Understanding Transit Travel Behavior: Value added by Smart Cards

by

Saumya Gupta

Bachelor of Technology in Civil Engineering
Indian Institute of Technology, Bombay, 2004

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for the Degree of

Master of Science in Transportation

at the

Massachusetts Institute of Technology

June 2006

Signature of Author: ______________________________

Department of Civil and Environmental Engineering

Certified by: ______________________________

Mikel Murga
Lecturer of Civil and Environmental Engineering
Thesis Co-Supervisor

Certified by: ______________________________

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

Accepted by: ______________________________

Andrew J. Whittle
Chairman, Departmental Committee for Graduate Students

Understanding Transit Travel Behavior – Value added by Smart Cards

by
Saumya Gupta

Submitted to the Department of Civil and Environmental Engineering on May 26, 2006 in Partial Fulfillment of the Requirements for the Degree of Master in Science in Transportation

ABSTRACT

Travel behavior represents a particularly complex area of research in transportation given the interaction between transport supply characteristics and the user perceptions which guide his/her decisions. Thanks to the advent of automated data collection including smart cards, each customer transaction can now be recorded, providing a far more accurate, detailed and continuous stream of data for travel behavior analysis than the data previously available only from conventional travel surveys. This thesis explores the opportunities afforded by the data being generated by systems such as Automated Fare Collection (AFC), Automated Vehicle Location (AVL) and fare media in the form of smart cards, focusing on the Chicago Transit Authority and its smart card known as the Chicago Card. The processing of this data allows insights on transit user's decisions concerning the access to the transit system both at the origin and destination of the trip, together with the actual path choice and travel behavior dynamics.

The knowledge gained on transit users' behavior could lead to important policy recommendations such as the value of consolidating or adding routes and services. In this context, the penetration of Chicago Card users among all system riders has been examined to determine how representative is the behavior of current card users. These card holders could also serve as a de facto longitudinal panel to gauge reactions to changes in the transit Level of Service (LOS). This thesis also prepares the ground for future path-choice modeling.

The thesis presents an initial analysis of some basic travel parameters including frequency, time consistency, access distances, and route variability both for the first and the return trip among consistent users, as recorded during two weeks in September 2004 and September 2005. Given the exploratory nature of this thesis, several examples of actual trips have been visualized in order to develop some working hypotheses on travel behavior that might drive future research in this area. Finally, the thesis also discusses some of the potential data issues (reliability, synchronization and integrity between AFC and AVL records) involved in these analyses.

Thesis Co-Supervisor: Mikel Murga
Title: Lecturer of Civil and Environmental Engineering

Thesis Co-Supervisor: Nigel H. M. Wilson
Title: Professor of Civil and Environmental Engineering
Director, Master of Science in Transportation Program

TABLE OF CONTENTS

TABLE OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisors Mikel Murga and Prof. Nigel H. M. Wilson, to have given me the opportunity to work on this interesting topic. Mikel's unparalleled enthusiasm and expertise in TransCAD saw me through the unnerving and complex phase of analyses. His support kept me going and exploring new applications of smart cards. While I was engrossed in analysis and could not foresee the importance of "keep writing" philosophy, Mikel's discipline got me to the stage of completing the thesis on time. I credit his meticulous overseeing, immense patience and dedication during the daily meetings which greatly improved and enriched my thesis. At this point, I cannot fail to mention that his computer savvy ways and insistence on video conference kept the thesis alive and nurturing even while he was in Spain.

I owe a lot of my understanding of the thesis topic to Prof. Nigel H. M. Wilson. I am greatly thankful for his invaluable inputs, for his extreme patience with my writing, and critical feedback which has evolved the thesis to its present form. His guidance has rescued me from getting digressed time and again. His insistence on quality work has seen through many stages of analysis and write-ups. In retrospect this was a difficult phase of learning and self-improvement but I believe that this will be helpful in my future life.

While working on my thesis, it was very helpful to have words of encouragement from Fred Salvucci. I would also like to thank my lab-mates – Giorgia, Joanne, Hong Yi, Chris, Tara, Mike, Lavanya, and Drew, for their faith and encouragement which kept me going. I am also thankful to John Attanucci to have gone through the initial phases of my research and steered it in the right direction. Ginny has always been there to help us – no matter what, you go to her and it gets resolved! Kudos to her invaluable assistance.

I would also like to express gratitude to my parents and sisters, whose constant encouragement kept me going. It was extremely difficult in the end, but their boost saw me through it.

Last but not the least, I acknowledge the support of my boyfriend, Saurabh Gupta all the while I was working on my thesis. His contributions to the thesis are invaluable. He has helped me with the analysis, with the interpretation of the results, and put all this on paper. He has gone through many sleepless nights of reading and revising my drafts... This thesis would have never been completed on time without his encouragement. I thank him dearly and sincerely.

1 Introduction

This thesis examines smart cards[1] as a valuable new data source for transit research. We can gain additional benefits by linking smart card data with the Automated Vehicle Location[2] (AVL) data. The thesis, therefore, develops tools to extract information from smart cards and link it with the AVL data; uses this information to better understand travel behavior; and recommends future work that can expand its applicability.

1.1 Motivation

Travel behavior has been one of the most important as well as intriguing areas of research in transportation. It represents a particularly complex topic given the interaction between transport supply characteristics and the user perceptions which guide their decisions[3-5]. With the advent of automated data collection systems and specifically smart cards, individual transactions can now readily be recorded[1], providing a more accurate, detailed and continuous stream of data for travel behavior analysis, than the data previously available only from conventional travel surveys[6].

The introduction of automated data collection systems into the public transport sector has substantially improved the quantity and quality of available data for varied applications[7, 8]. One such application, which is very valuable for service design and planning by transit agencies, is the estimate of passenger volumes between a given origin and destination, commonly referred to as the 'OD matrix' [8, 9]. Conventionally the estimates of travel volumes have relied heavily on data from passenger surveys and manual counts which are expensive to collect, and process, and subject to bias and error [8, 10]. Incorporating smart card information to estimate O-D matrices can be used to counter these problems[7, 8]. Linking smart card data with the census and the AVL data offers a wide range of research applications. These are summarized in Table 1-1 [11]

Table 1-1: Potential uses of a linked AFC-user database [11]

Market Research and Service Planning
Analysis of demographics (e.g. age, gender, income) of riders by route/station
Analysis of travel characteristics (e.g. frequency of use, transfer patterns, etc.) of riders by route/station
Analysis of travel by riders with particular demographics
Analysis of travel by riders with particular travel characteristics
Analysis of demographics of riders making particular trip patterns
Analysis of travel characteristics of riders making particular trip patterns
Analysis of the spatial coverage of transit network system
Analysis of changes in travel patterns over time by people with particular demographics
Analysis of changes in travel patterns over time by people with particular travel characteristics
Analysis of the demographics of riders by time of day
Analysis of the transportation characteristics of riders by time of day
Analysis of demographics of riders using particular services (e.g. express, limited stop, night owl)
Identification of individuals for detailed survey or focus groups
Development of a mailing list for public meeting notices
Travel Demand Forecasting
Provision of a large sample transit "travel diary", including demographic data
Study of travel changes as reactions to fare changes (elasticity) by demographics
Operations
Development of a mailing list for service change announcements
Development of an e-mail list for delays and emergency detours
Pricing/Fare Policy
Analysis of complete trip-making patterns to evaluate new fare products
Evaluation of the feasibility of a trip frequency-based discount or a "guaranteed best fare" policy
Analysis of travel by fare category
Study of price-elasticities by demographic characteristics
Marketing
Identification of distinct market segments of riders
Targeting of marketing information to the most appropriate users
Identification of market segments with low penetration
Use of demographic database to conduct targeted surveys
Advertising
Development of route/station demographic profiles to identify target locations for particular advertisers and set advertising rates accordingly

An earlier work by Utsunomiya et al.[6] establishes smart cards as a powerful research tool. Her research used Chicago Card information to analyze walk access distance, frequency and consistency of daily travel patterns, and the variability of behavior by residential area. These analyses of variation in boarding location and bus route describe the initial path-choice decision of the customers. This thesis, which focuses on the potential of smart cards to analyze travel behavior, derives its inspiration from Utsunomiya's work.

We will explore and implement methodologies to extract useful information from smart cards and use this information to understand better transit customer travel behavior. Amongst different forms of automated data available [Automated Fare Collection (AFC) [1, 12], Automated Passenger Counts (APC) [13], Automated Voice Annunciation System (AVAS) [14, 15], Automated Vehicle Location (AVL) [2]], we will utilize AFC and AVL systems for our data requirements and processing. This data will be used in conjunction with the Geographical Information System (GIS) network modeling capabilities of TransCAD to analyze and understand origin-destination travel patterns for the smart card users.

1.2 Objectives and scope of the thesis

In this thesis we address two major objectives as outlined below:

1. Develop methodologies for extracting useful information on transit travel behavior from smart cards
2. Use this information to understand transit travel behavior

We meet these objectives by using CTA as our case study and analyzing the Chicago Card and AVL data. As an outcome of this research we confront and address the challenges of extracting, processing, and linking the data; gain insights into travel characteristics of Chicago's smart card holders and make interesting observations which can in the future be developed into full fledged research projects.

In this thesis we will describe the different kinds of analyses that can be conducted using automated data collection systems. In addition, the travel characteristics for customers based on their first trip of the day are determined; and path-choice characteristics for a small sample of customers are studied by manual analysis. This thesis therefore prepares the ground for future path-choice modeling by identifying and understanding the travel behavior characteristics of frequent and consistent customers.

1.3 Organization of the thesis

The remainder of this thesis comprises four chapters. Chapter 2 introduces the various automated data collection systems that are the focus of this research. It also describes the CTA transit network. Chapter 3 discusses the methodologies for extracting useful data from the Smart Cards. It describes a methodology for linking the AFC-AVL data for determining bus boarding locations and the processing approach for calculating the access distance. In Chapter 4, we apply the methodologies that have been developed in Chapter 3 to the CTA transit network to understand the travel characteristics of various CTA transit users. While the aggregate analysis carried out in this chapter leads to important insights into current transit performance, a microscopic analysis of path choice by a small sample of users prepares the ground for future path choice modeling. Chapter 5 summarizes the results of this research and makes suggestions for future work.

2 Examining Travel behavior using Smart Cards

This chapter introduces the various automated data collection systems that have been the focus of this research. We begin this chapter with a review of smart cards (Section 2.1), followed by a review of related automated data collection techniques – Automated Fare Collection (AFC) and Automated Vehicle Location (AVL) (Section 2.2). The CTA network, which forms the case study for our research, is described in Section 2.3.

2.1 Smart Cards

Smart cards are plastic cards (like a credit card) with an embedded integrated circuit for storing information. Since the introduction of Automated Fare Collection (AFC) system in the public transport sector [1], the use of smart cards as fare payment media is becoming increasingly popular.

Smart cards are extremely simple to use - load the card with monetary value via kiosks at transit stations or over the Internet, and then pass the card over a sensor at a station, and the fare is paid[1]. This greatly enhances customer convenience by allowing easy and quick processing[1]. Because the information for each smart card transaction can be stored, an agency gets access to a continuous stream of data on user transactions. If the penetration of smart cards is high and uniform across the transit agency's network, this data can be a vital complement to, or even substitute for the manually conducted surveys typically used to learn about transit customers.

Amongst AFC fare media, smart card data offers the following additional advantages from the research perspective:

1. Every smart card has a unique serial number identified with an individual which permits us to observe his/her travel patterns[7].
2. The long life of smart cards gives us an opportunity to observe individual travel behavior over a longer period of time. Longevity of smart cards also makes them suitable to be used as a form of longitudinal panel data.
3. For most transit agencies, smart card registration requires a customer to enter his name, billing address, shipping address, phone number, card pin and e-mail address. The address information gives us accurate origin information directly for most customers although some customers will provide a non-home address as the contact address.

2.2 Automated Data Collection Systems

Smart cards are one element in a type of Automated Data Collection System[16], the Automated Fare Collection (AFC) systems. In this section we introduce AFC system more formally as well as Automated Vehicle Location (AVL) system which is another important element used in this research to obtain full value from smart card behavioral analysis.

2.2.1 Automated Fare Collection Systems

In the 1970s, the concept of the magnetic stripe ticket with electronic entrance/exit gate was introduced, with the hope of solving some of the problems inherent in traditional cash-based fare collection systems. Although these magnetic stripe-based systems do address some of the inherent problems with the cash-based system, they are costly to maintain and do not solve the regional interoperability challenge. To address these costs and the interoperable network issues, the smart card has become the logical replacement for the magnetic stripe ticket. The business case for smart card-based automated fare collection is robust and based on several factors, including reduced maintenance costs, faster throughput for passengers through turnstiles, fraud reduction, more flexible fare schemes, specific event schemes and system management information. In its full implementation, the smart card can allow customers to travel seamlessly between modes – subway, rail, bus, taxi, ferry – using the same payment mechanism.

Automated Fare Collection[17] and smart cards address several of the key issues facing transit today [12]:

- **Increasing customer convenience and satisfaction**. Finding ways to make it easier for people to use transit will in turn increase ridership and generate more revenue. The smart card can speed passenger's boarding process, reduce times in queues and allow for easier access to multiple forms of transit using the same payment mechanism.

- **Reducing costs of collecting fares**. Studies show that transit agencies spend on average 15 cents out of every $1 just collecting fare revenue [12]. Automating this labor intensive and time consuming process can help drive down ticketing costs by at least 50%, improve traffic-pattern monitoring, minimize cash handling and decrease errors – all while increasing customer satisfaction.
- **Lowering lost revenues due to fraud**. Losses from passengers traveling without a ticket can greatly cut into the bottom line[18]. With Automated Fare Collection, passengers whose smart card payments are not up-to-date are not allowed access to the transit system.

2.2.2 Automated Vehicle Location Systems

Automatic Vehicle Location (AVL) [2] is a computer-based vehicle tracking system. For transit, the actual real-time position of each vehicle is determined and relayed to a control center. Actual position determination and relay techniques vary, depending on the needs of the transit system and the technologies employed. The dominant AVL technology deployed today is GPS, representing close to 75% of all systems deployed [19]. According to a 2002 U.S. D.O.T report, 172 transit agencies in the U.S. had deployed AVL systems [20].

Dead reckoning is sometimes used to supplement the GPS system. Both systems transmit location information while one system serves to compensate for the other when a bus reaches an area where one of the systems is ineffective. In particular, tall buildings and underpasses make dead reckoning an effective complement to GPS. The combination of the two systems also permits a high degree of accuracy in locating vehicles[2]. These two techniques are briefly described below:

a) Dead-reckoning

Dead reckoning is among the oldest navigation technologies [2]. Dead reckoning sensors can measure distance and direction from a fixed point (under the most basic setup, an odometer and compass could be used to calculate position from a specific stop on a route). Typically, these systems act as a backup to another AVL system. This relatively inexpensive system is self-contained on the bus. This system has a number of drawbacks including:

- Uneven surfaces and hills can compromise the positioning information
- Should the vehicle leave a fixed route, its location will no longer be known since there will be no waypoints off the fixed route
- Accuracy degrades with distance traveled, and regular recalibration is required (for example, tire circumference changes with wear).

b) Global Positioning System (GPS)

Due to the shortcomings of other AVL technologies, GPS has become the most popular system for new installations over the last few years [2]. GPS utilizes the signals emitted from a network of 24 satellites, which are picked up by a receiver placed onboard the bus. The satellite system covers almost all of North America, eliminating the need to place transmitters/receivers along any route. The existence of the satellite system means that the main cost for the agencies result from purchase of the GPS receivers and the equipment needed to transmit to communication control center. While the U.S. military, which oversees the satellite system, has limited the accuracy of the system in the past, it is now allowing more accurate readings. The accuracy and low cost of GPS makes it the most appealing, though it too has some problems. Foliage, tall buildings, and tunnels can temporarily block the satellite signal, and at times satellite signals do not reach specific locations. Typically dead reckoning is used in conjunction with GPS to fill in such gaps.

Recent evidence indicates that AVL technology is leading to significant gains in productivity as well as increases in transit ridership[19]. AVL technology coupled with effective control strategies allows improved schedule adherence and timed transfers, more accessible passenger information, increased availability of data for transit management and planning, and efficiency/productivity improvements in transit services. AVL also creates many possibilities for ITS systems integration including: providing transit buses with traffic signal priority; incorporating transit information in traveler information systems; developing multi-application electronic payment systems and using buses to automatically communicate traffic speeds.

2.3 CTA Transit Network

The CTA operates the nation's second largest public transportation system covering the city of Chicago and 40 surrounding suburbs [21] (Figure 2-1). On an average weekday, nearly 1.5 million rides are taken on the CTA. About 1 million passenger trips a day are taken on the bus system which is served by approximately 2,000 buses that operate over 151 routes and 2,273 route miles. All the buses are AVL enabled, with dead reckoning used as a supplement the GPS system. CTA's 1,190 rapid transit cars operate over seven routes and 222 miles of track serving about 500,000 customer trips each day at 144 stations [21].

CTA's Automated Fare Collection (AFC) system, which was installed in 1997, replaced the CTA's tokens with magnetic-stripe stored-value cards printed on plastic. In 1999, a small pilot program ("Chicago Gold") for disabled riders in 1999 led to universal installations of smart card readers in 2000. In August 2000, a more extensive pilot program distributed stored-value "Chicago Cards" to volunteers. The pilot program was expanded system-wide in November 2002. The Chicago Card Plus, which could link directly to the credit or debit card account of a transit user, was introduced in January 2004[22]. There is an active effort towards getting the cards registered as this permits CTA to track user movements through the system, thus providing the kind of data which is examined within the thesis. The information required for registration of a Chicago Card includes name, billing address, shipping address, phone number, card pin and e-mail address.

CTA smart cards account for 16% of all the system rides (10% on bus system, 27% on rail system) and 29.2% of first ride market (20% on bus system, 41% on rail system) share for March 2006[23]. These numbers clearly show that smart cards are not yet ubiquitous throughout the CTA system. Studies[24] also reveal that the geographical and modal distribution of smart cards is not uniform. Smart cards are more widespread in the affluent northern residential areas of Chicago[25] and are much more popular amongst rail users than bus users.

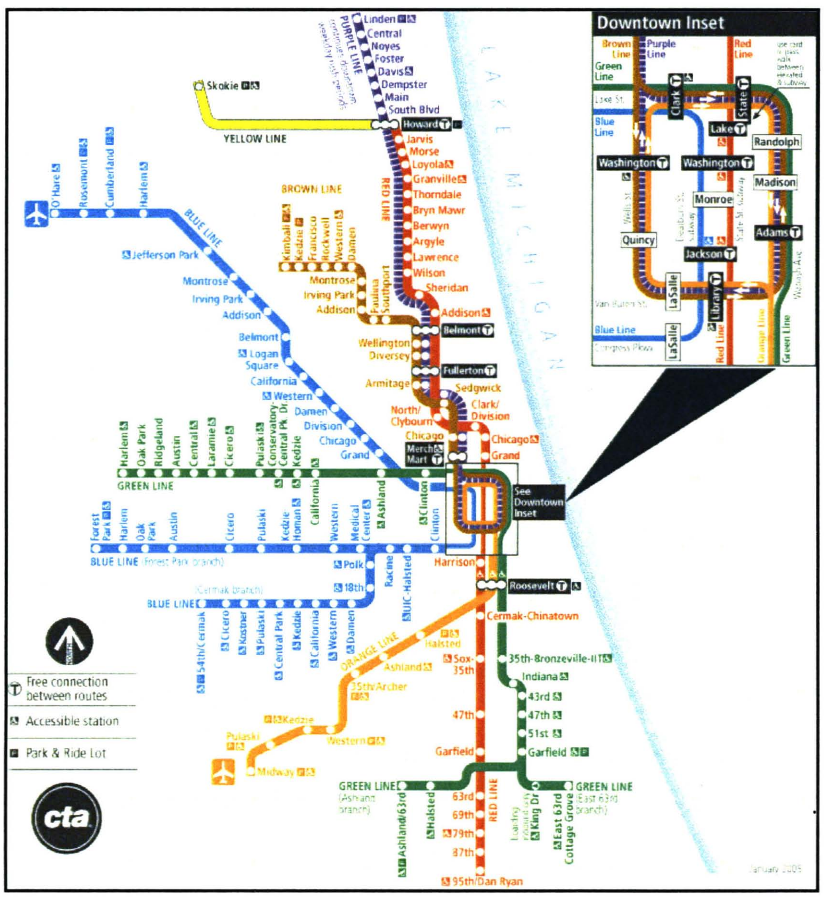


Figure 2-1: CTA's rail system [26]

At current levels of penetration, which does not provide a fully representative customer base, smart card information cannot be the sole basis for service design and planning decisions in Chicago. Given this limitation this thesis proposes tools which open up new ways to examine and analyze data. The methods proposed in this thesis only become more valuable as smart card penetration increases and usage becomes more prevalent.

3 Methodology for Extracting Travel Information

This chapter outlines different methodologies which have been developed to extract useful information from smart cards with the aim of improving our understanding of travel behavior of transit users. Section 3.1 describes the method used to identify the boarding stop for bus customers. Section 3.2 defines the method used to compute access distances from home to the first CTA transit service used daily. Finally Section 3.3 discusses the problem of inferring the destination of a trip.

3.1 Determining boarding location

The information about a customer's boarding location can be extracted from smart card transactions. This section develops methodologies for inferring the boarding location of a customer using AFC to AVL linking. Smart Card transactions are made whenever a Chicago Card user enters an AFC gate at a rail station or uses a fare-box on a bus. While it is straight forward to determine rail boarding location, inferring the bus boarding locations is not trivial. For locating the bus stop at which a transaction is made, one must be able to locate the position of the bus at the time transaction was made. The time which gets recorded when the users swipes the smart card in the bus is the AFC time, and the time which is recorded when the bus passes through a stop is the AVL time. Therefore, by AFC-to-AVL linking we can determine the bus stop boarding location.

3.1.1 AFC-to-AVL linking for inferring bus boarding locations

Figure 3-1 shows the different tables, their data fields and the match required to infer the bus boarding locations. We start with the AFC records, locate the bus where the transaction occurred and then perform multiple matches with tables associated with AVL data to get the bus stop descriptions in the form of cross streets. These cross streets are locations where the AFC transaction took place. The AFC-to-AVL linking is a multi-table match and a missing or erroneous link anywhere during the match process will prevent inference of the stop location. Such situations may arise due to a wrong or incomplete login by the driver and/or due to incoherent look-up tables.

This algorithm for AFC-to-AVL linking consists of three steps:

- Locating the bus which the customer boarded,
- Selecting the AVL time for this bus that corresponds to the AFC transaction,
- Finding the location of the bus at the matched AVL time.

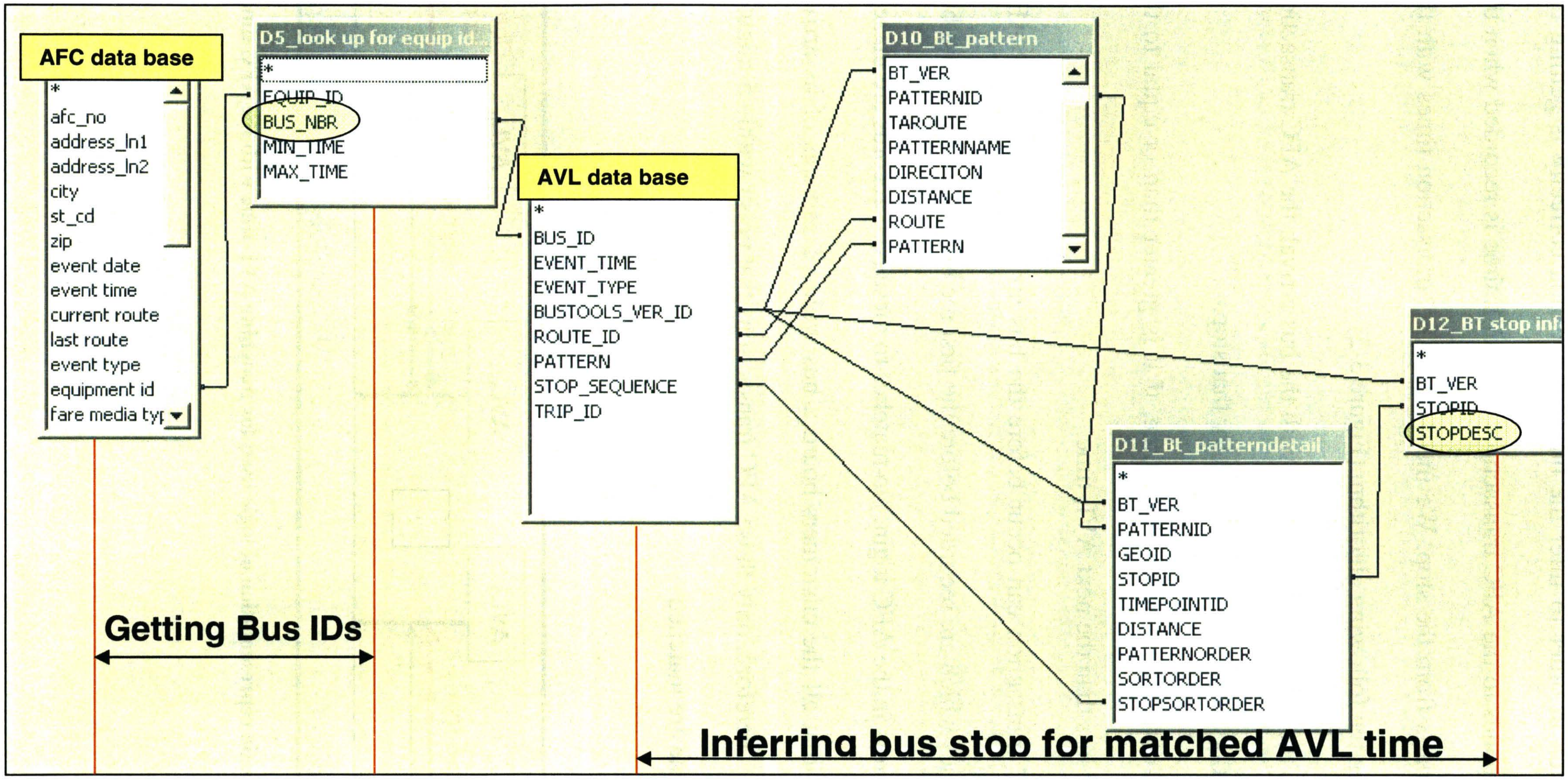


Figure 3-1: Multi-table match to infer bus stops

The most critical match in order to infer the bus stop location correctly is getting the AVL time that corresponds to the AFC transaction. The AVL time is recorded when the bus arrives at and departs from the stop. We match the AFC transaction times with the arrival AVL time using the following algorithm (Figure 3-2).

- Since we have the AVL time associated with the bus arrival, the AFC transaction for a stop must occur after this AVL time for that stop.
- An AFC time corresponds to an AVL time, if it is greater than or equal to the AVL time and less than the next AVL time.
- Since no AFC transaction can occur before the bus arrives at a stop, any AFC transaction times which are recorded before the first recorded AVL time for a bus are taken as errors in the AFC login. No matches to these AFC records are made.
- It is expected that all the customers board a bus within 5 minutes of its arrival time at a stop. Therefore, only those AFC transactions that occur within 5 minutes of the AVL time are matched.

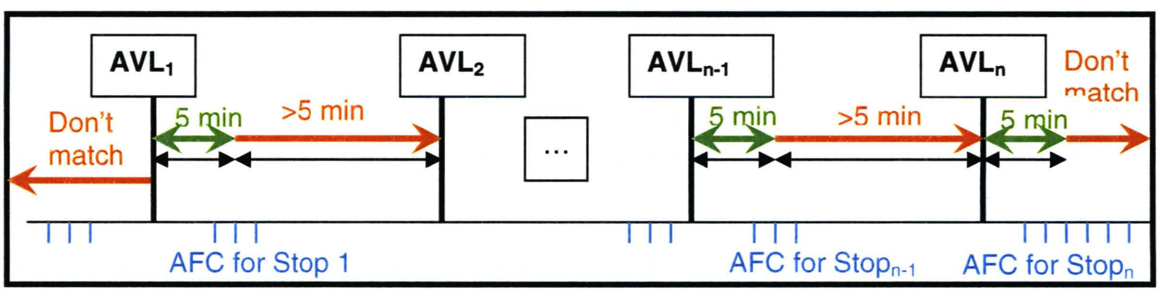


Figure 3-2: Diagrammatic representation of logic used for matching AVL time with the AFC time

In addition to inferring the boarding stop, other useful information such as the distribution of inter AVL times for a bus, distribution of time elapsed between an AVL event and its corresponding AFC events etc. can be obtained from this analysis.

3.2 Access distance Analysis

One of the important parameters affecting the choice of boarding location is the access distance[27, 28]. In this thesis we compute the access distances in TransCAD by measuring the actual distance between the origin node of a user and the station location. In earlier work by Utsunomiya[6], access distances were approximated as distances between nearest nodes to the origin and station location. In this thesis, we calculate the exact access distance shown by the solid line in Figure 3-3 compared with the nearest node distance[6] represented by the dotted line. This is achieved by using geocode and tag features of TransCAD[29].

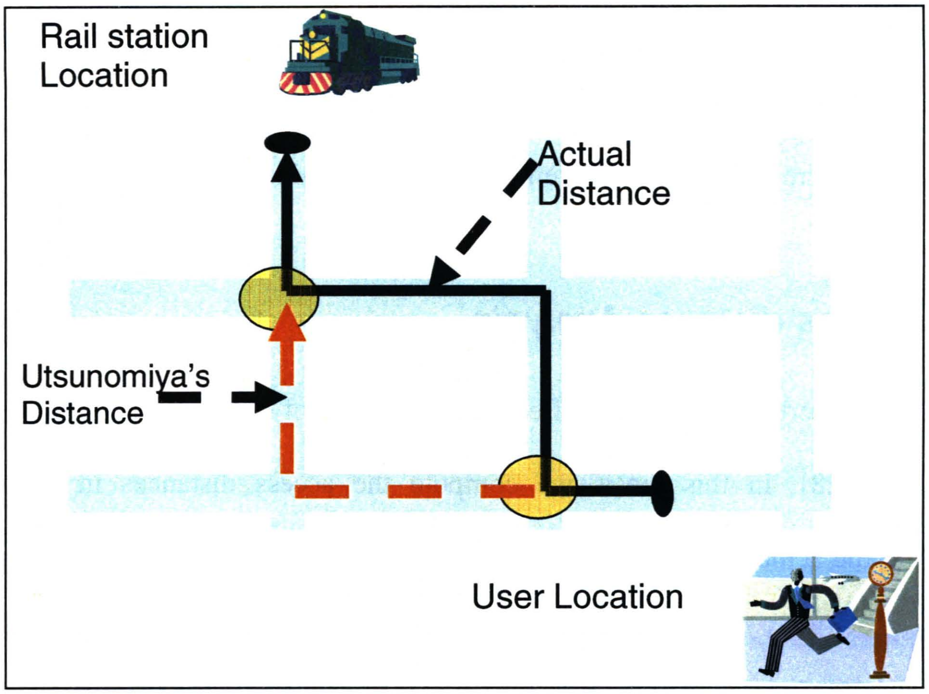


Figure 3-3: Schema showing the actual and nearest node access distance

3.3 Inferring destination ends of a user

For path choice analysis it is important to infer the destination end of the trip. However, since the exit transaction is not recorded, it is not trivial to determine the destination. In this section we describe our approach to infer the destination end of the work trip. While, the current implementation of this procedure is manual, the possibility of automating this step exists.

The procedure we have used to infer the destination end of the first trip involves manually analyzing the daily trip itineraries of customers according to the following steps:

1. If there is significant time difference between the first and the second trip of the day, then the second trip boarding location may well be the destination of the initial trip, or at least in the vicinity of the destination.

2. As a confirmation of whether it is indeed the destination, the frequency of the second boarding location is examined. If a customer repeatedly boards at the same second boarding location, then this confirms that the he is indeed making daily trips to a destination in the vicinity of that station.

Another refinement which can be applied to infer destinations systematically is based on the following logic:
For those customers who make only two trips in a day, the second trip boarding location is very likely to be the destination end of the first trip. And the trip made at the second boarding location is likely to be the return home trip.

For this latter case, we isolated all such trip pairs for the week of analysis and inferred the second boarding location to be in the vicinity of the destination of their first trip of the day.

3.4 Conclusions

In this chapter, we first presented a methodology to infer the bus boarding location based on linking of AFC and AVL records. An approach to calculate the exact access distances for a transit user was also presented. Finally, a methodology to infer the destination end of a user is presented.

4 Application to Chicago Transit Authority

In this chapter, we explore the travel behavior of users of the Chicago Transit Authority by extracting useful information from Chicago Cards using the methodologies described in chapter 3. In Section 4.1 a process for selecting and classifying frequent CTA users by first mode of travel is presented with the results obtained. In Section 4.2 the results of the process to determine boarding location are presented. Section 4.3 presents the classification of CTA Chicago Card customers by mode taken. Section 4.4 presents the results of the analysis of access behavior variability. Section 4.5 presents the results of the analysis of access distance. Section 4.6 presents the results of the analysis of path choice. Finally, Section 4.7 presents the results of the analysis of Chicago Card penetration.

4.1 Selecting regular and consistent users

In order to understand path-choice behavior[4], we have chosen a sample of transit users who make frequent and consistent trips between a given origin and a destination. There is no easy way to infer the destination end of trip for a customer since there are no system exit transactions given the flat fare system in place at the CTA. In this section, we will develop a methodology to select those users whose first trips of the day are expected to be work trips based on their frequent and time consistent travel habits. For these trips the destination can be assumed to be the same for a user.

Our sample customers for the path-choice study should exhibit the following travel characteristics:

a) **Frequent-weekday** – The first prerequisite for a customer is that he should be a **frequent** CTA user, so that we can trace his trips over multiple days. In addition we can expect that a frequent user knows the system well and makes rational path-choice decisions. We use the **weekday** trips made by the user in order to avoid the different behavior typical on weekends. For this thesis, a customer who uses the transit system for 3 or more weekdays is defined as a frequent weekday customer.

b) **Time Consistent** – Since we want to select those users who make regular work trips, we enforce the condition that they should also be time consistent, which means that they access the transit system at approximately the same time each day. For this thesis, a frequent weekday customer who accesses the transit system within a time window of 30 minutes for at least 3 weekdays is defined as time consistent.

The rationale for defining these two conditions will be developed in the following two sub-sections. By enforcing these two conditions, our sample will consist of customers who are frequent and 30 minute time consistent on their first trip for at least 3 week days during the week used for analysis. This knowledge should prove important for CTA as commuting trips mean repeated business which should be at the core of every marketing effort.

4.1.1 Frequent Weekday Customers

In this section users are classified based on frequency of use. The distribution of customers based on frequency of use for two weeks in September 2005 and September 2004 is shown in Table 4-1 and Table 4-2 respectively. There are 57140 and 42094 frequent weekday customers for September 2005 and September 2004 representing 68% and 71% of all Chicago Card (Plus) customers respectively. While the absolute numbers of customers have significantly increased in September 2005, the frequency of use distribution is similar for the two years. Therefore, we can say that most of the growth in Chicago cards has likely occurred in the already existing segment of users implying that either the current marketing effort has focused on this user segment or the current value proposition of Chicago card is attractive principally to these users. Clearly, an effort in marketing and redesigning the value proposition of Chicago cards is required so that additional market segments could be captured to expand the current customer base. It will be interesting to see whether the customer base has changed after 2005 with the recent fare restructuring of Chicago cards aimed at increasing the share of Chicago Card users.

Table 4-1: Classification of CC customers by frequency of weekday use (Sep 2005)

Frequency of weekday use	# CC customers	% total
1	15493	19%
2	10989	13%
3	10729	13%
4	15393	18%
5	31018	37%
Total	83622	100%

Table 4-2: Classification of CC customers by frequency of weekday use (Sep 2004)

Frequency of weekday use	# CC customers	% total
1	9869	17%
2	7656	13%
3	8097	14%
4	11625	19%
5	22372	38%
Total	59619	100%

4.1.2 Time Consistent Customers

After identifying the pool of frequent users, we will check if these users are time consistent in accessing the system. For a given customer, this involves evaluating the time difference between the minimum and the maximum times for his first trips of the day and checking if this difference is less than the defined threshold for time consistency [15 minutes, 30 minutes, 45 minutes, 1 hr, etc].

The methodology is illustrated for a representative 5 day frequent customer in Figure 4-1. This customer has 5 first trips of the day which are recorded at 8:03:43, 7:51:58, 8:06:35, 8:01:49, and 8:03:33 am on Monday, Tuesday, Wednesday, Thursday, and Friday respectively. Figure 4-1 shows the first trip times arranged in ascending order. We now evaluate the time difference between the minimum and the maximum access times, which are 7:51:58 and 8:06:35 on Tuesday and Wednesday respectively. The difference between these two times is 14 minutes 37 seconds; therefore, this customer is **time consistent within a 15 minute interval**.

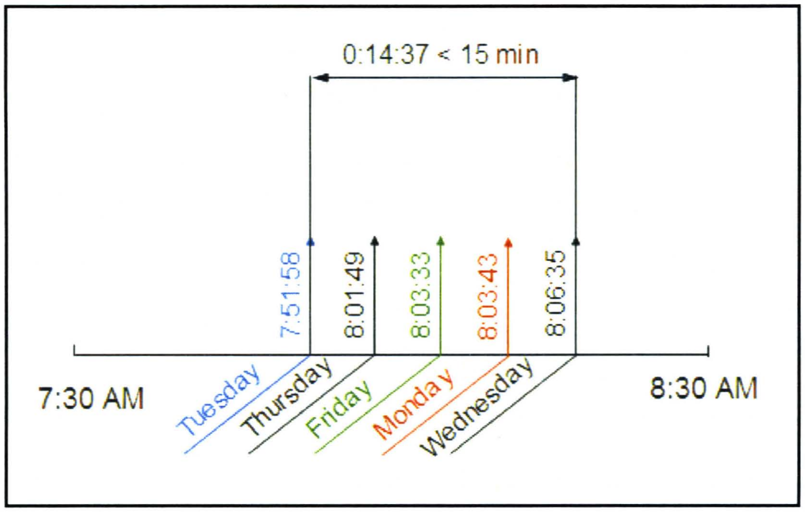


Figure 4-1: First trip 15 minute time consistency definition

A 5 day frequent customer can be 3, 4 or 5 day time consistent. To check if a 5 day frequent customer is time consistent, we start by checking if the customer is 5 day time consistent by comparing the first and the fifth times[1]. If the customer is not 5 day time consistent, we check if he is 4 day time consistent. For this we have to compare the time difference between the first-fourth and second-fifth times. If the customer fails to be time consistent in either of the checks, we check if he is 3 day time consistent. Here we have 3 time combinations to check for. These are first-third, second-fourth, third-fifth. If this customer fails 3 day time consistency test as well then he is not a time consistent customer.

This methodology was used to determine the level of time consistency of the frequent weekday CTA customers. This methodology has been implemented in MATLAB

[1] Times here refer to times of the first trip of day, arranged in ascending order; so that first time corresponds to the earliest first trip time and fifth time refers to the latest first trip time for a 5 day frequent customer.

(Appendix 2). The results for first trip time consistency for different time thresholds for frequent customers for September 2005 and September 2004 are shown in Figure 4-2 and Figure 4-3 respectively.

From Figure 4-2, we see that 59% of frequent customers are time consistent within 15 minutes and 71% of frequent customers are time consistent within 30 minutes. We also see a striking similarity in the level of time consistency for the September 2005 and September 2004 data. It should be noted that the fraction of 15 minute time consistent users decreases as the frequency of use decreases. For a 3 day frequent user, being 15 minute time consistent requires him to satisfy a very restrictive condition of being consistent for all the 3 days that he accesses the system. In addition, it is less likely that the 3 trips that a 3-day frequent user makes are work trips and are therefore less likely to be 15 minutes consistent.

The 30 minute time threshold to classify customers as time consistent was selected over 15 minutes, because 15 minutes may be too restrictive given service frequency. For example, bus riders may arrive at a bus stop consistently but may not be able to board at a consistent time because of variable bus arrivals. A threshold of 30 minutes seems reasonable to prevent quite consistent bus riders from being excluded from the analysis. In addition, since we are trying to capture a time threshold within which work trips are made, it must represent the flexibility an employee enjoys in reaching his work place which could be 30 minutes for most places. Also, as seen from Figure 4-2, using 30 minutes as the threshold increases the percentage of qualifying Chicago Card users to 71% from 59% in 2005, an increase of almost 7000 users.

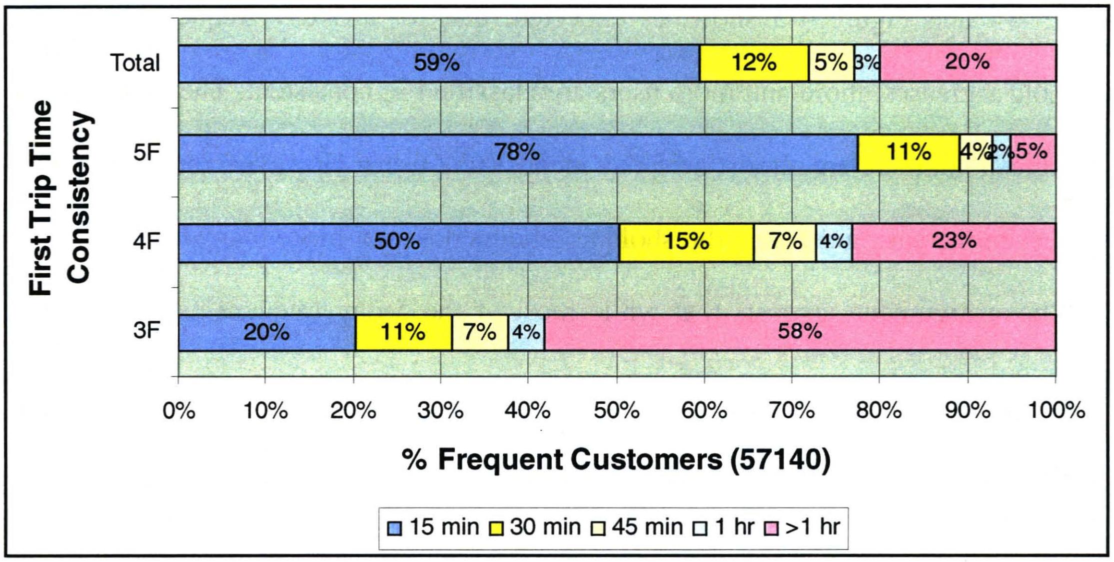


Figure 4-2: First trip time consistency for frequent customers (September 2005)

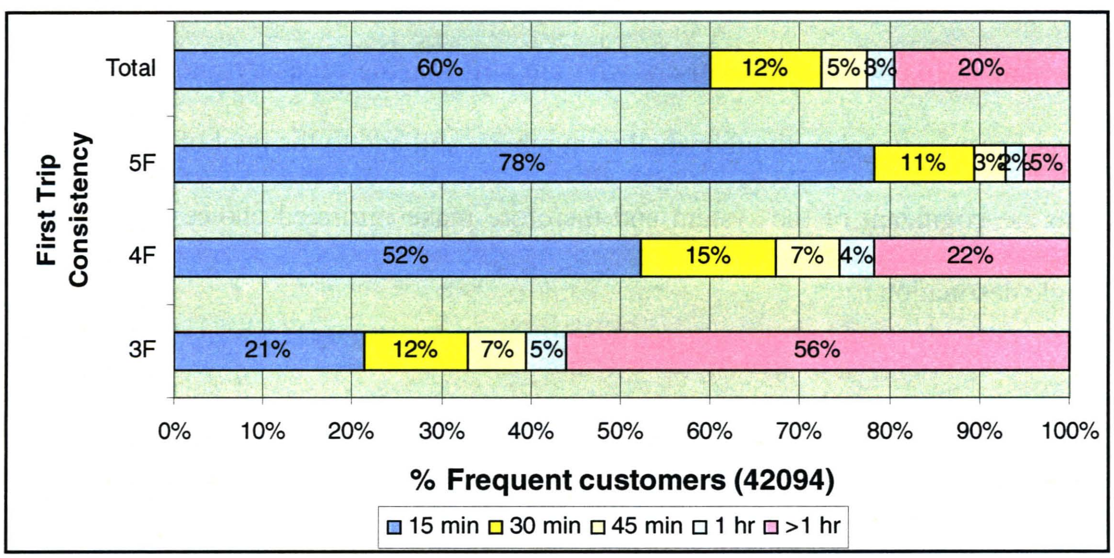


Figure 4-3: First trip time consistency for frequent customers (September 2004)

Figure 4-4 and Figure 4-5 show the mix of types of consistencies for different time windows for September 2005 and September 2004 data respectively. These figures demonstrate striking similarity in the distribution for the two years of data which could again be the result of Chicago Card growth among a similar base of customers.

Figure 4-4 and Figure 4-5 show, as expected, that for a 5 day frequent user, as the threshold increases, more and more users are classified as consistent. Therefore, with a 2 hour threshold, there are almost no 5 day inconsistent users. However, for 3 day frequent users, even increasing the time threshold to 2 hours does not make many users consistent. The above argument suggests that while most of the 5 day frequent users are making work trips, the trips taken by many 3-day frequent users may not necessarily be work trips.

Regardless, there are many 3-day frequent users who are 30 minutes time consistent for all the three days. Being 3 day consistent for all the 3 days is a restrictive condition and this sample likely consists of users who are still making regular trips albeit for only 3 days. These users, even though small in numbers, still add to the pool of those customers, who are cognizant of the system and therefore make informed choices, and travel to a single destination.

Figure 4-4 and Figure 4-5 also reveal the gamut of sociological patterns in terms of travel habits, open to different interpretations: 5-day work week or less, trip chaining on just one day per week etc. Frequency of usage is also a first step towards how to turn actual numbers of trips into numbers of users, which eventually could serve to establish percentages of population being served. The point is that we tend to look at transit commuters as a single category when in fact this graph shows ample diversity and thus, provides us with another way to examine the nature of the Chicago Card customers being served by CTA.

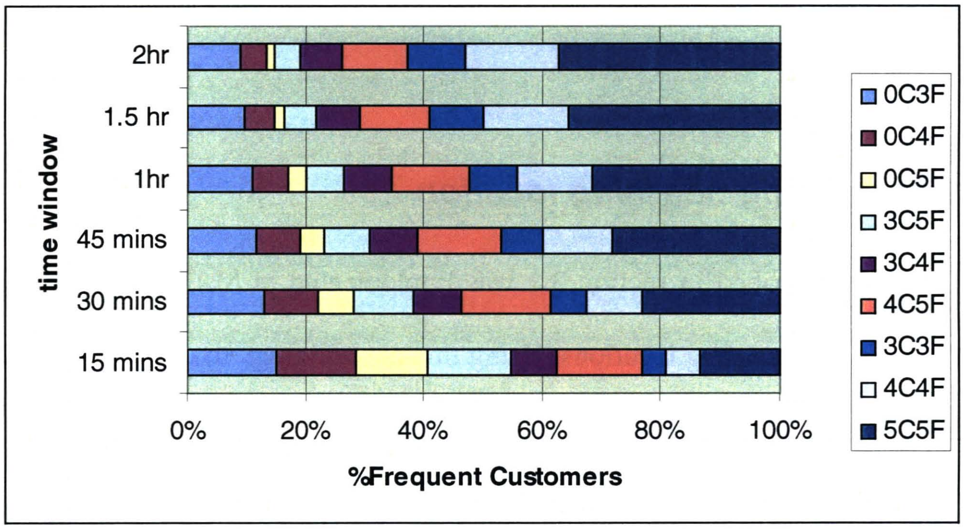


Figure 4-4: Consistency distribution for different time thresholds (September 2005)

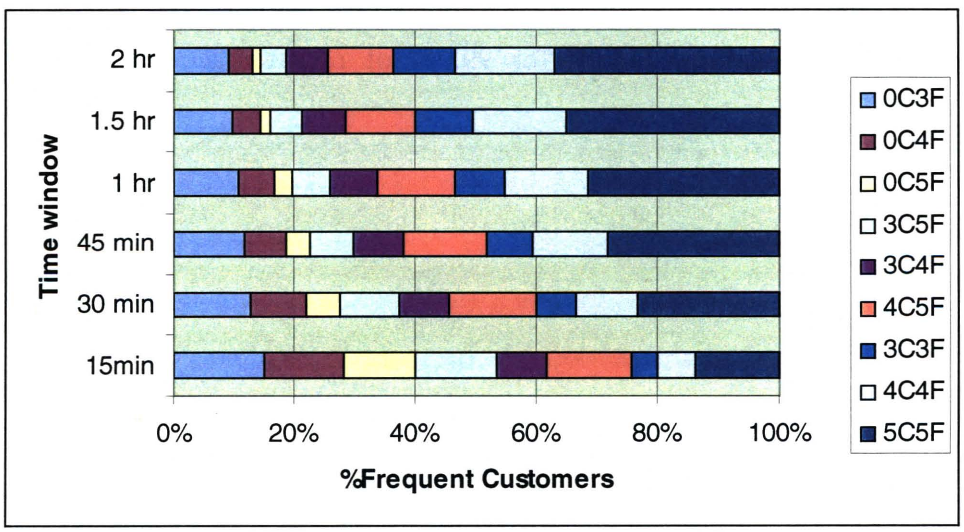


Figure 4-5: Consistency distribution for different time thresholds (September 2004)

After filtering for **frequency of use (3 days minimum) and 30 minute time consistency**, our sample comprises 41043 (72% of frequent customers) for the September 2005 analysis. All subsequent analysis in this chapter, unless otherwise stated, is done using this pool of **41043 frequent and 30 minute time consistent** customers for **September 2005.**

4.2 Determining boarding location

The algorithm described in section 3.1 was implemented in MATLAB (Appendix 1) to determine the bus boarding locations. Using this methodology to capture the critical AVL time that corresponds to the AFC transaction, we were able to match the time for more than 99% of AFC records within 5 minutes of an AVL time for September 12, 2005 (Figure 4-6). This verifies the assumption used for inferring bus boarding location that all the customers board a bus within 5 minutes of its arrival time at a stop. Finally, we were able to identify boarding stations for 99.9% of rail transactions and 86% of bus transactions.

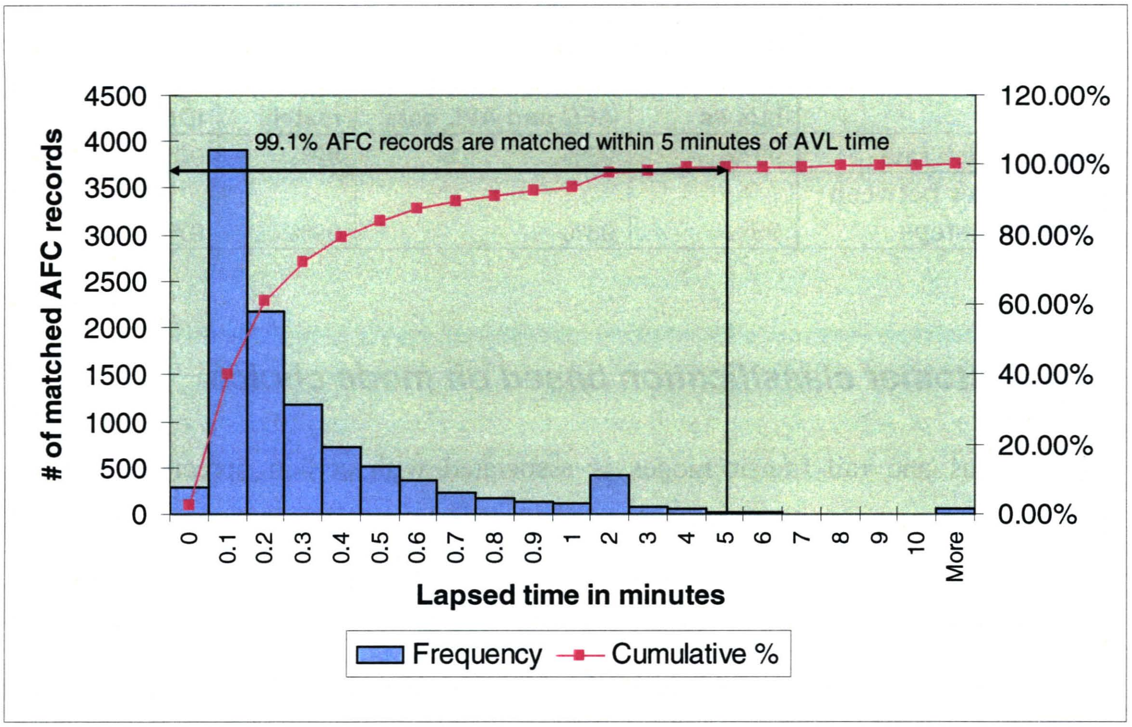


Figure 4-6: Distribution of time difference between the AVL time and matching AFC times for September 12, 2005

The match-rate for bus boarding locations at different levels of matching is shown in Table 4-3. While our methodology of AFC-AVL linking gives us a high match rate of 99%, the total fraction of boarding stations that are successfully inferred is lower due to other steps with lower success rates. The two steps which provide a lower match rate are (1) identifying the bus numbers from AFC records (low match rate of 95%) and, (2) obtaining Stop IDs from Pattern IDs (match rate of 96%). The match rate in these steps is contingent on the correct login by the operator and the current low match rates are likely to be the result of operator login error. Improvement in login procedures should increase the identification rate of bus boarding locations, and thus, the value of this type of analysis.

Table 4-3: Match rate for identifying bus stops using AFC-AVL linking

	Identifying bus #s	Records for common buses in AFC and AVL data	AVL time match	Pattern IDs	Stop IDs
Overall match rate	95%	93%	92%	90%	86%
Match rates between successive steps	95%	98%	99%	97%	96%

4.3 Customer classification based on mode choice

Each of bus and rail transit modes is associated with certain perceived and actual characteristics. A higher density of bus stops, traffic lights and congestion, slow boarding and fare collection can lead to a perception that the bus mode is slower. Trains are perceived to be quieter, faster, more reliable and less polluting than buses. Since, the characteristics of these two modes are different their user base is expected to have different characteristics as well. For example, since the distance between adjacent rail stations is typically larger than for bus stops, the variation in boarding locations on first trips for rail users is likely to be smaller than for bus users. Similarly, we might expect greater variability in choice of routes for bus users because multiple routes sometimes serve the same destination. It is therefore important to classify customers by mode choice in analyzing their behavior.

In addition, in traditional modeling, modal split relies heavily not only on the level of service of the travel options, but on characteristics of the user himself. Therefore, splitting customers by mode choice might be a first step in understanding user preferences for a given mode. Furthermore, the variation in mode usage directly implies path choice; categorizing users by mode choice is an easy way to determine this variation.

We categorize our sample of frequent and time consistent customers based on mode choice (bus, rail or both) on their first trip using the following definitions:

1. **Rail**: These customers use rail for all weekday first trips
2. **Bus**: These customers use bus for all weekday first trips
3. **Mixed**: These customers use a combination of both rail and bus for their weekday first trips

The distribution of users among these three categories is shown in Table 4-4. This distribution holds only for those CTA customers who use Chicago Cards and travel frequently and consistently and is not representative of the entire CTA customer base, where bus ridership far exceeds rail ridership.

Table 4-4: Chicago Card customer distribution by mode (September 2005)

User Type	# Customers	%
Rail	23842	58%
Bus	13218	32%
Mixed	3983	10%
Total	41043	100%

4.4 System access variability

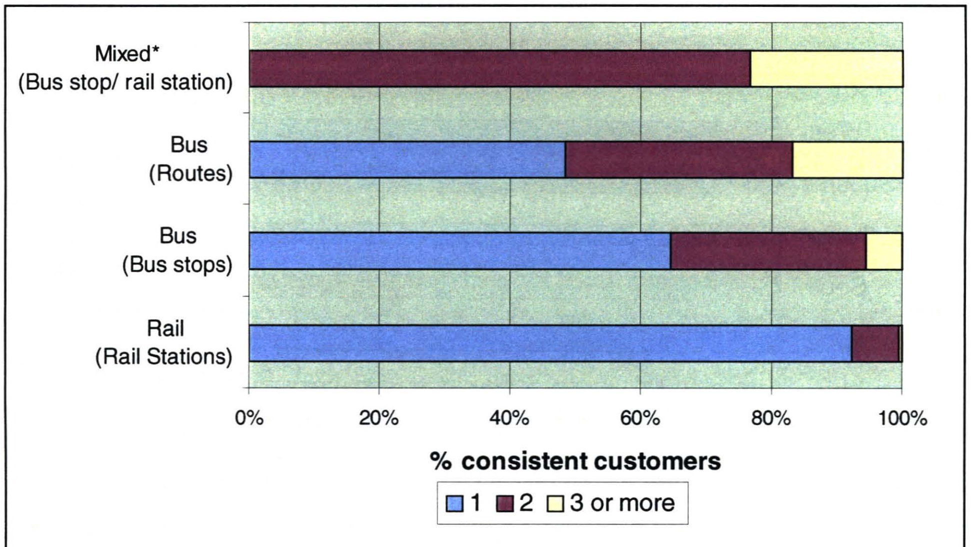
Variation in boarding location and route selected for traveling between a given origin and destination is a reflection of path-choice by a customer. We therefore analyze which customers demonstrate path-choice at the home end by examining their variation in accessing the transit system - boarding location and route for first trips of the day. The definitions for these two components of system access variability are as follows:

1. **Variation in boarding location:** This is the number of different boarding locations[2] (rail stations for rail users, bus stops for bus users, and rail stations and bus stops for mixed users) used by a frequent and time consistent customer on first trips during the week of analysis.
2. **Variation in route choice:** Defined for a bus or mixed user, this is the number of different routes used by a frequent and time consistent customer on first trips during the week of analysis.

Figure 4-7 shows these variations for different types of users – bus, rail, mixed. We see that the behavior of rail users is much less variable, with more than 90% of these customers using only a single rail station for their first transit trips. Bus users demonstrate greater variability with respect to both stop and route choice. Route variability is higher than stop variability which may be due to several reasons:

- multiple routes serving the same origin stop and the same destination, and,
- wrong manual login on the AFC system

[2] While counting the number of different boarding locations, we considered only those bus stops or rail stations whose locations could be inferred



*On the left margin the headings represent the user type (rail, bus, mixed) and labels in braces represent the variability that we are addressing (e.g. rail stations for rail users, or bus stops for bus uses, or routes for bus users, etc.)

Figure 4-7: Route and boarding location variability by different user types

To better understand system access variability by bus users, we analyze route and bus stop variation for different geographical areas as shown in Figure 4-8 and Figure 4-9 respectively.

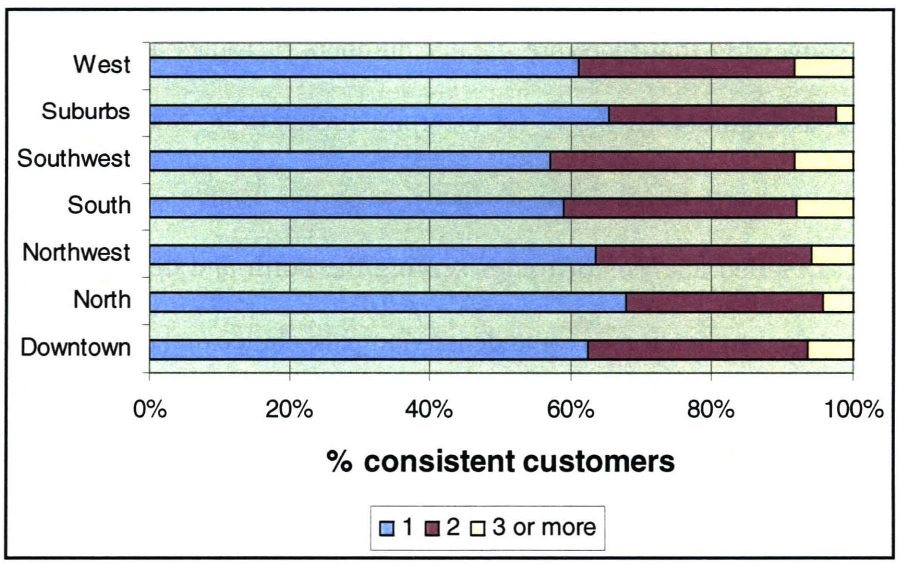


Figure 4-8: Bus stop variability for bus users by region

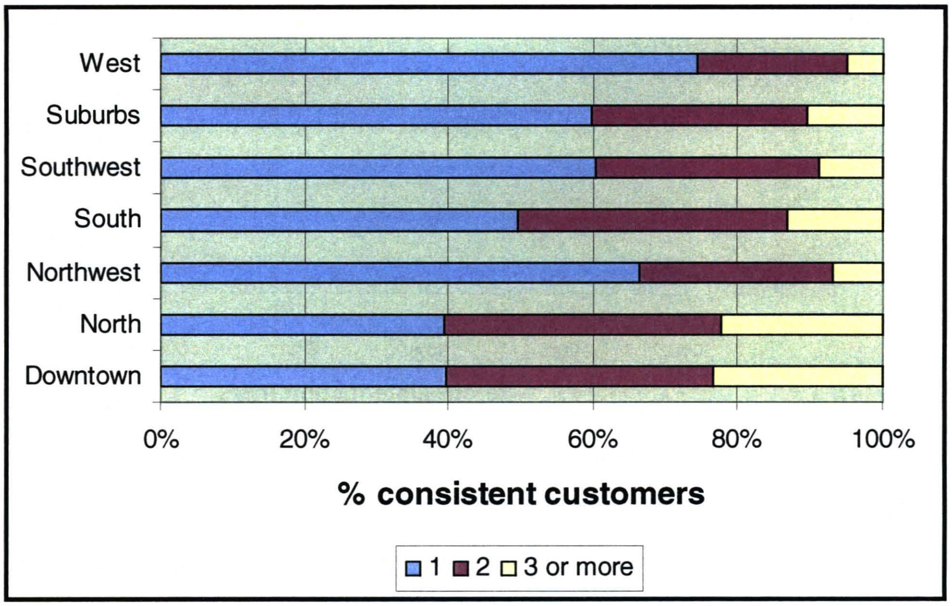


Figure 4-9: Route variability for bus users in different regions

Figure 4-8 shows that the bus stop variability is similar across regions[30]with about 60% of the bus users boarding at the same stop everyday, while almost 30% use two different stops. However, in Figure 4-9 we see that the route variability for bus users differs significantly across regions. Bus users in the north and downtown areas exhibit high variability, suggesting that there are greater numbers of routes available in these areas. For example, as Figure 4-10 shows for a certain stretch of N. Sheridan road, a CTA bus rider at any bus stop might have as many as 5 bus routes available to take to his destination. Detailed analysis of routes serving the north and downtown areas can help us understand the implication of route density on path choice and customer satisfaction.

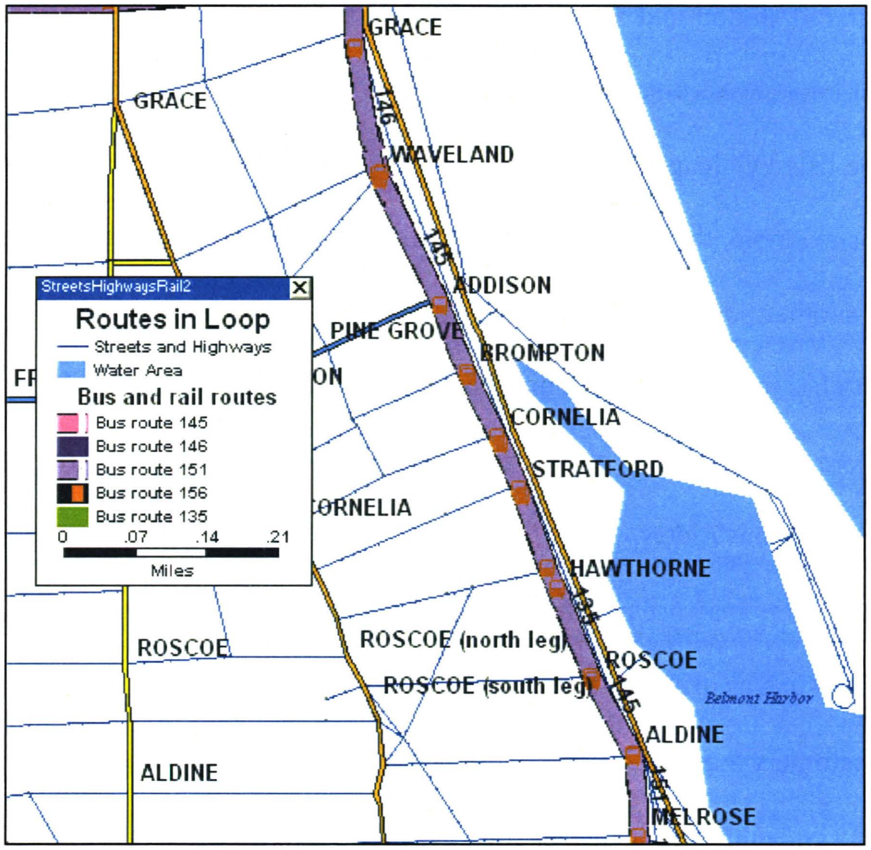


Figure 4-10: Bus routes on N. Sheridan road

4.5 Access distance

Walk access distance is an important characteristic of transit services. Along with other service characteristics such as in-vehicle time, frequency, and auto availability, it is an important determinant of customers' mode and path choices. We therefore compute access distance for CTA customers using the methodology described in Section 3.2 and show those results in this section.

The access distance distributions for rail users on their first trips for September 2004 data are shown in Figure 4-11. We see that about 33% of the users travel more than a mile to get to their rail station and about 22% travel more than 2 miles. Some of the plausible reasons for long access distances are[6]:

- The listed Chicago Card billing address may not be the customer's home address. For example, the registered billing address may be an office address, an address of another person who pays the bill (e.g. parents), or an old address for a customer who has moved after obtaining the card but didn't update the address.
- There may be errors in the AFC data attributable to missing transactions and/or incorrect bus routes.
- Customers may use a different transportation mode (other than walk) to access the CTA such as drive or ride, or commuter rail or taxi.
- Customers may not be accessing the system from their home in some cases – they may be using transit for one-way trips to home and not from home, or may not be beginning their trip from home on particular days (e.g. sleepovers)

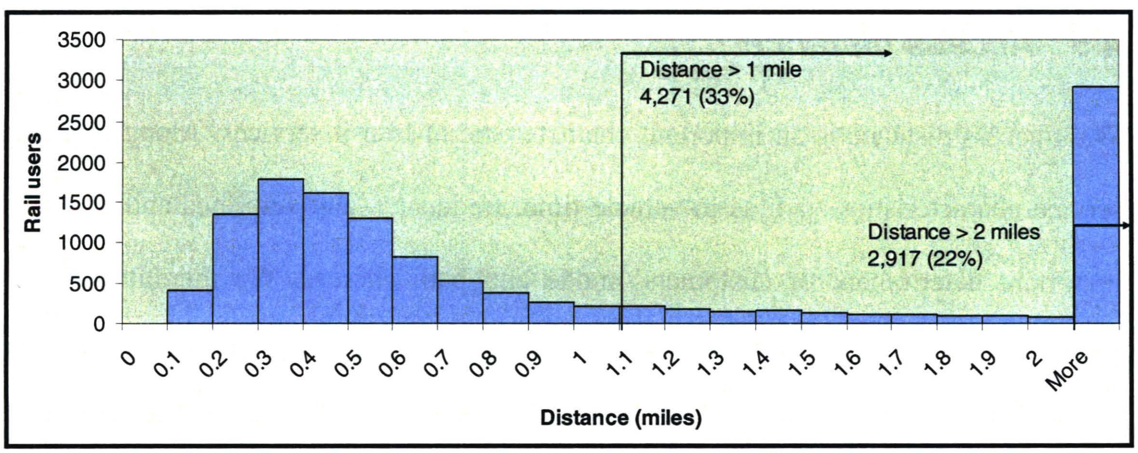


Figure 4-11: Access distance distribution for rail users

Another interesting aspect of access behavior is its variation among residential areas. In Figure 4-12, we plot the cumulative distribution of walk access distance for rail customers by residential area. The distribution of walk access distance for rail differs among residential areas as it is shortest for customers in downtown and largest for customers in southwest. For example, in the downtown area, 33% of the customers have access distance of 0.20 miles or less, while for the southwest region the same 33% is at 0.4 miles. Lower cumulative access distance reflects higher station densities in the Loop and north areas. These differences might result from the combined effect of residential density and station spacing.

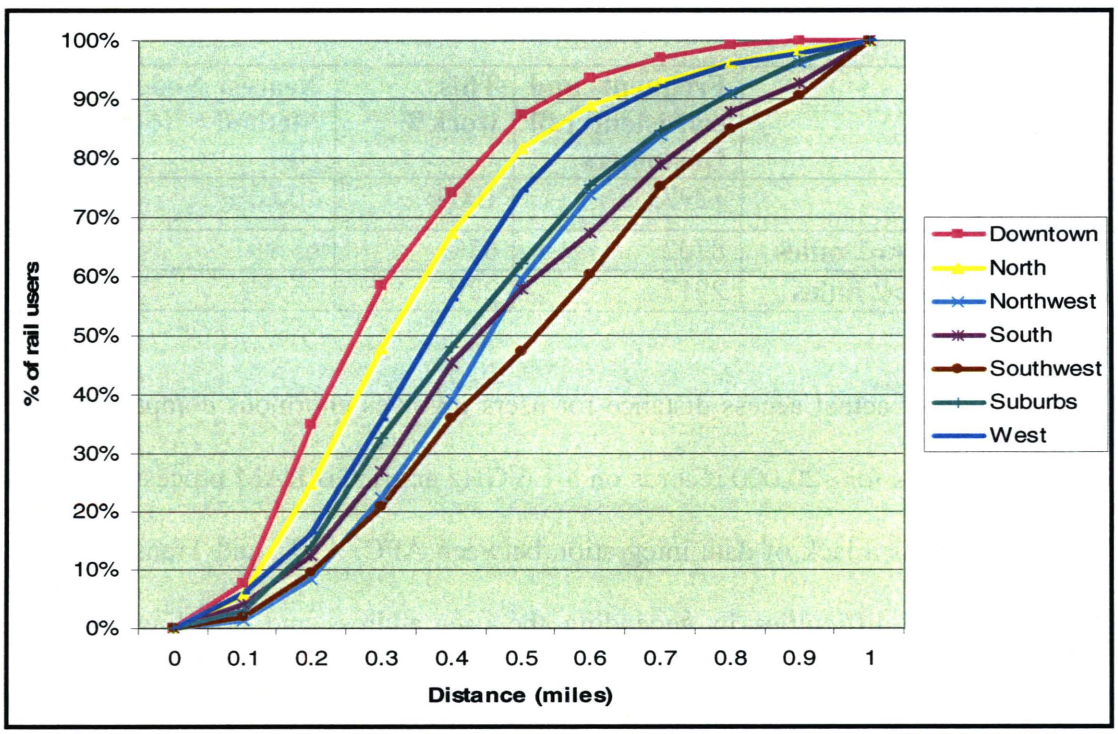


Figure 4-12: Cumulative distribution of rail access distance by residential area

Comparing these results for rail access distance with those obtained from nearest node distance computation approach[6] (see Table 4-5) we see that the results are similar. This may be because we are computing averages for large samples and so the effect of over estimation and under estimation with the nearest node method tends to cancel out. In addition, Table 4-5 shows that a greater percentage of frequent and time consistent Chicago Card rail customers have access distance of a mile or less, than for those of a broader category of Chicago Card rail customers. This suggests that people taking less frequent or less time consistent CTA rail trips may be willing to take longer trips to access CTA rail service.

Table 4-5: Access distance results with actual and nearest node methods

	Frequent and consistent rail Customers	This work%	Nearest Node Method [6] (%)	Rail Customers [6]
Total analyzed	12973	100%	100.0%	31250
Access distance <=1 miles	8702	67%	65.6%	20500
Access distance >2 miles	2917	22%	25.5%	7967

Determination of actual access distance for users involves enormous computation times (as long as 4 days for ~20,000 records on a 1.6 GHz and 1 GB RAM processor) and data cleaning. There is a lack of data integration between AFC, AVL, and TransCAD model files, leading to difficulties in geocoding the user address and boarding location in TransCAD. While a substantial effort in cleaning the data (correcting the address locations for customers and rail stations), improves the results, all the addresses could still not be geocoded.

Because the slight improvement in results is associated with high computation costs for the access distance methodology, the following discussion on bus access distance relies on Utsunomiya's[3] results using the nearest node distance computation methodology[6].

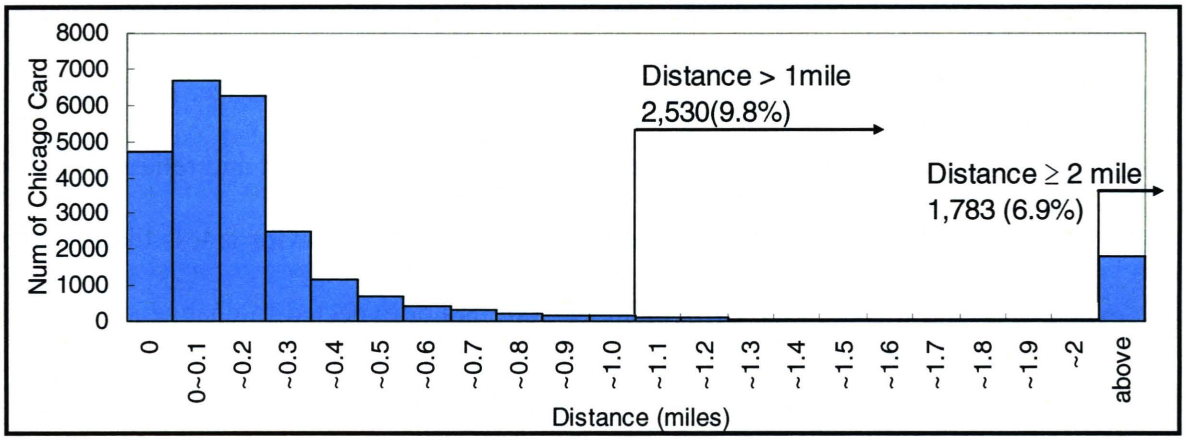


Figure 4-13: Access distance distribution for bus users[6]

Figure 4-13 shows the access distributions for bus users on their first trips for the September 2004 data [6]. A number of 'zero' access distance cases are seen because the customer address sometimes share its nearest street intersection node with the node for the nearest bus stop. Comparing Figure 4-11 and Figure 4-13 we see that access distance varies significantly by mode. This difference can be explained by the differences in network and access point densities between bus and rail. In Chicago the rail network is only 12% [6] the size of the bus network and average rail station spacing is significantly higher than for bus. According to the CTA Service Standards [6] rail stations are ideally located about 0.5 miles apart, whereas bus stops are located at major cross-street intersections at about 0.125 miles spacing. The difference in bus and rail access distance

[3] The results and the accompanying analysis on bus access distances are taken from Utsunomiya[6].

may also reflect a difference in the perceived generalized cost between bus and rail – customers might be prepared to walk further to access rail service because of the overall time components to their destination.

In general, how far a customer is willing to walk to ride transit depends both on the density of access points and on the attractiveness of the transit service. For example, a customer may be prepared to walk further if service is frequent, fast and reliable. Without any significant changes in the network, the access distance behavior is less likely to vary over a year and the analysis for September 2005 is not expected to change significantly from that of September 2004.

4.6 Path choice analysis for a smaller sample

The variability in route and boarding location amongst different types of Chicago Card users, classified as bus, rail and mixed, demonstrates path-choice by these customers. In this section, we analyze the path-choice characteristics for a smaller sample of customers for their daily commute between home and the Loop as the general destination.

CTA's extensive rail and bus service in the north lake shore area offers its residents several reasonable alternative paths to choose from when traveling between their homes and destinations in the Loop. This fact has led us to select a sample of customers living on or around Belmont Street. Frequent and consistent customers living within a buffer of 0.2 miles along Belmont Street between Belmont Red Line Station and Chicago River were selected as shown in Figure 4-14. These customers are categorized on the basis of their first trip mode choice as rail, bus and mixed and are represented by dots of different colors.

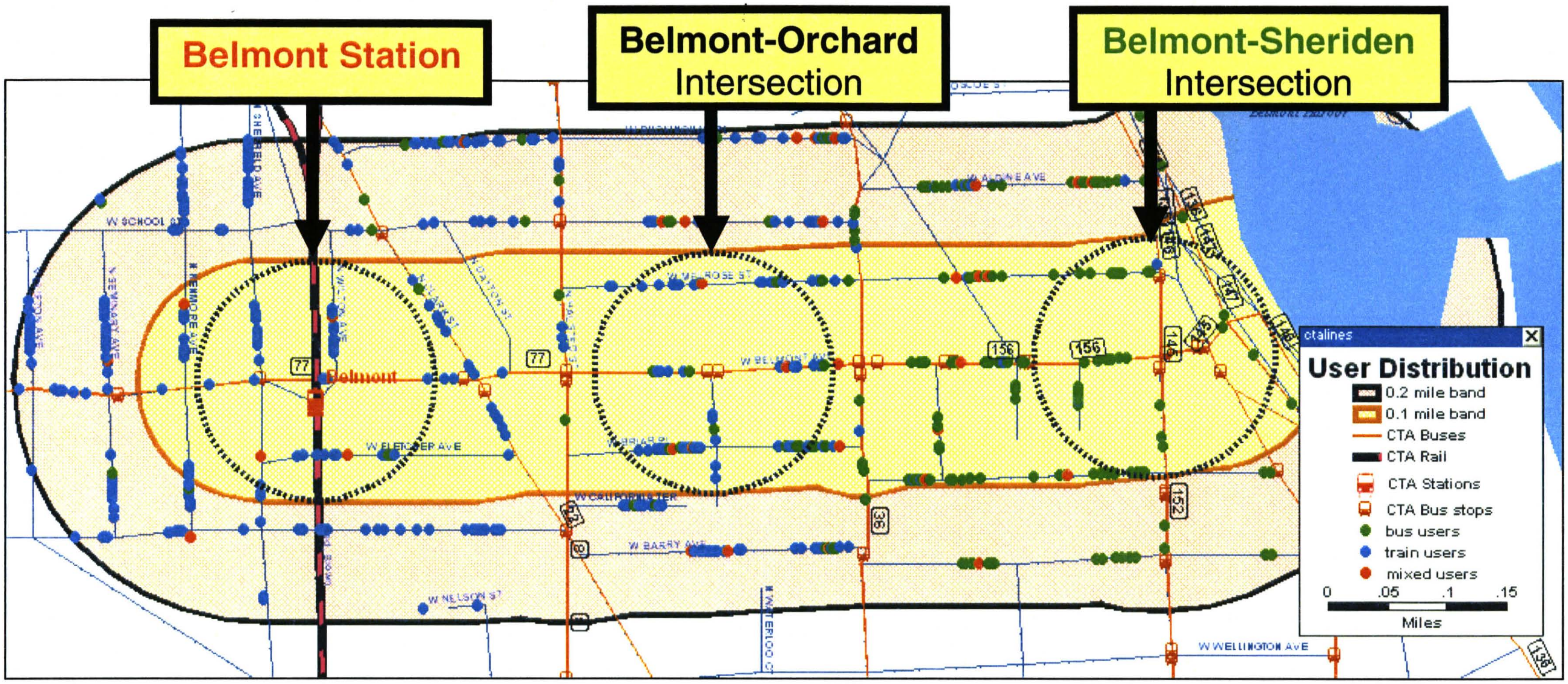


Figure 4-14: Belmont Street Sample

Table 4-6 shows the distribution of different types of users within this corridor. It is interesting to note that while the fraction of bus and rail users is similar (45% and 48% respectively), mixed users are smaller in number (7%). This homogeneity appears justifiable given the comparable level of service for the bus and rail modes in this area. This homogeneity is, however, not manifested at a microscopic level where the geographical position of a user strongly influences his mode choice.

Table 4-6: Distribution of customers on Belmont Street corridor within 0.2 miles buffer

User Type	# Customers	% customers
Bus	658	45%
Rail	704	48%
Mixed	99	7%
All users	**1461**	**100%**

In the following sections, we have identified access distances and other factors that could affect the mode choice decision of these users. We have also examined the variation in mode choice for the return trips and for all trips made over the week of analysis. In the last part of this section, microscopic analyses, that can be used to extract path-choice parameters at disaggregate level, have been conducted.

4.6.1 First trip mode choice

Although the Chicago Card users on Belmont Street presumably belong to a similar socio-economic make-up and most of them work in the downtown, they use different modes to commute as shown by Figure 4-14. In this section we will identify factors that may govern their mode choice for the first trip of the day.

a) Effect of access distance

The importance of access distance in determining mode choice is evident from rail user's dot density near the Belmont Station and bus user's dot density near the Sheriden area in Figure 4-14. In order to test the sensitivity of proximity to a rail station, we have selected a sample of customers living within 0.1 mile of Belmont station, Belmont-Orchard intersection and Belmont-Sheriden intersection as shown in Figure 4-15. The modal distribution for these customers is shown in Table 4-7.

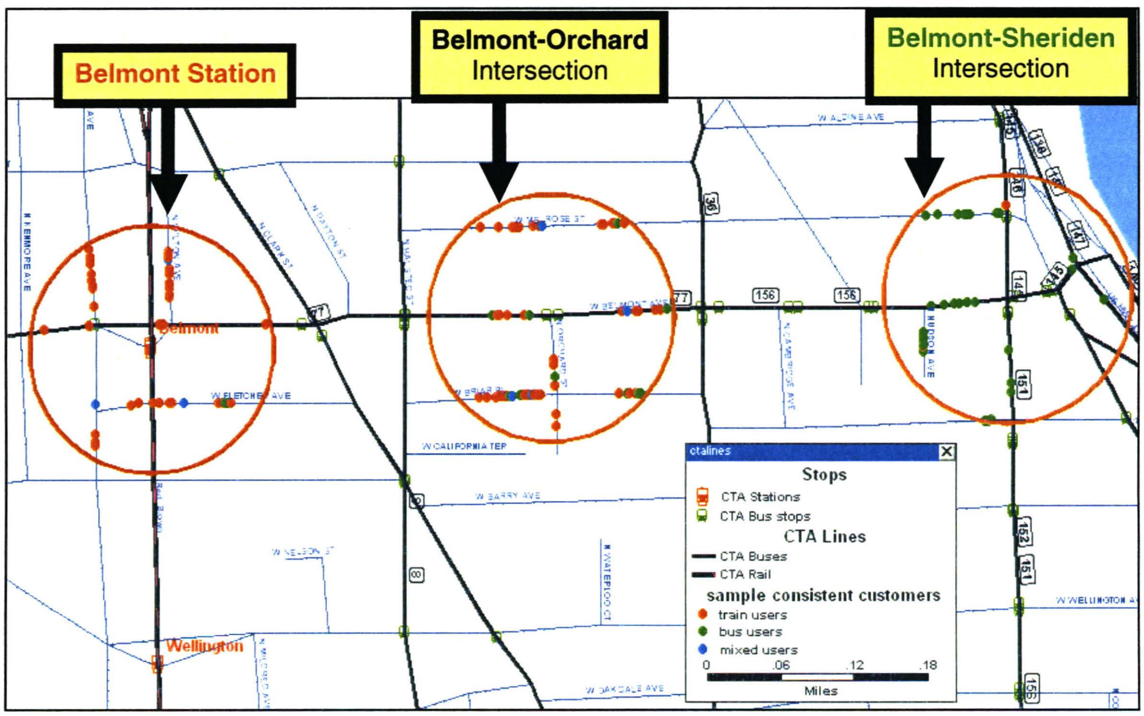


Figure 4-15: Sample customers within 0.1 mile buffer on Belmont Street

Table 4-7: Distribution of customers along Belmont Street, within a 0.1 miles buffers around intersections

Intersection	# bus users	# mixed users	# rail users	# Total customers
Belmont Station	2 (3%)	4 (5%)	73 (92%)	79
Belmont Orchard	20 (18%)	9 (8%)	83 (74%)	112
Belmont Sheriden	170 (85%)	10 (5%)	21 (10%)	201

These numbers suggest that as the distance from rail station increases, access distance strongly affects the mode chosen by a customer, given that:

- Customers who live within 0.1 mile of Belmont Station (79 in number) are predominantly rail users (92%).
- Those who live within 0.1 mile of the Belmont Sheriden intersection (201 in number) are predominantly bus users (85%).

With the above reasoning, users who live close to Belmont-Orchard intersection, midway between Sheriden and Belmont station, are expected to show mixed usage of bus and rail. The distribution of customers around the Belmont-Orchard intersection however shows that the great majority of users in this area are rail users (74% from Table 4-7) with an average access distance of approximately 0.326 miles[4]. Clearly, factors other than access distance at the origin affect a user's mode choice decision. These factors can include wait time, in-vehicle travel time to destination, en-route transfers, etc. The effect of these factors on mode choice decision is quantified in the following section.

b) Effect of generalized cost

The generalized cost includes all the factors that shape the modal decision by incorporating all the time and monetary terms included in a trip with each term weighted

[4] This distance is in line with the rail access distance distribution shown in Figure 4-11.

by the customer's perceptions. A list of those factors include access time, wait time, in-vehicle travel time, transfer time, egress time and fares.

For generalized cost computations in TransCAD for the sample users shown in Figure 4-15, time values were converted to their monetary equivalent assuming $12/hr as the value of time. We computed the generalized cost incurred when traveling from locations near the three points of interest [Belmont station, Belmont-Orchard, Belmont-Sheriden] to any other node within a polygon drawn around the downtown area. The spatial variation of transit generalized costs in equivalent dollars is shown in Figure 4-16, Figure 4-17, and Figure 4-18 respectively for Belmont Sheriden, Belmont Station and Belmont-Orchard. It is to be noted that the time components of the generalized cost have not been weighted with factors representing the user's perception, as they are not available.

For each of the three different origin points we observe different patterns of generalized cost distribution. In particular, for the case of trips starting at Belmont-Sheriden (Figure 4-16), the existence of express bus service, translates into pockets of lower cost around the express bus service stops. These maps visualize the different accessibility from each point based on the different transit level of service (LOS) available from each location.

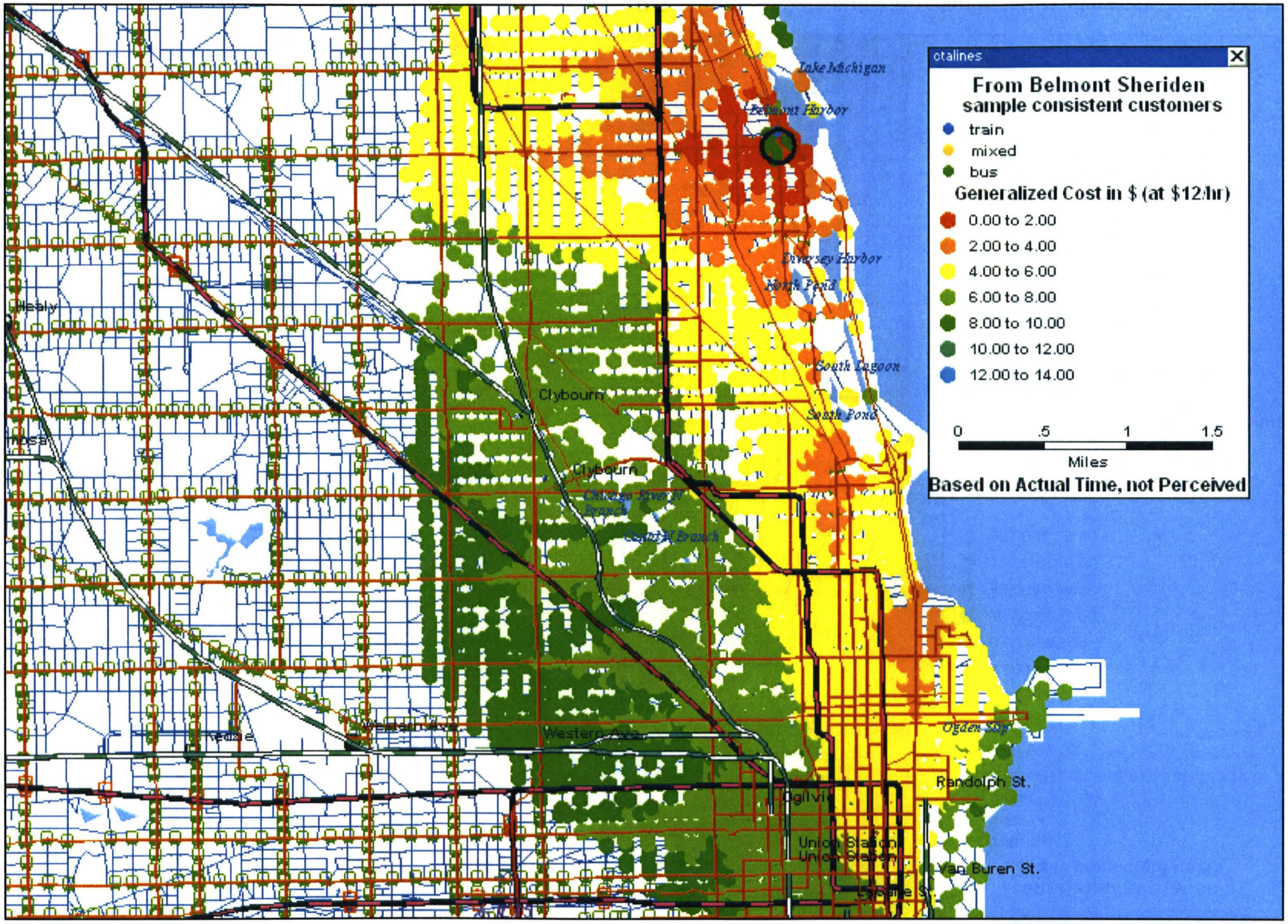


Figure 4-16: Generalized cost for customers near Belmont-Sheriden (predominantly bus users)

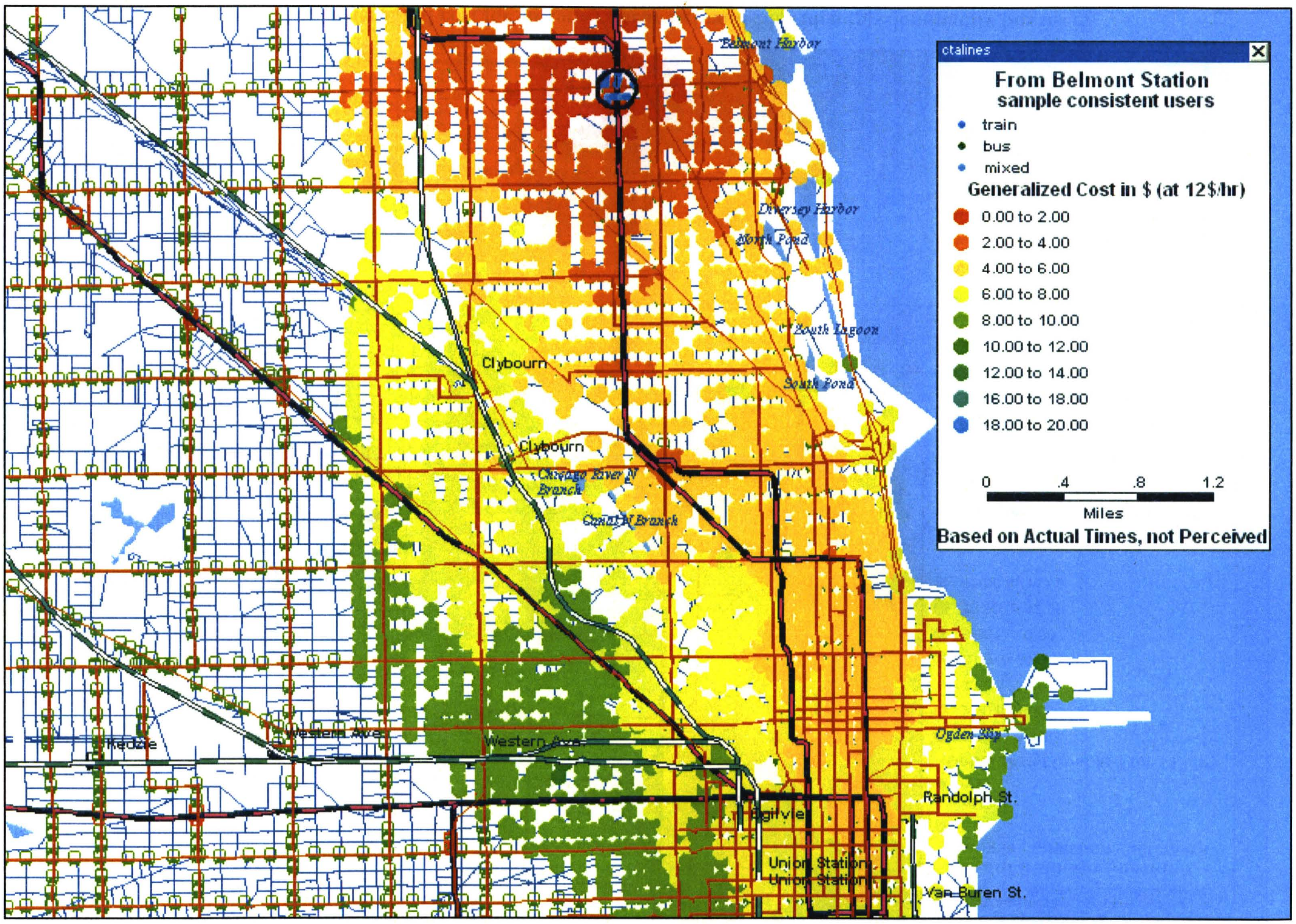


Figure 4-17: Generalized cost for customers near Belmont Red Line Station (predominantly rail users)

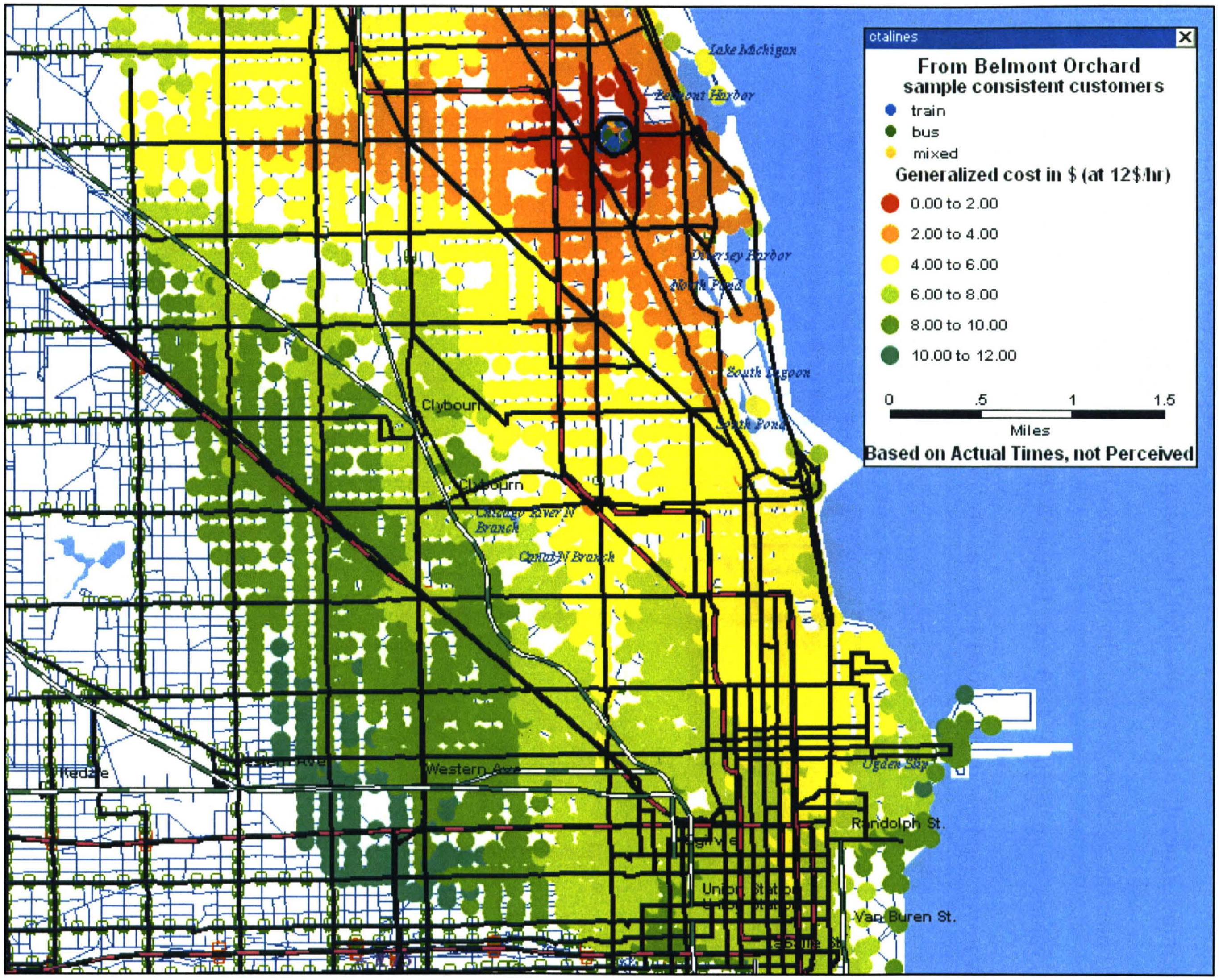


Figure 4-18: Generalized cost for customers near Belmont-Orchard

In order to determine, whether the behavior of the transit users in Belmont-Orchard could be predicted based on generalized cost, the generalized cost for users near Belmont-Orchard has been estimated separately for rail and bus. While Figure 4-18 plots the minimum generalized cost obtained by using either bus or rail, Figure 4-19 and Figure 4-20 plot the generalized costs by using only bus or only rail respectively. By comparing the values from Figure 4-19 and Figure 4-20 we find that the generalized cost for a downtown trip falls in the broad range of \$6-\$8, regardless of the mode used.

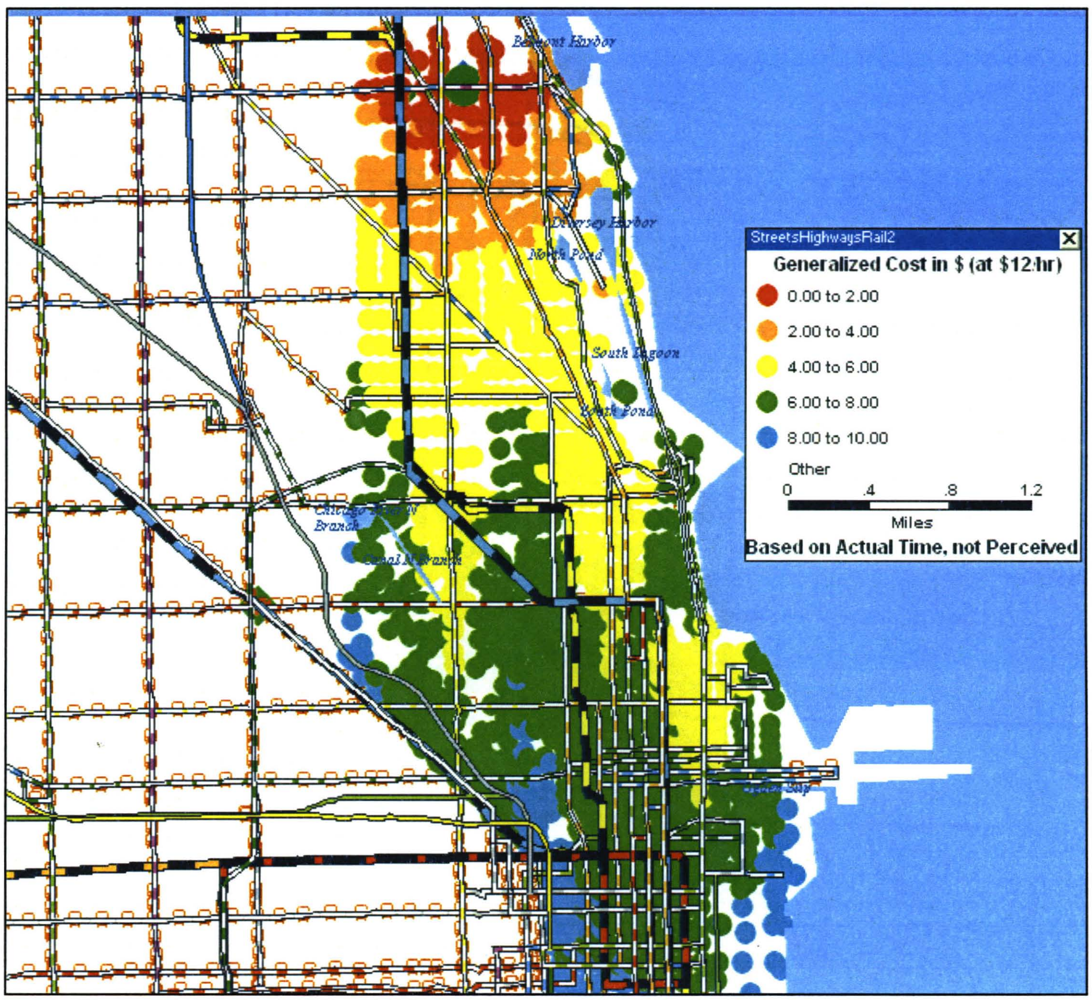


Figure 4-19: Generalized cost for customers near Belmont-Orchard (using bus mode)

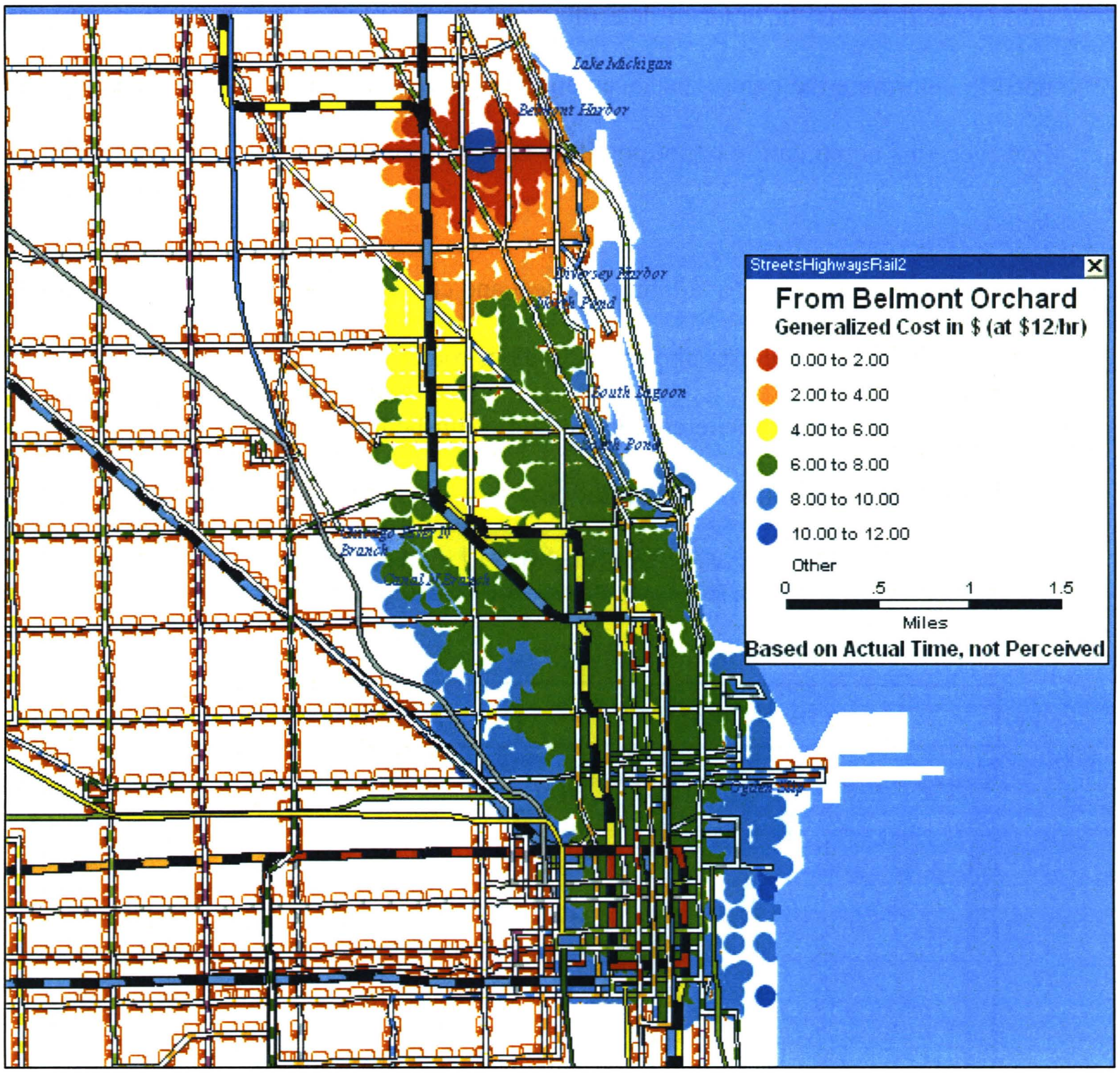


Figure 4-20: Generalized cost for customers near Belmont-Orchard (using rail mode)

Upon further examination of exact values of generalized cost using either of the modes (Figure 4-21) we find that the values are indeed very similar. This suggests that although some of these numeric differences might explain mode choice, there can be other factors in addition to access distance and generalized cost that could affect the mode choice decision. One of the reasons for this high number of rail users in Belmont-Orchard area could be that the destinations of these users are better served and easily accessible by rail as compared to bus. In order to determine whether this hypothesis is correct, we would need to determine the egress distances on the destination side. With the current data sets however, this computation is not possible. Other reasons for a higher rail preference can be:

1. Less crowding and availability of seat and comfort,
2. Ignorance by the user about multiple bus routes that serve his destination,
3. Preference for a sheltered rail station (vs. unsheltered bus stops),
4. Availability of kiosks at rail stations,
5. Greater perception of reliability of the rail mode[5], etc.

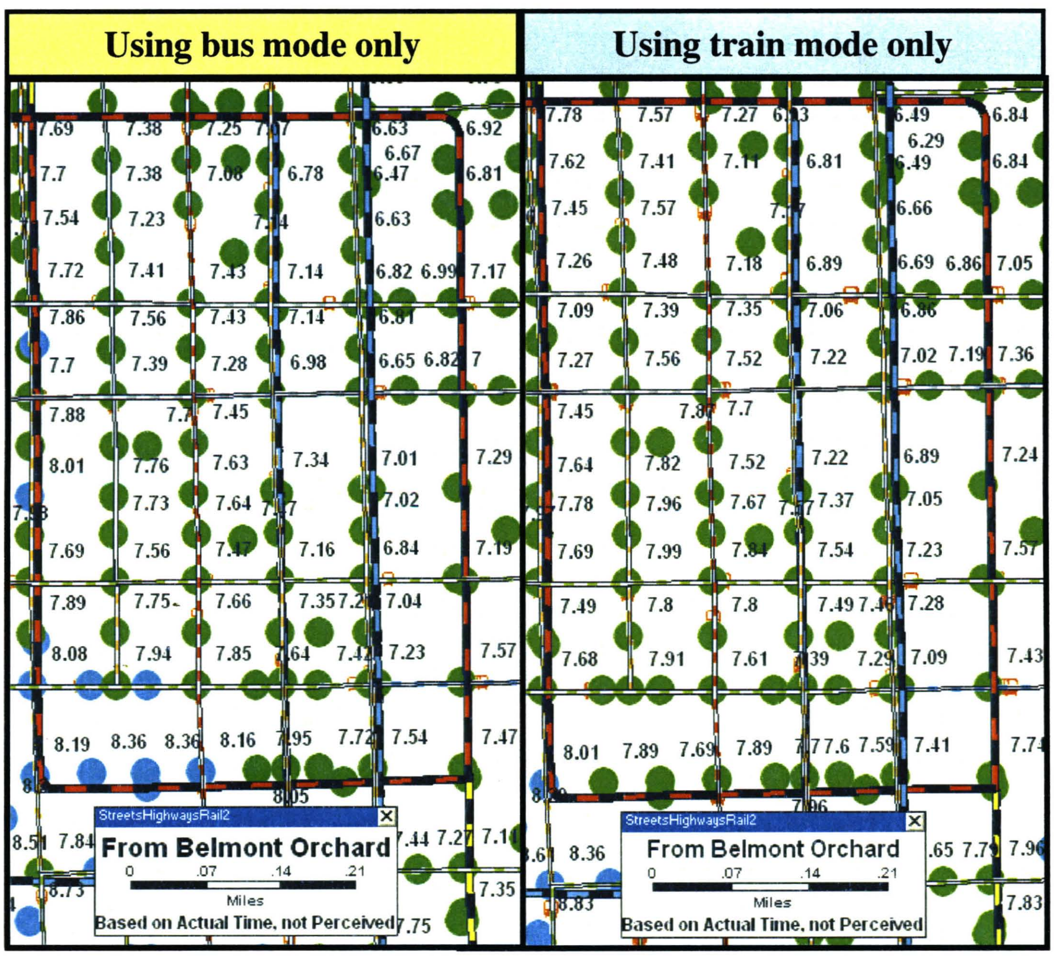


Figure 4-21: Generalized cost from Belmont-Orchard to the Loop using different modes

4.6.2 Mode choice variation for return trips

In order to examine if users who predominantly use one mode on their first trip, use the same mode for the return trip, we have examined their return trip mode choice. Customers who are predominantly rail users living in the vicinity of Belmont Station and those who are predominantly bus users living in the vicinity of Sheriden Street are chosen. Halsted is chosen as the boundary for rail users on the east, while Broadway is taken as the western limit for bus users as shown in Figure 4-22. The distribution of sampled customers by mode choice for their first trip is shown in Table 4-8.

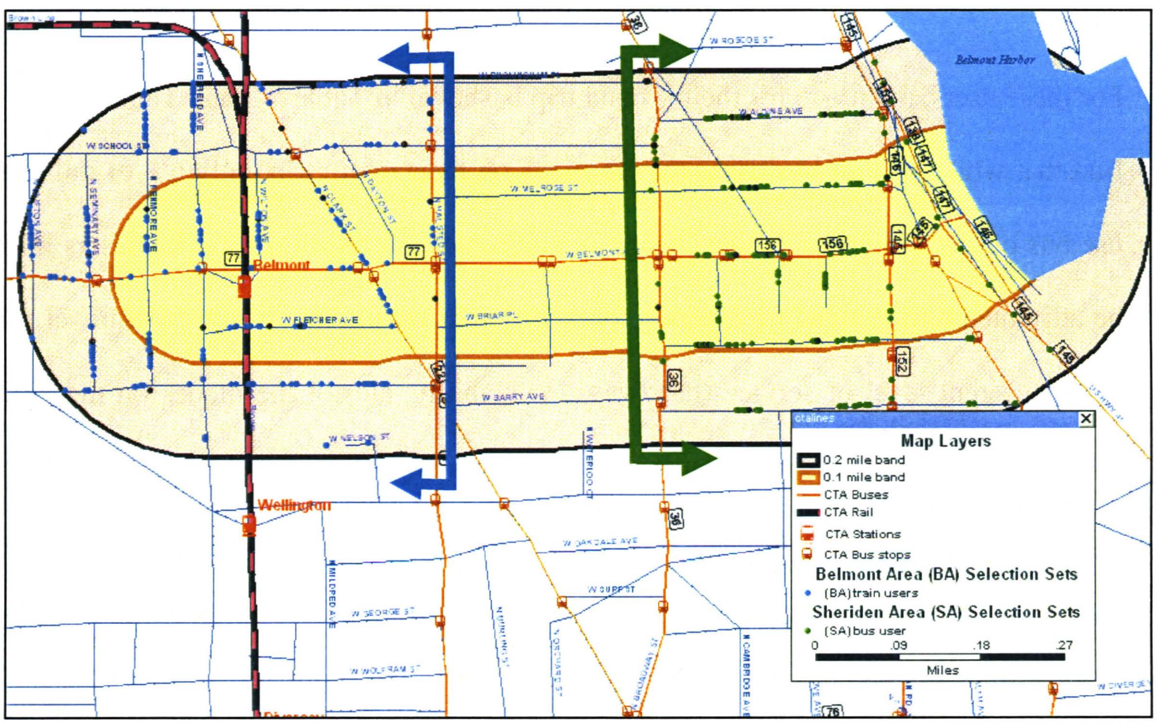


Figure 4-22: Sample customers for return trip mode choice

Table 4-8: Distribution of sample customers for their return trip mode choice

User Type	SA[5]	BA[6]
bus	560	32
rail	105	401
mixed	57	15
total	**722**	**448**

We have inferred the return trip boarding location for the 560 bus users from the Sheriden area and for the 401 rail users from the Belmont Station area using the methodology described in Section 3.3. Out of 560 bus users in Sheriden Area, 481 made only 2 trips on at least one weekday. Out of 401 train users in Belmont Station Area, 376 made only 2 trips on at least one weekday.

For these users, the mode for their second trip is shown in Table 4-9 and Table 4-10. We find that while 30% of bus users change mode on their return trip, only 6% of rail users use bus on their return trip. The change in mode for their return trip for bus users might be attributed to a chained trip purpose or to different sensitivity to aspects of travel time or to different level of service (headways, availability, etc) of the modes on the return trips. These numbers also suggest that for corridors with dominant modes and services, customers are less likely to change their mode of travel. In order to test this claim, we have examined the mode choice variation for all the trips made over the week of analysis in the following section.

---

[5] SA = Sheriden Street Area
[6] BA = Belmont Station Area

Table 4-9: Mode choice on the return trip for Sheriden area bus users

User classification (second trip)	# users	# trips	% users	% trips
bus[7]	340	1018	71%	70%
rail[8]	47	106	10%	7%
mixed[9]	94	324	20%	22%
Total	**481**	**1448**	**100%**	**100%**

Table 4-10: Mode choice on the return trip for Belmont Station area rail users

User classification (second trip)	# users	# trips	%users	% trips
Rail	354	1159	94%	94%
Bus	2	2	1%	0%
Mixed	20	68	5%	6%
Total	**376**	**1229**	**100%**	**100%**

4.6.3 Mode choice variation for all trips

To test that customers are less likely to change their mode of travel for corridors with dominant modes and services, we have studied their mode choice for all trips for the week of analysis. For this hypothesis, our sample comprises the customers living within a buffer of 0.2 miles along Belmont Street between Belmont Red Line Station and Chicago River as shown in Figure 4-23.

[7] Those users who use only bus on all their return trips
[8] Those users who use only rail on their return trips
[9] Those users who use both rail and bus on their return trips

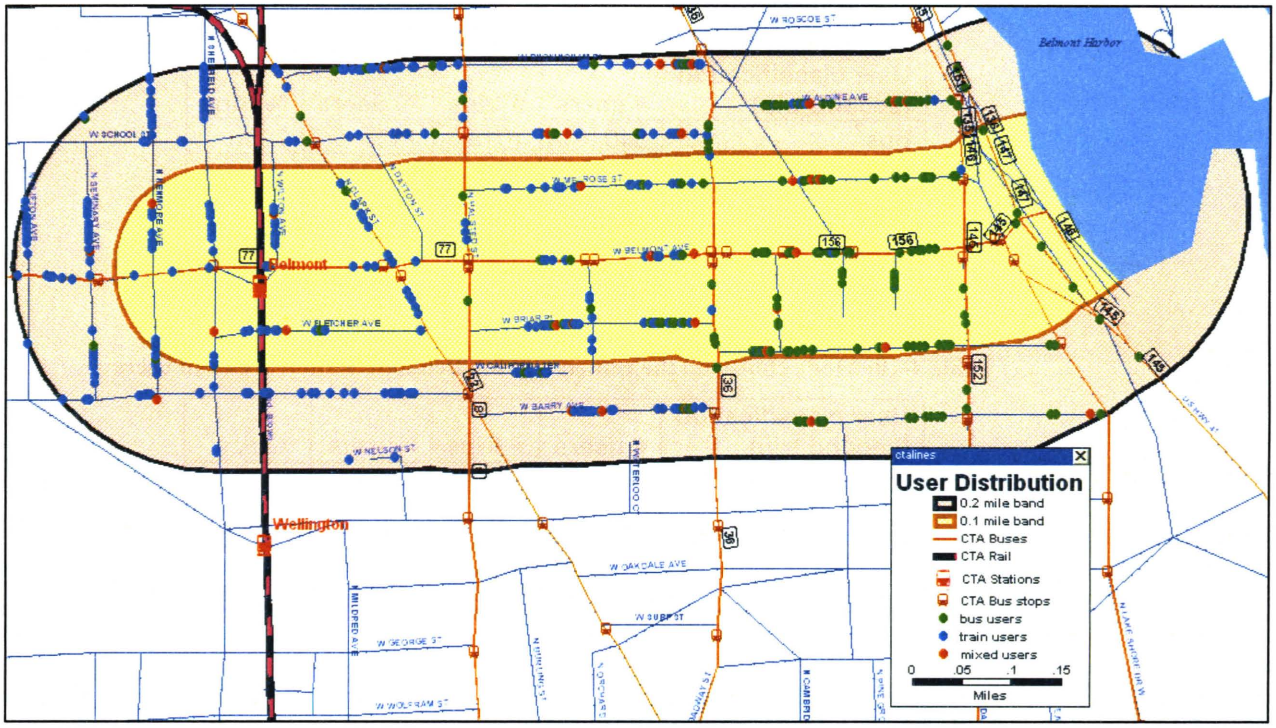


Figure 4-23: Sample customers for mode choice variation on all trips of the week

Table 4-11: Distribution of sample customers for mode variation on all trips of the week

User Type	# Customers
Bus	658
Rail	704
Mixed	99
All users	**1461**

Based on the mode choice for all their daily trips, we have further classified customers as:

- All Bus – Those who used only bus for all their daily trips during the study week
- All Train – Those who used only rail for all their daily trips during the study week
- Mixed – Those who used both bus and rail, for their daily trips during the study week

The distribution of customers based on mode-choice cross-classification for both their first trips and for all their daily trips is shown in Table 4-12. An important result is that the preference for the mode used on their first trip is higher for rail users than for bus users. About 47% of bus users (based on first trip classification) use only bus for all their daily trips, while 71% of rail users (based on their first trip classification) use only rail for all their daily trips.

Table 4-12: Distribution of customers based on mode choice on their first and daily trips

	User type (all daily trips)			
User type (first trip of day)	All Bus	All rail	Mixed	# customers
Bus	47%	0%	53%	658
Rail	0%	71%	29%	704
Mixed	0%	0%	100%	99
All users	**21%**	**34%**	**44%**	**1461**

The distribution of users, who always use one mode on their first trip and change their mode on other trips of the day, is not representative of how often do they actually use the other mode. To represent this, we have determined the percentage of trips made by the two modes as shown in Table 4-13.

Table 4-13: Distribution of all week trips based on mode choice

User Type (first trip of day)	% train trips	% bus trips	Total Trips
Bus	18.8%	81.2%	7297
Rail	92.1%	7.9%	7107
Mixed	58.9%	41.1%	1163
All users	**55.3%**	**44.7%**	**15591**

Table 4-12 and Table 4-13 show that while 53% of first trip bus users also make rail trips, these trips constitute a small fraction (19%) of their total trips. Table 4-13 for rail users shows an even stronger uniformity, as although 29% of first trip rail users also use bus, their bus trips constitute less than 8% of their total CTA trips. Based on these numbers we can accept the hypothesis that customers are less likely to change their mode of travel for corridors with dominant modes and services.

Several factors associated with time of the trip, with trip purpose, and with perception of level of service can better explain the true reasons for this travel decision. In the next section we have attempted a microscopic analysis to analyze these factors.

4.6.4 Microscopic analysis for path-choice

In this section we have mapped daily trips to extract path-choice parameters for a reduced sample of seven customers – 4 rail users and 3 bus users, who live within 0.1 mile of Belmont-Orchard street intersection and commute to the Loop. Through these examples we have tried to examine the path-choice behavior of customers and tried to find a visual correlation between their route-choice, access distance, travel times and generalized cost. The daily trips for the sample users along with a brief description of their travel habits are tabulated in Appendix 3. As an illustration we show the results for a rail user in this section.

We have extracted the path-choice parameters according to the following methodology:

1. Infer the boarding locations for all the daily trips

2. Map the user location, boarding location and chosen route on CTA's TransCAD model
3. Estimate the access distance to the boarding location and the, generalized cost for the path chosen

For a rail user living in Belmont Orchard area, we inferred his boarding locations for daily trips as shown in Table 4-14. We next mapped the user location, boarding location and chosen route on CTA's TransCAD model (Figure 4-24) to estimate the access distance to the boarding location and the, generalized cost for the path chosen.

Table 4-14: Daily trips for sample rail user

Consistency	Frequency	Serial #	Date	Time	Boarding Location	Mode	Route
3	4	1112690200	9/14/2005	7:53:48 AM	Belmont_Red	rail	1057
3	4	1112690200	9/14/2005	5:24:09 PM	LaSalle_Van Bur	rail	1130
3	4	1112690200	9/15/2005	7:49:20 AM	Belmont_Red	rail	1057
3	4	1112690200	9/15/2005	5:34:49 PM	LaSalle_Van Bur	rail	1130
3	4	1112690200	9/15/2005	6:08:09 PM	BELMONT + SHEFFIELD	bus	77
3	4	1112690200	9/16/2005	8:09:13 AM	Belmont_Red	rail	1057
3	4	1112690200	9/16/2005	2:28:17 PM	State/Lake	rail	1126
3	4	1112690200	9/16/2005	2:50:41 PM	BELMONT + SHEFFIELD	bus	77

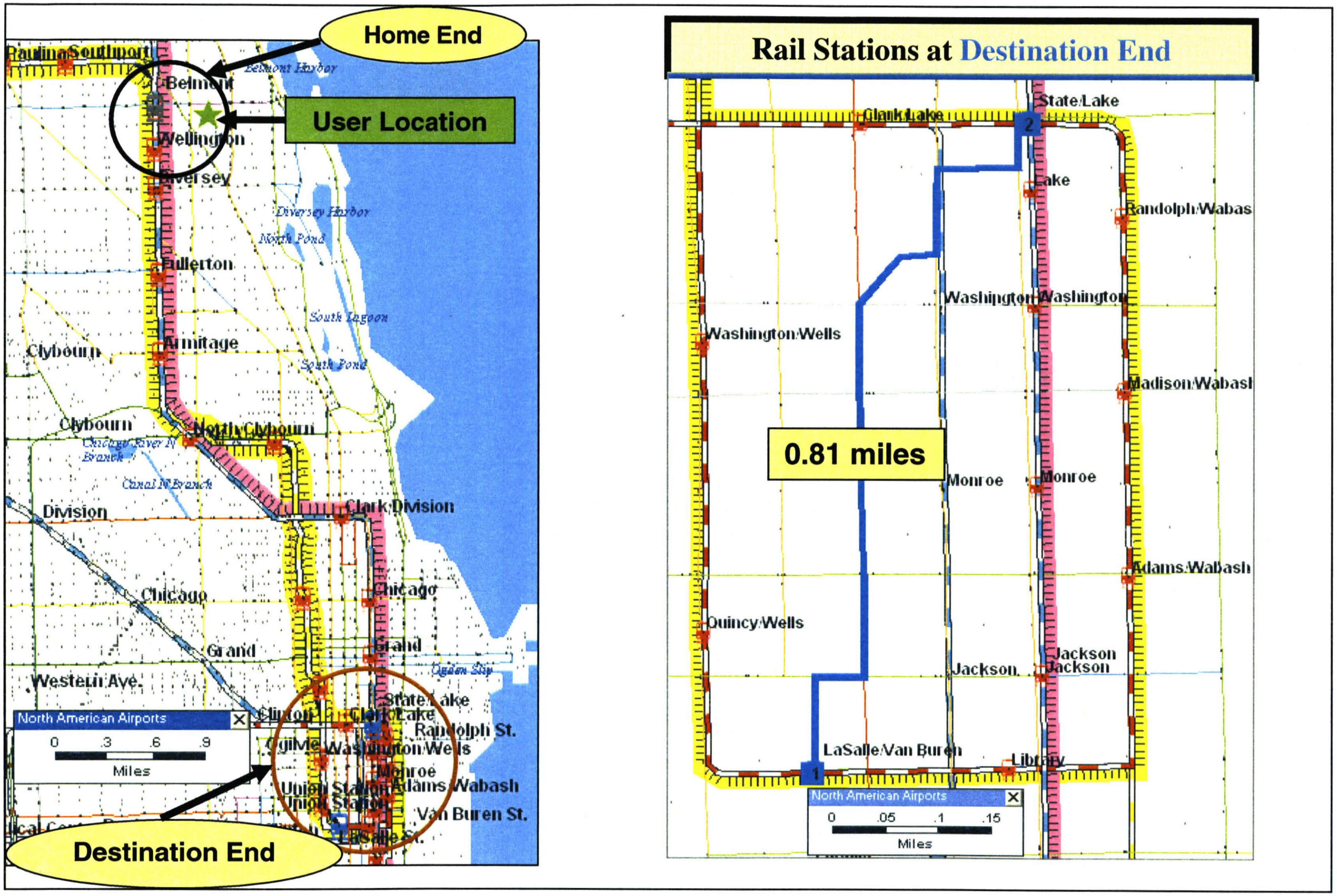


Figure 4-24: Rail sample user profile

After this microscopic analysis we were able to infer the following travel characteristics for this user:

- He is a 3 day consistent rail user type
- He uses **Belmont station** as the only boarding station at the home end
  - Access distance from home to Belmont Station = 0.4 miles
- He uses 2 different boarding stations on the return trip – **LaSalle/VanBuren, State/Lake** located at a distance of 0.81 miles of each other

Generalized cost[10] (GC) for trips between his boarding location at home and the destination end are shown in Table 4-15.

Table 4-15: Generalized cost for rail sample user

From	To	Route	Fare ($)	IVTT (min)	Wait Time (min)	Egress Time (min)	# Stops	Time Cost ($)	GC ($)
Belmont Station	LaSalle/ VanBuren	Brown Line	1.75	19.5	2.0		10	4.30	6.05
Belmont Station	State/Lake	Red Line	1.75	13.5	2.0	3.0	6	3.70	5.45
LaSalle/ VanBuren	Belmont Station	Purple Line	1.75	19.5	5.0		10	4.90	6.65
State/Lake	Belmont	Brown Line	1.75	17.5	2.5		9	4.00	5.75

4.6.5 Conclusions

In this section, we have analyzed trip patterns, including boarding location at home and at the destination end (for the work return trip), route variability, mode-choice, etc., for a sample of customers on the Belmont Street corridor. Mode choice has also been examined for all other daily trips. We determined that the boarding location at the home

[10] Generalized cost refers to actual cost obtained by actual times and not the factored times

end is significantly affected by access distance for users who are in the vicinity of a bus/rail station, and showed how path-choice is determined by the generalized cost. In addition to the factors outlined above, we have hypothesized other factors which could play a role in path choice decision. The extraction of these parameters and the systematizing of the return trip boarding location can pave the way to potential path-choice modeling applications for smart card holders. We have developed methodologies and showed results to analyze mode choice for the return trip. This microscopic study not only gives us more insight into path-choice characteristics of customers, but also tells us how powerful the AFC, AVL and TransCAD tools can be to analyze such behavior.

4.7 Chicago Card Penetration

In order to see how far the conclusions that are revealed by these analyses of smart cards can be generalized, it is important to know how representative the smart cards are of the total user population. In this section we propose an index called "Chicago card penetration" which gives a better indication of the levels of penetration of the smart cards. This also finds application in areas ranging from determining factors which affect an individual's decision to purchase a Chicago Card to setting up marketing targets for Chicago Cards.

4.7.1 Definition of 'Chicago card penetration' rate

CTA's card registration policy provides the home address for most registered customers. Spatial distribution of the smart cards however, is not a good parameter to measure penetration as it is not normalized with respect to the number of transit users. Frequency and time consistency analysis (Section 4.1) for CTA smart card customers underlines the significance of the commuting market for the current CTA Chicago card penetration. From CTA's perspective, it is important to know the number of Chicago Card customers among those CTA users who use transit for work trips. Therefore, the spatial distribution of smart cards is normalized with respect to the number of residents who use transit for work trips according to the latest census. Mathematical formulation of this proposed index – **Chicago card penetration (CCP)** is as follows:

$$\text{Chicago card penetration (CCP)} = \frac{\text{\# Chicago card customers (in a particular TAZ)}}{\text{\# Transit commuters (in a particular TAZ)}}$$

The '# Chicago cards customers' can be determined from the CTA's registration data and the '# Transit commuters' can be obtained from Census data[31] (CTPP[11] 2000). The unit of analysis is the traffic analysis zone (TAZ) which is a small geographical area most commonly used in conventional transportation planning models. The size of a zone varies, but for a typical metropolitan area, a TAZ represents about 3000 people. The number of transit commuters for each of the TAZs in Chicago is known from the CTPP 2000 data[31].

One of the issues in this analysis is that we are combining September 2004 smart card data with data for the number of transit commuters from CTPP 2000[31]. The time inconsistency between the two datasets could have been problematic. However this is less

[11] CTPP stands for Census Transportation Planning Package

likely to be an issue because – transit commute ridership is not expected to change significantly over few years, and we are conducting a relative analysis as opposed to an absolute analysis.

4.7.2 Geographical factors affecting Chicago card penetration

Geographical distribution plots of Chicago Card Penetration help us easily identify the zones having potential transit commuters but low smart card penetration; it can be used to study post-marketing effects on smart card penetration and assess the success of marketing efforts.

As an illustration, the Chicago Card Penetration rate for September 2004 is shown in Figure 4-25. To avoid problems with very small numbers, CCP is shown only for TAZs with at least 5 Chicago Cards. Spatial distribution of the CCP with respect to rail lines is also shown in Figure 4-25. Overlaying CCP onto the map of Chicago, demonstrates the following:

1. Highest Chicago card penetration rates are in the north and north-west areas as well as in areas at the outer ends of the Green line (west) and Orange Line.
2. In the north-west, zones of high Chicago card penetration follow the Blue Line (to O'Hare).

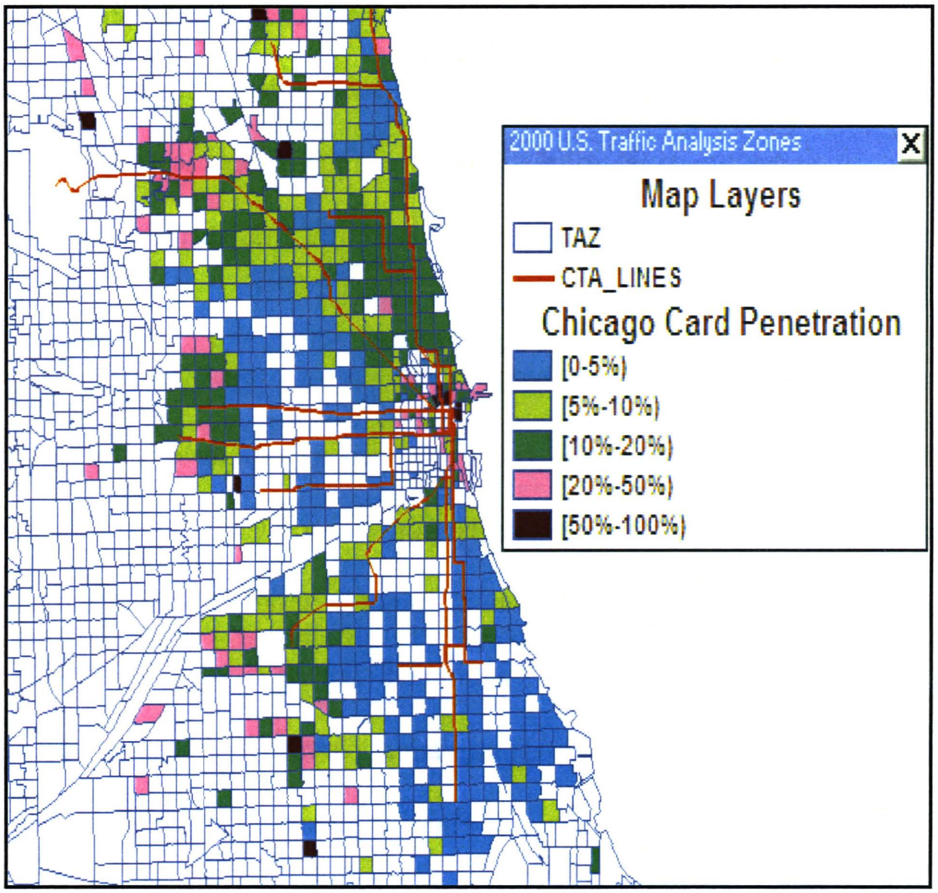


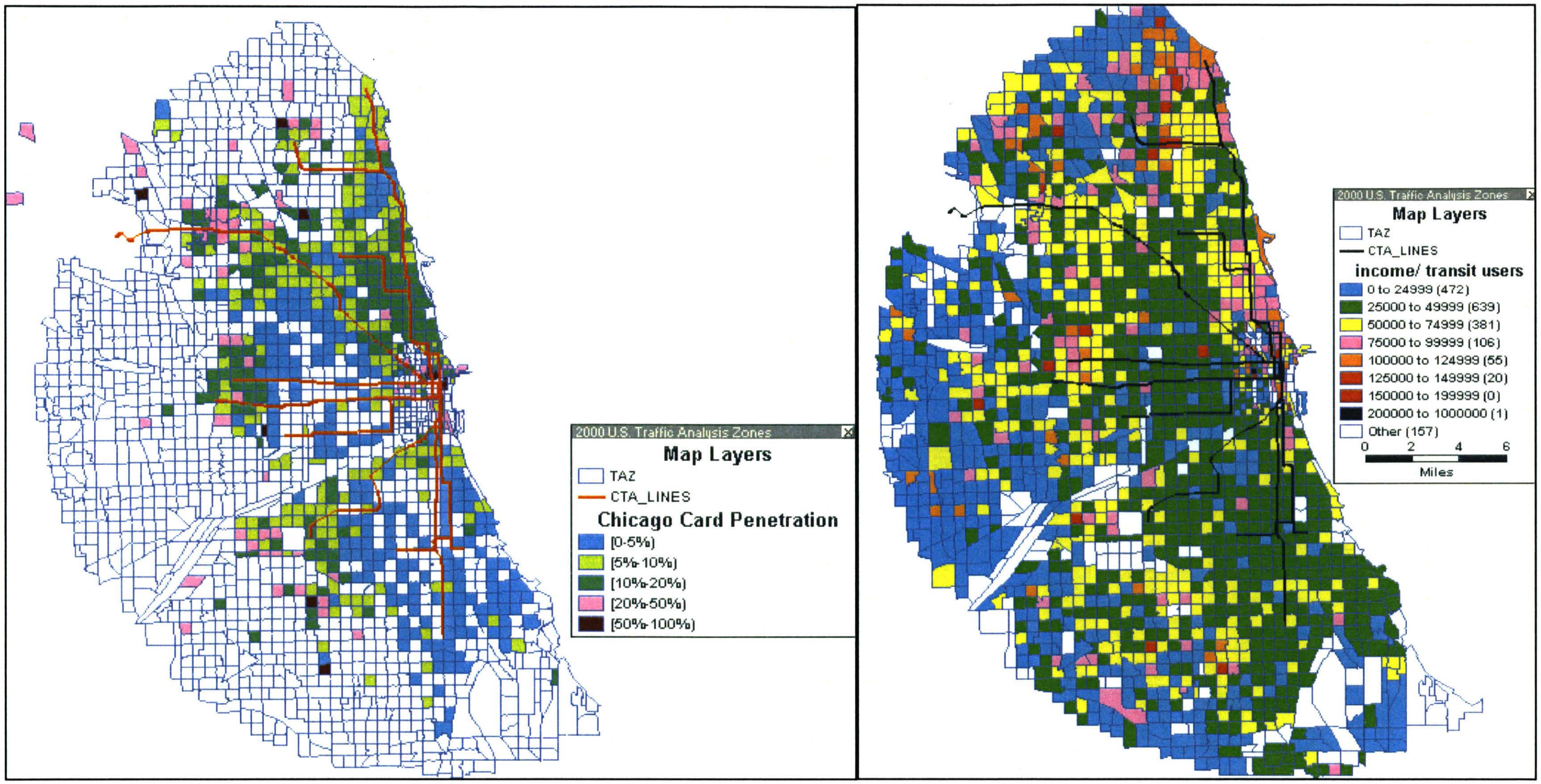
Figure 4-25: Chicago card penetration (CCP) (September 2004)

4.7.3 Socio-demographic factors affecting Chicago card penetration

In order to promote increased market penetration of Chicago Cards, one should determine the factors that govern an individual's decision to purchase these cards. In addition to geographical factors, socio-demographic characteristics (such as income level, auto ownership, household size, age, etc.) could have an impact. A correlation of Chicago Card Penetration with socio-demographic factors could help us understand the kind of people attracted to Chicago Cards.

Figure 4-26 compares the spatial distribution of Chicago Card Penetration with the household income of transit users, to see if there is an evidence of visual correlation. The two plots show apparent spatial correlation between the income levels and Chicago Card Penetration along the Red Line (to Howard) in north, the west end of the Green line (Lake) and Blue Line (O'Hare).

From the previous chapters we also know that there is certain bias towards rail among Chicago Card users which differs markedly from general CTA customers. One working hypothesis is that living close to rail stations might be more expensive and attract higher income families. Beyond this, we attempted linear regression of Chicago Card Penetration with various other socio-demographic parameters but these models however did not show any significant correlation. The likely reason for poor correlation could be low and non-uniform penetration of smart cards in Chicago. CTA smart cards account for 16% of all the system rides (10% on bus system, 27% on rail system) and 29.2% of the first ride market (20% on bus system, 41% on rail system) share for March 2006[23], which shows that smart cards are not yet ubiquitous throughout the CTA system. Besides, analysis reveals that the geographical and modal distribution of smart cards is not uniform[25], and that smart cards are more widespread in the affluent northern regions of Chicago[23, 25].



a) (b)

Figure 4-26: Chicago Card penetration and household income of transit users

4.7.4 Conclusions

We used September 2004 data to establish "Chicago Card Penetration" as a useful measure for understanding smart card take up rates. Its spatial distribution readily identifies the areas of high transit ridership for work trips with poor smart card usage, and can be used to study post-marketing effects on smart card penetration and assess the success of such marketing efforts. Another advantage of using Chicago card penetration is the ability to study the possible correlation between smart card usage and users' socio-demographic characteristics.

As an illustration we have shown the spatial distribution of Chicago card penetration with rail transit lines in Chicago. By comparing it with the spatial distribution of household income levels of transit users we show evidence of smart card usage being affected by income levels.

5 Summary and Future Work

This thesis has examined smart cards as a data resource for transit research applications. In particular, it develops methodologies to extract useful information from smart cards and applies it to better understand transit travel behavior. Chicago Transit Authority (CTA) has been used as a case study focusing on analysis of Chicago Card transactions for two weeks in September 2004 and September 2005. In this chapter we summarize the findings of the research and its application to CTA (Section 5.1), and suggestions for future work (Section 5.2).

5.1 Summary

In this section we summarize the results of the thesis in two sections. These sections are devoted to methodologies that have been developed and the results of applying them to the CTA transit network.

5.1.1 Methodologies developed

We developed methodologies to accomplish the following:

1. Select regular and consistent customers who make work trips
2. Determine boarding location
3. Determine access distance
4. Examine path-choice behavior at a fine level of detail

A methodology was developed to select those users whose first trips of the day are expected to be work trips, based on their frequency and time consistency travel habits. While the method to check for frequency of weekday use is simple, the method to check for time consistency is more complex. For the latter, codes were developed in MATLAB to establish the distribution of time consistent customers for different time thresholds (Appendix 2).

While inferring rail boarding locations from the AFC transactions is easy, determining the bus stop location is non-trivial and entails the linking of AFC data and AVL data. Amongst the different steps involved in the linking process, the most critical is to get the AVL time that corresponds to the AFC transaction time. We have developed an efficient algorithm to find this time match and implemented a code in MATLAB which gives the matched AVL time for each of the AFC transactions (Appendix 1). The lack of success in matching some records, which represents 0.01% for the rail and 14% for bus, might be attributed in the latter case, to improper login by the bus operator.

Determination of access distance has been attempted in the past [6] using distance between the nearest nodes at the origin and the destination ends. In this thesis a new methodology has been developed to find the actual access distance between the origin and the destination stops using TransCAD's GIS capabilities. The new method however requires enormous computation times (as long as 3-4 days for ~20,000 records on a 1.6

GHz processor with 1 GB of RAM) and data cleaning because of the lack of integration between the AFC and TransCAD model datasets.

5.1.2 Findings from CTA's transit network

Having developed these tools to examine the smart card transaction data they were applied to CTA's Chicago Card (CC) data base for two weeks in September 2005 and 2004 to learn about the travel behavior of its smart card holders. We have found out that out of 83,622 card holders examined for the week of September 2005, 68% are frequent weekday customers and 72% of these frequent customers are also 30 minute time consistent on their first trip of the day. These numbers are strikingly similar in the weekday-frequency and time consistency distribution of 2004 card holders. Analogously to September 2005 data, for the week of September 2004, we could find that 71% of card holders are frequent and 72% of these are also 30 minute time consistent. This underlines the significance of the commuting market for the current CTA Chicago card penetration.

This pool of frequent and time consistent customers from the September 2005 database was further analyzed with respect to their travel characteristics. The boarding locations for their AFC transactions were successfully inferred for 99.9% of rail transactions and 86% of bus transactions. These CTA customers were then classified based on their mode choice for the first trip of the day. This set of CTA customer were classified on this basis as 58% rail users, 32% bus users and 10% mixed users. Given that bus comprise two-thirds of CTA's daily rides this finding strongly suggests that Chicago Card take up is strongly biased towards rail users.

For these different types of users (rail, bus, mixed) the variability in boarding location and route for their first trips was examined. The behavior of rail users was found to be significantly less variable than bus users with more than 90% of these customers using only a single rail station for their morning transit access needs. On the other hand, for bus users, there is significant variation in boarding location, and still higher variability in route selection. From a regional analysis standpoint, it was found that while the stop variability is similar across regions, bus users demonstrate different route variability by region. Bus users residing in the north and downtown regions exhibit the highest route variability, because of the greater number of routes serving these areas. There may be different hypothesis to explain these variations ranging from the grid-like structure of the bus network, to the fact that some users may opt for the first bus to arrive independently of their route, given that more than one route serves the same general destination area.

Among frequent and time consistent rail customers of the September 2004 database, that is 12,973 users, we also established the actual access distance. We can conclude that 67% of these customers have a minimum access distance of less than a mile with an average value of 0.38 miles.

By showing the travel characteristics of sample users living on Belmont Street we have also shown the potential of smart cards for path-choice analysis. For a small sample of users we have shown that by using smart card data with the Geographical Information System (GIS) network modeling capabilities of TransCAD, we can observe path-choice characteristics such as the access distance and the generalized cost of their chosen alternative. The systematic identification of the return trip boarding location and the extraction of the travel parameters, can pave the way to potential path-choice modeling applications for smart card holders. By examining the Belmont area in more detail, we have been able to establish variations of the return trip in terms of modal change when compared to the first trip. This modal change occurs for 30 % of bus users, while only for 6% of rail users. This finding requires a more microscopic approach as shown for the set of examples contained in Appendix 3. This modal change could have potential policy implications in terms of the service design for each mode. Furthermore, the observed behavior might suggest trip chaining guiding the mode choice or different user perceptions and expectations for the return trip.

Finally, we used September 2004 to establish an index "Chicago Card Penetration", as a useful measure for understanding smart card growth. This level of penetration has been derived by tying smart card data with the census data, by using the number of work trips reported for a given Traffic Analysis Zone (TAZ). Its spatial distribution readily identifies the areas of high transit ridership for work trips, but with poor smart card usage. The recent changes in fares at CTA will hopefully provoke a higher penetration, thus reinforcing the usefulness of the smart cards to better understand travel patterns of its clientele as recommended by this thesis.

5.2 Future Work

Based on the research conducted for this thesis, there is clearly considerable potential for improving and understanding transit travel behavior through the analysis of smart card transaction data in conjunction with GIS tools. This section is divided into the methodology section and the findings section.

5.2.1 Methodologies for future research

1. Systematizing the identification of return trip boarding location: Transit travel behavior analysis can greatly benefit if the identification of the trip end can be systematized. This can be done by using all trips by the user as an input. It will help in determining the origin-destination pairs at a disaggregate level so that a rigorous path choice analysis can be performed by comparing the parameters of the chosen path with those of the reasonable alternatives.
2. Assessing transfer penalties for transit users: Using the trip itineraries for a user, we can determine his boarding location and route choice for inter-modal and intra-bus transfers. We can then plug in boarding location and route characteristics to find transfer penalties implied in the users decision [32, 33].
3. Examining all the trips to establish activity based tour data by recording times and boarding locations along the day.

5.2.2 Applications for future research

1. In the section 4.1, we have determined that only 31% of the 3-day frequent users are time consistent within 30 minute time threshold and even increasing the time threshold to 2 hours does not make many users consistent. With the aim of saving computation effort future research on work trips can therefore focus on a sample consisting of at least 4-day frequent and at least 3-day 30 minute time consistent customers
2. Use smart cards to complement longitudinal panels given that it will focus on the revealed preference and not just on the stated preference: We can use smart card data to observe the travel patterns of users over different periods of time. In particular this can be done to measure people's reactions to system changes in terms of level of service, new fares, etc. without incurring heavy survey costs.
3. Establish good regression models that explain smart card penetration based on socio-demographic factors.
4. Propose measures to increase card penetration[34], in line with recent fare changes at CTA.
5. Analyze trips other than work trips: In this research we have focused primarily on the travel characteristics of work trips. Interesting insights could also be gained by conducting a similar analysis for all other trip purposes.
6. Find other interesting travel characteristics like trip chaining, weekend travel behavior, analyses for different fare categories, etc.

Bibliography

1. Fleishman, D., et al., *Multipurpose Transit Payment Media.* TCRP Report, 1998. **32**.

2. Loukakos, D., *Automatic Vehicle Location*. 2003, http://www.calccit.org/itsdecision/serv_and_tech/Automatic_vehicle_location/automatic_vehicle_location_summary.html.

3. Palma, A.d. and N. Picard, *Route choice decision under travel time uncertainty.* Transportation Research Part A: Policy and Practice, 2005. **39**(4): p. 295-324.

4. Ben-Akiva, M., et al. *Modelling inter urban route choice behaviour*. in *Ninth International Symposium on Transportation and Traffic Theory,*. 1984.

5. Avineri, E. and J.N. Prashker, *Sensitivity to travel time variability: Travelers' learning perspective.* Transportation Research Part C: Emerging Technologies, April 2005. **13**(2): p. 157-183.

6. Utsunomiya, M., J. Attanucci, and N. Wilson, *Potential Uses of Transit Smart Card Registration and Transaction Data to Improve Transit Planning.* TRB, 2005.

7. Zhao, J., N.H.M. Wilson, and A. Rahbee. *The Potential impact of Automated Data Collection on Urban Public transportation Planning*. in *Second Workshop on Schedule-Based Approach in Dynamic Transit Modelling, Theory and Applications*. 2005. Ischia, Italy.

8. Zhao, J., *The Planning and Analysis Implications of Automated Data Collection Systems: Rail Transit OD Matrix Inference and Path Choice Modeling Examples.*, in *Department of Civil and Environmental Engineering*. 2004, Massachusetts Institute of Technology,: Cambridge, MA.

9. Barry, J.J., et al., *Origin and destination estimation in New York City with automated fare system data.* Transportation Research Record, 2002. **02-1045**(n 1817): p. p 183-187.

10. Slavik, M., *Errors in Origin-Destination Surveys Done by Number-Plate Techniques.* Transportation Research Record, 1985. **1050**: p. pp46-52.

11. Transystems Corporation, Memorandum: CTA SmartCard Traveler Database Feasibility, September, 2003.

12. http://www.eds.com/news/features/2897/, *Mass Transit: Thinking Outside the Fare Collection Box.* 2005.

13. Strathman, J.G. and J.R. Hopper, *Evaluation of Automatic Passenger Counters: Validation, Sampling, And Statistical Inference.* Transportation Research Record, 1991(n 1308): p. p. 69-77.

14. C.T.A., *CTA AVAS.* 2006, http://www.transitchicago.com/news/archpress.wu?action=displayarticledetail&articleid=106283.

15. Peng, Z.-R. and O. Jan, *Assessing means of transit information delivery for advanced public transportation systems.* Transportation Research Record, 1999(n 1666): p. p 92-100.

16. Erik Sheridan Wile, *Use of Automatically Collected Data to Improve Transit Line Performance*, in *Department of Civil and Environmental Engineering*. September 2003, MIT: Cambridge.

17. Mathy, G., *Automated Fare Collection.* 2005, EDS.com.

18. http://www.eds.com/services/casestudies/downloads/transportlondon.pdf, *Best-in-class operational services take London travel into the 21st century.* EDS.com, 2005.

19. David Gillen, Elva Chang, and D. Johnson, *Productivity Benefits and Cost Efficiencies from ITS Applications to Public Transit: The Evaluation of AVL California PATH Workshop.* CALIFORNIA PATH PROGRAM, September 2000: p. 88.

20. U.S.D.O.T., *Advanced Public Transportation Systems Deployment in the United States*. June 2003, http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPTS_TE/13846.html.

21. C.T.A., *CTA Overview*. 2006, http://www.transitchicago.com/welcome/overview.html.

22. http://en.wikipedia.org/wiki/Chicago_card, *Chicago Card.* 2006, Wikipedia.com.

23. Busby, J., *Fare Media Summary - March 2006*. 2006, Fare Media Summary - March 2006, Chicago Transit Authority: Chicago, IL.

24. Gupta, S. *Understanding Transit Travel Behavior - Value added by Smart Cards.* in *Transport Chicago*. 2006. Chicago, IL.

25. Market Research, Strategic Planning Department, and Planning and Development Division, *2003 Chicago Card Customer Survey*. February 2004, 2003 Chicago Card Customer Survey, Chicago Transit Authority: Chicago, IL. p. 16.

26. C.T.A., *CTA Maps*. 2006, http://www.transitchicago.com/maps/systemmaps.html#a.

27. O'Sullivan and S.J. Morrall, *Walking Distance to and from Light-Rail Stations.* Transportation Research Record, 1996. **1538**: p. 19-26.

28. Alshalalfah, B. and A. Shalaby. *Relationship between Walk Access Distance to Transit and Socioeconomic, Demographic and Transit Service Characteristics.* in *84th Annual Meeting of the Transportation Research Board.* 2005. Washington, D.C.

29. TransCAD, *Transportation GIS Software, User Guide, Caliper*. 2000.

30. Northwest Research Group Inc., *Customer Satisfaction Survey*. 2002, Chicago Transit Authority 2001: Bellevue, WA. p. Technical Report MR02-05.

31. CTPP. <http://www.transtats.bts.gov/databases.asp?Mode_ID=3&Mode_Desc=Highway&Subject_ID2=0>, Census Transportation Planning Package (CTPP) 2000.

32. Guo, Z., *Assessment of the Transfer Penalty for Transit Trips in Downtown Boston--A GIS-based Disaggregate Modeling Approach*, in *Transportation*. 2003, MIT: Cambridge.

33. Guo, Z. and Nigel H.M. Wilson. *Assessment of the Transfer Penalty for Transit Trips: A GIS-based Disaggregate Modeling Approach*. in *TRB*. 2004. Washington D.C.

34. Yi, H., in *Department of Civil and Environmental Engineering*. 2006, Massachusetts Institute of Technology,: Cambridge, MA.

6 Appendix 1

This section has the MATLAB code that we used to match the AFC records with the AVL data. We passed the AFC and AVL records in following manner:

- The AFC and AVL files are passed as text files, space delimited
- AFC text file has three data fields – BUS #, AFC time in seconds, Serial #
- The AFC data is arranged by BUS # (ascending) and then by AFC time (ascending)
- The AVL text file has two data fields – BUS #, AVL time in seconds
- The AVL data is arranged by BUS # (ascending) and then by AVL time (ascending)
- We pass data for only those buses which match up in AFC and AVL data sets
- The data is passed for each day like September 12, September 13, etc

As an output, we get BUS #, AFC time (in seconds) and the matched AVL time (in seconds) written to an Excel worksheet. There can be three cases of a non-match, these are:

1. AFC record for a bus occurs before the minimum recorded AVL time for the bus. In such a case the code reports "-1" as matched AVL time.
2. AFC record for a bus occurs beyond 5 minutes of last AVL time for the bus. In such a case the code reports "-2" as matched AVL time.

3. AFC record for a bus occurs between the first and the last AVL times recorded for a bus; however the AFC record occurs after 5 minutes of the AVL time. In such a case the code reports "-3" as matched AVL time.

MATLAB code is as follows:

```
% Bus AVL times as column vectors
% This has (a) bus #, (b) AVL times
clear;

format long;
fid=fopen('avl_sep12.txt','r');
[avl,count] = fscanf(fid,'%f %f',[2 inf]);
avl=avl';
busno=1;
row=1;
lengthavl=(length(avl(:,1))-1)
for(i=1:lengthavl)
    if(avl(i+1,1)==avl(i,1))
        avlbusvector(row,busno)=avl(i,2);
        row=row+1;
    end
    if(avl(i+1,1)~=avl(i,1))
       avlbusvector(row,busno)=avl(i,2);
       avlbusvectorsize(busno)=row;
       row=1;
       busid(busno)=avl(i,1);
       busno=busno+1;

    end
    if(i==lengthavl)
        avlbusvector(row,busno)=avl(i+1,2);
       avlbusvectorsize(busno)=row;
       busid(busno)=avl(i,1);
    end
end
fclose(fid);

% Bus AFC times as column vectors
% This has (a) bus #, (b) AVL times, (c) Serial #

fid=fopen('afc_sep12.txt','r');
[afc,count] = fscanf(fid,'%f %f %f',[3 inf]);
afc=afc';
busno=1;
row=1;
lengthafc=(length(afc(:,1))-1)
for(i=1:lengthafc)
    if(afc(i+1,1)==afc(i,1))
        afcbusvector(row,busno)=afc(i,2);
```

```
        afcserialno(row, busno)=afc(i,3);
        rowcounter(row,busno)=row;
        row=row+1;
    end
    if(afc(i+1,1)~=afc(i,1))
       afcbusvector(row,busno)=afc(i,2);
       afcbusvectorsize(busno)=row;
       afcserialno(row, busno)=afc(i,3);
       rowcounter(row,busno)=row;
       row=1;
       busid(busno)=afc(i,1);
       busno=busno+1;

    end
    if(i==lengthafc)
        afcbusvector(row,busno)=afc(i+1,2);
        afcserialno(row, busno)=afc(i+1,3);
       afcbusvectorsize(busno)=row;
       busid(busno)=afc(i,1);
    end
end
fclose(fid);

% matching times for preceding avl
for (i=1:busno)
    for(j=1:afcbusvectorsize(i))
        if(afcbusvector(j,i)>=avlbusvector(1,i)&&
(afcbusvector(j,i)<(avlbusvector(avlbusvectorsize(i),i)+(5*60))))%check
in if afc is greater than the min avl, and less than the last avl +5
mins
            for(k=1:avlbusvectorsize(i)-1)
                if(avlbusvector(k,i)<=afcbusvector(j,i)                  &&
afcbusvector(j,i)<avlbusvector(k+1,i))%checkin if afc lies bw 2 avls
                    if  (afcbusvector(j,i)-avlbusvector(k,i)<=5*60)  %and
if it is within 5 mins(5*60 secs) from matched avl
                    avlmatch(j,i)=avlbusvector(k,i);
                    else avlmatch(j,i)=-3; %-3 indicates it has a match
but is beyond 5 mins
                    end
                end
            end

        elseif (afcbusvector(j,i)<avlbusvector(1,i))
                avlmatch(j,i)=-1; %-1 indicates afc is less than the
min reported avl arrival!
        elseif
(afcbusvector(j,i)>=(avlbusvector(avlbusvectorsize(i),i)+(5*60)))
                avlmatch(j,i)=-2; % -2 indicates afc occured even after
5 mins of last avl recorded time!
        end
    end
end


row=1;
for (i=1:busno)
```

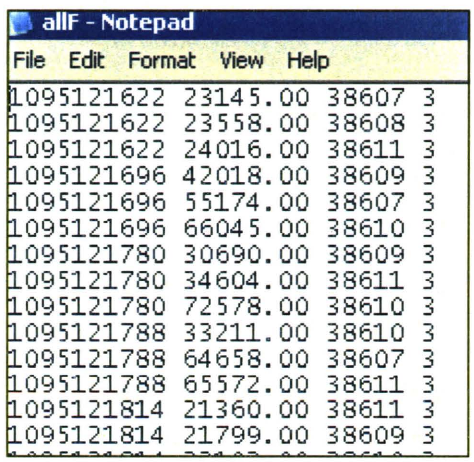
```
    for(j=1:afcbusvectorsize(i))
        matchlinearvector(row, 1)=busid(i);
        matchlinearvector(row, 2)=afcbusvector(j,i);
        matchlinearvector(row, 3)=avlmatch(j,i);
        row=row+1;
    end
end



%writing the linearized matched vector, which has, bus #, afc time and
matched avl time into an excel sheet
xlswrite('bus algo matlab check',matchlinearvector,'sep12', 'a2');
```

7 Appendix 2

This section has the MATLAB code that we used for time consistency analysis for frequent customers. We passed the data in following manner:

- We pass the data as text files
- The text file has four data fields – Serial #, First trip time (in seconds), date (as number), frequency of use
- The data is arranged by Serial # (ascending) and then by First trip time (ascending)

Print Screen of data file is shown in Figure A2- 1.



allF - Notepad

File Edit Format View Help

```
1095121622 23145.00 38607 3
1095121622 23558.00 38608 3
1095121622 24016.00 38611 3
1095121696 42018.00 38609 3
1095121696 55174.00 38607 3
1095121696 66045.00 38610 3
1095121780 30690.00 38609 3
1095121780 34604.00 38611 3
1095121780 72578.00 38610 3
1095121788 33211.00 38610 3
1095121788 64658.00 38607 3
1095121788 65572.00 38611 3
1095121814 21360.00 38611 3
1095121814 21799.00 38609 3
```

Figure A2- 1: Input text file

As an output, we get two files written to an EXCEL worksheet. These are:

1. First file contains the number of consistent customers for each of the frequency category and time threshold. Sample output is shown in Table A2- 1.

- In the output we don't get the names of the rows and columns. The columns denote the time threshold from 15 minutes to 2 hrs and beyond as shown in Table A2- 1.
- The rows denote different types of frequencies (3F, 4F and 5F)
- A cell denotes the number of consistent customers for time threshold denoted by the column and frequency type denoted by the row.

Table A2- 1: Sample output - counting the # of time consistent customers

	15 min	30 min	45 min	1 hr	1.5 hr	2 hr	>2 hr	Total
3F	2183	3344	4043	4491	5137	5517	5212	10729
4F	7734	10101	11171	11809	12561	12969	2424	15393
5F	24050	27598	28764	29381	29979	30285	733	31018
Total	33967	41043	43978	45681	47677	48771	8369	57140

2. Second file tags each of customer based on his type of consistency (0, 3, 4, 5 for 0C, 3C, 4C, 5C respectively) for different time thresholds. Sample output of this file is shown in Table A2- 2.
   - In the output we don't get the column names. The columns denote frequency type of the customer, serial #, time thresholds – 15 min to 2 hr for which type of consistency has been computed as shown in Table A2- 2.

Table A2- 2: Sample output – tagging frequent customers based on their time consistency

frequency	serial #	15 min	30 min	45 min	1 hr	1.5 hr	2 hr
3	1095121622	3	3	3	3	3	3
3	1095121696	0	0	0	0	0	0
3	1095121780	0	0	0	0	0	0
3	1095121788	0	0	0	0	0	0
3	1095121814	3	3	3	3	3	3
3	1095121836	0	0	0	0	0	0
3	1095121847	0	0	0	3	3	3
3	1095121858	0	0	0	0	0	0
4	1147839241	0	0	0	0	4	4
4	1147839253	4	4	4	4	4	4
4	1147839257	0	0	0	0	0	0
4	1147839287	3	4	4	4	4	4
4	1147839295	0	3	3	3	4	4
4	1147839297	0	0	3	4	4	4
5	1112707911	3	5	5	5	5	5
5	1112707915	4	4	4	4	5	5
5	1112707933	0	0	0	3	3	3
5	1112707937	3	3	4	5	5	5

MATLAB code is as follows:

This section describes the MATLAB codes that we have used to do the time consistency analysis for frequent customers.

```
% First trip times for frequent users as column vectors
clear;

format long;
fid=fopen('allF.txt','r');

[fr,count]=fscanf(fid,'%f %f %f %f',[4 inf]);
fr=fr';
userno=1;
row=1;
lengthfr=(length(fr(:,1))-1)
for(i=1:lengthfr)
    if(fr(i+1,1)==fr(i,1))
        frvector(userno,row+2)=fr(i,2);
        row=row+1;
    end
    if(fr(i+1,1)~=fr(i,1))
       frvector(userno,row+2)=fr(i,2);
       frvectorsize(userno)=row;
       row=1;
       frvector(userno,2)=fr(i,1);
       frvector(userno,1)=fr(i,4);
       userid(userno)=fr(i,1);
       userno=userno+1;
```

```
    end
    if(i==lengthfr)
        frvector(userno,row+2)=fr(i+1,2);
        frvector(userno,2)=fr(i,1);
        frvector(userno,1)=fr(i,4);
       frvectorsize(userno)=row;
       userid(userno)=fr(i,1);
    end
end
fclose(fid);


% Generating counts of consistent and inconsistent users
% Generating Time Consistency matrix for all users
min=[15 30 45 60 90 120];

for(k=1:length(min))
    check=min(k)*60;
    for (i=1:length(frvector)) %i controls the row position
        if (k==1)
        consistency(i,1)=frvector(i,1); % storing frequency
        consistency(i,2)=frvector(i,2); % storing serial #
        cons_tag(i,1)=frvector(i,1); % storing frequency
        cons_tag(i,2)=frvector(i,2); % storing serial #
        end
        if (frvector(i,1)==5)% checking consistency for 5F
            row =5;
            if( (frvector(i,row+2)-frvector(i,3))<=check)% row has the
number of first trips (5 for 5F...)
                consistency(i,k+2)=1;
                cons_tag(i,k+2)=5;
                else if ( (frvector(i,row+1)-frvector(i,3))<=check) %
last out
                        consistency(i,k+2)=1;
                        cons_tag(i,k+2)=4;
                    else if ( (frvector(i,row+2)-frvector(i,4))<=check)
% first out
                            consistency(i,k+2)=1;
                            cons_tag(i,k+2)=4;
                        else if ( (frvector(i,row)-
frvector(i,3))<=check) % second last out
                                consistency(i,k+2)=1;
                                cons_tag(i,k+2)=3;
                            else if ( (frvector(i,row+2)-
frvector(i,5))<=check) % second out
                                    consistency(i,k+2)=1;
                                    cons_tag(i,k+2)=3;
                                else if ( (frvector(i,row+1)-
frvector(i,4))<=check) % first and last out
                                        consistency(i,k+2)=1;
                                        cons_tag(i,k+2)=3;
                                    else consistency(i,k+2)=0;
                                        cons_tag(i,k+2)=0;
                                    end
                                end
```

```
                         end
                     end
                 end
         end
     end

     if (frvector(i,1)==4)% checking consistency for 4F
         row =4;
         if( (frvector(i,row+2)-frvector(i,3))<=check)% row has the
number of first trips (5 for 5F...)
             consistency(i,k+2)=1;
             cons_tag(i,k+2)=4;
             else if ( (frvector(i,row+1)-frvector(i,3))<=check) %
last out
                     consistency(i,k+2)=1;
                     cons_tag(i,k+2)=3;
                 else if ( (frvector(i,row+2)-frvector(i,4))<=check)
% first out
                         consistency(i,k+2)=1;
                         cons_tag(i,k+2)=3;
                     else consistency(i,k+2)=0;
                         cons_tag(i,k+2)=0;
                     end
                 end
         end
     end

     if(frvector(i,1)==3)% checking consistency for 3F
         row =3;
         if( (frvector(i,row+2)-frvector(i,3))<=check)% row has the
number of first trips (5 for 5F...)
             consistency(i,k+2)=1;
             cons_tag(i,k+2)=3;
             else consistency(i,k+2)=0;
                 cons_tag(i,k+2)=0;
         end
     end

    end
end


% counting consistent users
for (k=1:length(min))
    cons_count(1,k)=0;
    cons_count(2,k)=0;
    cons_count(3,k)=0;
    cons_count(4,k)=0;
    for (i=1:length(frvector))

        if(consistency(i,k+2)==1 && consistency(i,1)==3)
            cons_count(1,k)=cons_count(1,k)+1;
            else if(consistency(i,k+2)==1 && consistency(i,1)==4)
                    cons_count(2,k)=cons_count(2,k)+1;
                    else if(consistency(i,k+2)==1 &&
consistency(i,1)==5)
```

```
                                cons_count(3,k)=cons_count(3,k)+1;
                            end
                    end
            end
    end
    cons_count(4,k)=cons_count(1,k)+cons_count(2,k)+cons_count(3,k);

    if(k==length(min))
        cons_count(1,k+1)=0;
        cons_count(2,k+1)=0;
        cons_count(3,k+1)=0;
        cons_count(4,k+1)=0;
        for (i=1:length(frvector))
            if(consistency(i,k+2)==0 && consistency(i,1)==3)
                cons_count(1,k+1)=cons_count(1,k+1)+1;
                else if(consistency(i,k+2)==0 && consistency(i,1)==4)
                        cons_count(2,k+1)=cons_count(2,k+1)+1;
                        else if(consistency(i,k+2)==0 &&
consistency(i,1)==5)
                                cons_count(3,k+1)=cons_count(3,k+1)+1;
                            end
                    end
            end
        end
        cons_count(1,k+2)=cons_count(1,k)+cons_count(1,k+1);
        cons_count(2,k+2)=cons_count(2,k)+cons_count(2,k+1);
        cons_count(3,k+2)=cons_count(3,k)+cons_count(3,k+1);

cons_count(4,k+1)=cons_count(1,k+1)+cons_count(2,k+1)+cons_count(3,k+1)
;

        for (l=1:4)% moving across rows
            cons_count(l,length(min)+2)=0;
            for (j=length(min):length(min)+1)% moving across columns

cons_count(l,length(min)+2)=cons_count(l,length(min)+2)+cons_count(l,j)
;
            end
        end
    end

end

%writing consistency results in excel
xlswrite('consistency_05_new',cons_count,'Sheet2', 'c16');
xlswrite('consistency_05_new',cons_tag,'Sheet3', 'd3');
```

8 Appendix 3

For each of these seven sample users, the daily trips are enumerated in later part of the appendix and appear in same order as these samples.

RAIL USERS

Sample User 1. This customer uses rail to commute in both directions. His home and return trip locations are shown in Figure 8-1. His other characteristics include:

- 3 day consistent rail user
- Always boards at **Belmont station**
    - Home to Belmont Station access distance = 0.4 miles
- Uses rail for return trips
    - Uses 2 different stations on his return trip – **LaSalle/VanBuren, State/Lake**
    - Distance between two rail stations is 0.81 miles

Generalized cost[12] (GC) for trips between boarding location at home and destination end for this customer are shown in Table 8-1.

Table 8-1: Generalized cost for Sample User 1

From	To	Route	Fare ($)	IVTT (min)	Wait Time (min)	Egress Time (min)	# Stops	Time Cost ($)	GC ($)
Belmont Station	LaSalle/ VanBuren	Brown Line	1.75	19.5	2.0		10	4.30	6.05
Belmont Station	State/Lake	Red Line	1.75	13.5	2.0	3.0	6	3.70	5.45
LaSalle/ VanBuren	Belmont Station	Purple Line	1.75	19.5	5.0		10	4.90	6.65
State/Lake	Belmont	Brown Line	1.75	17.5	2.5		9	4.00	5.75

[12] Generalized cost refers to actual cost obtained by actual times and not the factored times

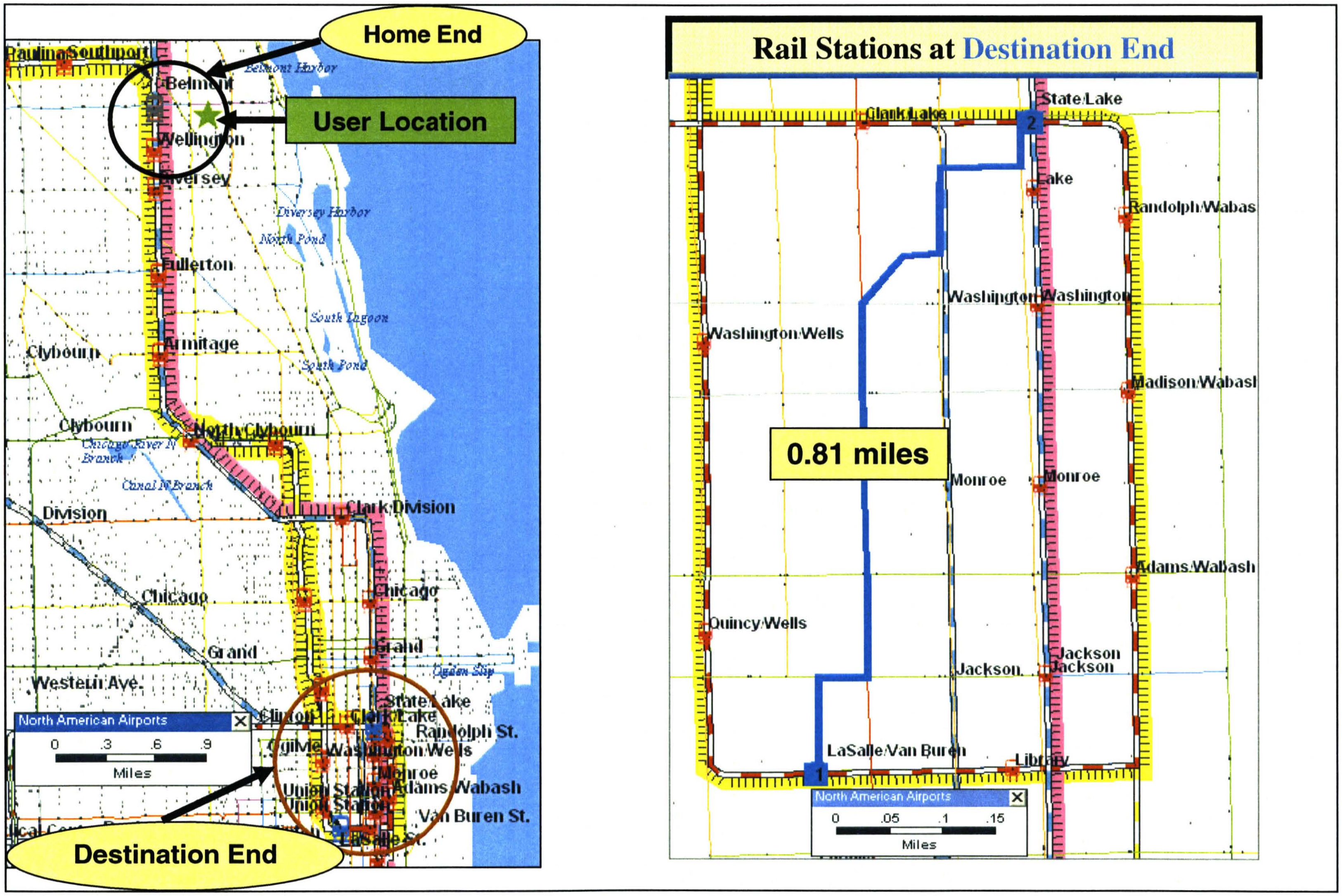


Figure 8-1: Sample User 1 profile

Sample User 2. This customer (Figure 8-2) also uses rail to commute in both directions, his other characteristics include:

- 5 day consistent rail user
- Always boards at **Belmont station**
  - o Access distance from home to Belmont Station = 0.2 miles
- Uses 2 rail stations on his return trips – **Quincy/Wells**, **Washington/Wells**
  - o Distance between two stations is 0.27 miles

Generalized cost for trips between boarding location at home and destination end for this customer is shown in Table 8-2.

Table 8-2: Generalized cost for Sample User 2

From	To	Route	Fare ($)	IVTT (min)	Wait Time (min)	# Stops	Time Cost ($)	GC ($)
Belmont Station	Washington/ Wells	Brown	1.75	16.5	2.0	8	3.70	5.45
Belmont Station	Quincy/ Wells	Brown	1.75	18.5	2.0	9	4.10	5.85
Quincy/ Wells	Belmont Station	Purple	1.75	18.5	7.5	9	5.20	6.95
Washington/ Wells	Belmont Station	Purple Line	1.75	16.5	7.5	8	4.80	6.55

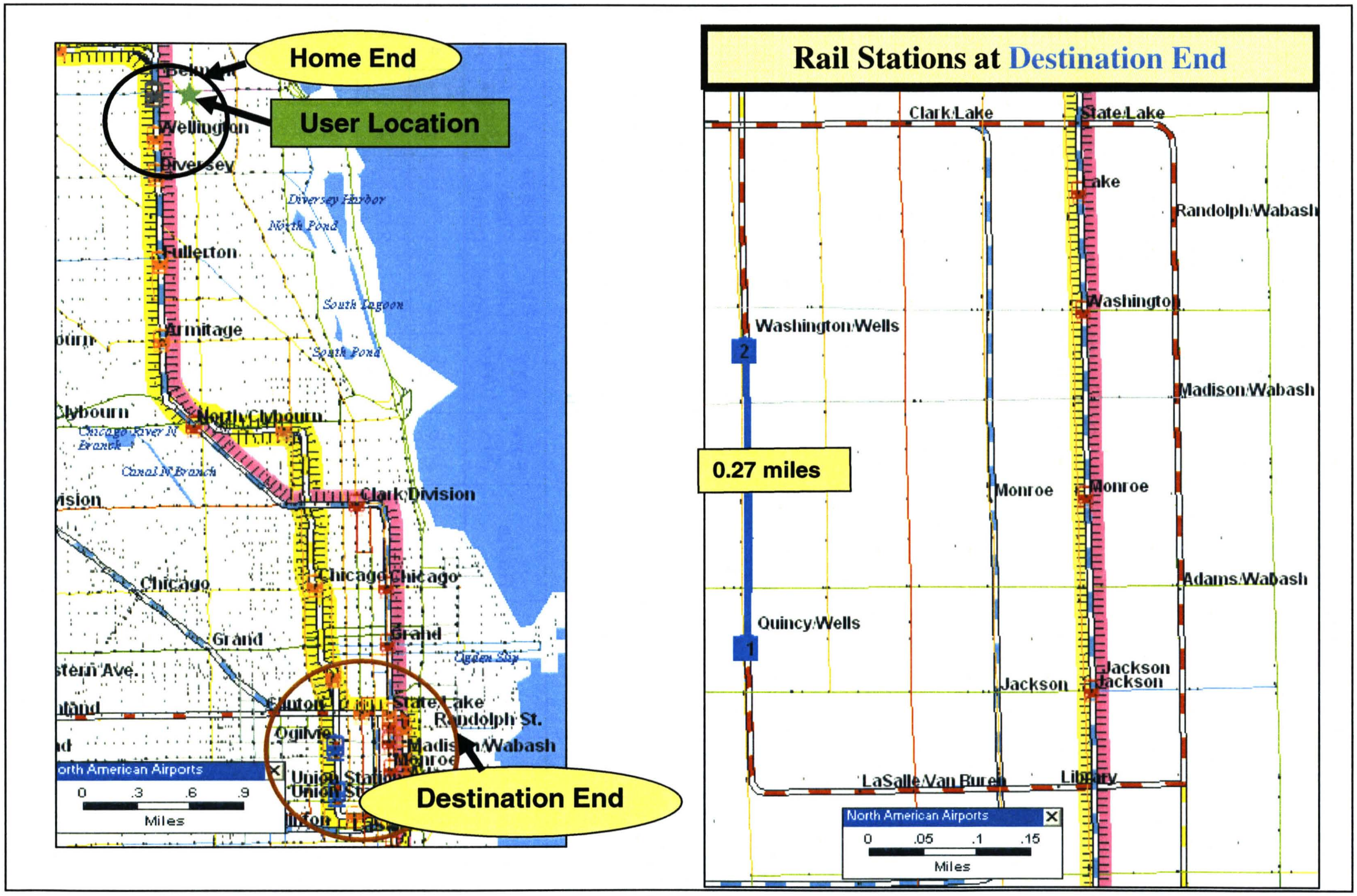


Figure 8-2: Sample User 2 Profile

Sample User 3. This customer (Figure 8-3) uses both bus and rail for his return trip, his other characteristics are:

- 4 day consistent rail user
- Always boards at **Belmont station**
    - o Access distance from home to Belmont Station = 0.2 miles
- Uses both rail and bus for return trips – **Quincy/Wells**, **Dearborn/Jackson** (Bus Route 22)
    - o Distance between 2 boarding locations is 0.3 miles
- Distance from home to nearest stop location of bus route 22 is 0.09 miles

Generalized cost for trips between boarding location at home and destination end for this customer is shown in Table 8-3.

Table 8-3: Generalized cost for Sample User 3

From	To	Route	Fare ($)	IVTT (min)	Wait Time (min)	Egress Time (min)	# Stops	Time Cost ($)	GC ($)
Belmont Station	Quincy/ Wells	Brown	1.75	18.5	2.0		9	4.10	5.85
Belmont Station	Dearborn/ Jackson	Red Line	1.75	15.1	2.0	2.0	9	3.82	5.57
Quincy/Wells	Belmont Station	Purple	1.75	18.5	5.0		9	4.70	6.45
Dearborn/ Jackson	Home Stop	Route 22	1.75	27.8	3.0		42	6.16	7.91

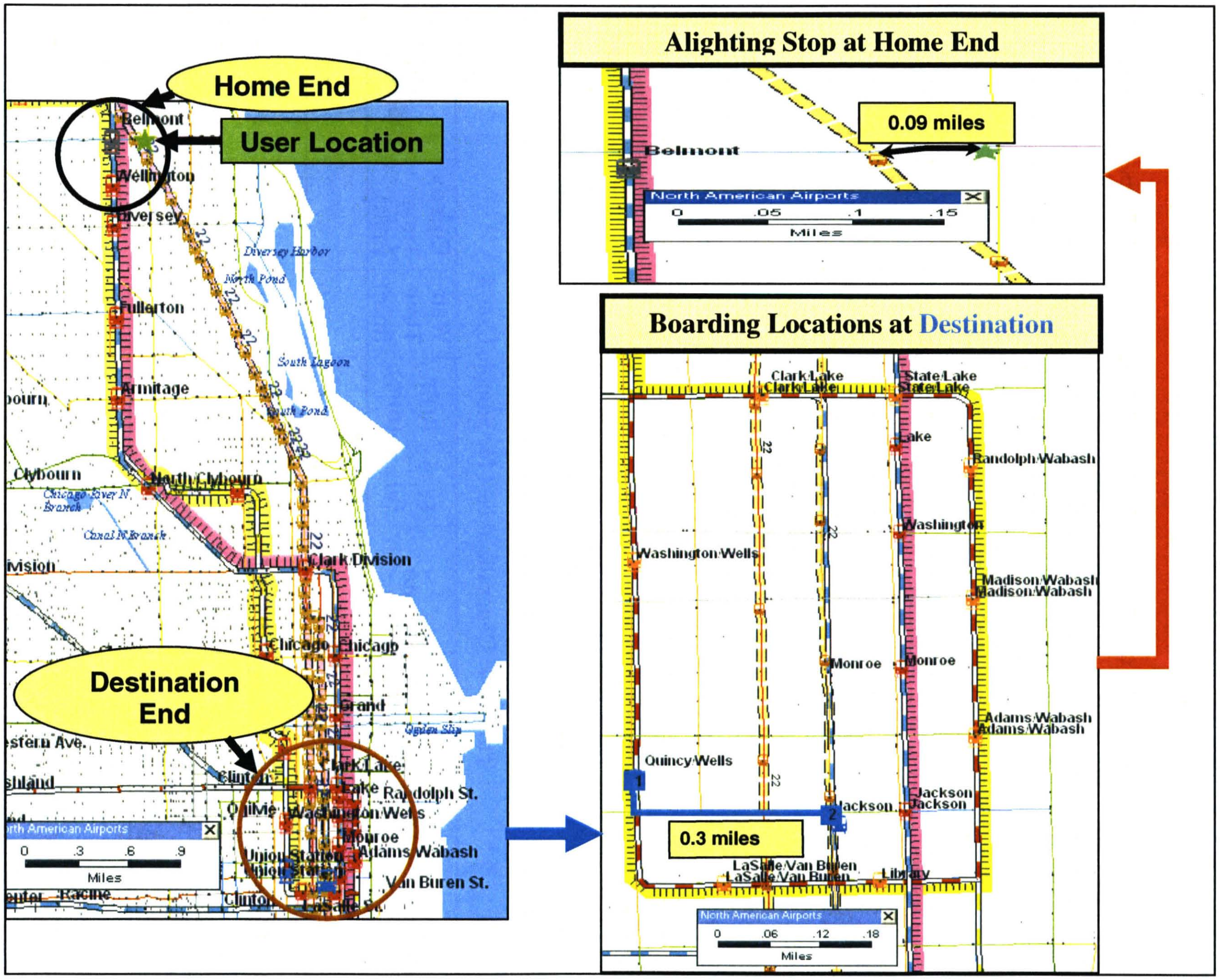


Figure 8-3: Sample User 3 Profile

Sample User 4. This rail user (Figure 8-4) predominantly uses bus on his return trips, his other trip characteristics include:

- 4 day consistent rail user
- Always boards at **Belmont station**
    - Access distance from home to Belmont Station =0.46 miles
- Uses both rail and bus for return trips – **SS Randolph/Washington, Michigan/South Water** (Bus Route 145)
    - Distance between 2 boarding locations is 0.5 miles
- Distance from home to nearest stop location of bus route 145 is 0.25 miles

Generalized cost for trips between boarding location at home and destination end for this customer are shown in Table 8-4.

Table 8-4: Generalized cost for Sample User 4

From	To	Route	Fare ($)	IVTT (min)	Wait Time (min)	Egress Time (min)	# Stops	Time Cost ($)	GC ($)
Belmont Station	SS Randolph/ Washington	Red Line	1.75	14.5	2.0		7	3.30	5.05
Belmont Station	Michigan/ South Water	Red Line	1.75	12.0	2.0	11.0	5	5.00	6.75
Michigan/ South Water	Home Stop	145, 148	1.75	11.6	4.0		11	3.12	4.87
SS Randolph/ Washington	Belmont Station	Red Line	1.75	16.5	2.0		7	3.71	5.46

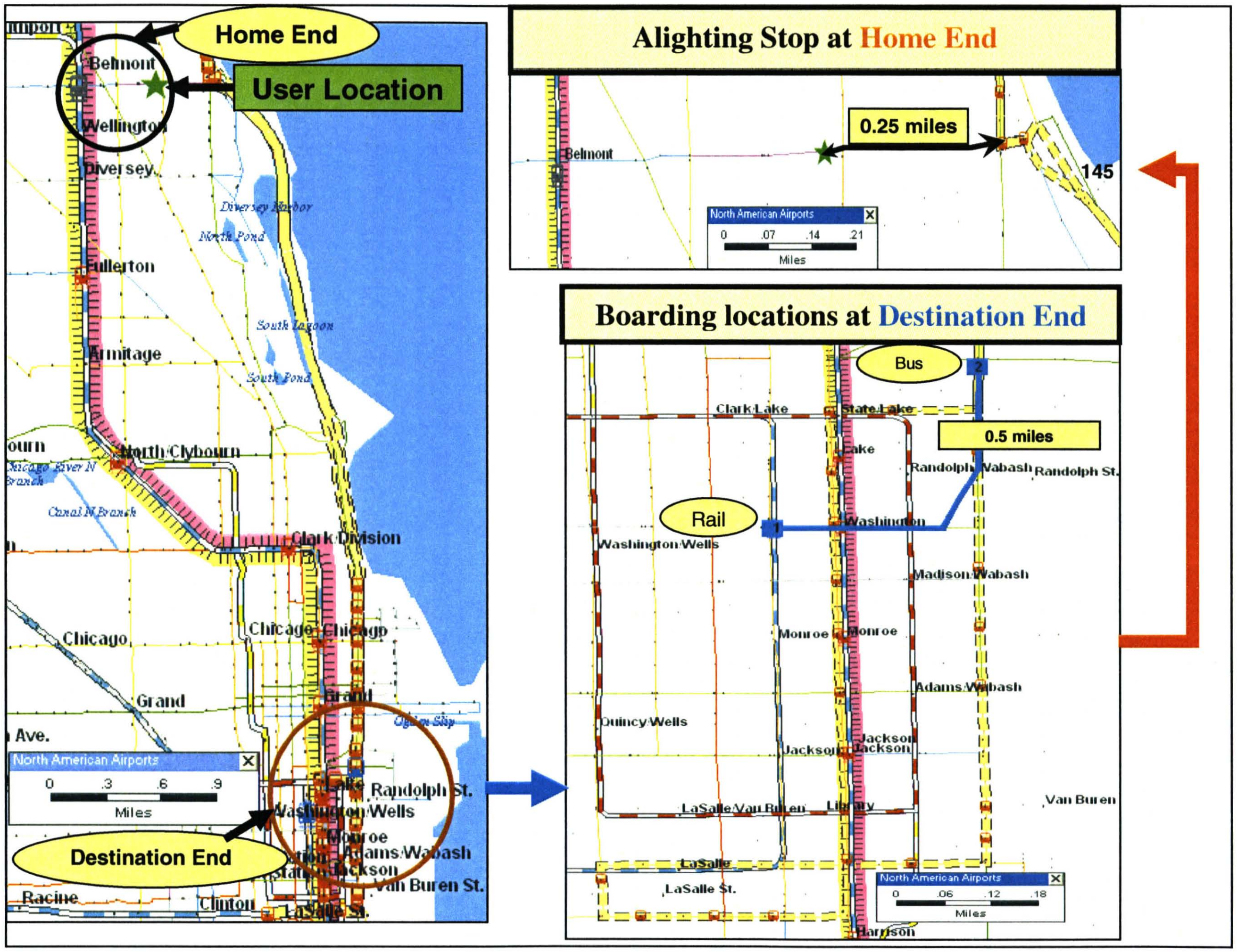


Figure 8-4: Sample User 4 Profile

BUS USERS

Sample User 5. This customer uses bus for both way trips. His home and return trip locations are shown in Figure 8-5. He demonstrates following travel behavior:

- 4 day consistent bus user
- At home end (first trip):
    a. boards at two bus stops – **Sheriden/Briar**, **Broadway/Belmont**
    b. uses three bus routes – **36, 143, 151**
    c. access distance from his home location to Sheriden/ Briar is 0.4 miles
    d. access distance from his home location to Broadway/ Belmont is 0.2 miles
- At destination end (return trip)
    a. boards at three bus stops – **Halsted/ North Avenue**, **La Salle/ Division, Michigan/ between Pearson and Chestnut**
    b. uses three bus routes – **8, 143, 156**
    c. distance between Halsted/ North Avenue, La Salle/ Division is 1.1 miles; distance between La Salle/ Division, Michigan/ between Pearson and Chestnut is 0.74 miles; distance between Halsted/ North Avenue, Michigan/ between Pearson and Chestnut is 1.8 miles
- At home end (return trip)
    a. Distance from home to nearest stop location of bus route 8 is 0.17 miles
    b. Distance from home to nearest stop location of bus route 156 is 0.07 miles

Generalized cost for trips between boarding location at home and destination end for this customer are shown in Table 8-5.

Table 8-5: Generalized cost for Sample User 5

From	To	Route	Fare ($)	IVTT (min)	Wait Time (min)	Access Time (min)	Egress Time (min)	# Stops	Time Cost ($)	GC ($)
Sheriden/ Briar	Sheriden/ Division	151	1.75	15.8	2.0			21	3.57	5.32
Broadway/ Belmont	Clark/ Division	36	1.75	17.0	2.5			24	3.89	5.64
Halsted/ North Avenue	Home Stop	8	1.75	10.1	4.0			16	2.83	4.58
Clark/ Division	Home Stop	156	1.75	20.1	2.5			25	4.53	6.28
Broadway/ Belmont	Clark/ Division	135, Alternate path	1.75	7.9	2.5	5.8	1.6	6	3.55	5.30

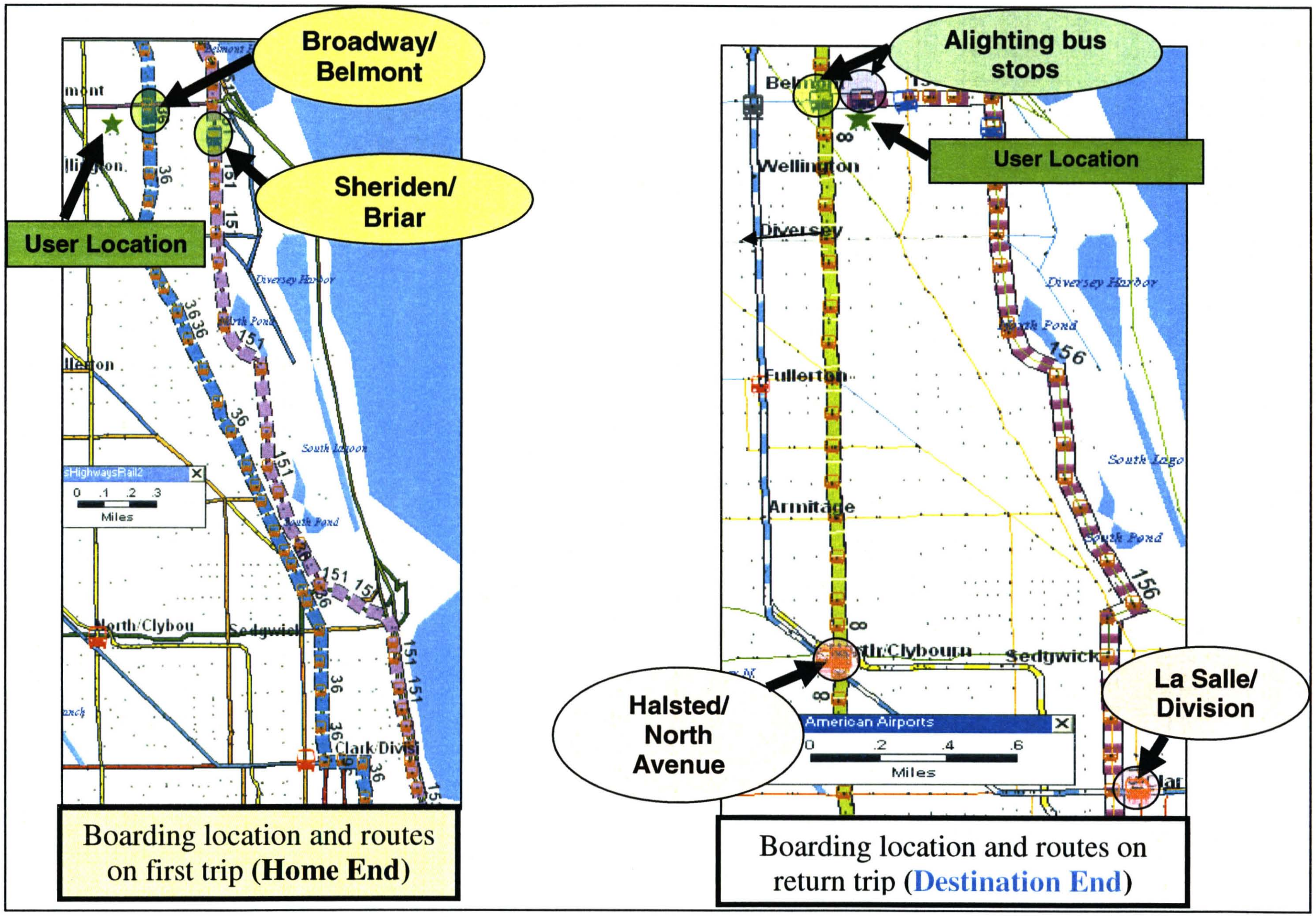


Figure 8-5: Sample User 5 profile

Sample User 6. This customer (Figure 8-6) uses both bus and rail on his return trips. His other travel characteristics are:

- 5 day consistent bus user
- At home end (first trip):
    a. boards at two bus stops – **Halsted/ Belmont**, **Halsted/ Clark**
    b. uses single bus route – **8**
    c. access distance from his home location to Halsted/ Belmont is 0.21 miles
    d. access distance from his home location to Halsted/ Clark is 0.37 miles
- At destination end (return trip)
    a. boards at two bus stops and a rail station – **Halsted/ Jackson**, **Halsted/ Monroe, Quincy/Wells**
    b. uses single bus routes – **8**
    c. distance between two bus stops is 0.18 miles
    d. distance between bus stops and rail station is approximately 0.7 miles

Generalized cost for trips between boarding location at home and destination end for this customer are shown in Table 8-6.

Table 8-6: Generalized cost for Sample User 6

From	To	Route	Fare ($)	IVTT (min)	Wait Time (min)	# Stops	Time Cost ($)	GC ($)
Halsted/Belmont	Halsted/Adams	8	1.75	21.7	4.0	35	5.14	6.89
Halsted/Jackson	Home stop	8	1.75	21.6	4.0	35	5.12	6.87
Quincy/Wells	Belmont Station	Purple	1.75	18.5	5.0	9	4.70	6.45

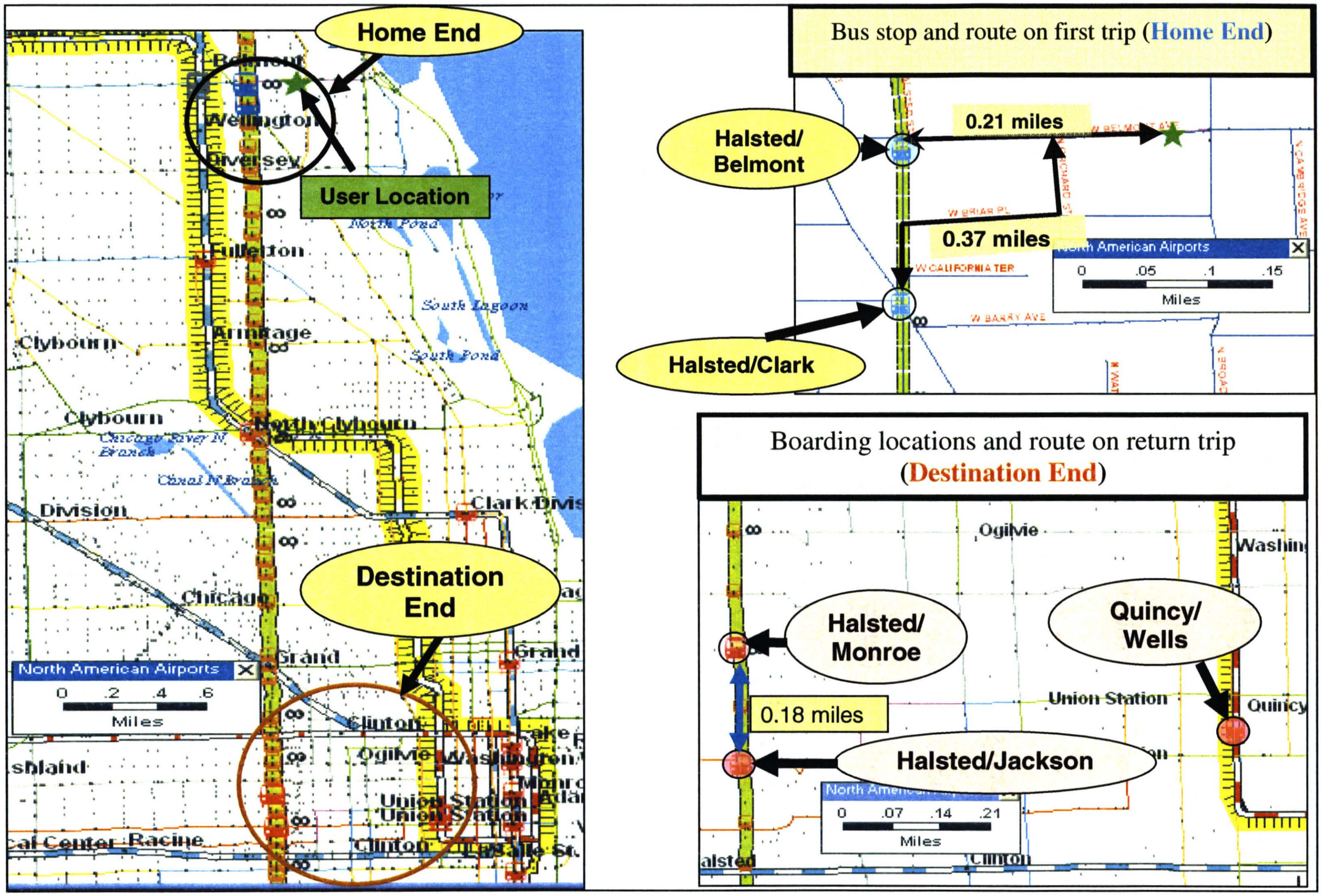


Figure 8-6: Sample User 6 profile

Sample User 7. This customer (Figure 8-7) demonstrates an interesting itinerary. He boards a bus to go to Belmont Rail Station and subsequently to his destination. He demonstrates the complex process of trip-making and its understanding. His travel behavior shows the following:

- 5 day consistent bus user
- At home end (first trip):
    a. Boards a bus to get to Belmont Rail station
    b. Boards at two bus stops – **Belmont/ Broadway**, **Belmont/ Pine Grove**
    c. Uses two bus routes – **77, 156**
- At destination end (return trip)
    a. Always uses bus back on return trip
    b. Uses different routes and stops
    c. He has a complex trip chain; he uses multiple routes (**29, 36, 145, 146**) from **State/Adams**, **Dearborn/ Adams** to go elsewhere

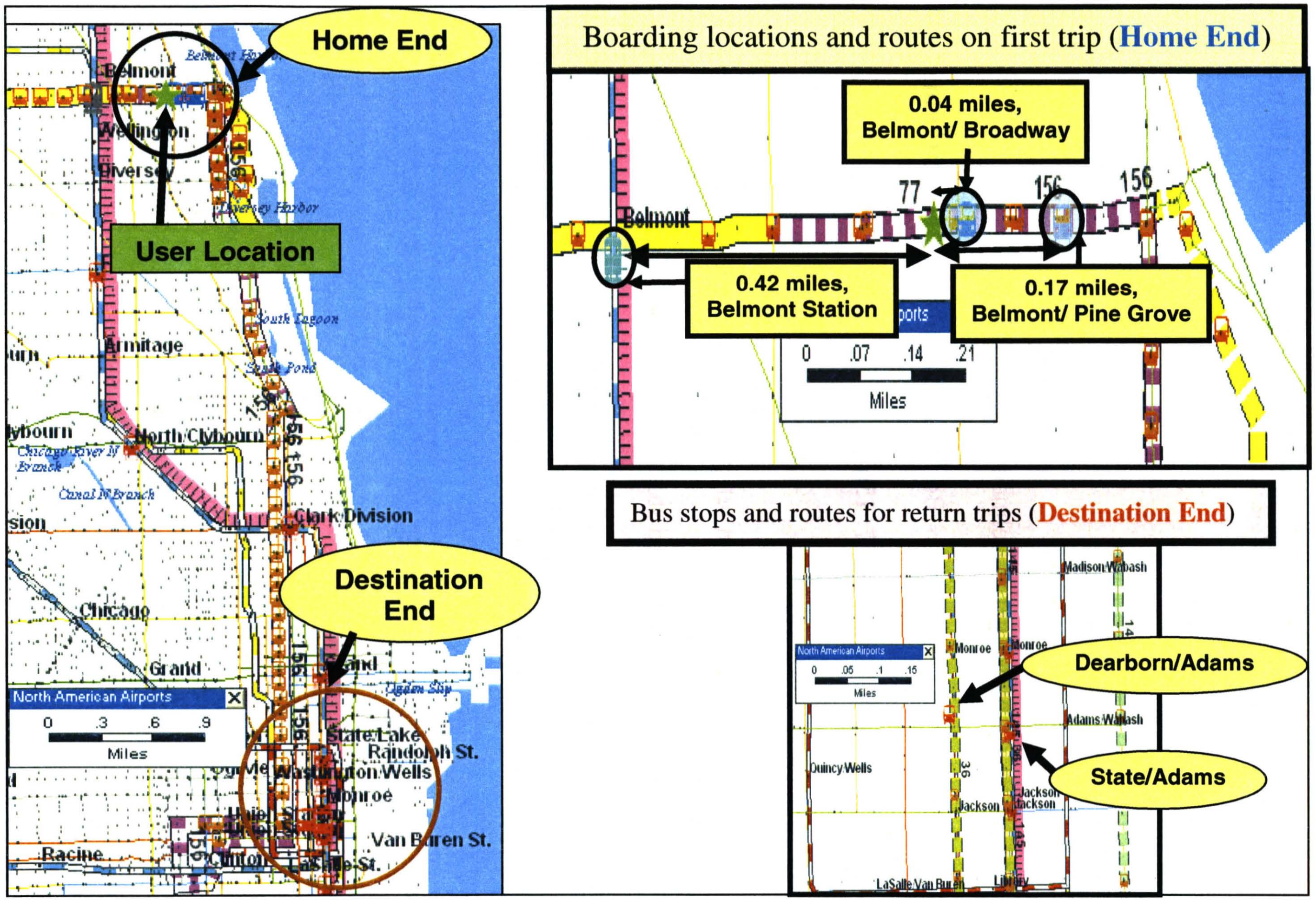


Figure 8-7: Sample User 7 profile

Rail Users

1. Rail users at both trip ends

Consistency	Frequency	Serial #	Date	Time	Boarding Location	Mode	Route
3	4	1112690200	9/14/2005	7:53:48 AM	Belmont_Red	rail	1057
3	4	1112690200	9/14/2005	5:24:09 PM	LaSalle_Van Bur	rail	1130
3	4	1112690200	9/15/2005	7:49:20 AM	Belmont_Red	rail	1057
3	4	1112690200	9/15/2005	5:34:49 PM	LaSalle_Van Bur	rail	1130
3	4	1112690200	9/15/2005	6:08:09 PM	BELMONT + SHEFFIELD	bus	77
3	4	1112690200	9/16/2005	8:09:13 AM	Belmont_Red	rail	1057
3	4	1112690200	9/16/2005	2:28:17 PM	State/Lake	rail	1126
3	4	1112690200	9/16/2005	2:50:41 PM	BELMONT + SHEFFIELD	bus	77

2.

Consistency	Frequency	Serial #	Date	Time	Boarding Location	Mode	Route
5	5	1147806835	9/12/2005	7:40:28 AM	Belmont_Red	rail	1057
5	5	1147806835	9/12/2005	5:31:19 PM	Quincy/Wells	rail	1131
5	5	1147806835	9/13/2005	7:44:05 AM	Belmont_Red	rail	1057
5	5	1147806835	9/13/2005	4:24:09 PM	Quincy/Wells	rail	1131
5	5	1147806835	9/14/2005	7:43:28 AM	Belmont_Red	rail	1057
5	5	1147806835	9/15/2005	7:37:02 AM	Belmont_Red	rail	1057
5	5	1147806835	9/15/2005	5:56:24 PM	Washington/Well	rail	1185
5	5	1147806835	9/16/2005	7:43:24 AM	Belmont_Red	rail	1057
5	5	1147806835	9/16/2005	4:15:38 PM	Quincy/Wells	rail	1131

3. Rail users, who sometimes use bus on return trips

Consistency	Frequency	Serial #	Date	Time	Boarding Location	Mode	Route

4	4	1112721100	9/13/2005	7:55:54 AM	Belmont_Red	rail	1057
4	4	1112721100	9/13/2005	5:04:04 PM	Quincy/Wells	rail	1131
4	4	1112721100	9/14/2005	8:02:15 AM	Belmont_Red	rail	1057
4	4	1112721100	9/14/2005	5:04:00 PM	Quincy/Wells	rail	1131
4	4	1112721100	9/14/2005	6:38:41 PM	Belmont_Red	rail	1057
4	4	1112721100	9/14/2005	9:39:15 PM	IrvingPark_Brwn	rail	1035
4	4	1112721100	9/15/2005	8:03:03 AM	Belmont_Red	rail	1057
4	4	1112721100	9/15/2005	5:14:38 PM	Quincy/Wells	rail	1131
4	4	1112721100	9/15/2005	7:01:31 PM	HALSTED + BELMONT	bus	8
4	4	1112721100	9/16/2005	8:19:04 AM	Belmont_Red	rail	1057
4	4	1112721100	9/16/2005	5:08:32 PM	DEARBORN + JACKSON	bus	22

4.

Consistency	Frequency	Serial #	Date	Time	Boarding Location	Mode	Route
4	5	1112721190	9/12/2005	8:26:13 AM	Belmont_Red	rail	1057
4	5	1112721190	9/12/2005	5:12:35 PM	MICHIGAN + SOUTH WATER	bus	145
4	5	1112721190	9/13/2005	8:22:41 AM	Belmont_Red	rail	1057
4	5	1112721190	9/13/2005	5:34:14 PM	MICHIGAN + SOUTH WATER	bus	148
4	5	1112721190	9/15/2005	8:24:45 AM	Belmont_Red	rail	1057
4	5	1112721190	9/15/2005	5:23:31 PM	SS Rand-Washing	rail	1183
4	5	1112721190	9/16/2005	8:23:54 AM	Belmont_Red	rail	1057
4	5	1112721190	9/16/2005	5:14:06 PM	MICHIGAN + SOUTH WATER	bus	145

Bus Users

1. Bus user, who uses bus only

Consistency	Frequency	Serial #	Date	Time	Boarding Location	Mode	Route
4	4	1112720724	9/12/2005	8:20:53 AM	SHERIDAN + BRIAR	bus	151

4	4	1112720724	9/12/2005	5:39:41 PM	HALSTED + NORTH AVENUE	bus	8
4	4	1112720724	9/13/2005	8:14:53 AM		bus	143
4	4	1112720724	9/13/2005	5:49:20 PM		bus	36
4	4	1112720724	9/14/2005	8:19:35 AM	SHERIDAN + BRIAR	bus	143
4	4	1112720724	9/14/2005	5:07:57 PM	LASALLE + DIVISION	bus	156
4	4	1112720724	9/15/2005	8:14:40 AM	BROADWAY + BELMONT	bus	36
4	4	1112720724	9/15/2005	5:52:54 PM	MICHIGAN + BETW. PEARSON & CHESTNUT	bus	143

2. Bus user, who also uses rail

Consistency	Frequency	Serial #	Date	Time	Boarding Location	Mode	Route
5	5	1147815512	9/12/2005	8:02:55 AM	HALSTED + BELMONT	bus	8
5	5	1147815512	9/12/2005	5:23:04 PM	HALSTED + JACKSON	bus	8
5	5	1147815512	9/13/2005	7:59:42 AM	HALSTED + BELMONT	bus	8
5	5	1147815512	9/13/2005	5:04:35 PM	HALSTED + MONROE	bus	8
5	5	1147815512	9/14/2005	8:16:08 AM	HALSTED + BELMONT	bus	8
5	5	1147815512	9/14/2005	5:05:07 PM	HALSTED + JACKSON	bus	8
5	5	1147815512	9/14/2005	7:54:39 PM	CLARK + BARRY/HALSTED	bus	22
5	5	1147815512	9/14/2005	10:35:53 PM	Clark/Division	rail	1060
5	5	1147815512	9/15/2005	7:52:51 AM	HALSTED + BELMONT	bus	8
5	5	1147815512	9/15/2005	5:18:22 PM	Quincy/Wells	rail	1131
5	5	1147815512	9/16/2005	7:53:29 AM	HALSTED + CLARK	bus	8
5	5	1147815512	9/16/2005	5:18:17 PM	Quincy/Wells	rail	1131
5	5	1147815512	9/16/2005	6:11:52 PM	CLARK + LINCOLN	bus	36

3. Bus user who has a complex rail link

Consistency	Frequency	Serial #	Date	Time	Boarding Location	Mode	Route
5	5	1147845244	9/12/2005	8:00:25 AM	BELMONT + BROADWAY	bus	77
5	5	1147845244	9/12/2005	8:08:35 AM	Belmont_Red	rail	1057

5	5	1147845244	9/12/2005	6:19:21 PM	DEARBORN + ADAMS	bus	36
5	5	1147845244	9/13/2005	8:02:58 AM	BELMONT + PINE GROVE	bus	77
5	5	1147845244	9/13/2005	8:11:44 AM	Belmont_Red	rail	1057
5	5	1147845244	9/13/2005	10:54:19 AM	STATE + ADAMS	bus	145
5	5	1147845244	9/13/2005	5:42:48 PM	STATE + VAN BUREN	bus	146
5	5	1147845244	9/13/2005	7:52:33 PM	BROADWAY + BUENA	bus	36
5	5	1147845244	9/14/2005	8:24:49 AM	BELMONT + BROADWAY	bus	77
5	5	1147845244	9/14/2005	8:29:28 AM	Belmont_Red	rail	1057
5	5	1147845244	9/14/2005	6:24:21 PM		bus	145
5	5	1147845244	9/14/2005	6:59:46 PM	BELMONT + SHERIDAN	bus	77
5	5	1147845244	9/15/2005	8:15:45 AM	BELMONT + BROADWAY	bus	156
5	5	1147845244	9/15/2005	8:21:52 AM	Belmont_Red	rail	1057
5	5	1147845244	9/15/2005	9:30:13 AM	STATE + ADAMS	bus	146
5	5	1147845244	9/15/2005	2:09:53 PM	STATE + ADAMS	bus	29
5	5	1147845244	9/15/2005	2:28:09 PM	STATE + WASHINGTON	bus	6
5	5	1147845244	9/15/2005	5:52:24 PM	STATE + ADAMS	bus	146
5	5	1147845244	9/15/2005	6:38:12 PM	BELMONT + SHERIDAN	bus	156
5	5	1147845244	9/16/2005	8:15:20 AM	BELMONT + BROADWAY	bus	77
5	5	1147845244	9/16/2005	8:19:20 AM	Belmont_Red	rail	1057
5	5	1147845244	9/16/2005	2:12:51 PM	STATE + ADAMS	bus	29
5	5	1147845244	9/16/2005	5:23:10 PM	LASALLE + QUINCY	bus	135
5	5	1147845244	9/16/2005	5:38:09 PM	STATE + LAKE	bus	145