The analysis of bus commuters’ travel characteristics using smart card data: the case of Shenzhen, China

Xinyi Shi
Ph. D Candidate, School of Transportation Engineering, Tongji University
Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University,
Shanghai, 201804, P.R. of China
Phone: (86)21-65988965, fax: (86)21-69589882, Email:xyshi888@163.com

Hangfei Lin*
Professor, School of Transportation Engineering, Tongji University
Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University,
Shanghai, 201804, P.R. of China
Phone: (86)21-65988965 fax: (86)21-69589882, Email:linhf@tongji.edu.cn

* Corresponding Author
Submission date: November, 14th 2013

Word Count: 4,424 (text) + 1500 (6 Figures) + 500 (2 Tables) = 6,424
Abstract

Smart card automatic fare collection (AFC) has increasingly become popular as they provide an efficient fare management. In most cities of China, the transit systems have separate AVL and AFC systems, not being connected with each other. This paper aims to analyze some characteristics of bus commuters based on the smart card data obtained in Shenzhen, China. Hence, we have to determine the boarding locations. We matched the transaction information with GPS data to predict the bus stops. And then bus commuters are defined to meet some rules. 13,626 O-D pairs among bus commuters were found for analysis of the characteristics. Most of bus commuters’ origins on weekdays focused on the residential zone, and destinations are in the city center in the morning. On weekends, there are more non-work-based activities. The three hours 7-9 a.m. 6-7 p.m. are the peak hours of public transportation. 78.14% of commuters take the same bus lines in morning and afternoon which indicated that bus lines offered to passengers are relatively few. The trip chain of the majority of bus commuters is “Bus-Bus” in a day. The transfers between bus and bus or metro are not many, and most of the transfer time is within 30 minutes which indicated a short travel time among commuters. The future work can concentrate on the accuracy improvement, the verification of result and research on the group of cardholders for more detailed analysis.
1. INTRODUCTION

Smart card automatic fare collection (AFC) has increasingly become popular as they provide an efficient fare management instead of the manual fare collection method. Data from smart card systems is considered to be very useful for transit planning. Moreover, the amount of data automatically collected is especially larger than that collected from surveys. In many developed countries, transit systems with automatic fare collection infrastructures are integrated with automatic vehicle location (AVL) technology. These systems record the time and location of the boarding transaction, providing a sophisticated description of the passenger activities within the transit network. But in China, the transit systems have independent AVL and AFC systems. Hence, finding the location of the boarding transaction is important and also difficult in data processing. Another problem is that cardholders do not need to tape cards when alighting buses, so we do not know their destinations as well. Once solving the problem of locating the stops, we can infer the destination stop and massive information can be obtained to help us for further analysis by the similar method. Therefore, this paper attempts to explore a method to determine the boarding location and then analyze the travel characteristics of bus commuters, including the bus commuters’ O-D pairs, based on the smart card data in Shenzhen, China. Commuters are the most important component of passenger flow during peak hours, so getting an accurate knowledge of their characteristics have great significance for the planning and management of the transportation system.

In this paper, we first provide background on the public transport system in Shenzhen and recent research on smart cards. The next part describes the method applied to determine the boarding bus stop and the main idea to estimate the commuter bus O-D pairs. In section 4, the case study and relevant results are introduced. The analysis of travel characteristics and discussions are presented in the following part. Finally, the conclusions are presented in section 6, and suggestions for further studies are provided.

2. BACKGROUND

2.1 Shenzhen, China

Shenzhen is a big city in the south of China, situated immediately north of Hong Kong, bordering Huizhou to the north and northeast and Dongguan to the north and northwest, shown in Figure 1. The area became China's first and one of the most successful—Special Economic Zones (SEZs). Shenzhen's modern cityscape is the result of the vibrant economy made possible by rapid foreign investment since the institution of the policy of "reform and opening" establishment of the SEZ in late 1979, before which it was only a small village. Shenzhen is now considered one of the fastest-growing cities in the world. (1) The municipality covers an area of 2,050 square kilometers including urban and rural areas, with a total population of 14.5 million in 2011. Figure 1 illustrates the location of Shenzhen. About 55% of daily trips in this city are dependent on the public transit system (9.9 million trips currently). There are 5 metro lines and over 830 bus lines in Shenzhen and bus trips account for 67.5% in daily public transit trips. Among the daily public transit trips, 21.4% is by Metro. (2) Obviously, the public bus system is the backbone of the public transit system and it serves nearly 7 million passengers per day.

![Figure 1: The map of Shenzhen.](image-url)
In 2004, Shenzhen introduced a smart card program and installed relevant equipment on buses. When boarding and alighting all the metro stations, passengers holding such cards can tape cards to the card readers. However, passengers only need to tape cards once when getting on buses not off because bus lines have the same price not distance-based. With smart cards, passengers have access to all the bus lines and metro in Shenzhen. When a smart card is placed within the range of a card reader, smart card transaction details (the card number, time, bus machine ID or metro station as well as other operational data) are created and stored in the on-board machine. The development of smart card technologies provides a good opportunity for researchers to analyze the data for transport planning.

### 2.2 Previous Work

The research on smart card technology applied to the public transit planning in these years. Some researchers indicated that using smart card data for transportation planning are meaningful and practical (3-6). Bagchi et al. (7) analyzed bus-to-bus interchanges of elderly passengers based on a sample of 98,000 journeys collected in Bradford and Southport, UK. Navick et al. (8) analyzed passenger miles, origin–destination (O-D) patterns based on the bus data in Los Angeles. The study suggested that the information where passengers get on and off is useful for effective planning. Barry and others (9) presented a method to estimate the destination location of New York City Transit’s automated fare collection system which is an entry-only system. There are two hypotheses used to estimate the alighting point: one is that passengers start their trip from the previous trip’s destination station; the other one is that the origin of the first trip is the one’s last trip’s destination. The research creates O-D trip tables as an input to a trip assignment model which is used to obtain passenger volumes on trains at peak hours and so on. Janine M. Farzin (10) outlined the process of creating an origin-destination matrix (ODM) in São Paulo. The study compared the results of ODM with prior household survey-based ODM results which suggested that the former data was better because of more comprehensive and detailed information.

In summary, previous work mainly focused on three aspects: the transfer problems, OD matrix estimation and travel behavior characteristics based on the data. The two systems (AVL and AFC) are independent in most cities in China, and the case study of Shenzhen is meaningful for the Chinese smart card analysis.

### 3. METHODOLOGY

It is acknowledged that to determine the bus stops where the cardholder board is the basis of the analyzing the smart card data. In this paper, we mainly use database technology to match the transaction information with the GPS data through the time of the transaction, and then get the latitude and longitude the place they board the bus from GPS data, and finally obtain the accurate bus stops by comparing the latitude and longitude with each bus’ latitude and longitude on bus stop table.

The data we need is composed of 4 types of tables: the smart card transaction data (the card number, time, machine ID or metro station of the transaction); bus line table (bus machine ID , bus line, the bus plate number); GPS data (latitude and longitude about the place the bus arrived returned every 20 seconds), and the bus stop table (details about latitude and longitude of every bus stop in Shenzhen).

The process is mainly divided into the following 3 steps and Figure 2 illustrates the steps.

**FIGURE 2 The steps of estimating the boarding bus stop.**

**Step1:** Determine the bus line and bus plate number through the 2 tables —smart card information and bus line information, linked the “Machine ID”. 
Step 2: Identify the latitude and longitude of transaction place. We can match the transaction time with the same time returned from the GPS data of the same bus. Practically, the time GPS conveyed will not be exactly the same as the transaction time because the GPS data is communicated with the master data about every 20 seconds. For example, a passenger taped the card at 8:35:04 a.m., but we can find 2 GPS records at 8:35:00 a.m. and 8:35:20 a.m. Hence, we think that if we can find the nearest GPS data time (8:35:00 a.m.), and that place is the nearest latitude and longitude from the exact place.

Step 3: Determine the bus station where the cardholder boards the bus by comparing the latitude and longitude of transaction place with that of all the bus stops in the same bus lines.

Let $x_G$, $y_G$ be the latitude and longitude we got from GPS, and the latitude and longitude of bus stop $i$ in bus line $A$ are $x_{Ai}$, $y_{Ai}$ ($i=1,2,…,n$). If we find that the latitude and longitude of Bus stop $B$ ($x_{AB}, y_{AB}$) satisfy the following formulation, we will conclude that Bus stop $B$ is the boarding place.

$$\sqrt{(x_G - x_{AB})^2 + (y_G - y_{AB})^2} = \text{Min} \{ \sqrt{(x_G - x_{Ai})^2 + (y_G - y_{Ai})^2} \}$$

After identifying the boarding stops, the next question is the place cardholders get off. As passengers, they only need to tape cards when they get on the bus, so it is hard to know their destinations. Since bus commuters are the most important component of passenger flow, we focus on them as the study objectives. Bus commuters refer to those who travel from home to work in morning peak and back from workplace to home in evening peak every weekday by bus, usually having at least two trips a day. Hence, it is obvious that origins of the majority of bus commuters’ last trip of the day are the same as destinations of the first trip of the day. Of course, transfer between trips should be considered. However, most of commuters have 2 trips one day from the data of Shenzhen case. So the transfer problem is not discussed in this paper.

Another question is that the information of the cardholders is unknown from the smart cards. Considering the characteristics of commuters, we define those who start the first and last trips at peak hours and behaving the same way for 5 weekdays as commuters.

4. THE DATA

The data contains one-week smart card transaction from 5th to 11th, March, 2012 in Shenzhen. Bus stop information and bus line table are also obtained. Nevertheless, we only obtain part of GPS data covered about 110 bus lines and 1800 buses. There are 827 bus lines and about 13,000 buses in the whole city of Shenzhen.

4.1 The Details Of Data

This part we show the example of the four tables to explain some fields.

- **Smart card transaction data**
  - Card_ID: represents the smart card number
  - Trade_Type: “21” indicates the metro boarding, “22” the metro alighting, “31” the bus boarding
  - Machine_ID: represents code of the machine on the vehicle

- **Bus line table**
  - The machine ID corresponds the bus line and bus plate number.

- **GPS data**
  - This table demonstrates the place, time and speed of a bus by GPS technology.

- **Bus stop table**
  - This table tells the longitude and latitude of every bus stop.

**FIGURE 3 The formation of the 4 data tables.**

- The smart card transaction data
- Card_ID: represents the smart card number
- Trade_Type: “21” indicates the metro boarding, “22” the metro alighting, “31” the bus boarding
- Machine_ID: represents code of the machine on the vehicle
- Bus line table
  - The machine ID corresponds the bus line and bus plate number.
- GPS data
  - This table demonstrates the place, time and speed of a bus by GPS technology.
- Bus stop table
  - This table tells the longitude and latitude of every bus stop.
4.2 The Description Of Data

The original dataset contains 39,646,483 records (29,997,328 on weekdays and 9,649,155 on weekends) in the week of March, 2012. There are about 2.1 million cards every weekday (1.8 million on weekend). Take Wednesday (7th, March) for example, the records is 6,033,183, including 1,065,256 metro boarding, 1,067,419 metro alighting, and 3,900,508 bus boarding. It is obvious that the number of bus boarding records is nearly 4 times as the metro boarding.

It is convinced that the number of metro records is in a more stable condition but only a slight reduction in the weekend according to Figure 4. But to see the bus records, the fluctuations is relatively large, continuously declined from over 4 million on Monday to less than 3.81 million on Friday, and the lowest point 2.78 million on Sunday. It is acknowledged that bus is the main mode of the public transport in Shenzhen city on weekdays.

Since the GPS data about 110 bus lines in Shenzhen are obtained, only a part of records we can know the exact bus station, about 300,000 records (200,000 cards) one day. Statistical result of the temporal distribution of bus passenger trips shows that the morning peak period is 7:00-9:00 and the evening peak period is 16:00-20:00. In order to find some bus commuter OD pairs, we use the idea of the trip chain which is that the stops where cardholders board in afternoon may be the destinations in morning. In spite of this relatively simple assumption, it reflects the travel behavior of the majority of bus commuters to some extent. And in accordance with the definition of the commuters, we can find part of commuter’s trip tracks.

Hence, the bus commuter’s smart cards are required to meet the following conditions:

a) The first and last trips of cardholders one day are bus trips and these trips’ GPS information is known.

b) The cardholders’ first and last of bus trips occur in the peak hours. (7:00-9:00 and 16:00-20:00 respectively)

c) Those commuters behave the same way as condition (a) and (b) for 5 weekdays.

There are about 13,600 cards meeting the three conditions. Though we know 200,000 smart cards’ bus station, we obtain only one trip information among most of the cards.

In view of the simple assumption of the bus commuters, we verify the idea based on the subway commuters. Since all the boarding and alighting metro data is definite in the AFC system, the process is simple. The result is that we find about 70,803 cards meeting the subway commuter rules which are far more than the records of bus commuters (13626). 91% of the cardholders’ first alighting the metro one day is the same station as boarding station in the last trip, which can guarantee the possibility and accuracy of our thinking to figure out the bus OD pairs.

5. ANALYSIS OF TRAVEL CHARACTERISTICS

Based on the 13,626 cards, we can know the tracks of those commuters. In the following part, several travel characteristics about those people will be discussed: the OD spatial distribution in the weekday and weekend,
the combination time group of the first trip and last trip occurred among single cards, the consistency of the bus lines commuters take in the morning and afternoon, and the statistics of the travel patterns of cardholders.

5.1 The Spatial Distribution Of Origins And Destinations

Figure 5 captures one residential area and the business center in Shenzhen to see the spatial distribution of the origins and destinations of those commuters in the first trip on weekdays and weekends. There are residential areas in the circle, and many office buildings are gathered in the oval area (the business center of Shenzhen). The figure shown in (a) represents those bus stops as the origin place in weekdays and those on weekends is revealed in (b). Figure 5(c) displays the destination bus stations of the commuters in weekdays and (d) reflects those stops on weekends. The data on weekend refers to the statistics those commuter cards we find in weekdays.

Clearly seen from (a) and (c), on weekday mornings, there are more origins within the circle, and fewer destinations in the oval zone, and the city center areas tend to be the destinations of commuters. It is concluded that most of the commuters travel back and forth between home and workplace, referring to the round and oval area in this figure respectively. The figure can confirm the estimation of OD pairs of the commuters as reasonable. However, the situation is totally different on weekends. The distribution of OD is relatively scattered, and the living zone still generates some trips, but the city center attracts fewer people. The reason is that there are more non-work-based activities on weekends. By comparison with the OD figure in weekdays and weekends, it reveals the travel behavior of commuters in Shenzhen.
5.2 The Combination Time Group of The First Trip And Last Trip Occurred Among Single Cards

FIGURE 6 The combination time group of the first trip and last trip occurred among single cards.

We choose those cardholders whose first trips are during the period between 7:00-9:00, and the last trips occurred between 16:00-20:00. Possibilities of the combination of the time belonging to single cards are limited. The label “7_18” in Figure 6 means the cardholder starts the first trip during 7:00 - 8:00, and start the last trip this day between 18:00 - 19:00. The figure shows that the proportion of the first trip during 7:00-8:00 and the last trip during 18:00-19:00 is the highest, more than 22%. The second group is “8_18”, accounting for about 16.7%. Other time combinations tend to be the average, unless 7_16, 8_16 are the least.

In summary, 7-9 a.m. and 6-7 p.m. are the important hours of bus commute traffic in Shenzhen.

5.3 The consistency of the bus lines commuters take

TABLE 1 The Consistency of the Commuters Taking the Bus Lines

<table>
<thead>
<tr>
<th>Bus Lines</th>
<th>Card Counts</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Same lines</td>
<td>10647</td>
<td>78.14%</td>
</tr>
<tr>
<td>Different lines</td>
<td>2979</td>
<td>21.86%</td>
</tr>
</tbody>
</table>

In the process of analysis, we raise the question whether those commuters take the same bus lines on and off work. The answer is that only 21.86% of cardholders take different bus lines. The reason could be that the passengers have several bus lines with the same origins and destinations, so they can choose any of them. Overall, 78.14% of passengers take the same bus lines in morning and afternoon which is quite high. The high percentage indicates that bus lines in Shenzhen offered to passengers are not enough, so commuters have to wait the only bus line.

5.4 Trip patterns

TABLE 2 The Trip Patterns

<table>
<thead>
<tr>
<th>Trip patterns</th>
<th>Counts</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-B</td>
<td>8265</td>
<td>62.86%</td>
</tr>
<tr>
<td>B-B-B-B</td>
<td>2141</td>
<td>16.28%</td>
</tr>
<tr>
<td>B-B-B</td>
<td>1309</td>
<td>9.96%</td>
</tr>
<tr>
<td>B-S-S-B</td>
<td>396</td>
<td>3.01%</td>
</tr>
<tr>
<td>others</td>
<td>1038</td>
<td>7.89%</td>
</tr>
</tbody>
</table>

Exploring the trip patterns of individual commuters is critical for public transit planning. We conclude several major trip chains found from 13,626 cards in Table 4. The series of alphabets constitute the trip patterns in a whole day which letter B represents bus and letter S means subway. Hence, “B-B” means the cardholders have only 2 bus trips one day and “B-S-S-B” represents that the passengers have 4 trips including 2 bus trips and 2 subway trips. The trip patterns, their frequencies and their percentage are showed in Table 2.

The majority of commuters have 2 bus trips one day, accounting for 62.86%. It is convenient for many passengers only to take the bus once to arrive at the destinations. And 16.28% of cardholders having 4 bus trips.
in a whole day mean that there may be a bus-bus transfer during one commute. We define those cards that the interval between 2 transaction records is less than 90 minutes as a transfer, because the interval contains the in-vehicle time of the first bus trip, transfer time and waiting time for another bus. The result is that 78.88% of cards satisfy the transfer rule. And the most of intervals are within half an hour. Hence, we confirm that most commuters’ “B-B-B-B” trip patterns contain transfer between buses and the transfer time is short.

As to the 3 bus trips (B-B-B) one day, one interval between 2 bus trips within 60 minutes accounts for about 75%. It is predicted that a transfer relationship exists between 2 trips. For example, there is a transfer between bus trips in mornings and only one bus trip in afternoons. The reasons why commuters don’t complete full trip chains are that they may take other faster transports like taxi to save time or go shopping or have some entertainment after work.

There are only a few people choose the “B-S-S-B” pattern who probably have bus-subway transfers. We find more than 85% of such cards pose a transfer relationship because the stations recorded in morning also exist in afternoon records. For example, one person takes bus on stop A and then enters the metro station B, finally alights the subway at station C in morning. In the journey back home, he steps the subway from station C to station B, and at last boards a bus from stop D to reach stop A. Therefore, the trip pattern ABC—CBD consists a full trip chain. We find the proportion of the time spent in the trips of “B-S” within 30 minutes is 76.18%. It is concluded that the commute time in Shenzhen is not long, but we should think why people would rather choose the “B-B” trip pattern than the “B-S”. Maybe the reason is that the accessibility of bus is better than subway.

The discussion above is about the commuters based on the 13,626 smart cards. To verify the validity of trip patterns of the commuters, we analyze the total database to see the trip patterns of the whole commuters. The percentage of “B-B” is about 56.14%; “B-B-B-B” is 13.02%; “S-S” is about 11.01%; “B-B-B” is 9.11%; “B-S-S-B” is 2.66%. Finally, other trip patterns sum up to 8.06%. Such percentage can be concluded that the trip patterns of the commuters from the total database are in good consistency with that of the bus commuters we choose. Because the commuters we analyzed focus on the bus commuters, the result does not contain the trip pattern “S-S” accounting for 11.01%. It is concluded that the 4 trip patterns discussed above is the main trip patterns among bus commuters in Shenzhen.

### 6. CONCLUSIONS

The database obtained from the smart card AFC system contains a great deal of information that is difficult to get through the traditional surveys. One essential piece of information is the locations where the passengers board the bus. This paper aims to find a method to locate the boarding bus stops and analyze the travel characteristics of bus commuters based on the GPS data, smart card data and other datasets obtained in Shenzhen. We have found more than 200,000 pieces of bus stop information and about 13,000 bus commuter O-D pairs. We may even achieve more if all the GPS data in the whole city can be obtained.

From the analysis, we conclude some characteristics of the bus commuters in Shenzhen.

1) Most of bus commuters’ origins on weekdays focused on the residential zone, and destinations are in the city center. On weekends, there are more non-work-based activities.

2) The three hours 7-9 a.m. 6-7 p.m. are the peak hours of public transportation.

3) 78.14% of commuters take the same bus lines in morning and afternoon which indicated that bus lines offered to passengers are relatively few.

4) The trip chain of the majority of bus commuters is “Bus-Bus” one day. The transfers between bus and bus or metro are not many, and most of the transfer time is within 30 minutes.

Based on the analysis of bus commuters’ travel characteristics and the method discussed before, it is meaningful to better plan transit system such as improving bus route efficiency, identifying new routes, and improving transit operation. Although smart card data has many advantages, still there exist some shortcomings. Smart card data cannot tell passengers’ identity, so it causes difficulties for further analysis, such as their detailed behavioral characteristics. Moreover, we only assumed some reasonable rules to predict those characteristics. Another problem is that the cardholders cannot fully represent all the bus passengers.

In summary, this paper locates the bus stop information and explored some characteristics of bus commuters in Shenzhen. The future work may concentrate on the accuracy improvement, the verification of result and research on the group of cardholders for more detailed analysis.

### REFERENCES