body heat versus the ambient temperature (Siemens VDO). Methods also exist to use video images taken inside the bus and count the number of passengers through image interpretation. (Wilson 2006)

1.3 Research Objectives

The ultimate goal of this research is to develop an algorithm to estimate a network-level bus passenger origin-destination matrix that can eventually be integrated with a rail OD model to create full inter-modal OD matrix. The immediate goal is to build a prototype algorithm to estimate a bus only OD matrix using raw data from Chicago Transit Authority (CTA). This algorithm will be network based, namely transfer flows will be included in the model, and the results will include all linked bus trips. The algorithm will be demonstrated through an application to the CTA bus network. This thesis intends to establish a process to estimate an OD matrix that can be automated for regular production use rather than just to obtain a set of OD matrices as a one-off exercise.

Estimating the rail system OD matrix is not described in this thesis. For detailed information on the process of estimating an entry-only rail system OD matrix, please see Zhao (2004) and Zhao, Wilson and Rahbee (2006).
1.4 Thesis Organization

This thesis is divided into eight chapters. The second chapter reviews related research. The third chapter presents an overall approach to estimating a bus OD matrix. Chapter 4 introduces the case study used in this thesis—the Chicago Transit Authority (CTA) bus network. Chapters 5 to 7 will describe in detail the OD matrix estimation application to the CTA bus network, with Chapter 5 describing data preparation, Chapter 6 discussing single bus route OD matrix estimation, and Chapter 7 discussing network level OD matrix estimation. Finally, Chapter 8 provides a summary of recommendations and suggests future research in this area.
Chapter 2

Literature Review

This literature review focuses on the problem of origin-destination (OD) estimation for public transport networks. Section 2.1 reviews transit OD estimation approaches based on traditional manual sources of data while Section 2.2 addresses the more recent research dealing with automatically collected data.

There is also an extensive literature on estimating origin-destination matrices for drivers on road networks (see for example Ianno 2002), but this is quite a different problem given the flexibility of drivers to select and change routes without the interchange penalty which characterize transit networks. For this reason this literature is not reviewed here.

2.1 OD Estimation Techniques Using Manual Data

Many research papers deal with various aspects of the estimation of transit passenger OD matrices. Most are focused either on surveys or on techniques to combine various sources of data to yield a passenger trip OD matrix in a cost-effective manner. The survey methodology per se is not the focus in this thesis, rather it is on the methods to combine different data sources, usually using a sample of passenger OD flows referred to as a seed matrix, and passenger boarding and alighting counts.

Ben-Akiva (1987) has listed various approaches that are either based on statistical methods with known properties, or based on more ad-hoc procedures. This thesis will focus on two specific approaches only, namely iterative proportional fitting (IPF) and maximum likelihood estimation (MLE).
2.1.1 Iterative proportional fitting (IPF)

The iterative proportional fitting (IPF) method is also known as Bregman's balancing method. Compared to other methods, this method is computationally easy (Navik and Furth 1994). The main difficulty with using the IPF technique is the probability that a significant number of potential OD pairs in a network have little or no travel between them. The matrix cells having zero OD flows due to the structure of the matrix are called structural zeros, and those due to the low sampling rate of low OD flows are called non-structural zeros or sampling zeros. (Mohri and Roark) If there are no structural zeros and if all data are consistent, IPF will yield a unique solution (Ben-Akiva 1987), but if the potential travel matrix has a lot of OD pairs with no observed trips, the IPF method will result in non-zero flows for only a portion of the OD pairs which do have flow.

The goal in IPF is to solve the following simultaneous equations:

\[ T_{ij} = a_i b_j t_{ij}, \forall i, j \] where \( T_{ij} \) is the total passenger flow between origin segment \( i \) and destination segment \( j \), and \( t_{ij} \) is the seed matrix flow between \( i \) and \( j \), \( a_i \geq 1 \) is the row factor, and \( b_j \geq 1 \) is the column factor. \hfill (2-1)

\[ \sum_i T_{ij} = M_j, \forall j \] where \( M_j \) is total alighting count for segment \( j \). \hfill (2-2)

\[ \sum_j T_{ij} = M_i, \forall i \] where \( M_i \) is total boarding count for segment \( i \). \hfill (2-3)

If \( a_i^k \) is the row factor at iteration \( k \), then the product of the row factors at all iterations will be the final row factor, \( a_i = \prod_k a_i^k \). Similarly, the product of the column factors at all iterations will be the final column factor. (Navick and Furth, 1994)

This method is usually used in the estimation of a single route OD matrix by combining a seed matrix with the total boarding and alighting counts. This method requires high quality total boarding and alighting counts as the control totals, and a small but representative seed matrix. The accuracy of the estimation depends on the reliability of the control totals. However, the
quality of the seed matrix is also very important to the final estimation result, with a highly biased seed matrix even worse than a null seed matrix or a gravity based trip distribution model. For further discussions please see Hsu (1985), Ben-Akiva (1987) and Navick and Furth (1994).

2.1.2 Maximum likelihood estimation (MLE)

This method is also referred to as Poisson Maximum Likelihood Estimator (PMLE) in Ben-Akiva 1987. The followings are assumptions are necessary for the formulation of PMLE:

1. Passengers arrive individually, and according to a Poisson process.
2. There is negligible bias in ride-check (APC) data.
3. Whether passengers choose to respond is a binomial process.
4. All passengers can board the first available bus.
5. The boarding and alighting counts are independent of each other.

Assumptions (1) to (3) are considered reasonable in most cases while the other assumptions will usually not be strictly satisfied, but Hsu has shown that the PMLE model is robust with respect to violation of these assumptions. He tested the model using simulation data under various degrees of independence between the boarding and alighting data, and found that it had no impact on the accuracy of the estimation. He also concludes that PMLE is preferred to the negative binomial MLE model which requires fewer assumptions but is more computationally complex. (Hsu 1985)

There are two types of independent observations:

1. The observation of a particular OD flow $Y_{ij}$: The probability of $Y_{ij}$ being observed is:

$$P(Y_{ij} = y_{ij} | \lambda_{ij}, P_{ij}) = e^{-\lambda_{ij} p_{ij}} \frac{\left(\lambda_{ij} p_{ij}\right)^{y_{ij}}}{Y_{ij}!}$$  (2-4)
where \( Y_{ij} \) is the passenger trip OD flow from origin stop i to destination stop j, \( \lambda_{ij} \) is the mean value of the Poisson process for passenger arrivals with origin i and destination j, and \( p_{ij} \) is the response rate (capture rate) or the percentage of passenger flows from i to j captured in the seed matrix.

(2) Observation of the boarding and alighting counts (\( x_{i*} \) and \( x_{*j} \))

We define \( x_{i*} = \sum_j y_{ij} \) as the total boarding at stop i, and \( x_{*j} = \sum_i y_{ij} \) as the total alighting at stop j. Therefore, \( x_{i*} \) and \( x_{*j} \) are also Poisson variables with the mean value being \( \lambda_{i*} = \sum_j \lambda_{ij} \) and \( \lambda_{*j} = \sum_i \lambda_{ij} \) respectively. So we have:

\[
P(x_{i*} | \lambda_{i*}) = e^{-\lambda_{i*}} \frac{(\lambda_{i*})^{x_{i*}}}{x_{i*}!},
\]

and

\[
P(x_{*j} | \lambda_{*j}) = e^{-\lambda_{*j}} \frac{(\lambda_{*j})^{x_{*j}}}{x_{*j}!}.
\]

The total probability or likelihood of observing all the data is:

\[
L = \prod_i \prod_j e^{-\lambda_{ij}} \frac{(\lambda_{ij})^{y_{ij}}}{y_{ij}!} \prod_i e^{-\lambda_{i*}} \frac{(\lambda_{i*})^{x_{i*}}}{x_{i*}!} \prod_j e^{-\lambda_{*j}} \frac{(\lambda_{*j})^{x_{*j}}}{x_{*j}!}
\]

(2-7)

Taking the natural log on both sides:

\[
LL = \sum_i \sum_j (-\lambda_{ij} p_{ij} + y_{ij} \ln \lambda_{ij} p_{ij}) + \sum_i (-\lambda_{i*} + x_{i*} \ln \lambda_{i*}) + \sum_j (-\lambda_{*j} + x_{*j} \ln \lambda_{*j}) + C
\]

(2-8)

Solve \( \max LL \), which is an unconstrained concave maximization problem.

Besides solving this problem numerically, we can also take derivatives on both sides over \( \lambda_{ij} \), and solve the following simultaneous equations:
\[
\frac{\partial LL}{\partial \lambda_{ij}} = -\lambda_{ij} + \frac{y_{ij}}{\lambda_{ij}} - 1 + \frac{x_{ij}}{\lambda_{ij}} - 1 + \frac{x_{ij}}{\lambda_{ij}} = 0, \forall i, j
\]  

(2-9)

(Based on Navick and Furth 1994, Ben-Akiva 1987, Ben-Akiva and Morikawa 1989, and Koutsopoulos 2006.)

MLE is very flexible and can be applied to estimate a single route OD matrix, a network level OD matrix, or transfer flows. It can be used to combine data from different sources with different reliability. However, PMLE does require large numbers of ride-check (or APC) observations so that the violation of the deterministic headway can be tolerated (Hsu 1985).

### 2.2 OD Estimation Using ADC Data Sources

With the increasing availability of ADC transit data, research has recently been completed using ADC data to estimate transit passenger OD matrix. Since different transit agencies have different data availability, and characteristics, these projects tend to have somewhat different foci and methodologies.

Differences exist between rail and bus systems, mainly because of the relative ease in determining origin station for the rail system and the fact that, for rail systems, it is easier to capture linked trips than unlinked trips. Therefore the methodology used to estimate the OD matrix for entry-only rail systems is quite different from the methodology proposed in this thesis to estimate the OD matrix for bus system. Rail systems with both entry and exit control provide origins and destinations from the farecard transaction data, and can provide a more accurate passenger trip OD matrix with simpler processing than an entry-only system. (Gordillo, 2006)

This section first reviews the trip-chaining method of destination inference used in entry-only rail system OD matrix estimation (Section 2.2.1). This method, with some adaptation, will be used to obtain the destination (or alighting) stop of a passenger trip recorded in farecard transaction data, which is one of the essential steps in obtaining a seed matrix. Section 2.2.2 will discuss another method used by Caliper Corporation to estimate bus system OD for New
York City Transit (NYCT). Besides these two cases, there is also work by Hoffman for Dublin bus using ticket data to estimate OD movements (O'Mahony).

2.2.1 Entry-only rail system destination inference

The fare collection systems for many older US rail networks including New York City Transit (NYCT) and Chicago Transit Authority (CTA) only have entry control, and the station where the passenger exits the system is not directly known from the available data. In order to infer the destination of a rail passenger, trip-chaining is used which is based on the following three major assumptions:

1. Travelers start their next trip at the destination station of their previous trip, or another station in close proximity.

2. Travelers end their last trip of the day at the station where they began their first trip of the day.

3. If a rail passenger uses a bus subsequently, the intersection between the bus and rail route indicates the rail trip destination. (Zhao 2004) (TransInfo 2004)

Therefore the destination of a rail system AFC trip is inferred to be the same as the origin station of the immediately subsequent transit trip (i.e. the "next trip" method). When the trip is the last trip of the day, the destination is inferred to be the origin station of the first trip of that day if this first trip is also a rail trip. This is referred to as the "last trip of the day" method. Special considerations are built into the destination inference methodology to use symmetric trip patterns and situations where several passengers share a single farecard. Further description of this prior research is found in Barry (2002), Rahbee (2002), Zhao (2004), and Zhao, Wilson and Rahbee (2006).

The next trip method and last trip of the day method together account for nearly 80% of the destinations inferred in the CTA study. These two methods are illustrated in Figure 2-1 which shows two rail trips by a single passenger, on the same day. We can infer the destination station of the first trip to be Lake/State which is the origin station of the next trip. We can also infer the destination of the second trip to be 87th Street which is the origin station of the
first trip of the day. The arrows indicate the passenger's change of location along this trip chain.

Figure 2-1 Trip-chaining Using CTA Example for a Single Passenger and Day

<table>
<thead>
<tr>
<th>Time</th>
<th>Mode</th>
<th>Rail Origin Station</th>
<th>Inferred Destination Station</th>
</tr>
</thead>
<tbody>
<tr>
<td>8:22:15</td>
<td>Red Line</td>
<td>87th (Red)</td>
<td>Lake/State (Red)</td>
</tr>
<tr>
<td>16:00:32</td>
<td>Red Line</td>
<td>Lake/State (Red)</td>
<td>87th (Red)</td>
</tr>
</tbody>
</table>

Since not all farecard trips can have their destinations successfully inferred, and also because farecards may not be used by all passengers, the resulting farecard trips with known origins and destinations need to be scaled to control totals such as total entry and exit counts to estimate a system-wide OD matrix for the rail system. The resulting OD matrix is at the network level for linked trips. The information contained is simply the linked entry and exit stations, without information about which line the passenger takes and where (s)he transfers. Therefore, for many applications which require unlinked trip information, there needs to be a further step to assign the OD flows to specific rail lines and transfer stations in the network.

2.2.2 Bus OD estimation without AVL data

The Automated Fare Collection (AFC) system of NYCT records data only when a passenger enters the subway system or boards a bus. The OD matrix estimation process for the subway system is as described in Section 2.2.1 above. However, for the bus system, the boarding stop of each farecard trip is unknown, and Caliper Corporation has designed a method to estimate bus OD matrix in New York City.
NYCT does not currently have an Automatic Vehicle Location (AVL) system installed on its bus fleet. Therefore, it is not possible to know the precise location of a bus at any point in time. However, based on scheduled run time, it is possible to know the general location of the bus along its route. Relying only on scheduled run time is certainly inadequate, and needs to be adjusted by the transfer information contained in the farecard (MetroCard). For example, based on scheduled run time, we expect bus A to be at transfer stop S1 at 8:00 am, but we see several passengers transferring from bus B to bus A at 7:50, and the only transfer point between these two routes is stop S1. Then we can infer that bus A was at stop S1 at 7:50am rather than 8:00. The time that bus A passes all the subsequent stops should also be adjusted by ten minutes for that trip. In this way, combining the scheduled running time of each vehicle trip with selected transfer transaction from the AFC system, an intensive automated processing method can be used to infer the boarding location for every AFC bus passenger on a trip-by-trip basis. The destinations can then be inferred using a trip-chaining method similar to the one for rail systems. This project is currently nearing completion.
Chapter 3

General Approach to Estimating Bus OD Matrix Using ADC

The best procedure to estimate the OD matrix will depend on the type and quality of data available; however, there is a general approach which will be described in this chapter. This general approach assumes that the agency has multiple ADC systems that can generate a set of sample boarding and alighting counts for all stops or segments (referred to as marginal values), a sample of passenger trips with known linked boarding and alighting stops or stop segments (referred to as the seed matrix), and transfer flow totals by route and direction. The general approach consists of the following three steps:

1. Process the raw ADC data to obtain a sample set of boarding and alighting counts which are usually provided by an APC system, a sample passenger trip OD matrix which will be our seed matrix and transfer flows which are generally provided by an AFC system. This first step is the most difficult to generalize since it is based on each agency's available data. In Section 3.1 below the discussion focuses on the major factors to consider. More detailed discussion in the context of case study of CTA bus system is contained in Chapter 5.

2. Combine the marginal values and the seed matrix using a statistical technique to produce a single route OD matrix for every bus route-direction combination (or “route-dir”). This step is discussed in Section 3.2.

3. Use the transfer flow totals from step 1 to link all the single route OD matrices. This step is discussed in Section 3.3.

It is important to note that this multi-step procedure provides valuable information about passenger travel patterns at each step. Even though the ultimate goal is the estimation of a network level OD matrix, most of the intermediate products such as the single route OD
matrices and the total transfer flows by route-dir have significant planning value to the agency individually.

3.1 Data Preparation

The first step in the general approach is to transform the raw data into the desired input required for the following steps. The data sources available in different transit agencies can be quite different. While the input to this step varies, the output should be (1) a set of sample boarding and alighting counts, (2) a set of sample passenger trips with OD determined (the seed matrix), and (3) transfer flow totals for all route-dir. Several factors need to be considered in this step.

The first factor is what types of data are available. For example, some agencies may have AVL data available, while others will not. The lack of certain types of data may seriously limit, or even make it impossible to estimate the required outputs with acceptable accuracy. For example, with AVL data, we can locate the boarding stop for a passenger trip. Without AVL data, it may still be possible to determine the general location where a passenger boarded the bus by identifying the location of the bus according to the schedule data which is adjusted by referencing transfer trips (NYCT 2003) (see Section 2.2.3). However, in this case, stop level accuracy will be compromised, and the resulting OD matrix is more likely to be credible at the route segment or zone level. Additionally, relying on schedule data implies that significantly more processing effort will be needed.

Another factor to be considered is the extent of data coverage for a particular type of data. If a particular type of data is only available for a small percentage of the total passenger (or vehicle) trips, we will need to decide on how to use this data.

For example, if the transit agency does not have APC equipment on all buses, do we need to scale the boarding and alighting counts to get total boarding and alighting counts? This depends on the requirements of the single route OD estimation method to be used.
If we need to scale the APC boarding and alighting counts to get the total boarding and alighting counts, we will need to define the scaling factor. In the case presented in this thesis, we will use AFC data to derive the scaling factor. Since the scaling factors are potentially different across the route-dirs and time periods, we need to estimate a scaling factor for each route-dir and time period.

During the data processing, an important factor to consider is the potential for bias, both in the raw data and as a result of the processing used. For example, when creating the seed matrix from entry-only AFC data, we need to infer the destination of trips. If a trip-chaining method of destination inference is used (see Rahbee 2004, Zhao 2004, Zhao, Wilson and Rahbee 2006) bias may be present in the resulting seed matrix. This is because passenger trips that are part of a regular trip pattern are more likely to have destinations successfully inferred. Passenger trips by frequent travelers are also more easily inferred, and they will have a larger influence in the seed matrix.

Whether biases like these will have a significant impact on estimated passenger trip patterns and the resulting OD matrix is difficult to determine deductively. Therefore, if such biases are potentially present, an effort should be made to verify the OD matrix result from other data sources. However, we should keep in mind that verifying the OD matrix result using manual data sources is difficult because the surveys tend to have comparatively larger statistical errors and biases.

### 3.2 Single Route OD Matrix Estimation

With the outputs from the previous step, we can combine the marginal values and the seed matrix to form a single route OD matrix for each route-dir. As described in Chapter 2, this step can be accomplished by applying either iterative proportional fitting (IPF) or maximum likelihood estimation (MLE). (Ben-Akiva, 1987)

#### 3.2.1 Iterative proportional fitting (IPF)

Using IPF to combine the marginal values and the seed matrix is computationally simpler. However, there is the added requirement that the marginal value be the total counts rather
than just a sample of the boarding and alighting counts. In other words, if IPF is chosen, we must scale the boarding and alighting counts from APC to get an estimate of the total boarding and alighting counts as discussed in Section 3.1.

To solve the system of equations (2-1) through (2-3), first, we sum over all rows, and obtain a row factor such that the row total multiplied by this row factor equals the marginal value for that row, in this case the total boarding counts for that row. Then we sum over all columns with each cell multiplied by corresponding row factors, and obtain a column factor such that the column total multiplied by this column factor equals the marginal value for that column, in this case the total alighting counts. We can repeat the process multiple times until we get convergence. If \( a_i^k \) is the row factor at iteration \( k \), then the product of the row factors at each iteration will be the final row factor, \( a_i = \prod_k a_i^k \). Similarly, the product of the column factors at each iteration will be the final column factor. (Navick and Furth, 1994)

Due to inconsistency of data from different sources, this system of equations can not always be solved while satisfying all the constraints. In these cases, the set of \( T_{ij} \), or the corresponding set of \( a_i \) and \( b_j \) that minimizes the total error will be used in the selected OD matrix.

3.2.2 Maximum likelihood estimation (MLE)

For MLE we usually need only a sample of the passenger OD flows (the seed matrix), and a sample of the boarding and alighting counts rather than total boarding and alighting counts. We need to slightly modify the MLE formulation shown in Section 2.1.3 to include variables representing the response rate \( p_{ij} \) and introduce the time period variables to represent the different sampling time periods for the different data sources. More specifically, the boarding and alighting counts obtained from APC data only represent observations during the time period equal to the sum of the vehicle headways with valid APC data. Though theoretically this headway should be the actual headway, the actual headway varies along the route. For simplicity the sum of scheduled headways for all vehicle trips with valid APC data is used instead.
We now re-write equation (2-4) in terms of the captured OD flow counts ($y_{ij}$):

Mean value of this Poisson process is the captured arrival rate for OD pair $i,j$: $\lambda_{ij}p_{ij}$

\[
P(y_{ij} | t, \lambda_{ij}, p_{ij}) = e^{-\lambda_{ij}p_{ij}t} \frac{(\lambda_{ij}p_{ij}t)^{y_{ij}}}{y_{ij}!}
\]

(3-1)

The log likelihood for $y_{ij}$ is then:

\[
L(y_{ij}) = -\lambda_{ij}p_{ij}t + y_{ij} \ln \lambda_{ij}p_{ij}t + C
\]

(3-2)

Where $t$ is the time period for data used in the study, and $C$ is a constant.

Similarly the observations of the boarding and alighting counts ($x_{i*}$ and $x_{*j}$) are obtained by rewriting equations (2-5) and (2-6):

\[
P(x_{i*} | t_{i*}, \lambda_{i*}) = e^{-\lambda_{i*}t_{i*}} \frac{(\lambda_{i*}t_{i*})^{x_{i*}}}{x_{i*}!}
\]

(3-3)

and

\[
P(x_{*j} | t_{*j}, \lambda_{*j}) = e^{-\lambda_{*j}t_{*j}} \frac{(\lambda_{*j}t_{*j})^{x_{*j}}}{x_{*j}!}
\]

(3-4)

Hence, $L(x_{i*}) = -\lambda_{i*}t_{i*} + x_{i*} \ln \lambda_{i*}t_{i*} + C$, and $L(x_{*j}) = -\lambda_{*j}t_{*j} + x_{*j} \ln \lambda_{*j}t_{*j} + C$

(3-5)

where: $t_{i*}$ is the combined scheduled headway for all vehicle trips with valid APC boarding data, and $t_{*j}$ is the combined scheduled headway for all vehicle trips with valid APC alighting data. Most likely $t_{i*}$ and $t_{*j}$ will be identical.

The total log likelihood is then:

\[
LL = \sum_{i} \sum_{j} (-\lambda_{ij}p_{ij}t + y_{ij} \ln \lambda_{ij}p_{ij}t) + \sum_{i} (-\lambda_{i*}t_{i*} + x_{i*} \ln \lambda_{i*}t_{i*}) + \sum_{j} (-\lambda_{*j}t_{*j} + x_{*j} \ln \lambda_{*j}t_{*j}) + C
\]

(3-6)

Since $p_{ij}$ is also unknown, it is impossible to estimate $\lambda_{ij}$ and $p_{ij}$ at the same time. By assuming that the response rate $p_{ij}$ is dependent only on the origin and destination location,
we can replace $p_{ij}$ by $p_i \cdot p_j$, and estimate $p_i$ and $p_j$ instead (Ben-Akiva, Macke and Hsu 1985):

$$LL = \sum_i \sum_j (-\lambda_{ij} p_i t + y_{ij} \ln \lambda_{ij} p_i p_j t) + \sum_i (-\lambda_{ji} t_i + x_i \ln \lambda_{ji} t_i) + \sum_j (-\lambda_{j} t_j + x_j \ln \lambda_{j} t_j) + C$$

(3-7)

To solve for $\max LL$, we set the derivatives to zero with respect to $\lambda_{ij}$, $p_i$ and $p_j$, and solve the resulting system of equations:

$$\frac{\partial LL}{\partial \lambda_{ij}} = -p_i p_j t + \frac{y_{ij}}{\lambda_{ij}} - t_i + \frac{x_i}{\lambda_{ij}} - t_j + \frac{x_j}{\lambda_{ij}} = 0, \ \forall i, j$$

(3-8)

$$\frac{\partial LL}{\partial p_i} = \sum_j (-\lambda_{ij} p_j t + \frac{y_{ij}}{p_i}) = 0, \ \forall i$$

(3-9)

$$\frac{\partial LL}{\partial p_j} = \sum_i (-\lambda_{ij} p_i t + \frac{y_{ij}}{p_j}) = 0, \ \forall j$$

(3-10)

The value of $\lambda_{ij}$ is the arrival rate of passengers with origin $i$ and destination $j$. We can multiply it by the time period of interest to obtain the total passenger OD flow for that time period.

There are other possible specification for $p_{ij}$, such as expressing $p_{ij}$ as a function of distance (Ben-Akiva and Morikawa, 1989), or treat it as uniform across all OD pairs. These specifications may be worth testing, but will not be included in this thesis.

3.3 Network Level Bus Passenger Linked Trip OD Matrix Estimation

Network level bus passenger trips can be defined in various ways, so we need to carefully define both “linked trip” and “unlinked trip”. An “unlinked trip” is used here to mean a trip between a passenger boarding and alighting on a single bus. A “linked trip” means one or more unlinked trips that are considered transfer trips by the transit agency that are between the
origin and the ultimate destination. Transfer trips seldom involve more than two unlinked trips, and all such unlinked trips must be within a specified time of each other. We can also specify that two trips on the same route, whether in the same direction or not, will not count as a transfer trip. With this definition, we can attempt to construct an OD matrix of linked trips for the bus system, using the boarding stop of the first trip in the series as the origin, and the alighting stop of the last trip as the destination of this linked trip. In this thesis, if a series of passenger trips are transfer trips as defined above, the first unlinked passenger trip is called the first trip, the second unlinked passenger trip in this series of trips is called the transfer trip (or first transfer trip), and if two transfers are allowed in the definition, the third unlinked trip in the series of trips will be called the second transfer trip. Also for each transfer, the unlinked trip that takes the passenger to the transfer stop is called the first leg of transfer and the unlinked trip that begins at the transfer stop is called the second leg of the transfer.

To transform the single-bus-route OD matrices we obtained in the previous step of the process into network level OD matrix, we need to use the total transfer counts which are usually obtained from farecard (AFC) data. The basic steps given the transfer flow totals are as follows: (1) obtain the origin distribution (for the first leg of the transfer trip) or the destination distribution (for the second leg of the transfer trip); (2) separate transfer flows from the single-bus-route matrices; (3) distribute the transfer flows to eligible origins and destinations.

Figure 3-1 illustrates how the non-transfer OD matrices of single routes and the transfer matrices between any two bus routes combine to form the network OD matrix. Each of the steps defined above is described in the following subsections.
3.3.1 Total transfer flows

Obtaining transfer flow totals is really part of data preparation, and it is included here because it is used only for network level OD matrix estimation.

AFC data is usually the basis for estimating transfer flow totals, and for network level OD matrix estimation, the resulting data has to reflect all transfer trips. Depending on data availability, we can either scale the aggregate transfer flows from AFC data to obtain the transfer flow totals, or estimate the transfer flow totals from other data sources.

Since the single route OD matrices obtained in the previous step are for route-dir, we also need to know the directions of both the first leg and the second leg of the transfer in order to form the network level matrix. The direction information may or may not be contained in the
AFC data, and it may be possible to obtain the vehicle direction by referring to the vehicle schedule records or AVL data.

3.3.2 Obtain the origin or destination distribution of the transfer flow

A. First leg of the transfer

For the first leg of the transfer, we know that the destination stop of this trip is the corresponding transfer stop. However, we do not know at which of the upstream stops these trips originate. There are two different methods to distribute the transfer flows among eligible origins.

Proportional distribution:

This method distributes the transfer flows to eligible origin stops proportional to the total flow from all eligible origin stops to the transfer stop. This method has the advantage of simplicity, and is illustrated in the following example.

Example:

We have the simple network shown in Figure 3-2, and we know that there are 2 transfers from Route #1 east to Route #2 north at stop 3. Figure 3-3 shows the partial OD matrices for Routes #1 East and #2 North obtained from previous steps. Transfers in other direction combinations between these two routes are not considered, but can be analyzed using the same method developed.

Figure 3-2 Example Schematic Network
With these input data, we need to assign the origins of the transfer flows. The eligible origin stops are Stop 1 and Stop 2 -- note the flow from Stop 3 to Stop 3 is zero. Using this method, the transfer flows that originate from Stop 1 and Stop 2 should have the same proportion, namely 3:2. Thus the transfer flow from Stop 1 to Stop 3 is \( 2 \cdot \frac{3}{3+2} = 1.2 \), and the transfer flow from Stop 2 to Stop 3 is \( 2 \cdot \frac{2}{3+2} = 0.8 \). Fractional flows are acceptable and can indicate probability of a certain OD flow.

**Modified IPF method:**

A more sophisticated method presupposes that the IPF method was used previously to obtain the single-bus-route matrices. In this method we need to modify the IPF method to include “pseudo-stops” representing transfers to a particular route-dir as a destination stop (Koutsopoulos 2006), as shown in Figure 3-4 for the same example.
The top table shows the standard OD matrix with the IPF results. The OD flows are obtained by combining the marginal value (total boarding and alighting counts) with a seed matrix. The OD flows ending at Stop 3 contain both the non-transfer flows (the flows with Stop 3 as the final destination of the linked trip) and the transfer flows.

The lower table in Figure 3-5 shows an OD matrix with an added column, labeled as “Transfer to Route # 2”. This is a pseudo-stop, and all passenger trips transferring to Route # 2 north will treat this pseudo-stop as their destination even though their actual (intermediate) destination is the transfer stop, Stop 3. The passenger trips that end at Stop 3 without transferring to other routes still have Stop 3 as their destination. The marginal value and the seed matrix used for IPF will all have to be adjusted by adding the pseudo-stop, and the OD matrix regenerated with the pseudo-stop present.

If a bus route can have transfers to multiple route-dir, then multiple pseudo-stops should be added with one for each potential transfer route/dir.

Comparison between proportional distribution and modified IPF:

Even though these two methods can both be used to obtain a distribution of origins or destinations of transfer flows, they are based upon different assumptions. Proportional
distribution assumes that the passenger OD behavior is the same for transfer passengers and non-transfer passengers. The modified IPF method does not require this assumption.

This assumption is not true in general because:
(i) The cost of a transfer trip is usually significantly lower, and this can create different behavior between transfer passengers and non-transfer passengers. For example, a passenger may prefer to walk for 10 minutes rather than paying the full fare for a bus ride, but if he only needs to pay the transfer fare, he may choose to take the bus instead.
(ii) For certain bus network structures, the OD pairs for transfer trips may be more limited than for non-transfer trips. For example, see Figure 3-6. This simple network contains bus route 1 and 2, going through ABCD and FBCE respectively. These two routes use the same type of bus and share the same stops between BC. Passengers can transfer freely between these two routes between BC. In this case, it is impractical to have a second leg of transfer trip with both origin and destination in between B and C because passengers can stay on the original route to reach their desired destination between B and C.

Figure 3-5 Example Network

A D

B C

E

F

B. Second leg of the transfer

On the other hand, for the second leg of the transfer, we know that their origins are the corresponding transfer stop, but we do not know at which of the downstream stops these trips end. The same two methods can be used to distribute the transfer flows among eligible destinations, but because of their similarity they will not be repeated here.
3.3.3 Non-transfer flow OD matrices

To construct a bus network level OD matrix, we need to separate the non-transfer flows from the single route OD matrices obtained from the steps in Section 3.3 above. The transfer flows contained there are simply segments of linked trips, and they are estimated in the next step of the procedure. This step excludes the transfer trips (both first and second legs of transfers) from the single route OD matrices leaving the trips that have no transfer, which can go directly into the network level OD matrix.

The implementation of this step depends on whether the proportional distribution method or modified IPF method was used in the previous steps. If the modified IPF method is chosen, we only need to remove the origins and destinations involving pseudo-stops from the resulting OD matrices, and what are left are non-transfer single route OD matrices.

If the proportional distribution method is used, we can simply subtract the transfer flows from the full OD flows to get the non-transfer trip OD matrix.

3.3.4 Transfer flow OD matrices

For many applications, the results from previous step showing the distribution of the origin and destination of the transfer trips may be adequate. However, there is one additional step needed to obtain the network level passenger trip OD matrix. In this step, we link the distribution of the origin of the first leg of the transfer with the destination of the second leg of transfer to obtain the OD flows. With these transfer flow matrices, we can combine them with the non-transfer flow single route OD matrices to form the network level OD matrix as shown in Figure 3-1.
Continuing with the previous example, Figure 3-7 shows the process of calculating an OD matrix for transfer flows. We know the marginal values and need to populate the OD cells. The total transfer trip boardings at Stop 1 (and Stop 2) of Route 1 are obtained as described in Section 3.3.1. Similarly the total transfer trip alightings at Stop 4 (and Stop 5) of Route 2 are obtained as described in Section 3.3.2. With the marginal values, we can estimate the OD flow (the cells in the table) by IPF with a null seed matrix.

It may be possible to obtain a seed matrix in this case from AFC data for high volume transfer points, but usually the transfer OD flows are very low. The resulting seed matrix will tend to be low quality, with many non-structural zeros, so the benefit of a seed matrix here is usually not enough to justify its use.
Chapter 4

Introduction to CTA Case Study

This Chapter introduces the case study application of the general methodology introduced in Chapter 3. Section 4.1 describes the CTA system itself and Section 4.2 describes the Automatic Data Collection systems used in the case study.

4.1 CTA System

The Chicago Transit Authority is the second largest US transit agency. The rail system serves about 500,000 passenger trips every day, and there are 1,190 rapid transit cars running on the eight routes and 222 miles of track with 144 stations. The rail system provides direct connection to the two major airports.

Besides the rail system, CTA operates approximately 2000 buses serving over 151 routes, and there are over 1 million passengers using the bus system every day. The CTA bus network is a grid system with parallel bus routes 0.5 to 1 mile apart. There are over 12,000 bus stops in the system, and there are usually 8 stops per mile. In 2005, the bus fare was $1.75, with an additional $0.25 for one transfer, which also covers a (free) second transfer. The transfer benefit is available between different modes. Passengers can board the bus using passes, magnetic transit cards, smart cards (Chicago Card or Chicago Card Plus), or by paying with cash. Cash paying transfer passengers obtain a magnetic transfer card.

However, starting in January 2006, a new fare policy created a fare differential based on fare media. The fares for those using Chicago Card or Chicago Card Plus are unchanged while fares for magnetic transit card remained at $1.75 for bus and $2.00 for rail. The cash fares are raised to $2.00 with the elimination of the transfer card. The significant impact of this fare policy change will be discussed in the case study.
Figure 4-1 CTA System Map
The data used in this case study is for five weekdays from September 12 to 16, 2005. To demonstrate the algorithm, all OD matrices are estimated for the morning peak time period of 6 to 9 am, and the network level OD matrix is estimated for a demonstration corridor of 6 routes only.

4.2 CTA ADC Systems

CTA has three main ADC systems: the Automated Fare Collection (AFC) system, the Automatic Vehicle Location (AVL) system, and the Automatic Passenger Count (APC) system. This section will describe the data available from these three systems.

4.2.1 Automatic Fare Collection (AFC) System

CTA has a flat fare system, so the fare collection system uses entry-only control and the CTA’s AFC system records transactions only when passengers enter the system. CTA’s AFC systems are not integrated with the AVL systems, so the data recorded does not have the stop ID for where the passengers paid their fares, and instead have a timestamp and the fare box equipment ID which can be linked to bus numbers using fare box equipment records. Except for cash payments, which do not create a record in the AFC data, all other fare transactions produce AFC records. However, the transfer card transaction records are different from other fare media types, which are considered farecards. Specifically, a farecard in this thesis is defined to be a magnetic transit card, Chicago Card or Chicago Card Plus, and any type of pass.

All fare media types can be used to board a bus (while rail station fare gates do not accept cash), but the farecard types are only available for purchase at limited locations. A passenger can only obtain Chicago Card or Chicago Card Plus at selected sales outlets, at CTA headquarters, or through the mail. Passes are more widely available and can be purchased at local grocery stores. Magnetic transit cards can be purchased at Automatic Vending Machines (AVM), which are available at every rail station. Passengers can add value to transit cards or Chicago Cards using AVM machines, or can add value to Chicago Card Plus online or automatically through credit cards when the value on the card drops below a threshold. Therefore, if a
passenger does not have any farecard type when he boards the bus, the only option he has is to pay with cash. Adding value is not possible at any bus stop or onboard any bus.

Before the fare policy change in January 2006, about 60% of bus passenger trips were made using various kinds of farecards, and are therefore captured in the farecard data.

(a) Data fields

A sample of the AFC data for farecard transactions is shown in Table 4-1 and consists of the following data fields:

(1) The Serial Number is a 10-digit number that uniquely identifies a farecard. This field is critical as it allows selection of transit trips by a particular passenger. We will assume in our analysis that one farecard is used by only one passenger although CTA allows multiple users to share the same farecard (except for passes) while still preserving the transfer benefits for every user. When this feature is used, additional processing is required to properly infer the destinations.

(2) The Time field contains timestamp information accurate to the second. This field is necessary to sort the transit trips (transactions) of a passenger in sequential order. As discussed in Section 2.2.1, sorting a passenger's transit trips in sequential order is one of the first steps in passenger trip destination inference.

(3) The Equipment ID contains an alphanumerical identifier for the fare box where the farecard transaction occurred, and can be mapped to the bus number with this fare box at any particular time. Thus the farecard transaction can be linked to a particular vehicle.

(4) The Transaction Type column shows whether this farecard transaction is a new purchase of a magnetic transit card, adding value, entering a particular type of gate, or insufficient value remaining. For the purpose of this thesis, the farecard transactions of interest fall into two categories: (1) entry into the system, and therefore the state of a passenger trip, or (2) a non-trip transaction at an AVM machine or some other equipment. The passenger trip transactions are going to be the main focus of our analysis. The non-trip transaction types
indicate the passengers' presence at the specified fare equipment at the recorded time. However, since fare equipment for non-trip transactions are located at a bus stop or onboard a bus, the only real use of these non-trip transaction types is when the passenger's trip pattern involves a rail trip.

(5) The Current Route field contains encoded route number for current bus route if this is a bus trip. If this is a rail trip or non-trip transaction, this field contains the rail station id.

(6) The Last Route field contains the information on the previous trip. If the previous trip is a bus trip, the bus route number is encoded here. If the previous trip is a rail trip, the entering station id is encoded here.

Table 4-1 Sample AFC Data

<table>
<thead>
<tr>
<th>Serial Number</th>
<th>Time</th>
<th>Equipment ID</th>
<th>Transaction Type (encoded)</th>
<th>CURRENT ROUTE (encoded)</th>
<th>LAST ROUTE (encoded)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1207814275</td>
<td>9/12/2005 12:33:19 AM</td>
<td>BTP06485</td>
<td>38</td>
<td>119</td>
<td>953</td>
</tr>
<tr>
<td>1208488781</td>
<td>9/12/2005 12:28:36 AM</td>
<td>BTP06485</td>
<td>38</td>
<td>119</td>
<td>1070</td>
</tr>
<tr>
<td>1207831931</td>
<td>9/12/2005 1:00:27 AM</td>
<td>BTP06485</td>
<td>38</td>
<td>119</td>
<td>766</td>
</tr>
<tr>
<td>1207762979</td>
<td>9/12/2005 12:28:24 AM</td>
<td>BTP06485</td>
<td>38</td>
<td>119</td>
<td>1063</td>
</tr>
<tr>
<td>1207736780</td>
<td>9/12/2005 1:28:43 AM</td>
<td>TT_02160</td>
<td>38</td>
<td>1081</td>
<td>81</td>
</tr>
</tbody>
</table>

(b) Location information

The AFC equipment onboard a bus cannot automatically obtain location information from the onboard AVL equipment. Determining the location information is straightforward for rail trips, as the entering rail station id is encoded in the Current Route column. However, determining the location of the bus stop where the fare transaction occurs is more complex, and depends on post-processing the CTA AVL system data which is described in Section 4.2.2 below.
(c) Transfer card data

Although recorded by the same fare equipment, there are differences between transfer card data and farecard data. The main difference is that transfer card data does not have a Serial Number, so it cannot be traced to a particular passenger. This means that the trip-chaining method of destination inference described in Section 2.2.1 cannot be used for transfer card data. Another relatively minor difference is that instead of a transaction type data field, we have an event type data field, which shows whether this transaction is the issuance of this transfer card or the first, or second, use of this transfer card (see Table 4-2). The issue of the transfer card with cash payment is not considered a transfer while the first or second use of the transfer card indicates a transfer.

Table 4-2 Sample Transfer Card Data

<table>
<thead>
<tr>
<th>EQUIPMENT ID</th>
<th>Time</th>
<th>EVENT TYPE</th>
<th>CURRENT ROUTE</th>
<th>LAST ROUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTP04977</td>
<td>9/12/2005 6:51:17 AM</td>
<td>1</td>
<td>22</td>
<td>1059</td>
</tr>
<tr>
<td>BTP05839</td>
<td>9/12/2005 4:31:04 AM</td>
<td>2</td>
<td>9</td>
<td>70</td>
</tr>
<tr>
<td>BTP05994</td>
<td>9/12/2005 12:58:56 AM</td>
<td>255</td>
<td>79</td>
<td>0</td>
</tr>
<tr>
<td>BTP04977</td>
<td>9/12/2005 8:51:17 AM</td>
<td>2</td>
<td>22</td>
<td>147</td>
</tr>
<tr>
<td>BTP06086</td>
<td>9/12/2005 1:08:11 AM</td>
<td>255</td>
<td>6</td>
<td>0</td>
</tr>
</tbody>
</table>

4.2.2 Automatic Vehicle Location (AVL) system

Every CTA bus is equipped with an Automatic Vehicle Location (AVL) system. Vehicle location tracking system is used to provide automated voice annunciation system via AVAS in compliance with the American with Disability Act (ADA) requirements. CTA’s AVL system is an off-line AVL system, with the onboard equipment recording the various vehicular events received from different sensors. The AVL equipment is linked to the odometer and a Global Positioning System (GPS) receiver unit and also has schedule and route data downloaded at the garage. The AVL equipment uses the GPS information together with the odometer and gyroscope plus route information to determine the correct bus stop. AVL equipment also records events such as operator log in, and start and end of the route. However, for this research, we focus only on the serviced stop records.

An AVL record contains the following six data fields:
(1) Time contains the timestamp of the event accurate to one second. For a serviced stop, this is the time when the bus door opens i.e. the arrival time at the stop.
(2) Bus ID contains the four digit bus number that uniquely identifies a vehicle.
(3) Route contains the alphanumeric route information.
(4) Trip ID contains the trip number information that is unique for any one-way vehicle trip.
(5) Direction contains the route direction of the bus at the stop.
(6) Stop Sequence and Route Pattern information that allow us to look up the stop name.

The AVL system also needs the route information data to identify the correct stop. In CTA, the route data is contained in several different tables. The BT_Pattern table contains the mapping information to look-up the system-specific pattern ID using route and route-specific pattern. The BT_PatternDetail table identifies the stop ID associated with the AVL serviced stop records given the pattern ID and stop sequence number. It also contains the heading and coordinates information of each stop. The stop descriptions are contained in the BT_StopInfo table.

4.2.3 Automatic Passenger Count (APC) system
The CTA Automatic Passenger Count (APC) system is fully integrated with the AVL system and hence shares most of the same data fields with the exception of the APC-specific data fields of passenger boarding and alighting counts. The sensor used to count passenger boardings and alightings uses double infrared beams. Based on the order the infrared beams are broken, the sensor determines whether the passenger is boarding or alighting. The sensor mounting locations vary by bus model, with the sensors on certain types of bus being more reliable and less likely to be out of alignment than others.

(a) Data description
As the sample data in Table 4-3 shows, the APC data contains the AVL data (Time, Bus ID, Route, Trip ID, Direction, and Stop), Boardings, the boarding counts for passengers entering through both the front and the rear doors, and Alightings, the alighting counts for both the front and the rear doors.
Table 4-3 Sample APC Data

<table>
<thead>
<tr>
<th>TIME</th>
<th>BUS ID</th>
<th>ROUTE</th>
<th>TRIP ID</th>
<th>DIRECTION</th>
<th>Boardings</th>
<th>Alightings</th>
<th>STOP</th>
</tr>
</thead>
<tbody>
<tr>
<td>9/12/2005</td>
<td>6734</td>
<td>20</td>
<td>15065450</td>
<td>East</td>
<td>0</td>
<td>3</td>
<td>WASHINGTON +</td>
</tr>
<tr>
<td>12:08:10 AM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>STATE</td>
</tr>
<tr>
<td>9/12/2005</td>
<td>6734</td>
<td>20</td>
<td>15065450</td>
<td>East</td>
<td>1</td>
<td>1</td>
<td>MADISON +</td>
</tr>
<tr>
<td>12:00:04 AM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PEORIA</td>
</tr>
<tr>
<td>9/12/2005</td>
<td>6734</td>
<td>20</td>
<td>15065450</td>
<td>East</td>
<td>0</td>
<td>0</td>
<td>WASHINGTON +</td>
</tr>
<tr>
<td>12:03:29 AM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>CANAL</td>
</tr>
<tr>
<td>9/12/2005</td>
<td>6729</td>
<td>20</td>
<td>15067244</td>
<td>East</td>
<td>0</td>
<td>1</td>
<td>WASHINGTON +</td>
</tr>
<tr>
<td>5:37:19 AM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LASALLE</td>
</tr>
</tbody>
</table>

(b) APC errors
The infrared beam technology used for passenger counting has a small but finite chance of undercounting the number of boardings or alightings. Over the length of a bus trip that can be over 10 miles, total boardings and alightings as recorded over the entire vehicle trip often differ by a small amount. As this difference is mostly due to undercounting, adjustments to boarding or alighting counts are needed to balance the total boarding and alighting counts. The adjustments will be described in detail in Chapter 5.

(c) APC coverage
Not every bus in CTA bus fleet is APC-equipped. Currently about 15% of buses are APC-equipped although this percentage will increase over time as CTA is committed to all new buses having APC equipment installed by the manufacturer.

At the route level, there are wide differences in the percentage of buses which are APC equipped. Certain routes with newer buses may have all APC-equipped buses while other routes may have less than 1% of the buses APC-equipped. In other words, although we can obtain total passenger boarding and alighting counts for certain routes from APC data alone, this is not possible for most routes.
4.3 Differences in Available Data between CTA Rail and Bus Systems

Previous work has been done to estimate the CTA rail system passenger trip OD matrix using ADC data (Zhao 2004, Zhao, Wilson and Rahbee 2006), but there has been no attempt at estimating the CTA bus system passenger trip OD matrix using ADC data. The bus OD matrix estimation methodology proposed in this thesis is quite different from the rail system methodology developed previously primarily because of the differences in available data. This section will summarize these differences.

As shown in Table 4-4, one notable difference is the availability of APC data for the bus system. As a data source independent from AFC data, the inclusion of APC data as an input should positively impact the estimation results. On the other hand, combining two independent data sources that are not always consistent requires more complex processing. Another major difference is that the origin of each farecard trip is recorded in the AFC data for rail trips, but location information requires data from the AVL system for bus. Additionally, the percentage of bus passenger trips that are made using farecards is lower than for rail. The rail trips recorded in AFC data are linked trips, and there needs to be additional steps to obtain unlinked trip information. On the other hand the bus trips recorded in AFC data are unlinked trips, and we need to use the transfer flow data to obtain a network level OD matrix or an OD matrix of linked passenger trips.

Table 4-4 Data Differences between CTA Rail and Bus

<table>
<thead>
<tr>
<th></th>
<th>Rail</th>
<th>Bus</th>
</tr>
</thead>
<tbody>
<tr>
<td>APC availability and coverage</td>
<td>None</td>
<td>&gt;15% of bus trips</td>
</tr>
<tr>
<td>Trip origin or Entry point</td>
<td>100% Known</td>
<td>Need AVL data (or GIS data) ~80% identifiable</td>
</tr>
<tr>
<td>AFC coverage</td>
<td>91%</td>
<td>About 60%</td>
</tr>
<tr>
<td>Other differences</td>
<td>Directly obtain linked trips</td>
<td>Directly obtain unlinked trips</td>
</tr>
</tbody>
</table>

There is one important note that due to the fare policy change mentioned in Section 4.1, the AFC coverage for both rail and bus has already increased dramatically from September 2005.
However, there will continue to be a small portion of bus passengers who will continue to use cash and therefore will not be captured by the AFC data.
Data preparation consists of collecting: (1) a set of boarding and alighting counts to develop marginal values, (2) a sample of passenger trips OD's as the seed matrix, and (3) transfer flow totals between all route-dir. We will defer the discussion of obtaining transfer flow totals to Chapter 7 when we describe the network level OD matrix estimation for the CTA bus system. This chapter will discuss the derivation of the marginal values and seed matrices.

This data preparation step can be divided into the following sub-tasks: segment definition (Section 5.1), boarding stop identification (Section 5.2), destination inference and seed matrix formation (Section 5.3), boarding and alighting count balancing (Section 5.4), and APC data scaling to get marginal values (Section 5.5).

5.1 Segment Definition
The CTA’s AVL and APC data are at the stop level. The AFC data can also be matched with AVL data to identify the transaction location at the stop level. However, some OD pairs have very low or no demand, and these low demand pairs may not be captured in the farecard data or in the seed matrices. Also for certain stops with low boarding or alighting counts, it is possible the APC data for these stops will contain zeros if the percentage of buses equipped with APC equipment for that route is low. Non-structural zeros in the data may cause problems when combining different data sources using the IPF method. At the same time, most transit applications do not require stop level passenger flows. Therefore, the level of detail used in this case study will be at the route segment level to achieve a balance between the accuracy and processing practicality. The various sub-tasks of data preparation will still be at stop level detail and then the data will be aggregated to the segment level, and for all subsequent steps, all analysis will be at the segment level.
The segments used in this thesis are defined as follows:

1. Each transfer stop is defined as a segment by itself. If two bus routes share the same street for a portion of their routes, only the first and the last common stops are considered transfer stops for the purpose of defining segments. Any stops that are within 111 meters or 364 feet (0.001 degree latitude) in Manhattan distance of each other are considered the same logical stop for this purpose.
2. All neighboring stops between two adjacent transfer stops are defined as a segment.
3. The stops between the route terminal (if not a transfer stop) and the next transfer stop are also defined as a segment.

Defined in this manner, there are two classes of segments: transfer segments and non-transfer segments. Treating transfer stops as distinct segments is mainly to prepare for the later development of the network level passenger OD matrix estimation process where the transfer segments will be shared by two or more different bus routes and transfer flows happen only at transfer segments. Figure 5-1 shows a schematic describing the different types of segments.

Figure 5-1 Segment Definition

It is also possible to combine the segments as defined here to form larger segments for the analysis if the data does not support this level of detail.
5.2 Boarding Stop Identification for AFC Trips

As mentioned previously, the CTA AFC data does not contain the boarding bus stop information. In order to obtain the boarding stop information for each farecard trip, we match available AFC data with AVL data to obtain the vehicle location at the time of a farecard transaction. The first step is to link the transaction fare box identification with the bus ID number. Then we can select the most likely records from the AVL records for that bus.

5.2.1 Look up bus number

The AFC data has the Equipment ID data field that identifies the fare box. However, fare boxes are not permanently mounted on particular buses, and can be moved from one bus to another. To translate the Equipment ID field into the unique 4-digit bus numbers, we created a table mapping Equipment ID and Time to Bus Number. Table 5-1 shown below is obtained from maintenance data by determining the earliest and latest time an Equipment ID is associated with a Bus Number. In this table, three Equipment IDs are shown, corresponding to one or more Bus Numbers. Columns “Min Time” and “Max Time” indicate the time range for which this association between Equipment ID and Bus Number is valid. For example fare box “BTP04047” is on bus 5473 throughout the whole week in September, while fare box “BTP04048” was initially on bus 6750 and then moved to bus 5438.

Table 5-1 Sample Equipment ID to Bus Number Look Up Table

<table>
<thead>
<tr>
<th>EQUIPMENT ID</th>
<th>BUS NBR</th>
<th>MIN TIME</th>
<th>MAX TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTP04047</td>
<td>5473</td>
<td>9/12/2005</td>
<td>9/16/2005 8:00:00 PM</td>
</tr>
<tr>
<td>BTP04048</td>
<td>5438</td>
<td>9/12/2005 6:48:03 PM</td>
<td>9/16/2005 11:00:00 PM</td>
</tr>
<tr>
<td>BTP04049</td>
<td>6743</td>
<td>9/15/2005 6:57:06 PM</td>
<td>9/16/2005 11:00:00 PM</td>
</tr>
</tbody>
</table>

5.2.2 Relationship between AFC and AVL records

AFC records for bus trips are created when passengers pay their fares, and AVL serviced stop records are created when the doors open at the stops to allow passengers to board or alight (a
"serviced" stop). The sequence of events is normally: (1) the bus stops creating an AVL record, (2) zero, one or more passengers board the bus creating AFC records, (3) the bus reaches the next serviced stop and this sequence is repeated. The relationship between an AFC bus trip record and corresponding AVL stop record normally has the following characteristics: (1) the matching AVL stop record precedes the AFC record; and (2) the AVL stop record is the closest to the AFC record in time.

The time difference between the AVL record and AFC record is the difference between when the bus stops and when each passenger pays. Usually there are very few, if any, passengers who pay after their bus has reached the next stop, and these passengers are not considered in this analysis. In some cases, a bus operator may open the bus doors to allow boarding while waiting at the terminal or a major time point, and in these cases some passengers will board the bus a few minutes after the bus doors open. However, if the time between bus arrival and passenger paying is longer than a few minutes, it is likely that this data is the result of some error and we do not rely on it to identify the boarding stop. Also in terms of processing, it is far easier if we only look at the AVL records that are within a short period before the AFC record, rather than looking at all preceding AVL records. Therefore, we only looked at AVL records that are within 5 minutes before the corresponding AFC record, selecting the latest AVL stop record among this group, and the stop identified in this AVL record is assigned as the stop where the AFC transaction occurs. If there is no valid AVL record within this time window, we consider the boarding stop for this AFC record unidentifiable and this AFC record is dropped.

Figure 5-2 shows the time distribution of the AVL records that immediately precede the AFC record of interest. We can see that the great majority of passengers pay within 90 seconds of doors opening and only a very small proportion (<2%) of passengers pay more than 300 seconds (5 minutes) after the doors open. Therefore, it is reasonable to limit the time window for matching AVL records to 5 minutes.
5.2.3 Boarding stop identification

With the Bus Number known, we can find the bus locations from the AVL data as described in Section 5.2.2. This step is logically straightforward, but involves several additional tables containing route structure data. 90% of the AFC record boarding stops were successfully identified using this method. The actual SQL queries in scripts for Oracle database are shown in Appendix C.

5.2.4 Identifying boarding stop based on transfer information

In those cases when the AVL data fail to identify the boarding stop, alternative method can be used based on the same assumptions as the trip-chaining method of destination inference discussed in Section 2.2.1. Specifically, we can identify the boarding stop of a bus trip as the unique intersection point of this bus route and the bus route or rail line used on the previous trip of this farecard. The assumption is that if there is a unique intersection, the transfer occurs at this intersection. Therefore, the condition for using this method to identify the boarding bus stop is: the intersection between the bus route and previous trip’s route or line is unique, and there is a previous transit trip record in the AFC data.
Because we already identified 90% of the passenger trips using the method described in Section 5.2.3, the additional contribution of using transfer information to identify boarding stop is less than 5%. Since this method requires trip-chaining data processing similar destination inference, it is implemented as part of this algorithm which is described in the next section.

5.3 Alighting Stop Inference

The destination stop inference method is based on the trip-chaining destination inference method used for entry-only rail system as described in Section 2.2.1. (See Barry 2002, Rahbee 2002, and Zhao 2004 for a more detailed description.) The basic trip-chaining method assigns the destination stop to be the origin of the next transit trip on the same farecard (see Section 5.3.1). However, the next transit trip may not be on the same route as the current trip, in which case we assign the destination of the first trip to be the stop on the current route that is closest to the origin of the next trip. This process is called destination reassignment and is discussed in Section 5.3.2.

5.3.1 Trip-chaining method

The major steps in automating the data processing required to implement this method are:

1. Sort the output data from the previous step (Section 5.2) (adding the rail trip farecard transaction data to use all transit trips) by Serial Number and Day and Time.
2. Assign a sequence number to each record.
3. Identify the Serial Numbers and Days that have only one trip, i.e. the passenger is making only one transit trip using this farecard, and delete these records as we can not infer destinations for them.
4. Isolate the first trip and last trip, and if they are both bus trips, assign the destination of the last trip to be the origin of the first trip (last trip of the day method, see Section 2.2.1).
5. Match each record with the next record.
6. Except for the last trip for each Serial Number and Day, assign the destination to be the origin of the next transit trip (next trip method, see Section 2.2.1).
7. Eliminate the rail trips added in step (1) above to leave bus trips only.
See Appendix C for all the SQL queries used to automate this process.

5.3.2 Destination reassignment

When inferring the destination of a trip, the next trip can be by rail, bus on a different route, or bus on the same route in either the same or the opposite direction. If the next trip is on the same bus route as the current trip, then we do not need to reassign the destination (unless this route runs on different streets in each direction).

If the next trip is a subway trip, then we need to find the bus stop on the current route that is closest to the subway station used for the next trip. This requires a look-up table that, for any bus route and rail station, gives the system Stop ID for the stop on given bus route that is nearest to the station. This look-up table is derived by calculating the distances between any rail station and any bus stop, eliminating the entries with Manhattan distances of more than 1110 meters or 3642 feet (0.01 degree latitude).

Similarly if the next trip is on a different bus route (or if the bus travels on different streets in each direction), we need to find the stop on the current bus route that is closest to the boarding stop of the next trip, and assign this stop on the current bus route to be the alighting stop. If the next trip is a bus trip on a different bus route (or different route-dir), it can either intersect the current bus route (Example A in Figure 5-4) or be parallel to the current bus route (Example B in Figure 5-4).

In Example A, as long as we know the current route and next route, Stop c, which is closest to next route, is the most likely alighting stop for current trip. Whether the boarding stop for the next trip is Stop a or Stop b does not matter: Stop c is going to be closer to any stop on the next route. In Example B in the same figure, on the other hand, whether the boarding stop of the next trip is Stop a or Stop b will matter. In this case, the most likely alighting stop is the stop nearest the boarding stop of next trip.
After this step, we have farecard trip data with both boarding and alighting stops consistent with the route information of the trip data, and we can then aggregate the trips by route and direction to form a single-bus-route seed matrix.

5.3.3 Seed matrix formation

The farecard bus trip data with the boarding and alighting stops identified can be aggregated to form the seed matrix. For our original sample of 2,736,454 farecard bus trips taken on five weekdays, we can identify the boarding stops of 2,484,166 trips (90% identification rate), and from these, we can infer the destination stops for 2,161,737 trips resulting in a 85% inference rate. After destination reassignment, a total of 1,713,463 trips remain (79% success rate). In the end, the seed matrix represents about 60% of all bus farecard trips. Considering on average that the farecard usage rate is estimated at about 60% for bus passenger trips, we will have about 36% of all passenger trips represented in the seed matrix. The sampling rate of
about 36% of all trips within a study time period is far higher than standard passenger survey typically achieves.

Table 5-2 shows the seed matrix for Route 20 eastbound in the morning peak period (6 to 9 AM) at the zonal level where each zone includes multiple segments.

Table 5-2 Seed Matrix for Route 20 Eastbound

<table>
<thead>
<tr>
<th>Route 20 Eastbound</th>
<th>Destination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austin-Cicero</td>
<td>Cicero-Kezie</td>
</tr>
<tr>
<td></td>
<td>32</td>
</tr>
<tr>
<td>Cicero-Kezie</td>
<td>55</td>
</tr>
<tr>
<td>Kezie-Ashland</td>
<td>46</td>
</tr>
<tr>
<td>Ashland-LaSalle</td>
<td></td>
</tr>
<tr>
<td>LaSalle-Columbus</td>
<td></td>
</tr>
</tbody>
</table>

Note: Data is the 6-9 am daily average based on Sep 12 to 16, 2005.

5.4 APC Data Cleaning
As discussed in Section 4.2.3(b), the APC boarding and alighting counts need to be adjusted so that the total boardings and total alightings over the entire length of a one-way bus trip are equal, i.e. no passengers remain on the bus before the one-way trip starts and no passengers are on the bus after the one-way trip ends. This is not always the case in reality, as there are some passengers boarding in the reverse direction close to one of the terminals to save some walking, to secure a seat, or to avoid the elements. Consideration of this terminal effect is beyond the scope of this thesis, and here we will make the simplifying assumption that there are no passengers on board before the first stop of the vehicle trip or after the last stop of the vehicle trip.
We need to sum the boarding and alighting counts by bus trip and by day and then adjust these counts to ensure their totals match by increasing the lower number. The difference to be adjusted should be distributed to all serviced stops for that vehicle trip. The assumption is that the higher the count, the more likely there is an undercount, and the adjustments should therefore be proportional to either the boarding or alighting counts at the stops depending on whether we are adjusting the boarding count or the alighting count. This is a simple way to correct errors as APC equipment is less likely to overcount than to undercount. This process will also help ensure internal data consistency since every boarding must correspond to an alighting during the same vehicle trip.

On occasion, the total boarding and alighting counts over a one-way vehicle trip can be very different, and this may indicate misalignment of one of the breaking beam sensors. The APC readings from this vehicle trip should be excluded, and this vehicle trip is not considered to have valid APC data. The threshold for exclusion is set when the difference is greater than 15% between the total boarding count and the alighting count over the one-way vehicle trip.

5.5 APC Data Scaling

If the Iterative Proportional Fitting (IPF) technique is to be used to combine the seed matrix and the marginal values, we will need this data scaling step to estimate the total boardings and alightings at each stop.

As discussed in Section 4.2.3(c), the percentage of APGequipped buses varies across CTA bus routes. Even for the few bus routes with all APGequipped buses, data scaling may still be necessary due to occasional equipment failures.

Figure 5-5 shows that the independent AFC and APC data sources provide us with the seed matrix and the scaling factors. To expand the APC boarding and alighting counts to estimate total boardings and alightings based on the independence assumption we have:

\[
P(AFC \mid APC) = P(AFC) \Rightarrow \frac{AFC_{APC}}{P_{APC}} = \frac{AFC_{Total}}{P_{Total}}
\]  

(5-1)
where $P_{\text{Total}}$ is the total unknown passenger count for a route-direction combination,

$P_{\text{APC}}$ is the cumulative APC passenger boarding count recorded for the same route-direction combination,

$AFC_{\text{APC}}$ is the total farecard trip counts on the APC vehicle trips for the same route-direction combination,

$AFC_{\text{Total}}$ is the total farecard trip counts for all vehicle trips for the same route-direction combination.

Thus:

$$\frac{P_{\text{APC}}}{P_{\text{Total}}} \approx \frac{AFC_{\text{APC}}}{AFC_{\text{Total}}} \quad (5-2)$$

Alternatively, we can consider APC data to be a series of ride-checks, and it is possible to expand the APC data by assuming the vehicle trips with valid APC data to be a random sample of the vehicle trips. This method will require a large sample size, and will need significantly more APC data than the scaling method discussed above and used here.

Figure 5-5 Relationship between Seed Matrix and Original Data

*The percentages are based on 2005 data*
Now we have both the seed matrix derived in Section 5.4 and the marginal values from this section (an example is shown in Table 5-3). We are now ready to start the next step of Single-bus-route OD Matrix Estimation.

Table 5-3 Seed Matrix and Marginal Values for Route 20 Eastbound, 6-9AM

<table>
<thead>
<tr>
<th>Destination</th>
<th>beyond PINE</th>
<th>PINE-CICERO</th>
<th>CICERO-PULASKI</th>
<th>PULASKI-KEDZIE</th>
<th>KEDZIE-WESTERN</th>
<th>WESTERN-ASHLAND</th>
<th>ASHLAND-CANAL</th>
<th>CANAL-LASALLE</th>
<th>LASALLE-WABASH</th>
<th>WABASH-STETSON</th>
<th>beyond STETSON</th>
<th>Row Sum</th>
<th>Scaled Total Boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond PINE</td>
<td>4</td>
<td>13</td>
<td>9</td>
<td>7</td>
<td>8</td>
<td>5</td>
<td>6</td>
<td>4</td>
<td>0</td>
<td>2</td>
<td>61</td>
<td>218</td>
<td></td>
</tr>
<tr>
<td>PINE-CICERO</td>
<td>15</td>
<td>33</td>
<td>38</td>
<td>17</td>
<td>15</td>
<td>15</td>
<td>5</td>
<td>9</td>
<td>2</td>
<td>3</td>
<td>152</td>
<td>427</td>
<td></td>
</tr>
<tr>
<td>CICERO-PULASKI</td>
<td>9</td>
<td>41</td>
<td>24</td>
<td>28</td>
<td>19</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>3</td>
<td>143</td>
<td>359</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PULASKI-KEDZIE</td>
<td>5</td>
<td>11</td>
<td>16</td>
<td>11</td>
<td>6</td>
<td>8</td>
<td>1</td>
<td>2</td>
<td>60</td>
<td>137</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEDZIE-WESTERN</td>
<td>9</td>
<td>30</td>
<td>15</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>70</td>
<td>177</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WESTERN-ASHLAND</td>
<td></td>
<td>8</td>
<td>29</td>
<td>11</td>
<td>18</td>
<td>6</td>
<td>3</td>
<td>76</td>
<td>150</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASHLAND-CANAL</td>
<td></td>
<td>43</td>
<td>63</td>
<td>55</td>
<td>23</td>
<td>21</td>
<td>207</td>
<td>454</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CANAL-LASALLE</td>
<td></td>
<td>5</td>
<td>23</td>
<td>5</td>
<td>11</td>
<td>44</td>
<td>130</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASALLE-WABASH</td>
<td></td>
<td>2</td>
<td>10</td>
<td>11</td>
<td>23</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WABASH-STETSON</td>
<td></td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beyond STETSON</td>
<td></td>
<td>0</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Column Sum</td>
<td>4</td>
<td>28</td>
<td>50</td>
<td>91</td>
<td>88</td>
<td>101</td>
<td>139</td>
<td>106</td>
<td>134</td>
<td>58</td>
<td>60</td>
<td>839</td>
<td>2114</td>
</tr>
<tr>
<td>Scaled Total Alightings</td>
<td>7</td>
<td>88</td>
<td>247</td>
<td>135</td>
<td>337</td>
<td>275</td>
<td>206</td>
<td>216</td>
<td>354</td>
<td>181</td>
<td>68</td>
<td>2114</td>
<td></td>
</tr>
</tbody>
</table>

Note: Average of 5 weekdays of data. Each origin and destination shown is 4 segment units combined.
Chapter 6

CTA Single Bus Route OD Matrix Estimation

In this Chapter, we present the results of the Iterative Proportional Fitting (IPF) method for a single route OD matrix estimation in Section 6.1 and the results of the Maximum Likelihood Estimation (MLE) method in Section 6.2. Finally, the results from both methods are compared in Section 6.3.

6.1 Iterative Proportional Fitting (IPF)

The procedure for implementing IPF was implemented using Microsoft Excel. Migrating the implementation to Microsoft Access or SQL for production purposes is the natural next step although this is beyond the scope of this thesis.

Because the seed matrix is based on only a sample of total passenger trips, the sum of all cells of a seed matrix over a row (or column) should always be less than the total boarding count for that row (or column). Thus all column factors and row factors should be greater than 1. Occasionally a column factor or row factor will be less than 1 indicating inconsistency between the total boarding or alighting counts and the seed matrix. Due to the possibility of such inconsistency, extra constraints are added to the generic IPF process to enforce the data relationship of \( a_i = \prod_k a_i^k \geq 1 \). Whenever a column or row factor would have a value of less than 1, it is set to 1 instead.

In Table 6-1, the resulting column and row factors for Route 20 eastbound between 6 to 9 am are shown. The OD flows in the seed matrix (see Table 5-3) are scaled by these factors shown in Table 6-1, and the results form the estimated single-bus-route OD matrix.

We can see that most of the column and row factors have values greater than 1, and only some segments with relatively low boarding or alighting counts have values equal to 1 (where the
additional constraint is tight). This shows that the data inconsistency is a bigger potential problem when the absolute value of the counts are low, and the same sampling rate may result in lower level of precision in these low count cases.

Table 6-1 IPF OD Matrix for Route 20 Eastbound 6-9AM

<table>
<thead>
<tr>
<th>destination</th>
<th>beyond PINE</th>
<th>PINE-CICERO</th>
<th>CICERO-PULASKI</th>
<th>PULASKI-KEDZIE</th>
<th>KEDZIE-WESTERN</th>
<th>WESTERN-ASHLAND</th>
<th>ASHLAND-CANAL</th>
<th>CANAL-LASALLE</th>
<th>LASALLE-WABASH</th>
<th>WABASH-STETSON</th>
<th>beyond STETSON</th>
<th>Scaled total boardings</th>
<th>Row factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond PINE</td>
<td>7</td>
<td>45</td>
<td>54</td>
<td>14</td>
<td>14</td>
<td>48</td>
<td>17</td>
<td>11</td>
<td>7</td>
<td>12</td>
<td>1</td>
<td>3</td>
<td>218</td>
</tr>
<tr>
<td>PINE-CICERO</td>
<td>42</td>
<td>151</td>
<td>54</td>
<td>83</td>
<td>40</td>
<td>20</td>
<td>9</td>
<td>19</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>217</td>
<td>1.24</td>
</tr>
<tr>
<td>CICERO-PULASKI</td>
<td>40</td>
<td>59</td>
<td>115</td>
<td>74</td>
<td>27</td>
<td>12</td>
<td>18</td>
<td>12</td>
<td>4</td>
<td>359</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PULASKI-KEDZIE</td>
<td>7</td>
<td>47</td>
<td>38</td>
<td>14</td>
<td>10</td>
<td>17</td>
<td>3</td>
<td>2</td>
<td>137</td>
<td>1.15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEDZIE-WESTERN</td>
<td>42</td>
<td>81</td>
<td>21</td>
<td>9</td>
<td>17</td>
<td>7</td>
<td>3</td>
<td>177</td>
<td>1.28</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WESTERN-ASHLAND</td>
<td>23</td>
<td>42</td>
<td>22</td>
<td>43</td>
<td>18</td>
<td>4</td>
<td>150</td>
<td>1.32</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASHLAND-CANAL</td>
<td>70</td>
<td>132</td>
<td>147</td>
<td>73</td>
<td>33</td>
<td>454</td>
<td>1.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CANAL-LASALLE</td>
<td>13</td>
<td>76</td>
<td>19</td>
<td>22</td>
<td>130</td>
<td>1.87</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASALLE-WABASH</td>
<td>4</td>
<td>28</td>
<td>15</td>
<td>46</td>
<td>1.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WABASH-STETSON</td>
<td>12</td>
<td>2</td>
<td>13</td>
<td>1.90</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beyond STETSON</td>
<td>0</td>
<td>1</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scaled total alightings</td>
<td>7</td>
<td>88</td>
<td>247</td>
<td>135</td>
<td>337</td>
<td>275</td>
<td>206</td>
<td>216</td>
<td>354</td>
<td>181</td>
<td>68</td>
<td>2114</td>
<td></td>
</tr>
<tr>
<td>Column factors</td>
<td>1.00</td>
<td>2.20</td>
<td>3.74</td>
<td>1.16</td>
<td>3.85</td>
<td>2.13</td>
<td>1.10</td>
<td>1.43</td>
<td>1.81</td>
<td>2.16</td>
<td>1.60</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Please see Appendix A for the estimated single-bus-route OD matrices for both directions of Routes 9, 20, 49, 49X, and 126. The OD matrices for all route-dir are easily obtainable, but are not shown because the purpose of this thesis is simply to demonstrate the proposed methodology.
6.2 Maximum Likelihood Estimation (MLE)

The MLE formulation discussed in Section 3.3.3 was implemented on Route 20 eastbound using averages from five weekdays of data. In this section the results using the MLE technique are presented. Using five instead of one day of data may be taking full advantage of the data losing some variability. Improved estimation results might be obtained by using the disaggregate data at the individual day, or even the individual trip, level. In this way the full extent of variability in the data can be used to improve the estimation. This improvement may be a topic for future research.

During the 6 to 9AM period over five weekdays, there are a total of 178 bus trips with 63 having valid APC data. The boarding and alighting counts are summed by segment with the results shown in Table 6-2.

Table 6-2 Unscaled APC Boarding and Alighting Counts

<table>
<thead>
<tr>
<th>Segments</th>
<th>beyond PINE</th>
<th>PINE - CICERO</th>
<th>CICERO - PULASKI</th>
<th>PULASKI - KEDZIE</th>
<th>KEDZIE - WESTERN</th>
<th>WESTERN - ASHLAND</th>
<th>ASHLAND - CANAL</th>
<th>CANAL - LASALLE</th>
<th>LASALLE - WABASH</th>
<th>WABASH - STEWSON</th>
<th>beyond STEWSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Boardings</td>
<td>87</td>
<td>171</td>
<td>143</td>
<td>55</td>
<td>71</td>
<td>60</td>
<td>181</td>
<td>52</td>
<td>18</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Total Alightins</td>
<td>3</td>
<td>33</td>
<td>93</td>
<td>51</td>
<td>126</td>
<td>103</td>
<td>77</td>
<td>81</td>
<td>133</td>
<td>68</td>
<td>25</td>
</tr>
</tbody>
</table>

The input data also includes the seed matrix shown in Table 5-2. Since we are using data of 5 morning peak periods of 3 hours each, the total observation time for the AFC data and seed matrix is 15 hours. For Route #20, the total observation time for the APC data correspond to the sum of the scheduled headways for all bus trips having valid APC data, 5.31 hours in this case.

Segments with zero boarding or alighting counts are combined with neighboring segments to reduce processing difficulties. (See Section 3.3.3.) If the passenger arrival rate 𝛼 is close to zero, it needs special treatment in the optimization processing and combining very low volume segments eliminates this potential problem. We used line search techniques to numerically estimate the unknown arrival rates (\( \alpha \)) and sampling rates (\( p_i, p_j \)) due to its conceptual simplicity (Bertsimas 1997). Tables 6-3 and 6-4 show the estimation results for Route #20 eastbound.
### Table 6-3 Estimated Passenger Flows from MLE

<table>
<thead>
<tr>
<th>destination</th>
<th>beyond PINE</th>
<th>PINE - CICERO</th>
<th>CICERO - PULASKI</th>
<th>PULASKI - KEDZIE</th>
<th>KEDZIE - WESTERN</th>
<th>WESTERN - ASHLAND</th>
<th>ASHLAND - CANAL</th>
<th>CANAL - LASALLE</th>
<th>LASALLE - WABASH</th>
<th>WABASH - STETSON</th>
<th>beyond STETSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond PINE</td>
<td>12.1</td>
<td>36.7</td>
<td>37.5</td>
<td>15.3</td>
<td>48.3</td>
<td>26.4</td>
<td>15.1</td>
<td>13.9</td>
<td>3.6</td>
<td>4.6</td>
<td></td>
</tr>
<tr>
<td>PINE - CICERO</td>
<td>60.7</td>
<td>147.0</td>
<td>51.2</td>
<td>96.0</td>
<td>50.7</td>
<td>19.1</td>
<td>16.2</td>
<td>14.1</td>
<td>8.8</td>
<td>2.2</td>
<td></td>
</tr>
<tr>
<td>CICERO - PULASKI</td>
<td>83.2</td>
<td>62.6</td>
<td>112.0</td>
<td>70.6</td>
<td>30.9</td>
<td>12.5</td>
<td>18.5</td>
<td>6.3</td>
<td>2.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PULASKI - KEDZIE</td>
<td>16.2</td>
<td>53.9</td>
<td>41.8</td>
<td>12.8</td>
<td>11.0</td>
<td>12.4</td>
<td>3.9</td>
<td>2.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEDZIE - WESTERN</td>
<td>50.9</td>
<td>79.6</td>
<td>29.4</td>
<td>15.6</td>
<td>14.6</td>
<td>3.9</td>
<td>2.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WESTERN - ASHLAND</td>
<td>30.5</td>
<td>40.0</td>
<td>30.6</td>
<td>35.1</td>
<td>18.9</td>
<td>5.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASHLAND - CANAL</td>
<td>81.3</td>
<td>124.5</td>
<td>182.3</td>
<td>72.9</td>
<td>32.5</td>
<td>13.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CANAL - LASALLE</td>
<td>9.4</td>
<td>88.3</td>
<td>33.9</td>
<td>15.7</td>
<td>28.3</td>
<td>6.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASALLE - WABASH</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WABASH - STETSON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>11.8</td>
<td>2.4</td>
<td></td>
</tr>
<tr>
<td>beyond STETSON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

As shown in Table 6-4, the estimated sampling rates vary greatly with a maximum of 0.66 for OD pairs from Cicero-Pulaski to beyond Stetson, and a minimum of 0.1 for intra-segment travel within Wabash-Stetson. \( p_i \cdot p_j \) is analogous to the inverse of the product of the row and column factors for a particular OD pair obtained in the IPF method. The cause of this variability is uncertain, and should be examined in future research.

### Table 6-4 Estimated Seed Matrix Sampling Rates

<table>
<thead>
<tr>
<th></th>
<th>beyond PINE</th>
<th>PINE - CICERO</th>
<th>CICERO - PULASKI</th>
<th>PULASKI - KEDZIE</th>
<th>KEDZIE - WESTERN</th>
<th>WESTERN - ASHLAND</th>
<th>ASHLAND - CANAL</th>
<th>CANAL - LASALLE</th>
<th>LASALLE - WABASH</th>
<th>WABASH - STETSON</th>
<th>beyond STETSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p_i )</td>
<td>0.46</td>
<td>0.77</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
<td>0.90</td>
</tr>
<tr>
<td>( p_j )</td>
<td>1.00</td>
<td>0.62</td>
<td>0.53</td>
<td>0.52</td>
<td>0.56</td>
<td>0.47</td>
<td>0.47</td>
<td>0.35</td>
<td>0.25</td>
<td>0.25</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Note that in Table 6-4, \( p_i \) and \( p_j \) have to be used jointly, i.e. \( p_i \cdot p_j \) is the seed matrix sampling rate for OD pair \( i \) to \( j \).
6.3 Comparison of the Estimation Results

The hourly average passenger arrival rate can be readily translated into a passenger OD matrix, and we can compare it with the results from the IPF method shown in Table 6-1. In Table 6-5, we show both results in the same table for ease of comparison. To better represent the differences between these two results, we show the OD passenger flows as percentages of total passenger flows (shown in the lower right corner of the table) for Route # 20 eastbound. The results shown in parenthesis are those obtained using IPF.

Table 6-5 Comparison of MLE and IPF results

<table>
<thead>
<tr>
<th>origin</th>
<th>beyond PINE</th>
<th>PINE - CICERO</th>
<th>CICERO - PULASKI</th>
<th>PULASKI - KEDZIE</th>
<th>KEDZIE - WESTERN</th>
<th>WESTERN - ASHLAND</th>
<th>ASHLAND - CANAL</th>
<th>CANAL - LASALLE</th>
<th>LASALLE - WABASH</th>
<th>WABASH - STETSON</th>
<th>beyond STETSON</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond PINE</td>
<td>1% (1%)</td>
<td>2% (2%)</td>
<td>2% (2%)</td>
<td>1% (1%)</td>
<td>2% (2%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>10% (10%)</td>
</tr>
<tr>
<td>PINE - CICERO</td>
<td>3% (3%)</td>
<td>6% (8%)</td>
<td>2% (2%)</td>
<td>4% (4%)</td>
<td>2% (2%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>20% (21%)</td>
</tr>
<tr>
<td>CICERO - PULASKI</td>
<td>4% (3%)</td>
<td>3% (3%)</td>
<td>5% (5%)</td>
<td>3% (3%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>17% (17%)</td>
</tr>
<tr>
<td>PULASKI - KEDZIE</td>
<td>1% (1%)</td>
<td>2% (2%)</td>
<td>2% (2%)</td>
<td>1% (1%)</td>
<td>0% (0%)</td>
<td>1% (1%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>7% (6%)</td>
</tr>
<tr>
<td>KEDZIE - WESTERN</td>
<td>2% (2%)</td>
<td>3% (3%)</td>
<td>2% (2%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>8% (8%)</td>
</tr>
<tr>
<td>WESTERN - ASHLAND</td>
<td>1% (1%)</td>
<td>2% (2%)</td>
<td>1% (1%)</td>
<td>2% (2%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>7% (7%)</td>
</tr>
<tr>
<td>ASHLAND - CANAL</td>
<td>4% (3%)</td>
<td>5% (8%)</td>
<td>3% (4%)</td>
<td>1% (1%)</td>
<td>3% (1%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>6% (6%)</td>
</tr>
<tr>
<td>CANAL - LASALLE</td>
<td>0% (0%)</td>
<td>4% (4%)</td>
<td>1% (2%)</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>1% (1%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>2% (2%)</td>
</tr>
<tr>
<td>LASALLE - WABASH</td>
<td>1% (1%)</td>
<td>1% (1%)</td>
<td>0% (0%)</td>
<td>1% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>1% (1%)</td>
</tr>
<tr>
<td>WABASH - STETSON</td>
<td>1% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
</tr>
<tr>
<td>beyond STETSON</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
<td>0% (0%)</td>
</tr>
<tr>
<td>Total</td>
<td>1% (1%)</td>
<td>4% (4%)</td>
<td>12% (11%)</td>
<td>6% (6%)</td>
<td>16% (16%)</td>
<td>13% (13%)</td>
<td>10% (10%)</td>
<td>10% (10%)</td>
<td>17% (17%)</td>
<td>8% (9%)</td>
<td>3% (3%)</td>
<td>2198 (2114)</td>
</tr>
</tbody>
</table>

Note: OD flow as a percentage of total flows. Percentages are results from MLE method, and results from IPF method are shown in parenthesis. Rounding errors may be present.

As we can see, while the results are virtually identical in percentage terms, there is a significant difference in terms of total passengers (~4%). The total boarding and alighting counts used in the IPF method are only estimated values while IPF requires that the control totals be of a much higher quality than the seed matrix. The total boarding and alighting counts are estimated based on the assumption that the AFC and APC data are completely independent.
Any correlation between these data sources will cause biased estimates of the control totals. Using MLE, the resulting total passenger flow is that which is most likely to explain all the observed data, and the assumption of independence between AFC and AVL data is not required. The results are therefore theoretically superior. Therefore, using IPF can give accurate information on the relative weight of OD flows, but the absolute values of the OD flows will depend on the accuracy of the total estimated boarding and alighting counts.

This conclusion is consistent with Hsu’s findings that “when the relative quantity of ride-check data is much larger than (for example ten times)” the on-board survey data, the differences between the results of IPF and PMLE is negligible while “when the relative quantity of ride-check data is about the same” as the on-board survey, the differences between the results of IPF and PMLE can be significant. (Hsu 1985) In our case the seed matrix is about 36% of total flows and the boarding and alighting totals are about 30% of the vehicle trips for Route #20. As such, for this case and likely for all bus OD estimation purposes at the CTA, the MLE method will produce better overall estimates of the total OD flows than the IPF method.

Schwarcz (2004) estimated an OD matrix for CTA Route #20 using only the boarding and alighting counts data with a null seed matrix. Table 6-6 is a comparison between Schwarcz’s result and the resulting single-bus-route OD matrix from the IPF process used in this thesis. Since Schwarcz defines the morning peak period of 7-9:15AM instead of 6-9AM, a comparison of the absolute values of the OD flows is not very meaningful, so the OD flows as a percentage of the respective total passenger flows are shown instead. The OD flows in Schwartz’s result are shown in parenthesis.

These two results are generally similar, but there are also some significant differences. Notably, there are more shorter trips and fewer longer trips in the current result. This is a result of the fact that the actual passenger behavior as represented by the seed matrix includes more short trips than predicted by the simpler trip distribution model. The prevalence of shorter trips intuitively makes sense because Route #20 parallels both the CTA Green and Blue rail lines. Passengers traveling longer distances should favor the parallel rail lines, and are less likely to take the Route #20 bus because the travel time on the bus is longer. As a result,
using a representative seed matrix allows us to make more accurate estimation of the OD movements on the route.

Table 6-6 Route #20 OD Matrix Comparison with Prior Estimates

<table>
<thead>
<tr>
<th>Zones</th>
<th>AUSTIN-PULASKI</th>
<th>PULASKI-KEDZIE</th>
<th>KEDZIE-WESTERN</th>
<th>WESTERN-ASHLAND</th>
<th>ASHLAND-HALSTED</th>
<th>HALSTED-COLUMBUS</th>
<th>Total Boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUSTIN-PULASKI</td>
<td>13% (13%)</td>
<td>17% (14%)</td>
<td>5% (4%)</td>
<td>6% (9%)</td>
<td>0% (1%)</td>
<td>6% (6%)</td>
<td>48% (46%)</td>
</tr>
<tr>
<td>PULASKI-KEDZIE</td>
<td>2% (1%)</td>
<td>3% (1%)</td>
<td>2% (3%)</td>
<td>0% (1%)</td>
<td>1% (3%)</td>
<td>8% (9%)</td>
<td></td>
</tr>
<tr>
<td>KEDZIE-WESTERN</td>
<td></td>
<td>1% (0%)</td>
<td>3% (2%)</td>
<td>1% (0%)</td>
<td>3% (3%)</td>
<td>7% (5%)</td>
<td></td>
</tr>
<tr>
<td>WESTERN-ASHLAND</td>
<td>1% (1%)</td>
<td>0% (1%)</td>
<td>7% (7%)</td>
<td>8% (9%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASHLAND-HALSTED</td>
<td></td>
<td>3% (0%)</td>
<td>13% (15%)</td>
<td>16% (16%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HALSTED-COLUMBUS</td>
<td></td>
<td></td>
<td>13% (14%)</td>
<td>13% (14%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Alightings</td>
<td>13% (13%)</td>
<td>19% (15%)</td>
<td>9% (5%)</td>
<td>12% (15%)</td>
<td>4% (4%)</td>
<td>43% (48%)</td>
<td>2114 trips (1761 trips)</td>
</tr>
</tbody>
</table>

Note: Significant differences are highlighted in bold face.
Chapter 7

Network Level OD Matrix Estimation

The estimation of a partial network level OD matrix for CTA bus system closely follows the general procedures proposed in Section 3.4. This Chapter will discuss the actual estimation and the results using these general procedures. Section 7.1 will define the partial network (or demonstration corridor) analyzed in this thesis. Section 7.2 will discuss obtaining the transfer flow totals by route-dir using the available data from CTA. Sections 7.3 and 7.4 will describe the distribution of the origins and destinations of the transfer flow totals obtained by using both the proportional distribution and modified IPF methods discussed in Sections 3.4.3. Finally Section 7.5 will discuss the IPF technique used to obtain the network level OD matrix for a selected demonstration corridor.

7.1 Demonstration Corridor

To demonstrate the methodology of network level OD matrix estimation, we selected six bus routes in the western corridor of CTA bus network as shown in Figure 7-1. Route #20 runs east-west mainly along Madison Avenue with the eastern end in the Loop. Route #49 runs north-south mainly along Western Ave while Route #X49 is the express (limited-stop) variant of Route #49. Route #126 runs east-west mainly along Jackson Street, and is parallel to Route #20 about 3 blocks to the south. Route #20 and #49/X49 provide transfers at Western and Madison, and passengers can transfer between Route #126 and #49/X49 at Western and Jackson.
Only the transfer flows among these routes in this corridor are considered in this chapter.

### 7.2 Transfer Flow Totals

Our basic assumption in obtaining transfer flow totals is that all transfers are made using either farecards or transfer cards*, and virtually no passenger will pay cash for transfer trips due to the fare differentials. (See Section 4.1.1.) Therefore, we will obtain the transfer flow totals by aggregating the transfer trips recorded in farecard data and the transfer trips recorded in transfer card data.

#### 7.2.1 Transfer trips in farecard data

The transfer trips made using farecards are not specifically marked as transfer trips in the farecard data; therefore, we need to develop a definition of transfer trips to select those farecard trips that are either the first, second, or third legs of a linked trip. In the case of CTA,

*Transfer cards were in use in September 2005, when the data used in this analysis were recorded, but they were recently eliminated as part of the CTA's January, 2006 fare changes.
the reduced fare transfer trips are defined in the fare policy as a series of passenger trips involving no more than three unlinked trips, and all unlinked trips that are part of this series of transfer trips must start within two hours of each other. We will use this official CTA definition for consistency purposes.

Even though the route information is contained in the AFC data, it is not as reliable as the route information contained in AVL data. The AFC and AVL systems require separate operator logins, but the AFC login generally will not have an impact on the bus operation, whereas the AVL login can determine the route number, the destination signs and the stop announcement. Therefore, operators usually pay more attention to the AVL login.

The AFC data is matched with AVL data to obtain the route and direction information for farecard trips (see Sections 5.3.2 to 5.3.4 for a description of how the AFC data is matched with the AVL data). We then link each passenger trip with the prior trip, and apply the definition of the transfer trips to eliminate passenger trip pairs that are not true transfer trips as defined above. For the purpose of this analysis, we also eliminate the transfer trips involving routes outside the demonstration corridor. We can then aggregate the transfer flows by “transfer-from” route-dir and “transfer-to” route-dir for the transfer trips in the farecard data. There will be some transfer trips that cannot be successfully matched against the AVL data to identify the travel direction and those transfer trips are discarded for the purpose of this analysis. As a result, this aggregate result will only be a sample of all transfer flows although it is a very significant sample with a sampling rate of about 80% of all transfers made using farecards. To obtain the total transfer flows by route-dirs, we will need to use the control totals for all transfer trips contained in farecard data to scale up the aggregate transfer flows by route-dirs to make up for the small data loss.

7.2.2 Transfer trips in transfer card data
In comparison, obtaining the transfer trips from transfer card data is relatively straightforward. All trips recorded in the transfer card data satisfy the definition of transfer trips, and we simply need to aggregate them by the transfer-from and transfer-to route.
The route information contained in the transfer card data is of similar quality as the route information in the farecard data, but we can only match them to the AVL data for the second legs of the transfers where we have the equipment id and time of transaction. More critically, travel direction information is not directly available in either farecard or transfer card data, and can be obtained only by matching to the same bus AVL data. The information we have for the first leg of the transfer is limited to previous route information, and we can not determine which of the preceding transfer card transactions within a two hour window is from the same transfer card due to the lack of a serial number. It is not possible to match to the AVL data to obtain the route and direction information. In summary, we can obtain the route and direction information for transfer trips that are the second leg of transfers with a corresponding small loss of data, but for the first leg of those transfers, we can only directly obtain the route ID.

To obtain the direction information for the first legs of the transfers in aggregate, we will assume that the directional distribution is independent from the fare media type used, and is therefore the same for transfers using regular farecards and for transfers using transfer cards. The aggregate transfer flows from farecard data are calculated first to obtain the proportion of first legs of transfer flows in each travel direction. Then the directional distribution of the first leg of transfer flows contained in transfer card data is allocated in the same proportion.

For example, when we consider the transfer card data for transfers from Route # 20 to Route # 49 north (5 weekdays, 6-9am), we use the farecard data to estimate the number of transfers from Route # 20 east and west to Route # 49 north. From the farecard data, we have 61 passenger trips transferring from Route # 20 east to Route # 49 north, and 17 passenger trips transferring from Route # 20 west to Route # 49 north. From the transfer card data, we have 84 passenger trips transferring from both directions of Route # 20 to Route # 49 north. We can then estimate that the number of trips transferring from Route # 20 east to Route # 49 north is $84 \cdot \frac{61}{78} = 65.7$, and from Route # 20 west to Route # 49 north is $84 \cdot \frac{17}{78} = 18.3$. 

7.2.3 Transfer flow total

We combine the aggregate transfer flows by route-dirs from the two data sources, and capture the majority of transfer flows. However, to obtain the total transfer flows, we need to account for the transfer trips lost in the previous processing. To do this, we first obtain a control total of all transfer flows for trips using transfer cards, and another control total of all transfer flows for trips using farecards. These control totals are counts that are not affected by data loss from matching with AVL data. Then we scale the transfer flows obtained from farecards (Section 7.2.1) to the control total for farecard data, and scale the transfer flows obtained from transfer cards (Section 7.2.2) to the control total for transfer card data. The transfer flows from these two data sources are scaled separately because the transfer flows are processed differently and the data loss percentages are different. For the transfer flows from each data source, we assume that the probability of data loss is independent of the “transfer from” and “transfer to” routes, and a uniform scaling factor for each data source is sufficient.

In the example of Route #20 east transferring to Route #49 north using 5 weekday AM periods of data, we have a total of 61 transfer trips from farecard data, and a total of 65.7 transfer trips estimated from transfer card data. In aggregate there are a total of 107210 transfers using farecards, and a total of 102035 transfers using transfer cards. We can successfully identify both the transfer-from and transfer-to route-dir information for 85218 transfers from the farecard data, and 91392 transfers from the transfer card data. The scaling factors for transfer flow from farecard and transfer card data are 1.26 and 1.12 respectively, and applying these scaling factors, the total transfer flow from Route #20 east to Route #49 north is 150.5 total transfers. This translates into about 30 transfers on average between these two route-dirs during each 6-9am time period.

The transfer flows between all the route-dirs within the demonstration corridor are shown below in Table 7-1. This table shows that the major transfer flow direction in this corridor is from Route #20 or #126 eastbound to #49 or #X49 northbound. The transfer flows between Route #49 and Route #X49 are part of two separate trips or round trips, and are not considered linked trips for this analysis.
Table 7-1 Transfer Flow Totals by Route-dir

<table>
<thead>
<tr>
<th>From Route</th>
<th>From Direction</th>
<th>To Route</th>
<th>To Direction</th>
<th>5 Day Total</th>
<th>Daily Transfer</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>Eastbound</td>
<td>X49</td>
<td>Northbound</td>
<td>182</td>
<td>36</td>
</tr>
<tr>
<td>20</td>
<td>Eastbound</td>
<td>49</td>
<td>Northbound</td>
<td>174</td>
<td>35</td>
</tr>
<tr>
<td>X49</td>
<td>Northbound</td>
<td>20</td>
<td>Westbound</td>
<td>107</td>
<td>21</td>
</tr>
<tr>
<td>49</td>
<td>Northbound</td>
<td>20</td>
<td>Westbound</td>
<td>88</td>
<td>18</td>
</tr>
<tr>
<td>126</td>
<td>Eastbound</td>
<td>X49</td>
<td>Northbound</td>
<td>78</td>
<td>16</td>
</tr>
<tr>
<td>20</td>
<td>Eastbound</td>
<td>49</td>
<td>Southbound</td>
<td>77</td>
<td>15</td>
</tr>
<tr>
<td>49</td>
<td>Southbound</td>
<td>20</td>
<td>Westbound</td>
<td>67</td>
<td>13</td>
</tr>
<tr>
<td>126</td>
<td>Eastbound</td>
<td>49</td>
<td>Northbound</td>
<td>58</td>
<td>12</td>
</tr>
<tr>
<td>49</td>
<td>Northbound</td>
<td>20</td>
<td>Eastbound</td>
<td>57</td>
<td>11</td>
</tr>
<tr>
<td>49</td>
<td>Southbound</td>
<td>20</td>
<td>Eastbound</td>
<td>57</td>
<td>11</td>
</tr>
<tr>
<td>X49</td>
<td>Northbound</td>
<td>20</td>
<td>Eastbound</td>
<td>54</td>
<td>11</td>
</tr>
</tbody>
</table>

Note: only the transfers with daily flow greater than 10 are included.

7.3 Proportional Distribution

The next step is to determine the distribution of origin and destination for the transfer flow totals determined in the previous section. One of the methods as discussed in Section 3.4.3 is to distribute the origins or destinations proportionally based on the total flows for the OD pairs involving the transfer segments.

We illustrate the implementation using proportional distribution by taking the transfer flow from Route # 20 east to Route # 49 north as an example, and the transfer flows between other route-dirs of the demonstration corridor can be analyzed using the same procedure. The single route OD matrix for Route # 20 eastbound was shown in Table 6-1, and the single route OD matrix for Route # 49 north is shown in Table 7-2.
Table 7-2 Route # 49 North OD Matrix

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond 78TH</td>
<td>1</td>
<td>16</td>
<td>11</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>78TH - 63RD</td>
<td>8</td>
<td>29</td>
<td>53</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>63RD - 55TH</td>
<td>7</td>
<td>46</td>
<td>9</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>55TH - 47TH</td>
<td>41</td>
<td>39</td>
<td>10</td>
<td>15</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ARCHER - 26TH</td>
<td>6</td>
<td>12</td>
<td>22</td>
<td>7</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>10</td>
<td>2</td>
<td>16</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>26TH - 16TH</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>16</td>
<td>2</td>
<td>12</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16TH - HARRISON</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>13</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HARRISON - MADISON</td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>15</td>
<td>4</td>
<td>14</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>MADISON - GRAND</td>
<td>3</td>
<td>6</td>
<td>27</td>
<td>1</td>
<td>10</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>GRAND - DIVISION</td>
<td>2</td>
<td>61</td>
<td>14</td>
<td>60</td>
<td>11</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>DIVISION - ARMITAGE</td>
<td>57</td>
<td>24</td>
<td>113</td>
<td>10</td>
<td>7</td>
<td>5</td>
<td>10</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>ARMITAGE - DIVERSEY</td>
<td>10</td>
<td>82</td>
<td>26</td>
<td>12</td>
<td>16</td>
<td>11</td>
<td>10</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>DIVERSEY - ADDISON</td>
<td>36</td>
<td>22</td>
<td>19</td>
<td>11</td>
<td>23</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>ADDISON - MONTROSE</td>
<td>6</td>
<td>16</td>
<td>7</td>
<td>15</td>
<td>6</td>
<td>5</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>MONTROSE - LAWRENCE</td>
<td></td>
<td>6</td>
<td>5</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>LAWRENCE - FOSTER</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

The transfer point between these two routes is Madison and Western. The OD flows with destinations at this transfer point for Route # 20 eastbound are shown in Table 7-3, and the OD flows with origins at this transfer point for Route # 49 north are shown in Table 7-4. The average daily transfer flow of 30.1 (150.5 for 5 weekdays) is distributed in proportion to these
OD flows. The results for the transfer flow between Route #20 east and Route #49 north are shown in Tables 7-5 and 7-6 respectively.

**Table 7-3 Total OD Flows Ending at the Transfer Point**

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond PINE</td>
<td>KEDZIE - WESTERN</td>
<td>41</td>
</tr>
<tr>
<td>PINE - CICERO</td>
<td>85</td>
<td></td>
</tr>
<tr>
<td>CICERO - PULASKI</td>
<td>103</td>
<td></td>
</tr>
<tr>
<td>PULASKI - KEDZIE</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>KEDZIE - WESTERN</td>
<td>46</td>
<td></td>
</tr>
</tbody>
</table>

**Table 7-4 Total OD Flows Originating from the Transfer Point**

<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HARRISON - MADISON</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>MADISON - GRAND</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>GRAND - DIVISION</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>DIVISION - ARMITAGE</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>ARMITAGE - DIVERSET</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>DIVERSEY - ADDISON</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>ADDISON - MONTROSE</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>MONROSE - LAWRENCE</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>LAWRENCE - FOSTER</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>beyond FOSTER</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

**Table 7-5 Transfer-only Flows with Origins Distributed**

<table>
<thead>
<tr>
<th>From</th>
<th>Transfer To Route #49 North</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond PINE</td>
<td>4</td>
</tr>
<tr>
<td>PINE - CICERO</td>
<td>8</td>
</tr>
<tr>
<td>CICERO - PULASKI</td>
<td>10</td>
</tr>
<tr>
<td>PULASKI - KEDZIE</td>
<td>4</td>
</tr>
<tr>
<td>KEDZIE - WESTERN</td>
<td>4</td>
</tr>
</tbody>
</table>
7.4 Modified IPF
The modified IPF involves modifying the seed matrix and the boarding and alighting totals as described below.

7.4.1 Modifying seed matrix
The first step in using modified IPF method is to modify the seed matrix used for single-bus-route OD matrix estimation. For Route # 20 east, four pseudo-columns and four pseudo-rows are added representing transfers to Route # 49 and Route # X49 both north and south as shown in Table 7-7. Similarly the seed matrices for all the route-dir in the demonstration corridor need to be modified by adding the pseudo-columns and pseudo-rows.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>HARRISON - MADISON</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 7-6 Transfer-only Flows with Destinations Distributed
### Table 7-7 Modified Route #20 Seed Matrix with Pseudo-columns and Pseudo-rows for Transfers to Route #49

<table>
<thead>
<tr>
<th>Destination</th>
<th>beyond PINE</th>
<th>PINE-CICERO</th>
<th>CICERO-PULASKI</th>
<th>PULASKI-KEDZIE</th>
<th>WESTERN</th>
<th>Transfer to 49 North</th>
<th>Transfer to 49 South</th>
<th>Transfer to X49 North</th>
<th>Transfer to X49 South</th>
<th>WESTERN-ASHLAND</th>
<th>ASHLAND-CANAL</th>
<th>CANAL-LASALLE</th>
<th>LASALLE-WABASH</th>
<th>WABASH-STETSON</th>
<th>beyond STETSON</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond PINE</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>PINE-CICERO</td>
<td>10</td>
<td>19</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>CICERO-PULASKI</td>
<td>12</td>
<td>10</td>
<td>10</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PULASKI-KEDZIE</td>
<td>4</td>
<td>6</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>7</td>
<td>0</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>KEDZIE-WESTERN</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>8</td>
<td>8</td>
<td>2</td>
<td>5</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer from 49 North</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer from 49 South</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer from X49 North</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transfer from X49 South</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WESTERN-ASHLAND</td>
<td>2</td>
<td>8</td>
<td>5</td>
<td>11</td>
<td>0</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASHLAND-CANAL</td>
<td>20</td>
<td>18</td>
<td>57</td>
<td>5</td>
<td>13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CANAL-LASALLE</td>
<td>0</td>
<td>31</td>
<td>3</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LASALLE-WABASH</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WABASH-STETSON</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beyond STETSON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>

#### 7.4.2 Modifying boarding and alighting totals

The boarding and alighting totals are modified to include the total transfers in the pseudo-columns and pseudo-rows, and the boarding and alighting totals for the regular column or rows are also modified to exclude the total transfers to become the total non-transfer boarding and alighting counts as shown in Table 7-8.
Table 7-8 Boarding and Alighting Counts with Pseudo-entries

<table>
<thead>
<tr>
<th>Segment</th>
<th>beyond PIN</th>
<th>PINE-O'COER</th>
<th>O'COER-PULASKI</th>
<th>PULASKI-KEDZIE</th>
<th>KEDZIE-WESTERN</th>
<th>WESTERN-ASL</th>
<th>ASL-CANAL</th>
<th>CANAL-LASALLE</th>
<th>LASALLE-WABASH</th>
<th>WABASH-BETSON</th>
<th>BETSON-X 43</th>
</tr>
</thead>
<tbody>
<tr>
<td>Boarding</td>
<td>218</td>
<td>427</td>
<td>359</td>
<td>137</td>
<td>177</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>150</td>
<td>454</td>
</tr>
<tr>
<td>Alighting</td>
<td>6</td>
<td>83</td>
<td>232</td>
<td>127</td>
<td>317</td>
<td>19</td>
<td>8</td>
<td>21</td>
<td>3</td>
<td>258</td>
<td>194</td>
</tr>
</tbody>
</table>

7.4.3 Modified IPF results

Applying the IPF procedure, we obtain the modified single-bus-route OD matrices. Separating the transfer trips out from other passenger trips reduces the flow between some OD pairs (including pseudo columns and rows). The number of transfer trips can be very low for certain transfer directions. These factors combine to cause a non-structural zero problem in the resulting seed matrix (for example see Table 7-7), and the resulting OD matrix will have OD pairs that can not be estimated. As such there will be bias in the resulting OD matrix. The modified OD matrix for route # 20 east is shown in Table 7-9.
### Table 7-9 Modified IPF Method with Pseudo Columns and Rows

| destination | beyond PINE | PINE-CICERO | CICEROPULASKI | PULASKIKEDZIE | KEDZIEWESTERN | Transfer to 49 N | Transfer to 49 S | Transfer to X49 N | Transfer to X49 S | WESTERN ASHLAND | ASHLAND CANAL | CANAL LASALLE | LASALLE WABASH | WABASH STETSON | beyond STETSON | Sum | factor |
|-------------|-------------|-------------|---------------|---------------|---------------|----------------|----------------|----------------|----------------|----------------|---------------|--------------|--------------|---------------|---------------|-------------|-------|--------|
| beyond PINE | 11          | 28          | 34            | 12            | 45            | 3             | 3             | 0              | 0              | 24             | 16            | 13           | 14           | 7             | 5             | 214        | 3.19   |
| PINE-CICERO | 58          | 131         | 47            | 75            | 7             | 5             | 3             | 2              | 41             | 15             | 13            | 12           | 10           | 3             | 423         | 3.05   |
| CICERO-PULASKI | 65    | 56          | 104           | 6             | 4             | 10            | 3             | 55             | 21             | 10             | 14            | 2             | 3             | 354          | 2.36   |
| PULASKI-KEDZIE | 14    | 41          | 0             | 3             | 1             | 1             | 40            | 9              | 7              | 11             | 3             | 3             | 134          | 1.61   |
| KEDZIEWESTERN | 46    | 4           | 1             | 8             | 2             | 63            | 21            | 10             | 10             | 6              | 3             | 173          | 1.88         |              |             |
| Transfer from 49 N | 3  | 1           | 0             | 1             | 1             | 0             | 7             | 1              | 0              | 8              | 1             | 0             | 1            | 0             | 1.00        |
| Transfer from 49 S | 4  | 2           | 0             | 0             | 0             | 0             | 8             | 1.00           |               |                |               |               |               |               |             |
| Transfer from X49 N | 2  | 2           | 0             | 0             | 0             | 0             | 6             | 1.00           |               |                |               |               |               |               |             |
| Transfer from X49 S | 2  | 1           | 0             | 1             | 0             | 0             | 5             | 1.00           |               |                |               |               |               |               |             |
| WESTERNASHLAND | 24  | 35          | 37            | 34            | 7             | 9             | 146           | 3.18          |               |                |               |               |               |               |             |
| ASHLANDCANAL | 77  | 115         | 155           | 69            | 34            | 451           | 2.72          |               |                |                |               |               |               |               |             |
| CANALLASALLE | 1   | 73          | 39            | 10            | 124           | 2.32          |               |                |                |                |               |               |               |               |             |
| LASALLEWABASH | 14  | 24          | 4             | 43            | 1.90          |               |                |                |                |                |               |               |               |               |             |
| WABASHSTETSON | 8   | 1           | 9             | 2.49          |               |                |                |                |                |                |               |               |               |               |             |
| beyond STETSON | 1   | 0           | 0             | 1.00          |               |                |                |                |                |                |               |               |               |               |             |
| Sum         | 11  | 86          | 231           | 129           | 317           | 20            | 16            | 22             | 7              | 258            | 199           | 208           | 340          | 176          | 77          | 2096 |
| factor      | 1.00| 1.99        | 2.25          | 2.38          | 4.42          | 1.31          | 1.00          | 1.88           | 1.00           | 4.18           | 1.45          | 2.35          | 1.00         | 5.29         | 1.00        |

#### 7.4.4 Comparison with the results from proportional distribution

The major differences between the two methods are (1) the difference in basic assumptions, and (2) the effect of the seed matrix. The proportional distribution method assumes that the passenger OD behavior is the same for transfer passengers and non-transfer passengers while the modified IPF method does not require this assumption to be true (see Section 3.3.2). A seed matrix including transfer route-dirs as both pseudo-origins and pseudo-destinations...
provides a sample of actual passenger OD behavior. Under the modified IPF method, we can fully utilize this information to obtain an OD matrix that is closer to the actual passenger behavior rather than assuming the transfer flows follow exactly the same pattern as the total OD flows. On the other hand, we will need very large sample size in order to avoid the potential biases in the modified IPF method that are due to non-structural zeros.

The transfer flows from Route # 20 east to Route # 49 north obtained using the proportional distribution method and using the modified IPF method are shown in Table 7-10 below. There are similarities in the results, but significant differences also exist. The most notable is the flow from Pulaski-Kedzie transferring to Route # 49 north. Here, the modified IPF method gives a flow of zero because the seed matrix has a zero in this cell. To avoid this kind of problem, we will need to use significantly more data (for example, one month of data) or use another method to eliminate the non-structural zeros. Otherwise, the proportional distribution method is preferred over modified IPF method.

Table 7-10 Transfer Flow Results Comparison

<table>
<thead>
<tr>
<th>From</th>
<th>Proportional distribution</th>
<th>Modified IPF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Transfer to</td>
<td>To Route #49 North</td>
<td>To Route #49 North</td>
</tr>
<tr>
<td>beyond PINE</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>PINE - CICERO</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td>CICERO - PULASKI</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>PULASKI - KEDZIE</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>KEDZIE - WESTERN</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

7.5 Transfer Flow OD Matrices

For the transfer flows between every route-dir pair, we need to obtain a transfer flow matrix. The results obtained in Section 7.2 or 7.3 (for example Table 7-5 and Table 7-6) can be used as the marginal values of this IPF process. For the transfer flows from Route # 20 east to Route # 49 north, please see Table 7-11 below for the results of IPF with a null seed matrix. (The seed matrix here will likely have more non-structural zeros and its costs likely outweigh its advantages.)
Table 7.11 Transfer Flow OD Matrix

<table>
<thead>
<tr>
<th>Origin (on #20 east)</th>
<th>Harrison - Madison</th>
<th>Madison - Grand</th>
<th>Grand - Division</th>
<th>Division - Armitage</th>
<th>Armitage - Diversey</th>
<th>Diversey - Addison</th>
<th>Addison - Montrose</th>
<th>Montrose - Lawrence</th>
<th>Lawrence - Foster</th>
<th>Beyond Foster</th>
<th>Sum</th>
<th>Row Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond PINE</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0.61</td>
<td></td>
</tr>
<tr>
<td>PINE - CICERO</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>8</td>
<td>1.25</td>
</tr>
<tr>
<td>CICERO - PULASKI</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>1.52</td>
</tr>
<tr>
<td>PULASKI - KEDZIE</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0.69</td>
</tr>
<tr>
<td>KEDZIE - WESTERN</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>0.68</td>
</tr>
<tr>
<td>Sum</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>2</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>Column factor</td>
<td>0.23</td>
<td>0.66</td>
<td>0.62</td>
<td>1.66</td>
<td>0.42</td>
<td>1.54</td>
<td>0.49</td>
<td>0.23</td>
<td>0.27</td>
<td>0.22</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Rounding error may exist.

This transfer-flow-only OD matrix can be represented in graphical forms by the desire lines as shown in Figure 7-2. The horizontal bus route is Route #20, and the vertical one is part of Route #49. The square boxes represent the boundaries of segment groups analyzed. Transfer flows of less than 2 are not shown. We can see the transfer flow between the two closest segment groups is negligible, and this is explained by the facts that there are few boardings in the originating segment group and some passengers may walk to Route #49. The transfer flows between segment groups that are far apart are low because these long distance trips would take very long by bus, and the destination segment groups on Route #49 are well served by CTA rail system.
Figure 7-2 Transfer Flows from Route # 20 East to Routes # 49/X49 North

Note: The boardings and alightings are by stop and the transfer flows by segment groups indicated by black boxes.

After the estimation of transfer flow OD matrices between all route-dirs in the demonstration corridor, we can form the network level matrix by combining them with the non-transfer single-bus-route OD matrices to form a network level OD matrix for the demonstration corridor.
Chapter 8

Summary and Recommendations

This Chapter begins with a discussion of potential bias (Section 8.1) and then presents recommendations on the various options tested during the OD estimation process (Section 8.2). Topics for further research are then proposed (Section 8.3).

8.1 Further Discussion of Key Results

8.1.1 Implication of network level OD matrix estimation

To be able to estimate single-bus-route OD matrices is already a significant improvement over conventional practice, but the ability to estimate a network-level OD matrix provides significant additional value to a transit agency. The benefits of estimating a network-level OD matrix include but are not necessarily limited to:

1. The transit agency can identify the critical transfer points and the critical transfer directions where improvements of passenger transfer experiences such as coordinated transfer or improved stop locations to minimize transfer walking distance will have the greatest value.

2. Better understanding of the transfer flows can provide the information for better operations control strategies. For example, the decision to hold a bus should consider the likelihood that on-board passengers may need to make timed transfers downstream.

3. The availability of a network level OD matrix provides a more rational basis for network structure redesign and improvement.
8.1.2 Discussion of Potential Bias

The estimation processes in this thesis does not always preserve the randomness of the samples, and there is a potential that such loss of randomness may create bias. Normally, this would call for comprehensive validation in order to use this model in production. However, such validation is usually not practical, as an extensive OD survey costs a great deal in both money and in time (NYC spent $1.5 million and 2 years in 1990) (Barry 2004). In fact, the purpose of this thesis is to provide an alternative to agencies to conducting expensive and time consuming system-wide passenger surveys which are likely to exhibit their own substantial biases and/or inaccuracies.

Most agencies are unable or unwilling to devote the resources needed to obtain a system-wide OD matrix, and even if an agency has devoted such resources, the survey is almost certain to be less comprehensive and contain more bias than the OD estimation results following the process described in this thesis. The likelihood that a trip is represented in the seed matrix is related to factors such as a passenger's likelihood to use a farecard for more than one trip. Such factors are not believed to be strongly correlated to the passenger OD behavior. In addition, the comprehensiveness of the ADC system data (30-40% sample in the case of CTA) limits any potential bias in the results.

This bus OD matrix estimation process unlikely to produce perfect results, but it is far superior to the traditional bus OD matrix estimation process using manually collected data, and because of the relatively large sample sizes used is likely to exhibit less bias in the resulting OD matrix.

8.2 Recommendations

8.2.1 Advantage of MLE

As discussed in Section 6.3.3, MLE method provides significant advantage over IPF method because of its less stringent data requirements. If the control totals are estimated total boarding and alighting counts, the IPF method will treat them as data sources with a high degree of confidence. This may be a suitable assumption for the traditional way of estimating an OD matrix using manually collected data where the total boarding and alighting counts are
far more comprehensive and are subject to fewer biases than OD passenger surveys. When using ADC system data to estimate an OD matrix, on the other hand, the estimated total boarding and alighting counts are not far superior statistically to the seed matrix, and assuming the boarding and alighting counts are the results of observations of random events has a far better theoretical basis.

However, the IPF method is acceptable if there is a way to obtain the total boarding and alighting counts accurately, or if the objective is only to obtain the relative weights of different OD flows. Since the high sampling rate of seed matrix sufficiently captured the pattern of passenger OD behavior, the relative weights of OD flows are almost the same as estimated through MLE, and the computational simplicity will then become the determining factor.

8.2.2 When to use modified IPF for transfer flows

Due to the low transfer OD flows, modified IPF method is not recommended unless the non-structural zeros can be largely eliminated by using more days’ of data or by combining segments further into larger segments (reducing the geographic detail as a result). In those cases, modified IPF method will be superior because the actual passenger transfers are taken into consideration in the formation of a transfer seed matrix. However, in other cases, proportional distribution should be used to determine the origins and destinations of transfer flows.

8.3 Potential Future Research

8.3.1 Bus/Rail network OD estimation

For a multi-modal transit agency, a comprehensive multi-modal system-wide OD matrix will provide significant advantage over the separate rail system and bus system OD matrices. Rail system OD matrix estimation using ADC system data has been discussed in Barry et al (2002), Zhao (2004), Rahbee and Czerwinski (2004) and Zhao, Wilson and Rahbee (2006). To combine the rail system OD matrix with the bus system OD matrix requires the determination of the origins and destinations of the inter-modal transfer passenger flows. The inter-modal transfer flows are only previously considered in the context of the single mode OD matrix.
estimations. The determination of the origins and destinations of the inter-modal transfers should be the subject of future research as a natural extension of this research.

8.3.2 Validation design

Comprehensive validation of the estimation process in this thesis is impractical, but limited confirmations of the results of this estimation process can be carried out. Future research can be conducted to compare portions of the results of this estimation process with other available data.

8.3.3 Integrating smart card data to enhance the OD matrix estimation and interpretation

Smart card data can provide much more intensive information regarding the travel behavior of an individual passenger. It may be linked to the passenger's residential address and the agency may obtain the passenger's socio-economic information. Smart cards are kept much longer than other farecards, and can be used to study passenger travel behavior changes over time. However, in the US, the smart card users are often a very small proportion of all transit users and may be non-representative. Combining the more intensive smart card data with other more comprehensive ADC system data can provide a deeper understanding of transit passenger behavior and help to interpret the estimated OD matrix.

8.3.4 Using MLE technique at network level

Instead of the sequential estimation approach developed in this thesis, the network level OD matrix can be estimated directly based on all available data. The result of this direct approach would be the OD matrix that theoretically can best explain all the observed data.

8.3.5 Applying the methodology to other transit agencies

Because different transit agencies have different available data, the OD matrix estimation process described in this thesis needs to be adapted to utilize other agencies’ available data. Depending on the particular situations, there may be different challenges in applying the general estimation process described in this thesis, and further studies should be done to apply the methodology to other transit agencies.
References


Bertsimas, D and J. Tsitsiklis, Introduction to Linear Optimization, Athena Scientific, Belmont, MA, 1997,


Furth, Peter, Discussions, 2005-2006.


Rahbee, Adam and David Czerwinski, Using Entry-Only Automatic Fare Collection Data to Estimate Rail Transit Passenger Flows at CTA. *Proceedings of the 2002 Transport Chicago Conference*, 2002


TransInfo, LLC., CTA Rail Passenger Mile Estimates for NTD Reporting, November, 2004


Appendix

A. Single route estimation results for Routes # 20, # 49, # X49, # 126 of CTA
(All tables are daily average for 6-9AM estimated using 5 weekdays of data)

1. Route # 20 eastbound

<table>
<thead>
<tr>
<th>origin</th>
<th>beyond PINE</th>
<th>PINE-CICERO</th>
<th>CICERO-PULASKI</th>
<th>PULASKI-KEDZIE</th>
<th>KEDZIE-WESTERN</th>
<th>WESTERN-ASHLAND</th>
<th>ASHLAND-CANAL</th>
<th>CANAL-LASALLE</th>
<th>LASALLE-WABASH</th>
<th>WABASH-STETSON</th>
<th>beyond STETSON</th>
<th>Total Boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond PINE</td>
<td>7</td>
<td>43</td>
<td>54</td>
<td>14</td>
<td>49</td>
<td>18</td>
<td>8</td>
<td>13</td>
<td>1</td>
<td>4</td>
<td></td>
<td>223</td>
</tr>
<tr>
<td>PINE-CICERO</td>
<td>39</td>
<td>147</td>
<td>54</td>
<td>83</td>
<td>41</td>
<td>22</td>
<td>10</td>
<td>21</td>
<td>7</td>
<td>4</td>
<td></td>
<td>428</td>
</tr>
<tr>
<td>CICERO-PULASKI</td>
<td>39</td>
<td>58</td>
<td>114</td>
<td>74</td>
<td>28</td>
<td>13</td>
<td>19</td>
<td>14</td>
<td>4</td>
<td></td>
<td></td>
<td>363</td>
</tr>
<tr>
<td>PULASKI-KEDZIE</td>
<td>6</td>
<td>46</td>
<td>38</td>
<td>15</td>
<td>10</td>
<td>19</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>140</td>
</tr>
<tr>
<td>KEDZIE-WESTERN</td>
<td></td>
<td>41</td>
<td>79</td>
<td>22</td>
<td>10</td>
<td>18</td>
<td>8</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>181</td>
</tr>
<tr>
<td>WESTERN-ASHLAND</td>
<td></td>
<td></td>
<td>21</td>
<td>41</td>
<td>21</td>
<td>42</td>
<td>19</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>148</td>
</tr>
<tr>
<td>ASHLAND-CANAL</td>
<td></td>
<td></td>
<td>68</td>
<td>131</td>
<td>145</td>
<td>78</td>
<td>31</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
<td>453</td>
</tr>
<tr>
<td>CANAL-LASALLE</td>
<td></td>
<td></td>
<td>13</td>
<td>75</td>
<td>20</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>129</td>
</tr>
<tr>
<td>LASALLE-WABASH</td>
<td></td>
<td></td>
<td>4</td>
<td>25</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>41</td>
</tr>
<tr>
<td>WABASH-STETSON</td>
<td></td>
<td></td>
<td>7</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beyond STETSON</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>Total Alightings</td>
<td>7</td>
<td>82</td>
<td>240</td>
<td>132</td>
<td>333</td>
<td>271</td>
<td>208</td>
<td>216</td>
<td>356</td>
<td>183</td>
<td>86</td>
<td>2114</td>
</tr>
</tbody>
</table>
### 2. Route # 20 westbound

<table>
<thead>
<tr>
<th>origin</th>
<th>beyond Wabash</th>
<th>Wabash - State</th>
<th>State - Wells</th>
<th>Wells - Jefferson</th>
<th>Jefferson - Ashland</th>
<th>Ashland - Western</th>
<th>Western - Kedzie</th>
<th>Kedzie - Pulaski</th>
<th>Pulaski - Laramie</th>
<th>beyond Laramie</th>
<th>Total boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond Wabash</td>
<td>4</td>
<td>2</td>
<td>8</td>
<td>67</td>
<td>35</td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>128</td>
</tr>
<tr>
<td>Wabash - State</td>
<td>6</td>
<td>6</td>
<td>66</td>
<td>58</td>
<td>12</td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>173</td>
</tr>
<tr>
<td>State - Wells</td>
<td>20</td>
<td>196</td>
<td>128</td>
<td>16</td>
<td>10</td>
<td>6</td>
<td>10</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>390</td>
</tr>
<tr>
<td>Wells - Jefferson</td>
<td>68</td>
<td>116</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>212</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jefferson - Ashland</td>
<td>17</td>
<td>30</td>
<td>18</td>
<td>12</td>
<td>9</td>
<td>9</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ashland - Western</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>21</td>
<td>36</td>
<td>51</td>
<td>35</td>
<td>14</td>
</tr>
<tr>
<td>Western - Kedzie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>27</td>
<td>52</td>
<td>41</td>
<td>23</td>
<td>141</td>
</tr>
<tr>
<td>Kedzie - Pulaski</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>26</td>
<td>58</td>
<td>37</td>
<td>121</td>
<td></td>
</tr>
<tr>
<td>Pulaski - Laramie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>96</td>
<td>188</td>
<td>284</td>
<td></td>
</tr>
<tr>
<td>beyond Laramie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>91</td>
<td>91</td>
<td></td>
</tr>
<tr>
<td>Total alightings</td>
<td>4</td>
<td>9</td>
<td>34</td>
<td>398</td>
<td>354</td>
<td>90</td>
<td>105</td>
<td>159</td>
<td>263</td>
<td>377</td>
<td>1792</td>
</tr>
</tbody>
</table>
### 3. Route # 49 northbound

<table>
<thead>
<tr>
<th>Destination</th>
<th>beyond 78th</th>
<th>78th - 63rd</th>
<th>63rd - Garfield</th>
<th>Garfield - 47th</th>
<th>Archer - 26th</th>
<th>26th - 16th</th>
<th>16th - Harrison</th>
<th>Madison - Grand</th>
<th>Grand - Division</th>
<th>Division - Armitage</th>
<th>Armitage - Diversey</th>
<th>Diversey - Addison</th>
<th>Addison - Montrose</th>
<th>Montrose - Lawrence</th>
<th>Lawrence - Foster</th>
<th>beyond Foster</th>
<th>Total Boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond 78th</td>
<td>1</td>
<td>16</td>
<td>11</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>52</td>
</tr>
<tr>
<td>78th - 63rd</td>
<td>8</td>
<td>29</td>
<td>52</td>
<td>6</td>
<td>3</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>121</td>
</tr>
<tr>
<td>63rd - Garfield</td>
<td>7</td>
<td>46</td>
<td>9</td>
<td>5</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>6</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>94</td>
</tr>
<tr>
<td>Garfield - 47th</td>
<td>42</td>
<td>41</td>
<td>11</td>
<td>16</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>5</td>
<td>18</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>158</td>
</tr>
<tr>
<td>47th - Archer</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>9</td>
<td>7</td>
<td>6</td>
<td>20</td>
<td>4</td>
<td>32</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>94</td>
<td>4</td>
<td>0</td>
<td>94</td>
</tr>
<tr>
<td>Archer - 26th</td>
<td>5</td>
<td>16</td>
<td>6</td>
<td>7</td>
<td>5</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>63</td>
<td>1</td>
<td>0</td>
<td>63</td>
</tr>
<tr>
<td>26th - 16th</td>
<td>6</td>
<td>6</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>15</td>
<td>2</td>
<td>12</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>59</td>
<td>1</td>
<td>59</td>
</tr>
<tr>
<td>16th - Harrison</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>12</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Harrison - Madison</td>
<td></td>
<td></td>
<td>2</td>
<td>6</td>
<td>6</td>
<td>14</td>
<td>4</td>
<td>13</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Madison - Grand</td>
<td></td>
<td></td>
<td>3</td>
<td>6</td>
<td>25</td>
<td>1</td>
<td>9</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grand - Division</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>60</td>
<td>13</td>
<td>59</td>
<td>11</td>
<td>4</td>
<td>3</td>
<td>8</td>
<td>160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Division - Armitage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armitage - Diversey</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diversey - Addison</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addison - Montrose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montrose - Lawrence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lawrence - Foster</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beyond Foster</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total alightings</td>
<td>1</td>
<td>23</td>
<td>48</td>
<td>148</td>
<td>61</td>
<td>26</td>
<td>52</td>
<td>24</td>
<td>42</td>
<td>37</td>
<td>41</td>
<td>242</td>
<td>62</td>
<td>369</td>
<td>101</td>
<td>77</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Origin</td>
<td>Beyond 7th</td>
<td>63rd - 7th</td>
<td>Garfield - 47th</td>
<td>35th - 47th</td>
<td>35th - Garfield</td>
<td>16th - Garfield</td>
<td>47th - Garfield</td>
<td>63rd - 7th</td>
<td>Total Allights</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------------</td>
<td>------------</td>
<td>-----------------</td>
<td>-------------</td>
<td>----------------</td>
<td>----------------</td>
<td>----------------</td>
<td>------------</td>
<td>---------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foster</td>
<td>0</td>
<td>23</td>
<td>11</td>
<td>8</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lawrence</td>
<td>0</td>
<td>2</td>
<td>19</td>
<td>6</td>
<td>28</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lawmont</td>
<td>4</td>
<td>12</td>
<td>39</td>
<td>7</td>
<td>10</td>
<td>9</td>
<td>3</td>
<td>1</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montrose</td>
<td>5</td>
<td>13</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Addison</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diversey</td>
<td>9</td>
<td>81</td>
<td>15</td>
<td>4</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amelage</td>
<td>10</td>
<td>21</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Armour</td>
<td>11</td>
<td>61</td>
<td>15</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chicago</td>
<td>7</td>
<td>23</td>
<td>17</td>
<td>2</td>
<td>5</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>North</td>
<td>17</td>
<td>21</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>11</td>
<td>21</td>
<td>14</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foster</td>
<td>0</td>
<td>3</td>
<td>46</td>
<td>34</td>
<td>34</td>
<td>34</td>
<td>46</td>
<td>34</td>
<td>124</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4. Route #49 southbound
<table>
<thead>
<tr>
<th>Origin</th>
<th>Total</th>
<th>78th</th>
<th>78th - 59th</th>
<th>59th - 51st</th>
<th>51st - Archer</th>
<th>Archer - 28th</th>
<th>28th - 18th</th>
<th>18th - Grand</th>
<th>Grand - Division</th>
<th>Division - Fullerton</th>
<th>Fullerton - Belmont</th>
<th>Belmont - Irving Park</th>
<th>Irving Park - Leland</th>
<th>Leland - Foster</th>
<th>beyond Foster</th>
<th>Total alightings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foster</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>36</td>
</tr>
<tr>
<td>Leland - Pigott Park</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>Beyond</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>36</td>
</tr>
</tbody>
</table>

5. Route # X49 northbound
### Route # X49 Southbound

<table>
<thead>
<tr>
<th>Destination</th>
<th>Beyond Foster</th>
<th>Foster - Leland</th>
<th>Leland - Irving Park</th>
<th>Irving Park - Belmont</th>
<th>Belmont - Fullerton</th>
<th>Fullerton - Division</th>
<th>Division - Grand</th>
<th>Grand - Jackson</th>
<th>Jackson - 16th</th>
<th>16th - 26th</th>
<th>26th - Archer</th>
<th>Archer - 51st</th>
<th>51st - 59th</th>
<th>59th - 87th</th>
<th>Beyond 87th</th>
<th>Total Boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beyond Foster</td>
<td>0</td>
<td>3</td>
<td>17</td>
<td>43</td>
<td>12</td>
<td>11</td>
<td>14</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>113</td>
</tr>
<tr>
<td>Foster - Leland</td>
<td>4</td>
<td>8</td>
<td>24</td>
<td>9</td>
<td>12</td>
<td>11</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>76</td>
</tr>
<tr>
<td>Leland - Irving Park</td>
<td>1</td>
<td>11</td>
<td>8</td>
<td>7</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>43</td>
</tr>
<tr>
<td>Irving Park - Belmont</td>
<td>6</td>
<td>4</td>
<td>9</td>
<td>12</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>45</td>
</tr>
<tr>
<td>Belmont - Fullerton</td>
<td>4</td>
<td>33</td>
<td>18</td>
<td>6</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>76</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fullerton - Division</td>
<td>29</td>
<td>94</td>
<td>15</td>
<td>16</td>
<td>13</td>
<td>3</td>
<td>10</td>
<td>3</td>
<td>5</td>
<td>0</td>
<td>190</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Division - Grand</td>
<td>10</td>
<td>17</td>
<td>19</td>
<td>6</td>
<td>5</td>
<td>7</td>
<td>1</td>
<td>7</td>
<td>3</td>
<td>3</td>
<td>74</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grand - Jackson</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jackson - 16th</td>
<td>5</td>
<td>7</td>
<td>1</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>2</td>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16th - 26th</td>
<td>2</td>
<td>6</td>
<td>22</td>
<td>1</td>
<td>12</td>
<td>3</td>
<td>46</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26th - Archer</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Archer - 51st</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>18</td>
<td>13</td>
<td>36</td>
<td>5</td>
<td>72</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>51st - 59th</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>12</td>
<td>2</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>59th - 87th</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>53</td>
<td>25</td>
<td>78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beyond 87th</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total Alightings</td>
<td>0</td>
<td>8</td>
<td>27</td>
<td>84</td>
<td>37</td>
<td>101</td>
<td>165</td>
<td>51</td>
<td>61</td>
<td>46</td>
<td>23</td>
<td>80</td>
<td>26</td>
<td>140</td>
<td>44</td>
<td>893</td>
</tr>
</tbody>
</table>
7. Route # 126 eastbound

<table>
<thead>
<tr>
<th>origin</th>
<th>beyond Central</th>
<th>Central - Cicero</th>
<th>Cicero - Homan</th>
<th>Homan - California</th>
<th>California - Damen</th>
<th>Damen - Ogden</th>
<th>Ogden - Halsted</th>
<th>Halsted - Clinton</th>
<th>Clinton - Franklin</th>
<th>Franklin - Dearborn</th>
<th>Dearborn - Michigan</th>
<th>beyond Michigan</th>
<th>Total boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond Central</td>
<td>22</td>
<td>9</td>
<td>10</td>
<td>16</td>
<td>18</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>92</td>
</tr>
<tr>
<td>Central - Cicero</td>
<td>29</td>
<td>56</td>
<td>59</td>
<td>99</td>
<td>34</td>
<td>22</td>
<td>2</td>
<td>13</td>
<td>18</td>
<td>21</td>
<td>1</td>
<td>3</td>
<td>354</td>
</tr>
<tr>
<td>Cicero - Homan</td>
<td>66</td>
<td>93</td>
<td>166</td>
<td>46</td>
<td>14</td>
<td>8</td>
<td>20</td>
<td>18</td>
<td>53</td>
<td>1</td>
<td>485</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homan - California</td>
<td>46</td>
<td>66</td>
<td>20</td>
<td>20</td>
<td>10</td>
<td>8</td>
<td>12</td>
<td>17</td>
<td>15</td>
<td>1</td>
<td>196</td>
<td></td>
<td></td>
</tr>
<tr>
<td>California - Damen</td>
<td>33</td>
<td>20</td>
<td>12</td>
<td>12</td>
<td>12</td>
<td>16</td>
<td>15</td>
<td>23</td>
<td>1</td>
<td>1</td>
<td>132</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Damen - Ogden</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>12</td>
<td>12</td>
<td>13</td>
<td>14</td>
<td>3</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ogden - Halsted</td>
<td>2</td>
<td>11</td>
<td>28</td>
<td>22</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>85</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halsted - Clinton</td>
<td>4</td>
<td>7</td>
<td>6</td>
<td>12</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinton - Franklin</td>
<td>6</td>
<td>12</td>
<td>91</td>
<td>19</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>129</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Franklin - Dearborn</td>
<td>1</td>
<td>9</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dearborn - Michigan</td>
<td>4</td>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>beyond Michigan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total sightings</td>
<td>22</td>
<td>38</td>
<td>132</td>
<td>214</td>
<td>382</td>
<td>125</td>
<td>67</td>
<td>46</td>
<td>118</td>
<td>124</td>
<td>267</td>
<td>33</td>
<td>1570</td>
</tr>
<tr>
<td>-----------------------</td>
<td>-----------------</td>
<td>------------------</td>
<td>---------------</td>
<td>----------------</td>
<td>-----------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>----------------</td>
<td>-------------------</td>
<td>-------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>---------------</td>
</tr>
<tr>
<td>Beyond Michigan</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Michigan - State</td>
<td>17</td>
<td>16</td>
<td>109</td>
<td>106</td>
<td>27</td>
<td>10</td>
<td>53</td>
<td>32</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>State - Clark</td>
<td>4</td>
<td>27</td>
<td>52</td>
<td>17</td>
<td>5</td>
<td>13</td>
<td>7</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Clark - Wacker</td>
<td>9</td>
<td>34</td>
<td>9</td>
<td>4</td>
<td>17</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>19</td>
</tr>
<tr>
<td>Wacker - Clinton</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Jackson - Van Buren</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(on Clinton)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clinton - Halsted</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>24</td>
</tr>
<tr>
<td>(on Van Buren)</td>
<td>0</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Halsted - Damen</td>
<td>41</td>
<td>28</td>
<td>10</td>
<td>6</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>12</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>91</td>
</tr>
<tr>
<td>Damen - California</td>
<td>6</td>
<td>34</td>
<td>18</td>
<td>8</td>
<td>4</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>California - Homan</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>21</td>
<td>7</td>
<td>13</td>
<td>52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homan - Cicero</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>33</td>
<td>22</td>
<td>46</td>
<td></td>
<td>101</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cicero - Laramie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>20</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beyond Laramie</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total alightings</td>
<td>0</td>
<td>19</td>
<td>20</td>
<td>151</td>
<td>203</td>
<td>56</td>
<td>22</td>
<td>163</td>
<td>83</td>
<td>62</td>
<td>91</td>
<td>43</td>
<td>104</td>
</tr>
</tbody>
</table>
B. Transfer flow OD matrices between Routes # 20, # 49, # X49, # 126 of CTA
(Results are presented in the order of decreasing total transfer flow. Rounding error may be present.)

1. Route # 20 eastbound to Route # 49 northbound

<table>
<thead>
<tr>
<th>destination (on #49 north)</th>
<th>beyond PINE</th>
<th>PINE - CICERO</th>
<th>CICERO - PULASKI</th>
<th>PULASKI - KEDZIE</th>
<th>KEDZIE - WESTERN</th>
<th>Transfer alightings</th>
</tr>
</thead>
<tbody>
<tr>
<td>origin (on #20 east)</td>
<td>beyond PINE</td>
<td>PINE - CICERO</td>
<td>CICERO - PULASKI</td>
<td>PULASKI - KEDZIE</td>
<td>KEDZIE - WESTERN</td>
<td>Transfer boardings</td>
</tr>
<tr>
<td>beyond PINE</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>PINE - CICERO</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>CICERO - PULASKI</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>PULASKI - KEDZIE</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>KEDZIE - WESTERN</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Transfer alightings</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>8</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

2. Route # 20 eastbound to Route # X49 northbound

<table>
<thead>
<tr>
<th>origin</th>
<th>beyond PINE</th>
<th>PINE - CICERO</th>
<th>CICERO - PULASKI</th>
<th>PULASKI - KEDZIE</th>
<th>KEDZIE - WESTERN</th>
<th>Transfer boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond PINE</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>PINE - CICERO</td>
<td>0</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>CICERO - PULASKI</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>PULASKI - KEDZIE</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>KEDZIE - WESTERN</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Transfer alightings</td>
<td>2</td>
<td>2</td>
<td>11</td>
<td>6</td>
<td>5</td>
<td>2</td>
</tr>
</tbody>
</table>

103
3. Route # X49 northbound to Route # 20 westbound

<table>
<thead>
<tr>
<th>destination</th>
<th>Western - Kedzie</th>
<th>Kedzie - Pulaski</th>
<th>Pulaski - Laramie</th>
<th>beyond Laramie</th>
<th>Transfer boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond 78th</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>78th - 59th</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>59th - 51st</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>51st - Archer</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Archer - 26th</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>26th - 16th</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>16th - Jackson</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Jackson - Grand</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Transfer alightings</td>
<td>4</td>
<td>8</td>
<td>6</td>
<td>3</td>
<td>21</td>
</tr>
</tbody>
</table>

4. Route # 49 northbound to Route # 20 westbound

<table>
<thead>
<tr>
<th>destination</th>
<th>Western - Kedzie</th>
<th>Kedzie - Pulaski</th>
<th>Pulaski - Laramie</th>
<th>beyond Laramie</th>
<th>Transfer boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond 78th</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>78th - 63rd</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>63rd - Garfield</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Garfield - 47th</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>47th - Archer</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Archer - 26th</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>26th - 16th</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>16th - Harrison</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Harrison - Madison</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Transfer alightings</td>
<td>2</td>
<td>4</td>
<td>6</td>
<td>4</td>
<td>18</td>
</tr>
</tbody>
</table>
5. Route # 126 eastbound to Route # X49 northbound

<table>
<thead>
<tr>
<th>destination</th>
<th>Jackson</th>
<th>Grand - Division</th>
<th>Division - Fullerton</th>
<th>Fullerton - Belmont</th>
<th>Belmont - Irving Park</th>
<th>Irving Park - Leland</th>
<th>Leland - Foster</th>
<th>beyond Foster</th>
<th>Transfer boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond Central</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Central - Cicero</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Cicero - Homan</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Homan - California</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>California - Damen</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Transfer alightings</td>
<td>1</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>16</td>
</tr>
</tbody>
</table>

6. Route # 20 eastbound to Route # 49 southbound

<table>
<thead>
<tr>
<th>destination</th>
<th>26th - 16th</th>
<th>16th - Jackson</th>
<th>Jackson - Grand</th>
<th>Grand - Division</th>
<th>Division - Fullerton</th>
<th>Fullerton - Belmont</th>
<th>Belmont - Irving Park</th>
<th>Irving Park - Leland</th>
<th>Leland - Foster</th>
<th>beyond Foster</th>
<th>Transfer boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>beyond PINE</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>PINE-CICERO</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>CICERO-PULASKI</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>PULASKI-KEDZIE</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>KEDZIE-WESTERN</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Transfer alightings</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>103</td>
<td>15</td>
</tr>
</tbody>
</table>
7. Route # 49 southbound to Route # 20 westbound

<table>
<thead>
<tr>
<th>Origin</th>
<th>Ashland - Western</th>
<th>Western - Kedzie</th>
<th>Kedzie - Pulaski</th>
<th>Pulaski - Laramie</th>
<th>Beyond Laramie</th>
<th>Transfer Boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beyond Foster</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Foster - Lawrence</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Lawrence - Montrose</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Montrose - Addison</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Addison - Diversey</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Diversey - Armitage</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Armitage - North</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>North - Chicago</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Chicago - Madison</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Transfer alightings</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>13</td>
</tr>
</tbody>
</table>

8. Route # 126 eastbound to Route # 49 northbound

<table>
<thead>
<tr>
<th>Origin</th>
<th>Harrison - Madison</th>
<th>Madison - Grand</th>
<th>Grand - Division</th>
<th>Division - Armitage</th>
<th>Armitage - Diversey</th>
<th>Diversey - Addison</th>
<th>Addison - Montrose</th>
<th>Montrose - Lawrence</th>
<th>Lawrence - Foster</th>
<th>Beyond Foster</th>
<th>Transfer Boardings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beyond Central</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Central - Cicero</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Cicero - Homan</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Homan - California</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>California - Damen</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Transfer alightings</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>12</td>
</tr>
</tbody>
</table>

Other transfer flows are too low, and are omitted here.

106
C. SQL code used to in data preparation

REM *******************************************************************
REM CTA BUS OD Generation
REM ORACLE script
REM Version 2.0
REM Date: 06-06-06
REM Any question on this script, please email: a_cui@hotmail.com
REM This script may need to be modified to suit all set-up environments.
REM *******************************************************************

REM *******************************************************************
REM Part A. Seed matrix
REM Sub-part a. Stop identification
REM *******************************************************************

drop table equip_bus
/
create table equip_bus as
SELECT MIT_M8050.EQUIP_ID, MIT_M8050.BUS_NBR, Min(MIT_M8050.CREATEDLLT) AS Min_Time,
       Max(MIT_M8050.CREATEDLLT) AS Max_Time
FROM MIT_M8050
GROUP BY MIT_M8050.EQUIP_ID, MIT_M8050.BUS_NBR
/

drop table SeedM_1_bus_nbr_all
create table SeedM_1_bus_nbr_all
as SELECT MIT_m8054.SER_NBR, MIT_m8054.USE_LLT, MIT_m8054.CURRENT_ROUTE,
       MIT_m8054.LAST_ROUTE, MIT_m8054.TRANS_EVENT, MIT_m8054.EQUIP_ID,
       EQUIP_BUS.BUS_NBR, 1 bus_flag, 1 trip_flag,
       to_char(use_llt - NUMTODSINTERVAL(3, 'HOUR'), 'DD') as Dayl
FROM MIT_m8054, EQUIP_BUS
WHERE TRANS_EVENT In (36,37,38,39,43,45,46,47,48,49,97,98,99)
       AND EQUIP_BUS.Min_Time<=use_llt
       AND EQUIP_BUS.Max_time>use_llt
       and MIT_m8054.EQUIP_ID = EQUIP_BUS.EQUIP_ID
/

drop table SeedM_3_AFC_AVL_window
create table SeedM_3_AFC_AVL_window
as SELECT SeedM_1_bus_nbr_all.SER_NBR, SeedM_1_bus_nbr_all.USE_LLT,
       SeedM_1_bus_nbr_all.CURRENT_ROUTE, SeedM_1_bus_nbr_all.LAST_ROUTE,
       SeedM_1_bus_nbr_all.TRANS_EVENT, SeedM_1_bus_nbr_all.BUS_NBR,
       EVENT_TIME, bus_flag, trip_flag, Dayl
FROM SeedM_1_bus_nbr_all, bus_statehist
WHERE SeedM_1_bus_nbr_all.USE_LLT>=event_time
       AND SeedM_1_bus_nbr_all.USE_LLT<event_time+numtodsinterval(5, 'MINUTE')
       and SeedM_1_bus_nbr_all.BUS_NBR = BUS_ID
       and event_type in (3,5)
/

drop table SeedM_3_AFC_AVL_precede
create table SeedM_3_AFC_AVL_precede
as SELECT SER_NBR, USE_LLT,
       CURRENT_ROUTE, LAST_ROUTE,
       TRANS_EVENT, BUS_NBR, bus_flag, trip_flag, Dayl,
       max(EVENT_TIME) as max_t
FROM SeedM_3_AFC_AVL_window
  group by SER_NBR, USE_LLT,
             CURRENT_ROUTE, LAST_ROUTE,
             TRANS_EVENT, BUS_NBR, bus_flag, trip_flag, Day1
/
drop table SeedM_3_AFC_AVL_match
/
create table SeedM_3_AFC_AVL_match
as SELECT SER_NBR, USE_LLT,
             CURRENT_ROUTE, LAST_ROUTE,
             TRANS_EVENT, BUS_NBR, bus_flag, trip_flag, Day1,
             event_time, bustools_ver_id, event_type, route_id, pattern,
             stop_sequence, trip_id, dwell_time
FROM SeedM_3_AFC_AVL_precede,bus_state_hist
where event_time= max_t
  and bus_nbr = bus_id
  and (event_type =3 or event_type=5)
/
drop table SeedM_3_AFC_stop
/
create table SeedM_3_AFC_stop
as SELECT SER_NBR, USE_LLT,
             CURRENT_ROUTE, LAST_ROUTE,
             TRANS_EVENT, BUS_NBR, bus_flag, trip_flag, Day1,
             event_time, bustools_ver_id, event_type, a.route_id, p.direction,
             pd.stopid,
             stopdesc, tripid, dwell_time
FROM SeedM_3_AFC_AVL_match a, bt_pattern p, bt_patterndetail pd,
      bt_stopinfo s
where bustools_ver_id = p.bt_ver
  and p.bt_ver = pd.bt_ver
  and pd.bt_ver = s.bt_ver
  and a.pattern = p.pattern
  and p.patternid = pd.patternid
  and a.stop_sequence = pd.stopsortorder
  and pd.stopid = s.stopid
  and trim(a.route_id) = trim(p.route)
/
REM ********************************************************************
REM Sub-part b. Destination inference
REM Add rail transactions
REM ********************************************************************

drop table SeedM_4_1_rail_records
/
create table SeedM_4_1_rail_records
as SELECT MIT_M8054.SER_NBR, MIT_M8054.USE_LLT, MIT_M8054.EQUIP_ID,
        to_char(use_llt - NUMTODSINTERVAL(3, 'HOUR'), 'DD') as Day1,
        MIT_M8054.TRANS_EVENT,
        MIT_M8054.CURRENT_ROUTE, MIT_M8054.LAST_ROUTE,
        MIT_M8054.FARE_MEDIA_TYPE,
        EQUIP_ENTRANCE.ENTRANCEID
FROM MIT_M8054 INNER JOIN EQUIP_ENTRANCE ON MIT_M8054.EQUIP_ID = EQUIP_ENTRANCE.EQUIP_ID
/
drop table SeedM_4_2_rail_sta

108
create table SeedM_4_2_rail_sta
as SELECT SeedM_4_1_rail_records.SER_NBR,
SeedM_4_1_rail_records.USE_LLT,
SeedM_4_1_rail_records.EQUIP_ID, SeedM_4_1_rail_records.Day1,
SeedM_4_1_rail_records.CURRENT_ROUTE,
SeedM_4_1_rail_records.FARE_MEDIA_TYPE,
SeedM_4_1_rail_records.ENTRANCEID, to_number(STATION) AS Sta
FROM SeedM_4_1_rail_records INNER JOIN ENTRANCE_STATION
ON trim(SeedM_4_1_rail_records.ENTRANCEID) =
trim(ENTRANCE_STATION.ENTRANCE)
/
drop table SeedM_4_3_stop2railmod_l1
create table SeedM_4_3_stop2rail_mod_1
as SELECT Rail_STOP2STATION.STATIONID, BT_PATTERNDETAIL.STOPID,
to_number(DIST) AS Distance
FROM (BT_STOP INNER JOIN Rail_STOP2STATION ON BT_STOP.TAGEOID =
Rail_STOP2STATION.STOPID)
INNER JOIN BT_PATTERNDETAIL ON (BT_STOP.GEOID =
BT_PATTERNDETAIL.GEOID)
AND (BT_STOP.BT_VER = BT_PATTERNDETAIL.BT_VER)
WHERE (((BT_STOP.BTVER)=157))
GROUP BY Rail_STOP2STATION.STATIONID, BT_PATTERNDETAIL.STOPID,
to_number(DIST)
/
drop table SeedM_4_4_stop2rail_mod_2
create table SeedM_4_4_stop2rail_mod_2
as SELECT SeedM_4_3_stop2railmod_l.STATIONID,
Min(SeedM_4_3_stop2rail_mod_l.Distance) AS MinOfDIST
FROM SeedM_4_3_stop2rail_mod_1
GROUP BY SeedM_4_3_stop2railmod_l.STATIONID
/
drop table SeedM_4_5_stop2rail_mod
create table SeedM_4_5_stop2rail_mod
as SELECT SeedM_4_3_stop2rail_mod_1.STATIONID,
SeedM_4_3_stop2rail_mod_1.STOPID,
SeedM_4_3_stop2rail_mod_1.Distance
FROM SeedM_4_4_stop2rail_mod_2 INNER JOIN SeedM_4_3_stop2rail_mod_1
ON (SeedM_4_4_stop2rail_mod_2.STATIONID =
SeedM_4_3_stop2rail_mod_1.STATIONID)
AND (SeedM_4_4_stop2rail_mod_2.MinOfDIST =
SeedM_4_3_stop2rail_mod_1.Distance)
GROUP BY SeedM_4_3_stop2rail_mod_1.STATIONID,
SeedM_4_3_stop2rail_mod_1.STOPID,
SeedM_4_3_stop2rail_mod_1.Distance
/
drop table SeedM_4_6_rail_trip_stop
create table SeedM_4_6_rail_trip_stop
as SELECT SeedM_4_2_rail_sta.SER_NBR, SeedM_4_2_rail_sta.USE_LLT,
SeedM_4_2_rail_sta.EQUIP_ID, SeedM_4_2_rail_sta.Day1,
SeedM_4_2_rail_sta.TRANS_EVENT, SeedM_4_2_rail_sta.CURRENT_ROUTE,
SELECT SeedM_3_AFC_stop.SER_NBR, SeedM_3_AFC_stop.DAY1,
       Min(SeedM_3_AFC_stop.USE_LLT) AS MinOfUSE_LLT,
       Max(SeedM_3_AFC_stop.USE_LLT) AS MaxOfUSE_LLT
FROM SeedM_3_AFC_stop
GROUP BY SeedM_3_AFC_stop.SER_NBR, SeedM_3_AFC_stop.DAY1
/

/insert into SeedM_4c_single_trans

CREATE TABLE SeedM_5a_multi_trans_tmp
AS
SELECT * FROM SeedM_4d_single_trans
WHERE ((SeedM_4d_single_trans_join.MinOfUSE_LLT)=MaxOfUse_llt)
/

}
as SELECT SeedM_4d_single_trans_join.SERNBR,
         SeedM_4d_single_trans_join.USE_LLT,
         SeedM_4d_single_trans_join.CURRENT_ROUTE,
         SeedM_4d_single_trans_join.LAST_ROUTE,
         SeedM_4d_single_trans_join.TRANS_EVENT,
         SeedM_4d_single_trans_join.BUS_NBR,
         SeedM_4d_single_trans_join.BUS_FLAG,
         SeedM_4d_single_trans_join.TRIP_FLAG,
         SeedM_4d_single_trans_join.DAY1,
         SeedM_4d_single_trans_join.EVENT_TIME,
         SeedM_4d_single_trans_join.BUSTOOLS_VER_ID,
         SeedM_4d_single_trans_join.EVENT_TYPE,
         SeedM_4d_single_trans_join.ROUTE_ID,
         SeedM_4d_single_trans_join.DIRECTION,
         SeedM_4d_single_trans_join.STOPID,
         SeedM_4d_single_trans_join.STOPDESC,
         SeedM_4d_single_trans_join.TRIP_ID,
         SeedM_4d_single_trans_join.DWELL_TIME
FROM SeedM_4d_single_trans_join
ORDER BY SeedM_4d_single_trans_join.SER_NBR,
         SeedM_4d_single_trans_join.USE_LLT
/
drop sequence id_seq
/
CREATE SEQUENCE ID_seq
  START WITH 1
  INCREMENT BY 1
  maxvalue 100000000
  CYCLE
/
drop table SeedM_5a_multi_trans
/
create table SeedM_5a_multi_trans
as SELECT id_seq.nextval as ID, SER_NBR, USE_LLT,
         CURRENT_ROUTE, LAST_ROUTE,
         TRANS_EVENT, BUS_NBR,
         BUS_FLAG, TRIP_FLAG,
         DAY1, EVENT_TIME,
         BUSTOOLS_VER_ID, EVENT_TYPE,
         ROUTE_ID, DIRECTION,
         STOPID, STOPDESC,
         TRIP_ID, DWELL_TIME
FROM SeedM_5a_multi_trans_tmp
/
drop table SeedM_5a_multi_trans_tmp
/
drop table SeedM_6a_destination
/
create table SeedM_6a_destination
as SELECT SeedM_5a_multi_trans.*, SeedM_5a_multi_trans_1.STOPID AS Dest_stop,
         SeedM_5a_multi_trans_1.STOPDESC AS Dest_stop_desc,
         SeedM_5a_multi_trans_1.ROUTE_ID AS Dest_route
FROM SeedM_5a_multi_trans INNER JOIN SeedM_5a_multi_trans
SeedM_5a_multi_trans_1
ON (SeedM_5a_multi_trans.SER_NBR = SeedM_5a_multi_trans_1.SER_NBR)
   AND (SeedM_5a_multi_trans.DAY1 = SeedM_5a_multi_trans_1.DAY1)
WHERE (((SeedM_5a_multi_trans.ID)=seedM_5a_multi_trans_1.id-1))
/
drop table SeedM_6b_first_last_trip
/
create table SeedM_6b_first_last_trip
as SELECT SeedM_5a_multi_trans.SER_NBR, SeedM_5a_multi_trans.DAY1,
    Min(SeedM_5a_multi_trans.USE_LLT) AS first_trip,
    Max(SeedM_5a_multi_trans.USE_LLT) AS last_trip
FROM SeedM_5a_multi_trans
GROUP BY SeedM_5a_multi_trans.SER_NBR, SeedM_5a_multi_trans.DAY1
/
INSERT INTO SeedM_6a_destination (SER_NBR, USE_LLT, CURRENT_ROUTE,
    LAST_ROUTE, TRANS_EVENT, BUS_NBR, EVENT_TIME, EVENT_TYPE, BUSTOOLS_VER_ID,
    ROUTE_ID, TRIP_ID, DWELL_TIME, DIRECTION, STOPID, STOPDESC, BUS_FLAG,
    TRIP_FLAG, DAY1, ID, Dest_route, Dest_stop, Dest_stopdesc
SELECT SeedM_5a_multi_trans.SER_NBR, SeedM_5a_multi_trans.USE_LLT,
    SeedM_5a_multi_trans.CURRENT_ROUTE,
    SeedM_5a_multi_trans.LASTROUTE,
    SeedM_5a_multi_trans.TRANS_EVENT, SeedM_5a_multi_trans.BUS_NBR,
    SeedM_5a_multi_trans.EVENT_TIME, SeedM_5a_multi_trans.EVENT_TYPE,
    SeedM_5a_multi_trans.BUSTOOLS_VER_ID,
    SeedM_5a_multi_trans.ROUTE_ID,
    SeedM_5a_multi_trans.TRIP_ID, SeedM_5a_multi_trans.DWELL_TIME,
    SeedM_5a_multi_trans.DIRECTION,
    SeedM_5a_multi_trans.STOPID, SeedM_5a_multi_trans.STOPDESC,
    SeedM_5a_multi_trans.BUS_FLAG,
    SeedM_5a_multi_trans.TRIP_FLAG, SeedM_5a_multi_trans.DAY1, SeedM_5a_multi_trans.ID,
    SeedM_5a_multi_trans_1.ROUTE_ID,
    SeedM_5a_multi_trans_l.STOPID, SeedM_5a_multi_trans_l.STOPDESC
FROM SeedM_5a_multi_trans SeedM_5a_multi_trans_l
INNER JOIN (SeedM_5a_multi_trans INNER JOIN SeedM_6b_first_lasttrip
ON (SeedM_5a_multi_trans.SER_NBR =
    SeedM_6b_first_lasttrip.SER_NBR)
AND (SeedM_5a_multi_trans.USE_LLT =
    SeedM_6b_first_lasttrip.last_trip)
AND (SeedM_5a_multi_trans.DAY1 =
    SeedM_6b_first_lasttrip.DAY1))
ON (SeedM_5a_multi_trans_l.SER_NBR =
    SeedM_6b_first_lasttrip.SER_NBR)
AND (SeedM_5a_multi_trans_l.USELLT =
    SeedM_6bfirst_lasttrip.firsttrip)
AND (SeedM_5a_multi_trans_l.DAY1 =
    SeedM_6b_first_last_trip.DAY1)
/
REM ******************************************************
REM Sub-Part c. Destination reassignment
REM ******************************************************
drop table SeedM_7b_dest_reasignment_tmp
/
create table SeedM_7b_dest_reasignment_tmp
as SELECT SeedM_6a_destination.SER_NBR, SeedM_6a_destination.USE_LLT,
    SeedM_6a_destination.CURRENT_ROUTE,
SELECT SeedM_6a_destination.TRANS_EVENT, SeedM_6a_destination.BUS_NBR, SeedM_6a_destination.BUS_FLAG, SeedM_6a_destination.TRIP_FLAG, SeedM_6a_destination.DAY1, SeedM_6a_destination.EVENT_TIME, SeedM_6a_destination.BUSTOOLS_VER_ID, SeedM_6a_destination.EVENT_TYPE, SeedM_6a_destination.ROUTE_ID, SeedM_6a_destination.DIRECTION, SeedM_6a_destination.DELT_TIME, SeedM_6a_destination.ID, SeedM_6a_destination.Dest_stop, SeedM_6a_destination.Dest_stop_desc, SeedM_6a_destination.Dest_route, Min(GEO_2A_WALK_MATCH.DIST) AS MinOfDIST FROM SeedM_6a_destination INNER JOIN GEO_2A_WALK_MATCH ON (SeedM_6a_destination.Dest_stop = GEO_2A_WALK_MATCH.STOPID) AND (SeedM_6a_destination.DIRECTION = GEO_2A_WALK_MATCH.D2) AND (trim(SeedM_6a_destination.ROUTE_ID) = GEO_2A_WALK_MATCH.ROUTE2) AND (SeedM_6a_destination.BUSTOOLS_VER_ID = GEO_2A_WALK_MATCH.BT_VER) WHERE (((SeedM_6a_destination.CURRENT_ROUTE) < 1024)) GROUP BY SeedM_6a_destination.SER_NBR, SeedM_6a_destination.USE_LLT, SeedM_6a_destination.CURRENT_ROUTE, SeedM_6a_destination.LAST_ROUTE, SeedM_6a_destination.TRANS_EVENT, SeedM_6a_destination.BUS_NBR, SeedM_6a_destination.BUS_FLAG, SeedM_6a_destination.TRIP_FLAG, SeedM_6a_destination.DAY1, SeedM_6a_destination.EVENT_TIME, SeedM_6a_destination.BUSTOOLS_VER_ID, SeedM_6a_destination.EVENT_TYPE, SeedM_6a_destination.ROUTE_ID, SeedM_6a_destination.DIRECTION, SeedM_6a_destination.DELT_TIME, SeedM_6a_destination.ID, SeedM_6a_destination.Dest_stop, SeedM_6a_destination.Dest_stop_desc, SeedM_6a_destination.Dest_route
/
drop table SeedM_7b_dest_reasignment_tmp2
/
create table SeedM_7b_dest_reasignment_tmp2
as SELECT SeedM_7b_dest_reasignment_tmp.SER_NBR, SeedM_7b_dest_reasignment_tmp.USE_LLT, SeedM_7b_dest_reasignment_tmp.CURRENT_ROUTE, SeedM_7b_dest_reasignment_tmp.LAST_ROUTE, SeedM_7b_dest_reasignment_tmp.TRANS_EVENT, SeedM_7b_dest_reasignment_tmp.BUS_NBR, SeedM_7b_dest_reasignment_tmp.BUS_FLAG, SeedM_7b_dest_reasignment_tmp.TRIP_FLAG, SeedM_7b_dest_reasignment_tmp.DAY1, GEO_2A_WALK_MATCH.BT_VER, SeedM_7b_dest_reasignment_tmp.BUSTOOLS_VER_ID, SeedM_7b_dest_reasignment_tmp.EVENT_TYPE, SeedM_7b_dest_reasignment_tmp.ROUTE_ID, SeedM_7b_dest_reasignment_tmp.STOPDESC, SeedM_7b_dest_reasignment_tmp.TRIP_ID, SeedM_7b_dest_reasignment_tmp.DWELL_TIME,
SeedM_7b_dest_reasignment_tmp.ID,
SeedM_7b_dest_reasignment_tmp.Dest_stop,
SeedM_7b_dest_reasignment_tmp.Dest_stop_desc,
SeedM_7b_dest_reasignment_tmp.Dest_route,
Max(GEO_2A_WALK_MATCH.STOPID) AS new_dest_stop,
Min(GEO_2A_WALK_MATCH.DIST) AS minOfDist
FROM SeedM_7b_dest_reasignment_tmp INNER JOIN GEO_2A_WALK_MATCH
ON (SeedM_7b_dest_reasignment_tmp.MinOfDIST =
GEO_2A_WALK_MATCH.DIST)
AND (SeedM_7b_dest_reasignment_tmp.Dest_stop =
GEO_2A_WALK_MATCH.STOPID2)
AND (SeedM_7b_dest_reasignment_tmp.DIRECTION =
GEO_2A_WALK_MATCH.DIRECTION)
AND (trim(SeedM_7b_dest_reasignment_tmp.ROUTE_ID) =
trim(GEO_2A_WALK_MATCH.ROUTE))
GROUP BY SeedM_7b_dest_reasignment_tmp.SER_NBR,
SeedM_7b_dest_reasignment_tmp.USE_LLT,
SeedM_7b_dest_reasignment_tmp.CURRENT_ROUTE,
SeedM_7b_dest_reasignment_tmp.LAST_ROUTE,
SeedM_7b_dest_reasignment_tmp.TRANS_EVENT,
SeedM_7b_dest_reasignment_tmp.BUS_NBR,
SeedM_7b_dest_reasignment_tmp.BUS_FLAG,
SeedM_7b_dest_reasignment_tmp.DAY1,
SeedM_7b_dest_reasignment_tmp.BUSTOOLS_VER_ID,
SeedM_7b_dest_reasignment_tmp.ROUTE_ID,
SeedM_7b_dest_reasignment_tmp.DIRECTION,
SeedM_7b_dest_reasignment_tmp.STOPID,
SeedM_7b_dest_reasignment_tmp.STOPDESC,
SeedM_7b_dest_reasignment_tmp.TRIP_ID,
SeedM_7b_dest_reasignment_tmp.DWELL_TIME,
SeedM_7b_dest_reasignment_tmp.ID,
SeedM_7b_dest_reasignment_tmp.TRIP_FLAG,
SeedM_7b_dest_reasignment_tmp.DEST_STOP,
SeedM_7b_dest_reasignment_tmp.TRIP_STOP_DESC,
SeedM_7b_dest_reasignment_tmp.TRIP_ID,
SeedM_7b_dest_reasignment_tmp.DIRECTION,
SeedM_7b_dest_reasignment_tmp.DEST_ROUTE
/
drop table SeedM_7b_dest_reasignment
/
create table SeedM_7b_dest_reasignment
as SELECT SeedM_7b_dest_reasignment_tmp2.SER_NBR,
SeedM_7b_dest_reasignment_tmp2.USE_LLT,
SeedM_7b_dest_reasignment_tmp2.CURRENT_ROUTE,
SeedM_7b_dest_reasignment_tmp2.LAST_ROUTE,
SeedM_7b_dest_reasignment_tmp2.TRANS_EVENT,
SeedM_7b_dest_reasignment_tmp2.BUS_NBR,
SeedM_7b_dest_reasignment_tmp2.BUS_FLAG,
SeedM_7b_dest_reasignment_tmp2.TRIP_FLAG,
SeedM_7b_dest_reasignment_tmp2.DAY1,
SeedM_7b_dest_reasignment_tmp2.BT_VER,
SeedM_7b_dest_reasignment_tmp2.BUSTOOLS_VER_ID,
SeedM_7b_dest_reasignment_tmp2.EVENT_TYPE,
SeedM_7b_dest_reasignment_tmp2.ROUTE_ID,
SeedM_7b_dest_reasignment_tmp2.DIRECTION,
SeedM_7b_dest_reasignment_tmp2.STOPID,
SeedM_7b_dest_reasignment_tmp2.STOPDESC,
SeedM_7b_dest_reasignment_tmp2.TRIP_ID,
SeedM_7b_dest_reasignment_tmp2.DWELL_TIME,
SeedM_7b_dest_reasignment_tmp2.ID,
SeedM_7b_dest_reasignment_tmp2.Dest_stop,
SeedM_7b_dest_reasignment_tmp2.Dest_route,
SeedM_7b_dest_reasignment_tmp2.new_dest_stop,
BT_STOPINFO.STOPDESC AS new_dest_desc
FROM SeedM_7b_dest_reasignment_tmp2 INNER JOIN BT_STOPINFO
  ON (SeedM_7b_dest_reasignment_tmp2.BT_VER = BT_STOPINFO.BT_VER)
  AND (SeedM_7b_dest_reasignment_tmp2.new_dest_stop =
  BT_STOPINFO.STOPID)
/
REM *******************************************************
REM Part B. Transfer flow aggregation
REM *******************************************************
drop table Transfer_lc_AFC_trans_only_tmp
/
create table Transfer_lc_AFC_trans_only_tmp
as SELECT SEEDM_3_AFC_STOP.SER_NBR, SEEDM_3_AFC_STOP.USE_LLT,
SEEDM_3_AFC_STOP.CURRENT_ROUTE,
SEEDM_3_AFC_STOP.LAST_ROUTE, SEEDM_3_AFC_STOP.TRANS_EVENT,
SEEDM_3_AFC_STOP.BUS_NBR,
SEEDM_3_AFC_STOP.BUS_FLAG, SEEDM_3_AFC_STOP.TRIP_FLAG,
SEEDM_3_AFC_STOP.DAY1,
SEEDM_3_AFC_STOP.EVENT_TIME, SEEDM_3_AFC_STOP.BUSTOOLS_VER_ID,
SEEDM_3_AFC_STOP.EVENT_TYPE,
SEEDM_3_AFC_STOP.ROUTE_ID, SEEDM_3_AFC_STOP.DIRECTION,
SEEDM_3_AFC_STOP.STOPID,
SEEDM_3_AFC_STOP.STOPDESC, SEEDM_3_AFC_STOP.TRIP_ID,
SEEDM_3_AFC_STOP.DWELL_TIME
FROM SEEDM_3_AFC_STOP
ORDER BY SEEDM_3_AFC_STOP.SER_NBR, SEEDM_3_AFC_STOP.USE_LLT
/
drop sequence id_seq
/
CREATE SEQUENCE ID_seq
  START WITH 1
  INCREMENT BY 1
  MAXVALUE 100000000
  CYCLE
/
drop table Transfer_lc_AFC_trans
/
create table Transfer_lc_AFC_trans
as select id_seq.nextval transID, SER_NBR, USE_LLT, CURRENT_ROUTE,
LAST_ROUTE, TRANS_EVENT, BUS_NBR,
BUS_FLAG, TRIP_FLAG, DAY1,
EVENT_TIME, BUSTOOLS_VER_ID, EVENT_TYPE,
ROUTE_ID, DIRECTION, STOPID,
STOPDESC, TRIP_ID, DWELL_TIME
from transfer_lc_AFC_trans_only_tmp
/
drop table Transfer_ic_AFC_trans_only_tmp
/
drop table Transfer_leAFC_join
/
create table Transfer_le_AFCjoin
as SELECT Transfer_ic_AFC_transfer_only.SER_NBR,
Transfer_ic_AFC_transfer_only.DAY1,
Transfer_lc_AFC_transfer_only.USE_LLT AS FromTime, t2.USE_LLT AS ToTime,
t2.LAST_ROUTE, t2.CURRENT_ROUTE,
Transfer_lc_AFC_transfer_only.BUS_NBR AS FromBus_nbr,
t2.BUS_NBR AS ToBusnbr, Transfer_lcAFC_transfer_only.ROUTE_ID AS FromRoute_id,
t2.ROUTE_ID AS ToRoute_id,
Transfer_lc_AFC_transfer_only.DIRECTION AS FromDirection,
t2.DIRECTION AS ToDirection, t2.STOPID, t2.STOPDESC,
Transfer_lc_AFC_transfer_only.TransID AS Trans_id
FROM Transfer_lc_AFC_trans Transfer_ic_AFC_transfer_only INNER JOIN
Transfer lc_AFCtrans t2
ON (Transfer lc_AFC_transfer_only.SER_NBR = t2.SER_NBR)
AND (Transfer lc_AFC_transfer_only.DAY1 = t2.DAY1)
WHERE (Transfer lc_AFC_transfer_only.TransID=t2.transID-1)
/
drop table Transfer_lf_AFC_prune
/
create table Transfer_lf_AFC_prune
as SELECT Transfer_le AFC_join.SER_NBR, Transfer_leAFC_join.DAYl,
Transfer_le AFC_join.FromTime,
TransferleAFCjoin.ToTime, Transfer_le AFC_join.LAST_ROUTE,
Transfer_le AFC_join.CURRENT_ROUTE,
Transfer_le AFC_join.FromBus_nbr, Transfer_le AFC_join.ToBus_nbr,
Transfer_le AFC_join.FromRoute_id,
Transfer_leAFC_join.ToRoute_id,
Transfer_le AFC_join.FromDirection, Transfer_le AFC_join.ToDirection,
Transfer_le AFC_join.STOPID, Transfer_le AFC_join.STOPDESC,
Transfer_leAFC_join.Trans_id
FROM Transfer_le_AFC_join
WHERE (Transfer le_AFC_join.LAST_ROUTE)<1024 AND
((Transfer le_AFC_join.CURRENT_ROUTE)<1024)
AND (to_number(to_char(FromTime, 'HH24')) -
to_number(to_char(ToTime, 'HH24')) < 2))
/

REM **********************************
REM apply transfer info to seed matrix
REM useful for modified IPF
REM **********************************

drop table SEEDM_7C_first_leg_trans
/
create table SEEDM_7C_first_leg_trans
as SELECT SEEDM_7B_DEST_REASIGNMENT.SER_NBR,
SEEDM_7B_DEST_REASIGNMENT.USE_LLT,
SEEDM_7B_DEST_REASIGNMENT.CURRENT_ROUTE,
SEEDM_7B_DEST_REASIGNMENT.LAST_ROUTE,
SEEDM_7B_DEST_REASIGNMENT.TRANS_EVENT,
SEEDM_7B_DEST_REASIGNMENT.BUS_NBR,
SEEDM_7B_DEST_REASIGNMENT.BUS_FLAG,
SEEDM_7B_DEST_REASIGNMENT.TRIP_FLAG,
SEEDM_7B_DEST_REASIGNMENT.DAY1, SEEDM_7B_DEST_REASIGNMENT.BT_VER,
SEEDM_7B_DEST_REASIGNMENT.BUSTOOLS_VER_ID,
SEEDM_7B_DEST_REASIGNMENT.EVENT_TYPE,
SEEDM_7B_DEST_REASIGNMENT.ROUTE_ID,
SEEDM_7B_DEST_REASIGNMENT.DIRECTION,
SEEDM_7B_DEST_REASIGNMENT.STOPID,
SEEDM_7B_DEST_REASIGNMENT.STOPDESC,
SEEDM_7B_DEST_REASIGNMENT.TRIP_ID,
SEEDM_7B_DEST_REASIGNMENT.DWELL_TIME,
SEEDM_7B_DEST_REASIGNMENT.ID, SEEDM_7B_DEST_REASIGNMENT.DEST_STOP,
SEEDM_7B_DEST_REASIGNMENT.DEST_STOP_DESC,
SEEDM_7B_DEST_REASIGNMENT.DEST_ROUTE,
SEEDM_7B_DEST_REASIGNMENT.NEW_DEST_STOP,
SEEDM_7B_DEST_REASIGNMENT.NEW_DEST_DESC,
Transfer lf AFC_prune.ToRoute_id,
Transfer lf AFC_prune.ToDirection
FROM SEEDM_7B_DEST_REASIGNMENT LEFT JOIN Transfer lf AFC_prune
ON (SEEDM_7B_DEST_REASIGNMENT.USELLT =
Transfer lf AFC_prune.FromTime)
AND (SEEDM_7B_DEST_REASIGNMENT.SER_NBR =
Transfer lf AFC_prune.SER_NBR)
WHERE (((SEEDM_7B_DEST_REASIGNMENT.DAY1)='12'))
/
drop table SEEDM_7D_second_leg_trans
/
create table SEEDM_7D_second_leg_trans
as SELECT SEEDM_7C_first_leg_trans.SER_NBR,
SEEDM_7C_first_leg_trans.USELLT,
SEEDM_7C_first_leg_trans.CURRENT_ROUTE,
SEEDM_7C_first_leg_trans.LAST_ROUTE,
SEEDM_7C_first_leg_trans.TRANS_EVENT,
SEEDM_7C_first_leg_trans.BUSNBR,
SEEDM_7C_first_leg_trans.BUS_FLAG,
SEEDM_7C_first_leg_trans.TRIP_FLAG,
SEEDM_7C_first_leg_trans.DAY1, SEEDM_7C_first_leg_trans.BT_VER,
SEEDM_7C_first_leg_trans.BUSTOOLS_VER_ID,
SEEDM_7C_first_leg_trans.EVENT_TYPE,
SEEDM_7C_first_leg_trans.ROUTE_ID,
SEEDM_7C_first_leg_trans.DIRECTION,
SEEDM_7C_first_leg_trans.STOPID, SEEDM_7C_first_leg_trans.STOPDESC,
SEEDM_7C_first_leg_trans.TRIP_ID,
SEEDM_7C_first_leg_trans.DWELL_TIME, SEEDM_7C_first_leg_trans.ID,
SEEDM_7C_first_leg_trans.DEST_STOP,
SEEDM_7C_first_leg_trans.DEST_STOP_DESC,
SEEDM_7C_first_leg_trans.DEST_ROUTE,
SEEDM_7C_first_leg_trans.NEW_DEST_STOP,
SEEDM_7C_first_leg_trans.NEW_DEST_DESC,
SEEDM_7C_first_leg_trans.ToRoute_id,
SEEDM_7C_first_leg_trans.ToDirection,
Transfer lf AFC_prune.FromRoute_id,
Transfer lf AFC_prune.FromDirection
FROM SEEDM_7C_first_leg_trans LEFT JOIN Transfer lf AFC_prune
ON (SEEDM_7C_first_leg_trans.USE_LLT = Transfer_1f_AFC_prune.ToTime)
          AND (SEEDM_7C_first_leg_trans.SER_NBR = Transfer_1f_AFC_prune.SER_NBR)
/
REM ****************************
REM transfer counts from transfer card data
REM ****************************

drop table Transfer_2a_trans_bus
/
create table Transfer_2a_trans_bus
as SELECT EQUIP_BUS.BUS_NBR, MIT_M8056.ISSUE_USE_LLT,
to_char(ISSUE_USE_LLT, 'DD') AS Day1,
MIT_M8056.XFER_EVENT, MIT_M8056.CURRENT_ROUTE,
MIT_M8056.LAST_ROUTE
FROM MIT_M8056 INNER JOIN EQUIPBUS ON MIT_M8056.EQUIP_ID =
EQUIP_BUS.EQUIP_ID
WHERE (((MIT_M8056.ISSUEUSELLT)>=min_time And
(MIT_M8056.ISSUE_USE_LLT)<max_time)
          AND ((MIT_M8056.XFER_EVENT)=1 Or (MIT_M8056.XFER_EVENT)=2))
/
drop table Transfer/2btransprune
create
    table Transfer_2b_trans_prune
as SELECT Transfer_2a_trans_bus.BUS_NBR,
Transfer_2a_trans_bus.ISSUE_USE_LLT,
Transfer_2a_trans_bus.Day1, Transfer_2a_trans_bus.XFER_EVENT,
Transfer_2a_trans_bus.CURRENT_ROUTE,
Transfer_2a_trans_bus.LAST_ROUTE
FROM Transfer_2atransbus
WHERE (((Transfer_2a_trans_bus.CURRENT_ROUTE)<>last_route)
          AND ((tochar(ISSUE_USE_LLT, 'HH24'))<'09' And
(to_char(ISSUE_USE_LLT, 'HH24'))>='06'))
/
drop table Transfer_2c_AVL
/
create table Transfer_2c_AVL
as SELECT Transfer_2b_trans_prune.BUS_NBR,
Transfer_2b_trans_prune.ISSUE_USE_LLT,
Transfer_2b_trans_prune.XFER_EVENT,
Transfer_2b_trans_prune.CURRENT_ROUTE,
Transfer_2b_trans_prune.LAST_ROUTE, BUS_STATE_HIST.BUS_ID,
Max(BUS_STATE_HIST.EVENT_TIME) AS MaxOfEVENT_TIME
FROM Transfer_2b_trans_prune INNER JOIN BUS_STATE_HIST
ON Transfer_2b_trans_prune.BUS_NBR = BUS_STATE_HIST.BUS_ID
WHERE (((BUS_STATE_HIST.EVENT_TYPE)=3 Or (BUS_STATE_HIST.EVENT_TYPE)=5)
          AND ((BUS_STATE_HIST.EVENT_TIME)<issue_use_llt And
Issue_USE_LLT<event_time+numtodsinterval(5, 'MINUTE')))GROUP BY Transfer_2b_trans_prune.BUS_NBR,
Transfer_2b_trans_prune.ISSUE_USE_LLT,
Transfer_2b_trans_prune.XFER_EVENT,
Transfer_2b_trans_prune.CURRENT_ROUTE,
Transfer_2b_trans_prune.LAST_ROUTE, BUS_STATE_HIST.BUS_ID
/
drop table Transfer_2d_AVL_match
CREATE TABLE Transfer_2dAVLmatch
AS SELECT Transfer_2c_AVL.BUS_NBR, Transfer_2c_AVL.ISSUE_USE_LLT,
Transfer_2c_AVL.XFER_EVENT,
Transfer_2c_AVL.CURRENT_ROUTE, Transfer_2c_AVL.LAST_ROUTE,
Transfer_2c_AVL.BUS_ID,
BUS_STATE_HIST.EVENT_TYPE, BUS_STATE_HIST.BUSTOOLS_VER_ID,
BUS_STATE_HIST.ROUTE_ID,
BUS_STATE_HIST.PATTERN,  BUS_STATE_HIST.STOP_SEQUENCE,
BUS_STATE_HIST.TRIP_ID
FROM Transfer_2c_AVL INNER JOIN BUSSTATEHIST ON
(Transfer_2c_AVL.BUS_ID = BUS_STATE_HIST.BUS_ID)
AND (Transfer_2c_AVL.MaxOfEVENT_TIME = BUS_STATE_HIST.EVENT_TIME)
WHERE (((BUS_STATE_HIST.EVENT_TYPE)=3 OR (BUS_STATE_HIST.EVENT_TYPE)=5))
/
DROP TABLE Transfer_2etransstop
/
CREATE TABLE Transfer_2e_trans_stop
AS SELECT Transfer_2d_AVL_match.BUSNBR,
Transfer_2d_AVL_match.ISSUE_USE_LLT,
Transfer_2dAVL_match.XFEREVENT,
Transfer_2d_AVL_match.CURRENT_ROUTE,
Transfer_2d_AVL_match.LAST_ROUTE, Transfer_2d_AVL_match.BUSID,
Transfer_2d_AVL_match.EVENT_TYPE,
Transfer_2dAVL_match.BUSTOOLS_VER_ID,
Transfer_2dAVL_match.ROUTE_ID, Transfer_2d_AVL_match.PATTERN,
BT_PATTERN.DIRECTION, Transfer_2d_AVL_match.STOP_SEQUENCE,
Transfer_2d_AVL_match.TRIP_ID,
10 AS FromDirection
FROM ((Transfer_2dAVL_match INNER JOIN BT_PATTERN
ON (Transfer_2d_AVL_match.BUSTOOLS_VER_ID = BT_PATTERN.BT_VER)
AND (Transfer_2d_AVL_match.PATTERN = BT_PATTERN.PATTERN)) INNER
JOIN BT_PATTERNDETAIL
ON (BT_PATTERN.BT_VER = BT_PATTERNDETAIL.BT_VER)
AND (BT_PATTERNDETAIL.PATTERNID = BT_PATTERN.PATTERNID)
AND (Transfer_2d_AVL_match.STOP_SEQUENCE =
BT_PATTERNDETAIL.STOPSORTORDER)) INNER JOIN BT_STOPINFO
ON (BT_PATTERNDETAIL.BT_VER = BT_STOPINFO.BT_VER)
AND (BT_PATTERNDETAIL.STOPID = BT_STOPINFO.STOPID)
WHERE (((Trim(ROUTE_ID))=Trim(route)))
/
DROP TABLE Transfer_3a_all_trans
/
CREATE TABLE Transfer_3a_all_trans
AS SELECT Transfer_lf_AFC_prune.SERNBR, Transfer_lf_AFC_prune.ToTime,
Transfer_lf_AFC_prune.LAST_ROUTE,
Transfer_lf_AFC_prune.CURRENT_ROUTE,
Transfer_lf AFC_prune.FromBus_nbr,
Transfer_lf AFC_prune.ToBus_nbr,
Transfer_lf_AFC_prune.FromRoute_id,
Transfer_lf AFC_prune.ToRouteid,
Transfer_lf_AFC_prune.FromDirection,
Transfer_lf_AFC_prune.ToDirection
FROM Transfer_lfAFC_prune
WHERE ((to_char(ToTime, 'HH24')>='06' AND to_char(ToTime, 'HH24')<>'09'))
/
DROP TABLE Transfer_3b_trans_modified
create table Transfer_3b_trans_modified
as SELECT Transfer_2e_trans_stop.BUSNBR,
        Transfer_2e_trans_stop.ISSUE_USE_LLT,
        Transfer_2e_trans_stop.XFER_EVENT,
        Transfer_2e_trans_stop.CURRENT_ROUTE,
        Transfer_2e_trans_stop.LAST_ROUTE,
        Transfer_2e_trans_stop.BUS_ID,
        Transfer_2e_trans_stop.EVENT_TYPE,
        Transfer_2e_trans_stop.BUSTOOLS_VER_ID,
        Route_RouteID.ROUTE_ID AS FromRoute_ID,
        Transfer_2e_trans_stop.ROUTE_ID AS ToRoute_ID,
        Transfer_2e_trans_stop.PATTERN,
        Transfer_2e_trans_stop.DIRECTION,
        Transfer_2e_trans_stop.STOP_SEQUENCE,
        Transfer_2e_trans_stop.TRIP_ID,
        Transfer_2e_trans_stop.FromDirection
FROM Transfer_2e_trans_stop INNER JOIN Route_RouteID
ON Transfer_2e_trans_stop.LAST_ROUTE =
to_number(Route_RouteID.CURRENT_ROUTE)
/
drop table Transfer_3c_direction_split
/
create table Transfer_3c_direction_split
as SELECT Transfer_3a_all_trans.FromRoute_id,
        Transfer_3a_all_trans.FromDirection,
        Transfer_3a_all_trans.ToRoute_id,
        Transfer_3a_all_trans.ToDirection,
        Count(Transfer_3a_all_trans.SER_NBR) AS Count
FROM Transfer_3a_all_trans
GROUP BY Transfer_3a_all_trans.FromRoute_id,
        Transfer_3a_all_trans.FromDirection,
        Transfer_3a_all_trans.ToRoute_id,
        Transfer_3a_all_trans.ToDirection
/
drop table Transfer_3d_afc_both_dir_1
/
create table Transfer_3d_afc_both_dir_1
as SELECT Transfer_3c_direction_split.FromRoute_id,
        Transfer_3c_direction_split.ToRoute_id,
        Transfer_3c_direction_split.ToDirection,
        Sum(Transfer_3c_direction_split.Count) AS BothDir
FROM Transfer_3c_direction_split
GROUP BY Transfer_3c_direction_split.FromRoute_id,
        Transfer_3c_direction_split.ToRoute_id,
        Transfer_3c_direction_split.ToDirection
/
drop table Transfer_3d_trans_dir_split
/
create table Transfer_3d_trans_dir_split
as SELECT Transfer_3b_trans_modified.FromRoute_ID,
        Transfer_3b_trans_modified.FromDirection,
        Transfer_3b_trans_modified.ToRoute_ID,
        Transfer_3b_trans_modified.DIRECTION AS ToDirection,
        Count(Transfer_3b_trans_modified.ISSUE_USE_LLT) AS CountOfISSUE_USE_LLT
FROM Transfer_3b_trans_modified
WHERE (((Transfer_3b_trans_modified.FromRoute_ID)<>toroute_ID))
GROUP BY Transfer_3b_trans_modified.FromRoute_ID,
  Transfer_3b_trans_modified.FromDirection,
  Transfer_3b_trans_modified.ToRoute_ID,
  Transfer_3b_trans_modified.DIRECTION
/
drop table Transfer_3e_afc_dir_split_ra /
create table Transfer_3e_afc_dir_split_ra
as SELECT Transfer_3c_direction_split.FromRoute_id,
  Transfer_3c_direction_split.FromDirection,
  Transfer_3c_direction_split.ToRoute_id,
  Transfer_3c_direction_split.ToDirection,
  Transfer_3c_direction_split.Count,
  Transfer_3d_afc_both_dir_l.BothDir, Count / BothDir AS Split
FROM Transfer_3d_afc_both_dir_l1 INNER JOIN Transfer_3c_direction_split
  ON (Transfer_3d_afc_both_dir_l.FromRoute_id =
  Transfer_3c_direction_split.FromRoute_id)
  AND (Transfer_3d_afc_both_dir_l.ToRoute_id =
  Transfer_3c_direction_split.ToRoute_id)
  AND (Transfer_3d_afc_both_dir_l.ToDirection =
  Transfer_3c_direction_split.ToDirection)
/
drop table Transfer_3e_afc_dir_split_ra_t /
create table Transfer_3e_afc_dir_split_ra_t
as SELECT Transfer_3d_trans_dir_split.FromRoute_ID,
  Transfer_3e_afc_dir_split_ra.FromDirection,
  Transfer_3d_trans_dir_split.ToRoute_ID,
  Transfer_3e_trans_dir_split.ToDirection,
  Transfer_3e_afc_dir_split_ra.Count,
  Transfer_3e_afc_dir_split_ra.BothDir, 0.5 AS factor
FROM Transfer_3e_afc_dir_split_ra RIGHT JOIN
  Transfer_3d_transdir_split
  ON (Transfer_3e_afc_dir_split_ra.FromRoute_id =
  Transfer_3d_transdir_split.FromRoute_ID)
  AND (Transfer_3e_afcdirsplit_ra.ToRoute_id =
  Transfer_3d_transdirsplit.ToRoute_ID)
  AND (Transfer_3e_afc_dir_split_ra.ToDirection =
  Transfer_3d_transdir_split.ToDirection)
WHERE (((Transfer_3e_afc_dir_split_ra.BothDir) Is Null))
/
INSERT INTO Transfer_3e_afc_dir_split_ra (FromRoute_ID, FromDirection,
  ToRoute_ID, ToDirection, Count, BothDir, Split )
SELECT Transfer_3e_afc_dir_split_ra_t.FromRoute_ID, Route_DIR.DIRECTION,
  Transfer_3e_afc_dir_split_ra_t.ToRoute_ID,
  Transfer_3e_afc_dir_split_ra_t.Count,
  Transfer_3e_afc_dir_split_ra_t.BothDir,
  Transfer_3e_afc_dir_split_ra_t.factor
FROM Transfer_3e_afc_dir_split_ra_t INNER JOIN Route_DIR
  ON Transfer_3e_afc_dir_split_ra_t.FromRoute_ID =
  Route_DIR.ROUTE_ID
/
drop table Transfer_3f_trans_adjusted_dir /
create table Transfer_3f_trans_adjusted_dir
as SELECT Transfer_3d_trans_dir_split.FromRoute_ID,
       Transfer_3eafc_dir_split_ra.FromDirection,
       Transfer_3d_trans_dir_split.ToRoute_ID,
       Transfer_3d_trans_dir_split.ToDirection,
       Transfer_3d_trans_dir_split.CountOfISSUE_USELLT,
       Transfer_3eafc_dir_split_ra.Count,
       Transfer_3eafc_dir_split_ra.BothDir,
       Transfer_3eafc_dir_split_ra.Split,
       countofissue_use_ltt*Split AS NewCount
FROM Transfer_3eafc_dir_split_ra INNER JOIN
Transfer_3d_trans_dir_split
ON (Transfer_3eafc_dir_split_ra.ToDirection =
Transfer_3d_trans_dir_split.ToDirection)
AND (Transfer_3eafc_dir_split_ra.ToRoute_id =
Transfer_3d_trans_dir_split.ToRoute_ID)
AND (Transfer_3eafc_dir_split_ra.FromRoute_id =
Transfer_3d_trans_dir_split.FromRoute_ID)

/*
  INSERT INTO Transfer_3c_direction_split ( FromRoute_ID, FromDirection,
  ToRoute_ID, ToDirection, Count )
  SELECT Transfer_3f_trans_adjusted_dir.FromRoute_ID,
  Transfer_3f_trans_adjusted_dir.FromDirection,
  Transfer_3f_trans_adjusted_dir.ToRoute_ID,
  Transfer_3f_trans_adjusted_dir.ToDirection,
  Transfer_3f_trans_adjusteddir.NewCount
  FROM Transfer_3f_trans_adjusted_dir
*/

drop table Transfer_4a_aggregate
/
create table Transfer_4a_aggregate
as SELECT Transfer_3c_direction_split.FromRoute_id,
       Transfer_3c_direction_split.FromDirection,
       Transfer_3c_direction_split.ToRoute_id,
       Transfer_3c_direction_split.ToDirection,
       Sum(Transfer_3c_direction_split.Count) AS SumOfCount
FROM Transfer_3c_direction_split
GROUP BY Transfer_3c_direction_split.FromRoute_id,
       Transfer_3c_direction_split.FromDirection,
       Transfer_3c_direction_split.ToRoute_id,
       Transfer_3c_direction_split.ToDirection
/

drop table Transfer_5a_AFC_control_sort_t
/
create table Transfer_5a_AFC_control_sort_t
as SELECT MIT_M8054.SER_NBR, MIT_M8054.USE_LLT, MIT_M8054.TRANS_EVENT,
       MIT_M8054.CURRENT_ROUTE,
       MITM8054.LASTROUTE
FROM MITM8054
WHERE (((MIT_M8054.TRANS_EVENT) In
       (36,37,38,39,43,45,46,47,48,49,97,98,99))
       AND ((MIT_M8054.CURRENT_ROUTE)=1) AND
       ((to_char(USE_LLT, 'HH24'))< '09'
       AND (to_char(USE_LLT, 'HH24'))>='05'))
ORDER BY MIT_M8054.SER_NBR, MIT_M8054.USE_LLT
/
drop sequence id_seq
/
CREATE SEQUENCE ID_seq
START WITH 1
INCREMENT BY 1
maxvalue 100000000
CYCLE
/
drop table Transfer_5a_AFC_control_sort
/
create table Transfer_5a_AFC_control_sort
as SELECT SER_NBR, USE_LLT, TRANS_EVENT, CURRENT_ROUTE,
LAST_ROUTE, id_seq.nextval ID
FROM Transfer_5a_AFCcontrolsort_t
/
rem ALTER TABLE "TRANSFER 5A AFC CONTROL SORT"
rem MODIFY(current_route null, last_route NULL)
rem /
rem UPDATE Transfer_5a_afc_control_sort t
rem SET CURRENT_ROUTE =
rem (select to_number("Adj_Cur_Route")
rem from Route_adj
rem where to_number(route_adj.current_route) = t.current_route)
rem /
rem UPDATE Transfer_5a_afc_control_sort t
rem SET Last_ROUTE =
rem "AdjCur_Route"
rem from Route_adj
rem where to_number(route_adj.current_route) = t.last_route)
rem /
rem commit
rem /
drop table Transfer_5b_AFC_match
/
create table Transfer_5b_AFC_match
as SELECT Transfer_5a_AFCcontrolsort.SER_NBR,
Transfer_5a_AFCcontrolsort.USE_LLT AS FromTime,
Transfer_5a_AFCcontrolsort.CURRENT_ROUTE AS FCRoute,
Transfer_5a_AFCcontrolsort.LAST_ROUTE AS FLRoute,
Transfer_5a_AFCcontrolsort.ID, t2.USE_LLT AS ToTime,
t2.CURRENT_ROUTE AS TCRoute,
t2.LAST_ROUTE AS TLRoute
FROM Transfer_5a_AFCcontrolsort INNER JOIN
Transfer_5a_AFCcontrol_sort t2
  ON Transfer_5a_AFCcontrol_sort.SER_NBR = t2.SER_NBR
WHERE ((Transfer_5a_AFCcontrol_sort.ID)=t2.ID-1))
/
DELETE
FROM Transfer_5b_AFC_match
WHERE (((Transfer_5b_AFC_match.FCRoute)<>TLRoute))
/
drop table Transfer_5d_afc_prune
/
create table Transfer_5d_afc_prune
as SELECT Transfer_5b_AFC_match.SER_NBR, Transfer_5b_AFC_match.FromTime,
Transfer_5b_AFC_match.FCRoute,
Transfer_5b_AFC_match.ID, Transfer_5b_AFC_match.ToTime,
Transfer_5b_AFC_match.TCRoute
FROM Transfer_5b_AFC_match
WHERE (((Transfer_5b_AFC_match.FromTime)>totime - numtodsinterval(2, 'HOUR'))
AND ((Transfer_5b_AFC_match.FCRoute)<=TCRoute And
(Transfer_5b_AFC_match.FCRoute)<1024)
AND ((Transfer_5b_AFC_match.TCRoute)<1024) AND ((to_char(ToTime, 'HH24'))>='06' And (to_char(ToTime, 'HH24'))<'09'))
/
drop table Transfer_5e_trans
/
create table Transfer_5e_trans
as SELECT MIT_M8056.ISSUE_USE_LLT, MIT_M8056.XFER_EVENT,
MIT_M8056.CURRENT_ROUTE,
MITM8056.LASTROUTE
FROM MITM8056
WHERE (((MIT_M8056.XFER_EVENT)<=2))
/
drop table Transfer_5e_trans_prune
/
create table Transfer_5e_trans_prune
as SELECT Transfer_5e_trans.ISSUE_USE_LLT, Transfer_5e_trans.XFER_EVENT,
Transfer_5e_trans.CURRENT_ROUTE,
Transfer_5e_trans.LAST_ROUTE
FROM Transfer_5e_trans
WHERE (((Transfer_5e_trans.CURRENT_ROUTE)<>last_route
And (Transfer_5e_trans.CURRENT_ROUTE)<1024) AND
((Transfer_5e_trans.LAST_ROUTE)<1024)
AND ((to_char(ISSUE_USE_LLT, 'HH24'))>='06' And
(to_char(ISSUEUSE LLT, 'HH24'))<'09'))
/
ALTER TABLE Transfer_5e_trans
MODIFY(current_route null, last_route NULL)
/
UPDATE Transfer_5e_trans
SET Transfer_5e_trans.CURRENT_ROUTE = (select "Adj_Cur_Route"
from Route_adj
Where Transfer_5e_trans.CURRENT_ROUTE = Route_adj.CURRENT_ROUTE
)
/
UPDATE Transfer_5e_trans
SET Transfer_5e_trans.last_ROUTE = (select "Adj_Cur_Route"
from Route_adj
Where Transfer_5e_trans.last_ROUTE = Route_adj.CURRENT_ROUTE
)
/
drop table Transfer_5f_afc_control
/
create table Transfer_5f_afc_control
as SELECT Count(Transfer_5d_afc_prune.SER_NBR) AS CountOfSER_NBR,
Transfer_5dafc_prune.FCRoute AS Last_route,
Transfer_5d_afc_prune.TCRoute AS Current_Route
FROM Transfer_5d_afc_prune
GROUP BY Transfer_5d_afc_prune.FCRoute, Transfer_5d_afc_prune.TCRoute
/
drop table Transfer_5g_trans_control
/
create table Transfer_5g_trans_control
as SELECT Count(Transfer_5e_trans_prune.ISSUE_USE_LLT) AS Count
FROM Transfer_5e_trans_prune
/

REM ****************************
REM scale up the transfer counts to control totals
REM ****************************
drop table Transfer_6a_trans_sum
/
cREATE TABLE Transfer_6a_trans_sum
AS SELECT Sum(Transfer_3f_trans_adjusted_dir.NewCount) AS SumOfNewCount
FROM Transfer_3f_trans_adjusted_dir
/
drop table Transfer_6b_trans_scale_factor
/
cREATE TABLE Transfer_6b_trans_scale_factor
AS SELECT 1.3*count/SumOfNewCount AS Factor
FROM Transfer_6a_trans_sum, Transfer_5g_trans_control
/
drop table Transfer_7a_sum
/
cREATE TABLE Transfer_7a_sum
AS SELECT Sum(Transfer_3c_direction_split.Count) AS SumOfCount
FROM Transfer_3c_direction_split
/
drop table Transfer_7b_sum_control
/
cREATE TABLE Transfer_7b_sum_control
AS SELECT Count(Transfer_5d_afc_prune.SER_NBR) AS CountOfSER_NBR
FROM Transfer_5d_afc_prune
/
drop table Transfer_7c_afc_scale_factor
/
cREATE TABLE Transfer_7c_afc_scale_factor
AS SELECT 1.3*c.countofser_nbr/s.sumofcount AS factor
FROM Transfer_7a_sum s, Transfer_7b_sum_control c
/
drop table Transfer_8a_afc_dir
/
cREATE TABLE Transfer_8a_afc_dir
AS SELECT a.FromRoute_id, a.FromDirection, a.ToRoute_id, a.ToDirection,
a.Count*factor AS Count
FROM Transfer_3c_direction_split a, Transfer_7c_afc_scale_factor
/
drop table Transfer_8b_trans_dir
/
cREATE TABLE Transfer_8b_trans_dir
AS SELECT Transfer_3f_transadjusted_dir.FromRoute_ID,
Transfer_3f_transadjusted_dir.FromDirection,
Transfer_3f_transadjusted_dir.ToRoute_ID,
Transfer_3f_transadjusted_dir.ToDirection,
NewCount*Factor AS Count
FROM Transfer_6b_trans_scale_factor, Transfer_3f_trans_adjusted_dir
/
drop table Transfer_8c_all_transfer
/
125
create table Transfer_8calltransfer
as SELECT afc.FromRoute_ID, afc.FromDirection, afc.ToRoute_ID,
afc.ToDirection,
afc.count+trans.count AS Total
FROM Transfer_8a_aafcdir afc INNER JOIN Transfer_8b_transdir trans
ON (afc.FromRoute_ID = trans.FromRoute_ID) AND (afc.FromDirection = 
trans.FromDirection)
AND (afc.ToRoute_ID = trans.ToRoute_ID) AND (afc.ToDirection = 
trans.ToDirection)
/

REM **************
REM select the internal transfer flows in the demonstration corridor 
only
REM **************
drop table Transfer_9a_corridor
/
create table Transfer_9acorridor
as SELECT Transfer_8c_all_transfer.FromRoute_ID, 
Transfer_8call_transfer.FromDirection, 
Transfer_8c_all_transfer.ToRouteID, 
Transfer_8c_all_transfer.ToDirection, 
Transfer 8calltransfer.Total
FROM Transfer_8calltransfer
WHERE (((Transfer_8c_all_transfer.FromRoute_ID) In
('20', '126', '49', 'X49'))
AND ((Transfer_8c_all_transfer.ToRoute_ID) In
('20', '126', '49', 'X49')))
/

REM **********************
REM Part C. boarding and alighting counts from APC data
REM **********************

REM **********************
REM Identify trip id's with valid APC data
REM **********************
drop table Trip_id_1
/
create table Trip_id_1
as SELECT EVENT_TIME,
BUS_ID,
EVENT_TYPE,
BUSTOOLS_VER_ID,
ROUTE_ID,
PATTERN,
STOP_SEQUENCE,
RON + fon boarding,
Foff + roff alighting,
DWELL_TIME,
HEADING,
TRIP_ID
FROM bus_state_hist
WHERE event_type = 3 or event_type = 5
/
drop table Trip_id_2
/
Create table Trip_id_2
as SELECT BUS_ID, BUSTOOLS_VER_ID,
    to_char(event_time + NUMTODSINTERVAL(3, 'HOUR'), 'DD') day1,
    ROUTE_ID, Trip_id_1.PATTERN,
    TRIP_ID, Sum(boarding) AS SumOfon,
    Sum(alighting) AS SumOfoff, Count(TRIP_ID) AS divider
FROM Trip_id_1
GROUP BY BUS_ID, BUSTOOLS_VER_ID, ROUTE_ID, PATTERN,
    TRIP_ID, to_char(event_time + NUMTODSINTERVAL(3, 'HOUR'), 'DD')
/
drop table Trip_id_3_valid_APC
/
create table Trip_id_3_valid_APC
as SELECT BUS_ID, BUSTOOLS_VER_ID, ROUTE_ID, PATTERN,
    TRIP_ID, day1, SumOfon, SumOfoff, divider, SumOfon - SumOfoff AS Diff
FROM Trip_id_2
GROUP BY BUS_ID, BUSTOOLS_VER_ID, ROUTE_ID, PATTERN, TRIP_ID, day1,
    SumOfon, SumOfoff, divider, SumOfon-SumOfoff
HAVING TRIP_ID>0
    AND Trip_id_2.SumOfon>0
    AND SumOfoff>0 AND SumOfon-SumOfoff<0.15*SumOfon
    AND SumOfon-SumOfoff<0.15*SumOfoff
/
drop table pattern_stops
/
create table pattern_stops
as SELECT BT_PATTERN.BT_VER, BT_PATTERN.DIRECTION, BT_PATTERN.ROUTE,
    BT_PATTERN.PATTERN, BT_PATTERNDETAIL.STOPSORTORDER,
    BT_STOPINFO.STOPID,
    BT_STOPINFO.STOPDESC
FROM BT_PATTERN, BT_PATTERNDETAIL, BT_STOPINFO
WHERE BT_PATTERN.BT_VER=157
    and BT_PATTERNDETAIL.STOPID = BT_STOPINFO.STOPID
    AND BT_PATTERNDETAIL.BT_VER = BT_STOPINFO.BT_VER
    and BT_PATTERN.PATTERNID = BT_PATTERNDETAIL.PATTERNID
    AND BT_PATTERN.BT_VER = BT_PATTERNDETAIL.BT_VER
/
drop table Trip_id_Stops
/
create table Trip_id_Stops
as SELECT Trip_id_3_valid_APC.BUS_ID, pattern_stops.BT_VER,
    Trip_id_3_valid_APC.ROUTE_ID, Trip_id_3_valid_APC.PATTERN,
    Trip_id_3_valid_APC.TRIP_ID, day1, pattern_stops.DIRECTION,
    Trip_id_3_valid_APC.SUMOFON, Trip_id_3_valid_APC.SUMOFOFF,
    Trip_id_3_valid_APC.DIFF, pattern_stops.STOPSORTORDER,
    pattern_stops.STOPID, pattern_stops.STOPDESC
FROM Trip_id_3_valid_APC, pattern_stops
where Trip_id_3_valid_APC.BUSTOOLS_VER_ID = pattern_stops.BT_VER
    AND trim(Trip_id_3_valid_APC.ROUTE_ID) =
    trim(pattern_stops.ROUTE)
    AND Trip_id_3_valid_APC.PATTERN = pattern_stops.PATTERN
/
REM ***********************
REM Balance the on-off counts
REM **************************

drop table APC_matched_on_off_tmp
/
create table APC_matched_on_off_tmp
as SELECT Trip_id_Stops.BUS_ID, Trip_id_Stops.BT_VER,
       Trip_id_Stops.ROUTE_ID, Trip_id_Stops.PATTERN,
       Trip_id_Stops.TRIP_ID, Trip_id_Stops.DIRECTION, RON+fon AS
       boarding,
       FOFF+roff AS alighting, bus_state_hist.EVENT_TIME, STOP_SEQUENCE,
       Trip_id_Stops.STOPDESC, dayl
FROM Tripid_Stops, bus_state_hist
WHERE (EVENT_TYPE =3 or event_type = 5)
and Trip_id_Stops.BT_VER=BUSTOOLS_VER_ID
AND Trip_id_Stops.ROUTE_ID=bus_state_hist.ROUTE_ID
AND Trip_id_Stops.PATTERN=bus_state_hist.PATTERN
AND Trip_id_Stops.STOPSORTORDER(+)=bus_state_hist.STOP_SEQUENCE
AND Trip_id_Stops.TRIP_ID=bus_state_hist.TRIP_ID
and trip_id_stops.dayl = to_char(event_time + NUMTODSINTERVAL(3,
       'HOUR'), 'DD')
/
drop table APC_matched_on_off_tmp2
/
create table APC_matched_on_off_tmp2
as select Bus_id, bt_ver, Route_ID, pattern, trip_id, direction,
       sum(boarding) as sumOfBoarding, sum(alighting) as sumOfAlighting,
       dayl
from Apc_matchedonoff_tmp
group by Bus_id, bt_ver, Route_ID, pattern, tripid, direction, Dayl
/
drop table APC_matched_on_off_tmp3
/
create table APC_matched_on_off_tmp3
as select Bus_id, bt_ver, Route_ID, pattern, trip_id, direction,
       sumOfBoarding as sumOfOn, sumOfAlighting as sumOfOff,
       dayl, sumOfBoarding-sumOfAlighting as Diff
from APC_matched_on_off_tmp2
/
drop table APC_matched_on_off
/
create table APCmatchedonoff
as select a.bus_id, a.bt_ver, a.route_id, a.pattern, a.trip_id,
       a.sumOfOn,
       a.sumOfOff, a.Diff, Boarding, Alighting, event_time,stop_sequence,
       stopID,
       stopDesc, a.Dayl
from APC_matched_on_off_tmp3 a, APC_matched_on_off_tmp b
where a.bus_id = b.bus_id
and a.bt_ver= b.bt_ver
and a.route_id = b.route_id
and a.pattern = b.pattern
and a.trip_id = b.trip_id
and a.direction = b.direction
and a.Dayl = b.Dayl
/
drop table APC_adj_on_off
create table APC_adj_on_off
as SELECT APC_matched_on_off.BUS_ID, APC_matched_on_off.BT_VER,
APC_matched_on_off.ROUTE_ID, APC_matched_on_off.PATTERN,
APC_matched_on_off.TRIP_ID, APC_matched_on_off.DIRECTION,
APC_matched_on_off.EVENT_TIME, APC_matched_on_off.STOP_SEQUENCE,
APC_matched_on_off.STOPID, APC_matched_on_off.STOPDESC,
APC_matched_on_off.Sum0fon, APC_matched_on_off.SumOfoff,
APC_matched_on_off.Diff, APC_matched_on_off.boarding,
APC_matched_on_off.alighting,
boarding*(Abs(diff)-diff)*boarding/(2*sumofon) AS adj_on,
alighting+(Abs(diff)+diff)*alighting/(2*sumofoff) AS adj_off
FROM APC_matched_on_off
WHERE Abs(Diff)<0.15*SumOfon
   And Abs(Diff)<0.15*SumOfoff
ORDER BY APC_matched_on_off.TRIP_ID, day1
/
REM *******************
REM aggregate on-off counts by stop
REM ************************
drop table APC_3_sum_by_stop
/
create table APC_3_sum_by_stop
as SELECT APC_adj_on_off.BT_VER, APC_adj_on_off.ROUTE_ID,
APC_adj_on_off.DIRECTION, APC_adj_on_off.STOPID,
APC_adj_on_off.STOPDESC, Sum(APC_adj_on_off.adj_on) AS Total_on,
Sum(APC_adj_on_off.adjoff) AS Total_off
FROM APC_adj_on_off
WHERE to_char(EVENT_TIME, 'HH24')<='09'
   And to_char(APC_adj_on_off.EVENT_TIME, 'HH24') >= '06'
GROUP BY APC_adj_on_off.BT_VER, APC_adj_on_off.ROUTE_ID,
APC_adj_on_off.DIRECTION, APC_adj_on_off.STOPID,
APC_adj_on_off.STOPDESC
/
drop table APC_3_sum_by_sSeg
/
create table APC_3_sum_by_sSeg
as SELECT APC_3_SUM_BY_STOP.BT_VER, APC_3_SUM_BY_STOP.ROUTE_ID,
APC_3_SUM_BY_STOP.DIRECTION, GEO_8d_segmentID.UniqueSegmentID,
GEO_8d_segmentID.UniqueSegmentDesc, GEO_8d_segmentID.SegmentDesc,
GEO_8d_segmentID.SegmentID, Sum(APC_3_SUM_BY_STOP.TOTAL_ON) AS on_count,
Sum(APC_3_SUM_BY_STOP.TOTAL_OFF) as off_count
FROM APC_3_SUM_BY_STOP, GEO_8d_segmentID
where APC_3_SUM_BY_STOP.STOPID = GEO_8d_segmentID.STOPID
    AND APC_3_SUM_BY_STOP.DIRECTION = GEO_8d_segmentID.DIRECTION
    AND APC_3_SUM_BY_STOP.ROUTE_ID = GEO_8d_segmentID.ROUTE
    AND APC_3_SUM_BY_STOP.BT_VER = GEO_8d_segmentID.BT_VER
GROUP BY APC_3_SUM_BY_STOP.BT_VER, APC_3_SUM_BY_STOP.ROUTE_ID,
APC_3_SUM_BY_STOP.DIRECTION, GEO_8d_segmentID.UniqueSegmentID,
GEO_8d_segmentID.UniqueSegmentDesc, GEO_8d_segmentID.SegmentDesc,
GEO_8d_segmentID.SegmentID
/
drop table APC_adj_on_off
/

create table APC_adj_on_off
as SELECT APC_matched_on_off.BUS_ID, APC_matched_on_off.BT_VER,
        APC_matched_on_off.ROUTE_ID, APC_matched_on_off.PATTERN,
        APC_matched_on_off.TRIP_ID, APC_matched_on_off.DIRECTION,
        APC_matched_on_off.EVENT_TIME, APC_matched_on_off.STOP_SEQUENCE,
        APC_matched_on_off.STOPID, APC_matched_on_off.STOPDESC,
        APC_matched_on_off.DAY1, APC_matched_on_off.SUMOFON,
        APC_matched_on_off.SUMOFOFF, APC_matched_on_off.DIFF,
        boarding + (Abs(diff) - diff) * boarding/(2*sumofon) AS adj_on,
        alighting + (Abs(diff) + diff) * alighting/(2*sumofoff) AS adj_off
FROM APC_matched_on_off
WHERE (Abs(Diff)<0.15*SumOfon And (Abs(Diff))<0.15*SumOfoff))
ORDER BY APC_matched_on_off.TRIP_ID
/
drop table APC_3_sum_by_stop
/
create table APC_3_sum_by_stop
as SELECT APC_adj_on_off.BT_VER, APC_adj_on_off.ROUTE_ID,
        APC_adj_on_off.DIRECTION,
        APC_adj_on_off.STOPID, APC_adj_on_off.STOPDESC, APC_adj_on_off.DAY1,
        Sum(APC_adj_on_off.ADJ_ON) AS Total_on,
        Sum(APC_adj_on_off.ADJ_OFF) AS Total_off
FROM APC_adj_on_off
WHERE to_char(APC_adj_on_off.EVENT_TIME, 'HH24') <'09'
    And to_char(APC_adj_on_off.EVENT_TIME, 'HH24') >= '06'
GROUP BY APC_adj_on_off.BT_VER, APC_adj_on_off.ROUTE_ID,
        APC_adj_on_off.DIRECTION,
        APC_adj_on_off.STOPID, APC_adj_on_off.STOPDESC, APC_adj_on_off.DAY1
/
REM **********************
REM scale APC data by AFC data
REM **********************

drop table APC_factor_afc_sum
/
create table APC_factor_afc_sum
as SELECT ROUTE_ID, DIRECTION, Count(q.SER_NBR) AS all_afc, dayl
FROM (SELECT SeedM_3_AFC_stop.SER_NBR, SeedM_3_AFC_stop.USE_LLT,
        SeedM_3_AFC_stop.CURRENT_ROUTE,
        SeedM_3_AFC_stop.LAST_ROUTE,
        SeedM_3_AFC_stop.TRANS_EVENT, SeedM_3_AFC_stop.BUS_NBR,
        SeedM_3_AFC_stop.Bustools_VER_id,
        SeedM_3_AFC_stop.EVENT_TIME,
        SeedM_3_AFC_stop.TRIP_ID, SeedM_3_AFC_stop.ROUTE_ID,
        SeedM_3_AFC_stop.DIRECTION,
        to_char(event_time + NUMTODSINTERVAL(3, 'HOUR'), 'DD')
FROM SeedM_3_AFC_stop
WHERE to_char(SeedM_3_AFC_stop.USE_LLT, 'HH24') >= '06'
    And to_char(SeedM_3_AFC_stop.USE_LLT, 'HH24') < '09'
) q
GROUP BY ROUTE_ID, DIRECTION, Day1
/
drop table APC_factor_afc_on_apc
/ create table APC_factor_afc_on_apc
as SELECT ROUTE_ID, DIRECTION, Dayl, Count(SER_NBR) AS On_apc
FROM (SELECT q.SER_NBR, q.USE_LLT, q.CURRENT_ROUTE, q.LAST_ROUTE,
        q.TRANS_EVENT,
        q.BUS_NBR, q.EVENT_TIME, q.TRIP_ID, q.ROUTE_ID, q.DIRECTION,
        q.Dayl
FROM Trip_id_3_valid_APC INNER JOIN (SELECT SeedM_3_AFC_stop.SER_NBR,
        SeedM_3_AFC_stop.USE_LLT,
        SeedM_3_AFC_stop.CURRENT_ROUTE,
        SeedM_3_AFC_stop.LAST_ROUTE,
        SeedM_3_AFC_stop.TRANS_EVENT,
        SeedM_3_AFC_stop.BUS_NBR,
        SeedM_3_AFC_stop.Bustools_VER_id,
        SeedM_3_AFC_stop.EVENT_TIME,
        SeedM_3_AFC_stop.TRIP_ID,
        SeedM_3_AFC_stop.ROUTE_ID,
        SeedM_3_AFC_stop.DIRECTION,
        to_char(event_time + NUMTODSINTERVAL(3, 'HOUR'),
        'DD') Dayl
FROM SeedM_3_AFC_stop
WHERE to_char(SeedM_3_AFC_stop.USELLT, 'HH24')>='06'
And to_char(SeedM_3_AFC_stop.USE_LLT,
        'HH24')<>'09'
) q
ON (Trip_id_3_valid_APC.ROUTEID = q.ROUTE_id)
AND (Trip_id_3_valid_APC.TRIPID = q.TRIP_ID)
and (trip_id_3_valid_apc.dayl = q.dayl)
GROUP BY ROUTE_ID, DIRECTION, Dayl
/
drop table APC/scalefactor-by-dir
create table APC_scale_factor_by_dir
as SELECT APC_factor_afc_on_apc.ROUTEID,
        APC_factor_afc_on_apc.DIRECTION, APC_factor_afc_on_apc.Dayl,
        all_afc/on_apc AS factor
FROM APC_factor_afc_on_apc INNER JOIN APC_factor_afc_on_apc
ON (APC_factor_afc_on_apc.ROUTE_ID = APC_factor_afc_on_apc.ROUTE_ID)
AND (APC_factor_afc_on_apc.DIRECTION =
        APC_factor_afc_on_apc.DIRECTION)
and (apc_factor_afc_sum.Day1 = apc_factor_afc_on_apc.Day1)
where all_afc <= 10*onapc
/
drop table APC_4_scaled
create table APC_4_scaled
as SELECT APC_3_sum_by_stop.BTVER, APC_3_sum_by_stop.ROUTE_ID,
        APC_3_sum_by_stop.DIRECTION, APC_3_sum_by_stop.STOPID,
        APC_3_sum_by_stop.STOPDESC, APC_3_sum_by_stop.Day1,
        total_on*factor AS Scaled_on,
        total_off*factor AS Scaled_off
FROM APC_3_sum_by_stop INNER JOIN APC_scale_factor_by_dir
ON (APC_3_sum_by_stop.ROUTE_ID = APC_scale_factor_by_dir.ROUTE_ID)
AND (APC_3_sum_by_stop.DIRECTION =
APC_scale_factor_by_dir.DIRECTION)
and (apc_3_sum_by_stop.Day1 = apc_scale_factor_by_dir.Day1)
/
drop table APC_4d_scaled_Seg
/
create table APC_4d_scaled_Seg
as SELECT APC_4_scaled.BT_VER, APC_4_scaled.ROUTE_ID,
APC_4_scaled.DIRECTION,
Sum(APC_4_scaled.Scaled_on) AS SumOfScaled_on,
Sum(APC_4_scaled.Scaled_off) AS SumOfScaled_off,
GEO_8D_SEGMENTID.SEGMENTID, GEO_8D_SEGMENTID.SEGMENTDESC
FROM APC_4_scaled INNER JOIN GEO_8D_SEGMENTID
ON (APC_4_scaled.STOPID = GEO_8D_SEGMENTID.STOPID)
AND (APC_4_scaled.DIRECTION = GEO_8D_SEGMENTID.DIRECTION)
AND (trim(APC_4_scaled.ROUTE_ID) =
trim(GEO_8D_SEGMENTID.ROUTE))
AND (APC_4_scaled.BT_VER = GEO_8D_SEGMENTID.BT_VER)
GROUP BY APC_4_scaled.BT_VER, APC_4_scaled.ROUTE_ID,
APC_4_scaled.DIRECTION,
GEO_8D_SEGMENTID.SEGMENTID, GEO_8D_SEGMENTID.SEGMENTDESC
/
drop table APC_5_scale_factor_5days
/
create table APC_5_scale_factor_5days
as select route_id, direction, sumon_apc/sumofallafc as factor
from (
    SELECT APC_FACTOR_AFC_SUM.ROUTE_ID, APC_FACTOR_AFC_SUM.DIRECTION,
    Sum(APC_FACTOR_AFC_SUM.ALL_AFC) AS SumOfALL_AFC,
    Sum(APC_FACTOR_AFC_ON_APC.ON_APC) AS SumOfON_APC
    FROM APCFACTORAFCONAPC RIGHT JOIN APCFACTORAFCSUM
    ON (APC_FACTOR_AFC_ON_APC.DIRECTION =
    APC_FACTOR_AFC_SUM.DIRECTION)
    AND (APC_FACTOR_AFC_ON_APC.ROUTE_ID =
    APC_FACTOR_AFC_SUM.ROUTE_ID)
    GROUP BY APC_FACTOR_AFC_SUM.ROUTE_ID,
    APCFACTORAFCSUM.DIRECTION
)
/
drop table APC_4b_scaled_5_day
/
create table APC_4b_scaled_5_day
as SELECT APC_3_SUM_BY_STOP.BT_VER, APC_3_SUM_BY_STOP.ROUTE_ID,
APC_3_SUM_BY_STOP.DIRECTION, APC_3_SUM_BY_STOP.STOPID,
APC_3_SUM_BY_STOP.STOPDESC, Sum(TOTAL_ON/factor) AS Boarding,
Sum(TOTAL_OFF/factor) AS Alighting,
APC_5_SCALE_FACTOR_5DAYS.FACTOR
FROM APC_5_SCALE_FACTOR_5DAYS RIGHT JOIN APC_3_SUM_BY_STOP
ON (APC_5_SCALE_FACTOR_5DAYS.ROUTE_ID =
APC_3_SUM_BY_STOP.ROUTE_ID)
    AND (APC_5_SCALE_FACTOR_5DAYS.DIRECTION =
    APC_3_SUM_BY_STOP.DIRECTION)
GROUP BY APC_3_SUM_BY_STOP.BT_VER, APC_3_SUM_BY_STOP.ROUTE_ID,
    APC_3_SUM_BY_STOP.DIRECTION, APC_3_SUM_BY_STOP.STOPID,
    APC_3_SUM_BY_STOP.STOPDESC, APC_5_SCALE_FACTOR_5DAYS.FACTOR
/
drop table APC_4C_SCALED_5_DAY_seg
/ create table APC_4C_SCALED_5_DAY_seg
as SELECT APC_4B_SCALED_5_DAY_seg.BT_VER, APC_4B_SCALED_5_DAY_seg.ROUTE_ID,
APC_4B_SCALED_5_DAY_seg.DIRECTION, APC_4B_SCALED_5_DAY_seg.STOPID,
APC_4B_SCALED_5_DAY_seg.STOPDESC, APC_4B_SCALED_5_DAY_seg.BOARDING,
APC_4B_SCALED_5_DAY_seg.ALIGHTING, APC_4B_SCALED_5_DAY_seg.FACTOR,
GEO_9C_MULT_SEG_3_BASE.MSEGID, GEO_9C_MULT_SEG_3_BASE.MSEGDESC
FROM APC_4B_SCALED_5_DAY
INNER JOIN GEO_9C_MULT_SEG_3_BASE
ON (APC_4B_SCALED_5_DAY_seg.BT_VER = GEO_9C_MULT_SEG_3_BASE.BTVER)
AND (APC_4B_SCALED_5_DAY_seg.DIRECTION = GEO_9C_MULT_SEG_3_BASE.DIRECTION)
AND (APC_4B_SCALED_5_DAY_seg.STOPID = GEO_9C_MULT_SEG_3_BASE.STOPID)
WHERE Trim(ROUTE_ID)=Trim(route)
GROUP BY APC_4B_SCALED_5_DAY_seg.BT_VER, APC_4B_SCALED_5_DAY_seg.ROUTE_ID,
APC_4B_SCALED_5_DAY_seg.DIRECTION, APC_4B_SCALED_5_DAY_seg.STOPID,
APC_4B_SCALED_5_DAY_seg.STOPDESC, APC_4B_SCALED_5_DAY_seg.BOARDING,
APC_4B_SCALED_5_DAY_seg.ALIGHTING, APC_4B_SCALED_5_DAY_seg.FACTOR,
GEO_9C_MULT_SEG_3_BASE.MSEGID, GEO_9C_MULT_SEG_3_BASE.MSEGDESC
/
drop table APC_5_OD_MSeg
/
create table APC_5_OD_MSeg
as SELECT APC_4C_SCALED_5_DAY_seg.BT_VER,
APC_4C_SCALED_5_DAY_seg.ROUTE_ID,
APC_4C_SCALED_5_DAY_seg.DIRECTION,
Sum(APC_4C_SCALED_5_DAY_seg.BOARDING) AS SumOfBOARDING,
Sum(APC_4C_SCALED_5_DAY_seg.ALIGHTING) AS SumOfALIGHTING,
APC_4C_SCALED_5_DAY_seg.MSEGID, APC_4C_SCALED_5_DAY_seg.MSEGDESC
FROM APC_4C_SCALED_5_DAY_seg
GROUP BY APC_4C_SCALED_5_DAY_seg.BT_VER,
APC_4C_SCALED_5_DAY_seg.ROUTE_ID,
APC_4C_SCALED_5_DAY_seg.DIRECTION, APC_4C_SCALED_5_DAY_seg.MSEGID,
APC_4C_SCALED_5_DAY_seg.MSEGDESC
/
D. Main Software Tools Used

Access 2002 SP3, XP, Published by Microsoft
Excel 2002 SP3, XP, Published by Microsoft
Matlab 7.0.4, Published by MathWorks
Oracle Database 9.2.0, Published by Oracle
TransCAD 4.8, Published by Caliper Corp