Special Issue on
Collaboration and Urban Transport

IMPROVEMENT OF AN OPTIMAL BUS SCHEDULING MODEL BASED ON TRANSIT SMART CARD DATA IN SEOUL

Kihwan NAM*, Myungkeun PARK

College of Business Management Engineering Department, Korea Advanced Institute of Science and Technology, Seoul, Republic of Korea

Abstract. This study was initiated with a goal of improving the bus scheduling model using the past data of "smart card". Traffic congestion level of Seoul is keep aggravating and it also has negative influence on air pollution and our health. Additionally, this heavy traffic causes high congestion costs. The continuous quantitative growth of the public transportation system brings the necessity of its efficient operation system for its future qualitative growth. The improvement of operation system is necessary also to improve public transportation operation cost efficiency of Seoul. In other words, the systematic planning is necessary for maximizing passengers' satisfaction level and the public transportation operation cost efficiency of Seoul. The current allocation interval of Seoul bus system is designed based on the empirical data of the past, which is incapable of immediate response to rapidly changing passenger demands. This research analyses passengers' behaviour and makes a proposal for the traffic network operation by analysing the "traffic card (smart card) big data", which comes from over 90% of the passengers so as to be flexible in dealing with rapid changes in demand.

Keywords: optimal public transport scheduling, smart card data, passenger time analysis, waiting time analytical model, moving time analysis model.

Introduction

As the living standards are improving due to the economic development of the modern society and the increase in the overall national income level, the traffic volume of the metropolitan area has been rapidly increased. However, this has caused various other side-effect problems such as traffic congestion and air pollution.

The contamination caused by the rapidly increased traffic in large metropolitan areas has direct adverse effects on human bodies, and also on environment, in a way such as increasing the probability of occurrence of bronchial diseases like asthma and rhinitis (Cesaroni *et al.* 2008). Especially, cases in Seoul, there is a higher level of air pollution due to a lot of traffic volumes. 70% of residents in Seoul view air pollution as the most serious environmental problem, and 53% of the population point out they should focus on reducing automobile emissions in the future (Ko 2009). The Ministry of Land, Infrastructure and Transport (2011) set a goal of reducing Green-House Gas (CHG) emissions by 34.3% compared to BAU by 2020. The Seoul Metropolitan Government also aims

to achieve the 40% reduction in CHG emissions by 2030 from the basic plan for eco-friendly energy in Seoul in 2009, compared to 1990. It is necessary to realize the need of reducing the CHG emissions and reducing the energy consumption from the government level to the individual level in Seoul. However, Seoul's energy demand trend has been steadily increased since 2005, and the trend is expected to continue until 2030. Therefore, the efficiency of the public transportation system, which accounts for 64.3% of the traffic volume, is expected to play a crucial role in the success or failure of energy demand and GHG emission reduction.

The problem of public transportation in larger metropolitan areas, especially buses, can be solved through the traffic demand management policy. Seoul and the surrounding metropolitan areas have high traffic volume that has reached to the saturation level. In the past, the infrastructural supply policy promoted personal traffic usage in order to facilitate using private cars, rather than pedestrians or public transportation. As a result, the traffic

congestion costs of Seoul has increased steadily. Therefore, to increase the efficiency of road space, policies for establishing the public transportation infrastructure such as bus and subway scheduling are promoted to increase the reliance on public transportation, aiming at the integration of the means to enable the activation of traffic cards and the transfers between each type of public transportation. However, as a result of the efforts to increase the traffic quantity, the current state of public transportation system is inefficient and the satisfaction level of passengers is continuously falling. Excessive bus operation causes environmental problems such as CHG emissions by excessive traffic and also decreases public transportation operation cost efficiency of Seoul.

Therefore, it is necessary to find the optimal bus dispatch interval that responds flexibly to the demand rather than the bus dispatch interval depending on experience. The process of planning the efficient operation of the public transportation network can be divided into four stages: route design, dispatch planning, vehicle assignment, and driver assignment (Ceder 2002). Among the four stages, the step that is determined to have the greatest impact on the traffic volume in Seoul is the stage of establishing a dispatch plan. This is the reason why so many studies have been conducted to design various models to derive the optimal dispatch interval. However, most of the previous researches are designed to predict the behaviour of passengers and to derive the results by simulation. In addition, the researches lack the focus on the integrated operating aspect, and instead their models are more focused on minimizing time.

These problems can be solved by the research using big data. Big data is defined as the data beyond the scope of the current system (Manyika *et al.* 2011), and it can be applied to the various fields, such as from the corporate marketing to the government policy establishment. In Korea, big data is actively utilized for a wide range of weather, welfare and transportation to enhance national competitiveness. Big data and data mining technologies are also used to reduce CHG emissions by efficiently improving the public transportation network. Traffic card data is generated every time, when passengers use it for any public transportation type. The data generated at a certain time is used for evaluating the efficiency of public transportation or analysing the behaviour of passengers through evaluating the efficiency of buses and subways.

The traffic card data generated in Seoul contains a lot of information. Since the traffic card must be in contact with the terminal even when getting off the bus for transferring between various means of public transportation, information on each travel section is stored. Also, the usage rate of traffic cards is high. In Seoul, the Seoul Metropolitan Government has promoted the use of traffic cards through a major public transport system reorganization since 2004 (Seoul Metro 2017). According to traffic statistics in Seoul, the traffic card usage rate of passengers using

public transportation is steadily increasing (100% subway in August 2012, 98.7% for buses and 53.5% for taxis). Passengers use traffic cards to generate data on more than 10 million traffic card transactions a day (Park, Lee 2007). The transaction card database of the Seoul Metropolitan Transportation Card contains information on boarding, departure time, destination, and transportation of passengers. The analysis of the traffic card transaction database is expected to contribute to the efficient management of the public transportation network, so it is used in a wide variety of research in the traffic geography field and the public transportation policy field to manage actual traffic data as database.

The purpose of this study is to re-evaluate the efficiency of public transportation, especially bus routes where air pollutants are relatively frequent, using the transaction data database of Seoul Metropolitan Buses. The bus operation plan consists of four steps (route design, dispatch planning, vehicle assignment, driver assignment) (Ceder 2002). The focus of the study is on a distribution plan that is directly related to GHG emissions among these four steps. In other words, the purpose of this study is to find the optimal bus dispatch time. The cost of passengers' time and bus operation varies according to the route of the vehicle in a certain period of time. Reducing the dispatch interval reduces passenger time costs (Bowman, Turnquist 1981), but increases overall bus operation costs (Jansson 1980). Therefore, it can be assumed that the bus dispatch interval that minimizes the total cost considering the time cost of the passenger and the bus operation cost is the optimum bus dispatch interval. If we find the dispatch interval that meets the demand by bus route and time zone, it will contribute to reducing the carbon emissions by reducing the overall traffic volume while improving the efficiency of passengers.

In order to optimize the bus dispatch time, the two factors that make up the total cost – passenger time and bus operation cost – must be considered at the same time. In this study, Yu *et al.* (2010), the passenger's waiting time and travel time are extracted and converted into cost units to calculate the passenger's time cost. In addition, the cost of bus operation according to the number of dispatches is derived by referring to the study by Park *et al.* (2008). By calculating the total cost plus these two variables and extracting the dispatch interval to minimize it, we propose the possibility of improving traffic congestion and air pollution by comparing the results with the current bus policy. We need to increase the degree of realism by applying actual data. Therefore, we conducted the analysis based on the transaction data of one week from March 10 to 16, 2013. Because the behaviour of passengers varies according to the day of the week and the time of day (Park, Lee 2007), we divide the weekly time into certain time zones and determine the optimal dispatch interval for each time zone.

1. Related works

1.1. Analysis of the relationship between public transportation and air pollution

A number of studies have been conducted on buses (Herndon *et al.* 2005; Puchalsky 2005; Ally, Pryor 2007; Chester, Horvath 2009) and trains (Puchalsky 2005; Messa 2006; Chester, Horvath 2009; Chester *et al.* 2013) and they quantified the amount of pollutants generated from buses and trains. Most of the policies were focused on the design of the railway, so there was not much research on the correlation between buses and air pollution. However, past studies have shown that buses also account for a large proportion of air pollution. Puchalsky (2005) compared the exhaust emissions of express bus systems and subway systems. Chester and Horvath (2009) followed the study with consideration of infrastructure construction, maintenance and repair process. Through research, it was found that the pollutants generated during the process of moving the passengers by the unit distance are often buses rather than trains. Furthermore, the relationship between CHG emissions and regional characteristics has been studied (Messa 2006; Chester, Horvath 2009; Cooney *et al.* 2013). The higher the complexity, the higher the CHG emissions. Therefore, it can be expected that air pollution will be reduced by optimizing the bus transportation network of Seoul, which contained more than 20% of the population of the Republic of Korea.

1.2. Transportation network planning

The process of designing a bus transportation network can be roughly divided into four stages: route design, dispatch planning, vehicle assignment, and driver assignment (Ceder 2002). In order to control the amount of traffic that's directly related to air pollution, it is necessary to optimize the dispatch schedule. Method of adjusting the bus dispatch interval has been studied using various models and constraints. Salzborn (1972) focused on minimizing the number of vehicles required as the means of establishing an optimal bus allocation plan. Among the various dispatch plans derived, we have chosen the alternative that minimizes user waiting time as an optimal alternative. Forbes *et al.* (1991) proposed a solution to the Multiple Deport Scheduling Problem (MDSP) using the linear relaxation method. We set the optimal dispatch interval to minimize the sum of variable costs and fixed costs required for bus operation. Dispatch time is also a large part of the process of determining the travel time of passengers (Constantin, Florian 1995). Constantin and Florian (1995) considered the fleet size as a constraint and derived the dispatch time that minimizes the expected travel and waiting time of the passenger. The sub-gradient algorithm is used to find the route spacing considering the route of passengers. In order to establish a more accurate cost model, both the operator's viewpoint and the passenger's viewpoint must be considered. Tom and Mohan (2003)

applied a model using genetic algorithms to derive optimal travel and dispatch intervals to minimize operating costs and total travel time for passengers.

1.3. Passenger benefits and operator cost analysis

The passenger cost is expressed by converting the waiting time and travel time into cost units. Yu *et al.* (2010) classified bus routes into two types in order to consider waiting time. Routes with high bus frequencies assumed that passengers would have a fixed chance of arriving at the bus stop, otherwise the passengers would be more likely to go to the bus stop as the bus arrival time approaches. Based on these assumptions, the total waiting time of the route was calculated as the product of the number of passengers waiting at each stop and the expected waiting time. The travel time can be obtained more easily. When the bus moves along the route, the travel time of all the passengers can be calculated using the travel time taken for each section between the stops and the number of passengers moved during that time. In this way, both latency and travel time can be obtained. However, considering the passengers' position, the waiting time for the bus arrival and the time spent traveling through the bus will be different. Ben-Akiva and Lerman (1985) conducted a study comparing the values of passenger latency and travel time. This study constructed a logistic model that shows the probability of a passenger choosing a certain means of travel, and derives the value of the passenger's waiting time and travel time. As a result, although the waiting time is shorter and the travel time is longer than the route with the shorter travel time, the preference of the route where people could get on the bus quicker was higher. It was found that passengers appreciate the value of the time spent waiting, which is estimated to be about twice as long as the travel time. Operational costs can be derived by using a hypothetical operating cost factor, which is the operating cost per unit distance (Park *et al.* 2008). In the process of deriving the operating cost factors, they considered the external and internal variables such as the space inside the bus, the speed of the vehicle, and accessibility.

1.4. Smart traffic card data analysis research

As Radio-Frequency IDentification (RFID) technology has developed, the payment system of public transportation has changed from cash to transportation cards. It is hoped that passenger information will be saved in the process of paying the fare with the traffic card, which will help to improve the service offered by the public transportation network. In response to this expectation, research has been carried out to improve the efficiency of the transportation network in various fields using smart traffic card data. Utsunomiya *et al.* (2006) found a problem with most public transports providing constant weekday bus schedules. By analysing customer usage patterns, the bus dispatch interval according to the usage pattern was determined for

each day of the week. Since traffic cards in some countries only store passenger information, an algorithm has been developed to infer missing information. Trépanier *et al.* (2007) developed an algorithm to predict the destination point by using data such as date, time, location, and card type. Most traffic card systems do not provide information about the destination point, so it is necessary to guess the most likely destination point through the comparison of the passenger's next riding point with similar past trip chains. This can identify the trip chain of passengers and also can be an index to evaluate the efficiency of the route. Their subsequent study (Trépanier *et al.* 2009) estimated bus operation efficiency by comparing route schedules and passengers' flight times. In this case, since the total boarding time at each stop takes a little time, data was classified by the first boarding passenger so that the boarding time of the passengers can be considered. A study on mechanisms to estimate insufficient information of traffic card data has also been conducted for subways (Munizaga, Palma 2012). It is necessary to develop an algorithm because information on the initial station of the passengers stored in the traffic card database can't provide the direction of the passengers. Algorithms were developed and applied to the subway system in such a way as to guess the direction and time of boarding through the passenger's next riding point and time, thereby making it possible to create a riding-destination matrix.

Summarized literature review is presented in the Table 1.

2. Model development

As mentioned earlier, one of the main causes of air pollution in large cities is the amount of traffic. Therefore, the greater the metropolitan area in Seoul that is covered by the transportation system, the more efficient it is. The most effective way to reduce traffic volume and reduce air pollution is to improve public transport efficiency. In particular, the role of the bus system that discharges CHG directly or indirectly is significant. In order to establish a bus transportation plan, it is necessary to go through four stages (Figure 1). First, the route must be designed according to the demand of each region and the appropriate amount of buses must be dispatched to the designed route. After that, the vehicle is assigned and the driver is assigned to each vehicle so that the bus can be operated (Ceder 2002).

The step that directly has impact on air pollution in the area where the bus transportation network is operated is the step, which involves preparing the dispatch schedule. Assuming that the demand can be met, the amount of CHG emissions can be minimized by operating a mini-
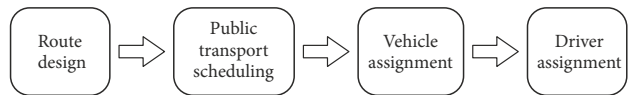

Figure 1. Public transport planning process

Table 1. Literature review

Analysis of the relationship between public transportation and air pollution	
Puchalsky (2005)	Comparison of exhaust emissions from bus express system and train system
Messa (2006)	Linking GHG emissions with local characteristics
Ally, Pryor (2007)	Correlation between bus and air pollution
Chester, Horvath (2009)	Consider infrastructure construction, maintenance and repair process study of exhaust emission
Chester, Horvath (2009)	Research on relationship between CHG emissions and technology development
Cooney *et al.* (2013)	A study on the change of CHG emissions by urban complexity
Research on transportation network planning	
Ceder (2002)	Bus transportation network design process classification and each stage model construction
Salzborn (1972)	Establish bus optimal dispatch plan from manager's point of view
Forbes *et al.* (1994)	Using linear mitigation, MDSP solution
Constantin, Florian (1995)	Derivation of dispatch time minimizing passenger's time
Tom, Mohan (2003)	Introduction of genetic algorithm, derivation of optimum movement path and displacement distance
Analysis of passenger benefits and operator cost analysis	
Yu *et al.* (2010)	Considering the time required for passengers,
Ben-Akiva, Lerman (1985)	Comparison of passenger waiting time and travel time value
Smart traffic card data analysis research	
Utsunomiya *et al.* (2006)	Analyse customers' usage patterns and determine dispatch interval
Trépanier *et al.* (2007)	Customer behaviour analysis and prediction of departure point
Trépanier *et al.* (2009)	Bus operation efficiency estimation
Munizaga, Palma (2012)	Development of algorithm for estimation of passenger's departure point in subway transportation network
Bagchi, White (2005)	Identification and classification of passenger behaviour, analysis of characteristics by group

mum number of buses. Therefore, it is important to find the appropriate number of buses for the designed routes.

In general, the buses in public transit systems run a certain route with a specified interval between each departure. Currently, the bus operation policy in Seoul sets the dispatch interval so that it can reflect the demand by such as setting the dispatch interval between the weekdays and weekends, and increasing the dispatch time in the case of commuting time where demand increases during the day. However, the dispatch interval system is based on the past experience rather than the actual usage history of a user. Therefore, even if there is a change in demand due to the different levels of congestion of a certain area, it will take a long time to reflect this in the dispatch interval. However, if traffic card transaction data can be used to immediately reflect fluctuating demand, faster demand forecasting can be possible and the benefits of passengers will possibly be increased by reflecting this in the dispatch plan.

2.1. Model building

If the additional number of buses are to be provided at a certain time for a certain route, the additional costs will be incurred from the operator's perspective. From the viewpoint of the passenger, utility is generated by reducing the time required. As a result, the two cost items have an opposite effect depending on the dispensing interval. Therefore, the dispatch interval that minimizes the total cost of both passenger and bus operation can be said to be the proper dispatch interval (Mohring 1972):

Total cost = Customer time cost + Bus cost. (1)

As can be seen from Equation (1), total cost is the sum of customer cost and bus operating cost. In this case, the bus operation cost can be expressed as the sum of the bus ownership cost including the labour cost, the expense, the management cost, and the operating cost including the fuel cost, the tire cost, and the component cost (Oldfield, Bly 1988):

Bus cost = Fixed cost + Operation cost. (2)

The cost of holding a bus varies depending on how many people are dispatched at the initial cost of one extra dispatch. In addition to the number of vehicles to be dispatched, the operating cost may vary depending on the distance travelled. Therefore, the cost of bus operation is proportional to the number of vehicles and the length of service routes.

2.2. Passenger time analysis

2.2.1. Waiting time analytical model

Yu *et al.* (2010) classified bus routes into two types. In this model, a route that has the dispatch interval of less than 10 min, is defined to have a small dispatch interval and if a route has the dispatch interval of larger than 10 min, it is defined to have a large dispatch interval. Passengers who want to take a route that has short interval do not consider the arrival time of the bus.

Therefore, it is assumed that the arrival pattern is at a certain probability until the bus arrives. On the other hand, passengers wanting to take a route with a large interval are trying to reduce the waiting time considering the arrival time of the bus. Therefore, the stop arrival pattern of passengers in this case is not constant but distributed according to a specific rule. In other words, a route where a bus arrives at the intervals of less than 10 min is defined as a case where the interval is small, and a case where the interval is large is defined as a case where the interval is larger than 10 min.

Small allocation interval passenger arrival model (f ≥ 6 veh/h). In the case (Figure 2) of the routes with small intervals, passengers are directed to the stop without taking into account the arrival time of the bus. Therefore, in this case, it can be inferred that the arrival probability of the passengers during the dispatch interval when the next bus arrives after the previous bus departs is constant.

Figure 2. Small allocation interval passenger arrival model

Therefore, the arrival probability of passengers can be expressed by the Equation (3). *H* [continuous] denotes that headway, and f_i [integer] denotes the frequency of i link:

$$f(x) = \begin{cases} \dfrac{1}{H/f_i}, & \text{for } 0 < x \leq H/f_i; \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

If the passengers arrive at the above probability, the average waiting time of the passengers is as follows:

$$\overline{x}_i = \frac{H/f_i}{2}. \quad (4)$$

Large allocation interval passenger arrival model (f < 6 veh/h). In the case of a route with a large interval, it is assumed that the passengers who want to ride know the bus schedule and are directed to the bus stop considering the arrival time of the bus to take. At this time, the passengers will be directed to the bus stop at the arrival time of the bus, so the probability of the passengers' arrival at the bus stop will be concentrated near the bus arrival time. Therefore, it can be assumed that customers arrive at a certain rate from ζ [continuous] minutes before the bus arrival time. Assuming that all customers can ride the bus, the arrival distribution of the passengers is presented in Figure 3.
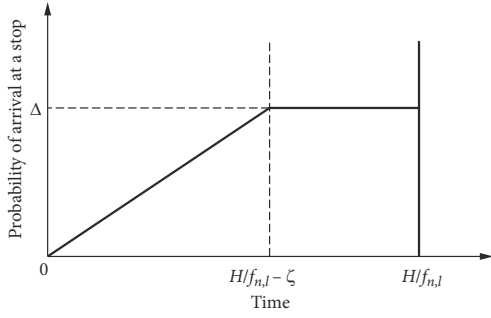
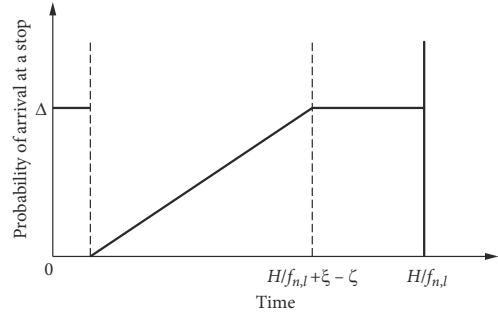Figure 3. Large allocation interval passenger arrival model A

Figure 4. Large allocation interval passenger arrival model B

Δ is the peak probability when the passengers arrive the most.

Theoretically, as shown in the graph above, passengers are concentrated on the arrival time of the bus, and all the passengers waiting at the bus stop will get on the bus, but the actual situation is slightly different. Considering the bus's time of arrival, there are some occasions where passengers miss the bus due to the little time difference, so to consider the situation and take into account the people who miss the bus, it is necessary to move the graph to the right by some time ξ [continuous].

In this case, the arrival density function of the passengers can be expressed as follows:

$$f(x) = \begin{cases} 0, & 0 < x \le \xi; \\ \dfrac{\Delta}{H/f_i}, & \xi < x \le H/f_i + \xi - \zeta; \\ \Delta, & H/f_i + \xi - \zeta < x \le H/f_i. \end{cases} \tag{5}$$

The graph, which considers this situation is presented in Figure 4.

Yu *et al.* (2010) sets the parameters $\zeta = 3.5$ min, $\xi = 0.5$ min, $\Delta = 0.2$ min in the Equation (5). Therefore, even when the dispatch time is large, the average waiting time can be calculated.

The total passenger waiting time at station j on route i is the product of the number of waiting passengers $b_{i,j}$ [integer] and the average waiting time $\overline{x}_i$ [continuous]:

$$t_{i,j}^w = b_{i,j} \cdot \overline{x}_i. \tag{6}$$

Therefore, the waiting time at each stop can be obtained by multiplying the average waiting time by the number of passengers waiting at each stop as shown in Figure 5.

Since there are two passengers waiting at the first stop and four passengers waiting at the second stop, the total

waiting time at each stop is calculated as $2 \cdot \overline{x}_i$, $4 \cdot \overline{x}_i$, respectively.

The total waiting time of route i can be obtained by multiplying the total waiting time at each stop by the dispense interval like shown below:

$$T_i^w = \sum_j f_i \cdot t_{i,j}^w. \tag{7}$$

2.2.2. Moving time analysis model

In order to calculate the total travel time of all passengers boarded on bus during the bus service segment, it is necessary to first divide each route by each segment and find out the boarding time in each segment. The total travel time in a certain segment can be calculated by multiplying the number of passengers that were on board on bus with the travel time in the segment (Table 2).

Table 2. Example of boarding section by passenger

Passenger	Boarding stop	Boarding time	Departing stop	Departing time
A	1	t_1	3	t_3
B	1	t_1	4	t_4
C	2	t_2	3	t_3
D	3	t_3	4	t_4

Suppose four passengers are boarded in a pattern like shown above. Then, for each section, the movement personnel and time are as presented in Figure 6.

As described above, it is possible to deduce travel time and the number of passengers traveling by each section using passengers' departure and their departure times.

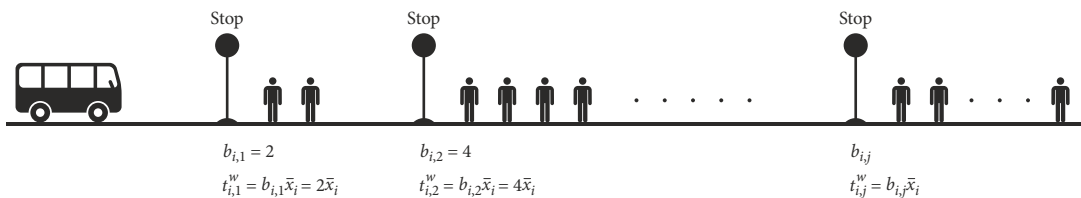The total boarding time for passengers using route i can be calculated by multiplying the bus travel time

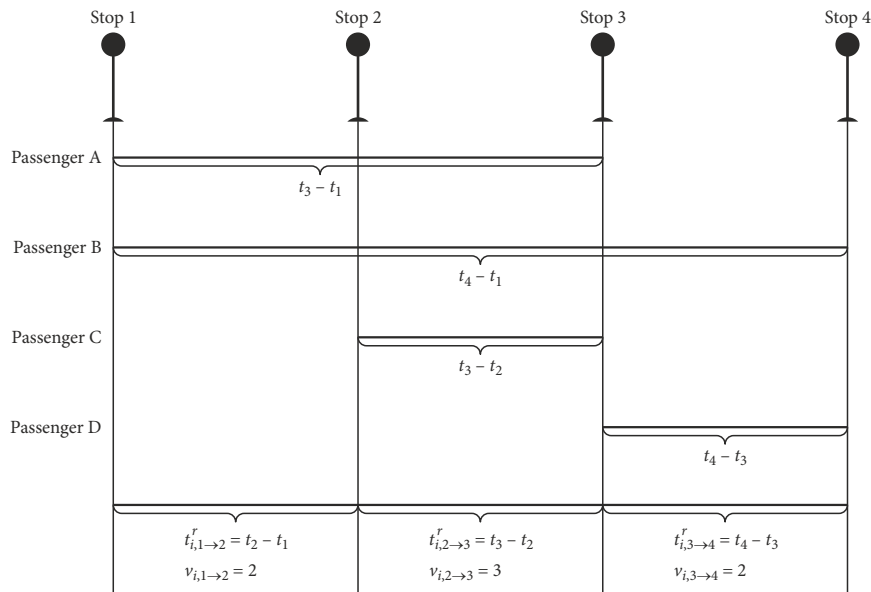Figure 5. Passenger waiting time model

Figure 6. Passenger travel time model

$t^r_{i,j \to j+1}$ [continuous] by the number of passengers $v_{i,j \to j+1}$ [continuous] travelled during this period plus the comfort index:

$$T^r_i = \sum_j w_{i,j \to j+1} \cdot t^r_{i,j \to j+1} \cdot v_{i,j \to j+1}, \qquad (8)$$

where: the comfort index is a variable that shows the level of crowdedness of the bus:

$$w_{i,j \to j+1} = \left(\frac{v_{i,j \to j+1}}{f_i \cdot V} \right)^2. \qquad (9)$$

The Equation (9) shows the number of passengers on a bus by dividing the number of passengers by the time of the bus. The actual data is applied by setting the V value to 0.25 (Yu *et al.* 2010).

2.2.3. Total cost analysis

The study by Ben-Akiva and Lerman (1985) demonstrates that the waiting time is associated with much higher cost than passenger travel time. Therefore, the weight on the waiting time is set as the proportional coefficient τ. The value of τ is set to 2 (Ben-Akiva, Lerman 1985). The floor passenger time, which is the sum of the waiting time and the moving time of the passenger, is expressed as the Equation (10):

$$\sum_i \left(T^w_i + T^r_i \right) = \sum_i \sum_j \tau \cdot f_i \cdot t^w_{i,j} + w_{i,j \to j+1} \cdot t^r_{i,j \to j+1} \cdot v_{i,j \to j+1}. \qquad (10)$$

The total travel time of passengers derived from the above formula must be converted into the cost units. In a study by Park *et al.* (2008), the cost conversion factor for passenger travel time is applied at 5235 won/h. In addition, since the bus cost factor of the bus is 309 won/veh·km and the cost of ownership cost is 13110 won, the cost required for one additional bus on any route is calculated by converting the travel distance of the route:

Total cost = Customer time cost + Bus cost =

$$\sum_i \left(T^w_i + T^r_i \right) \cdot 5325 \text{ won/h} +$$

Movement distance · 309 won/veh·km +

13110 won/veh. (11)

It is shown that the optimal dispense interval is when the total cost calculated using the process above is at the minimum cost. With this process, it is possible to find the proper dispatch interval of each time line of each route.

3. Results analysis

3.1. Data

From the March 4, 2013 to March 10, 2013, the model described above was combined with real-time traffic card transaction data for one week to derive the optimal dispatch interval. This traffic card transaction data contains all information of the customer. It has detailed information such as riding station, riding time, getting off station, getting off time, total number of passengers, total boarding time, total travel distance, transfer, vehicle ID and so on. Due to this rich information, it can be used effectively in the model of this study.

Since all traffic card transaction data in Seoul is too large data, bus No 104 is chosen as the bus that can average the data for most buses and represent it best. We analysed the boarding record of city bus No 104 (4×19 cemetery ~ Seoul Station) from 4:00 am (bus operation starting time on Sunday) to the midnight of a Sunday. Information from the card transaction data includes each passenger's boarding time, stopping point, exit time, stopping point, route information, and the like. Detailed Smart Traffic Card transaction data can be found in the Appendix.

3.2. Preliminary analysis

Basic data analysis was performed before detailed model analysis. In order to analyse passengers' and bus's traffic pattern, the number of passengers and bus traffic in each time zone were determined by time zone divided into two hour blocks (Table 3 and Figure 7).

As can be seen from the Table 3 and Figure 7, the number of passengers varies greatly over time. However, bus traffic does not change significantly except at dawn. Therefore, bus traffic needs to be adjusted to respond more flexibly to passenger demand.

Table 3. Bus traffic and number of passengers by time of day

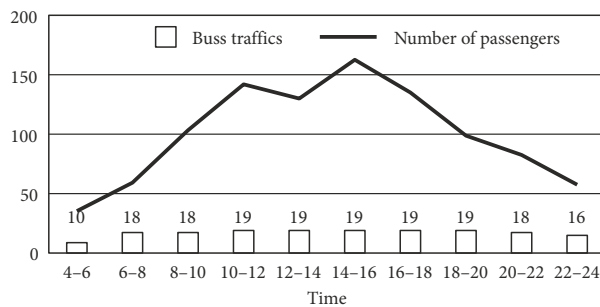| Slot	Bus traffic	Number of passengers
4…6	10	35
6…8	18	60
8…10	18	104
10…12	19	142
12…14	19	131
14…16	19	162
16…18	19	136
18…20	19	98
20…22	18	83
22…24	16	58

Figure 7. Number of passengers and buses passing by time

3.3. Model analysis

As we have seen above, we will analyse this study model to control the optimum bus dispatch that responds flexibly to passenger demand. Three time zones are selected from 10 time zones that are divided into two hour intervals from the smallest to the most frequent time zone. The time of the analysis covers three time zones from 4:00 to 6:00, from 14:00 to 16:00, and from 20:00 to 22:00. The number of passengers waiting at each stop, the number of passengers moved by each section, and the time required for each section are calculated as follows to obtain each passenger's required time (Table 4).

The pre-processing results are substituted into the model presented above, and each passenger's time is calculated. The estimated travel time of passengers shows a decreasing trend as the distance between each station decreases. However, the operating cost increases because smaller the intervals, more vehicles must be dispatched. As each bus is added, the travel cost is as high as the travel distance, and the holding cost is also increased. Here is an example of the calculation of the actual operating cost using the data of bus No 104.

The transit distance of main bus No 104 is 29.358 km and the bus operation cost per unit distance is 309 won. In addition, the cost of ownership is 1310 won per bus, so the operating cost of bus 104 is 22181.62 won per vehicle:

$Bus\ cost = 29.358\ \text{km} \cdot 309\ \text{won/veh} \cdot \text{km} +$
$13110\ \text{won/veh} = 22181.62\ \text{won}.$

Based on the above results, the passenger cost, operating cost, and the total cost of each dispatched vehicle were calculated for each time slot. The results for 20:00…22:00 hours, which can represent the entire pattern, are shown in the Table 5.

In the Table 5, the total cost trend shows that as the number of dispatched buses increases, the total cost decreases but later increases again after a certain number of

Table 4. Smart card data after pre-processing

vehc_id	ride_dtime	ride_sta_id	get on	get off	use_dist	use_seconds	interval	persons	required time
111749420	20130310041403	71613	18	25	4550	643	18…25	1	643
111749424	20130310041212	72239	8	18	3820	743	8…18	1	743
111749424	20130310042446	71613	18	28	6062	806	18…20	1	149
111749424	20130310042715	72012	20	29	5206	784	20…28	2	5256
111749424	20130310044619	72251	33	51	8778	1898	28…29	1	127
111749424	20130310044821	72208	34	55	9667	2153	33…34	1	122
111749424	20130310050028	72259	43	63	9736	2228	34…43	2	5816
111749426	20130310042026	8000880	7	18	4446	709	7…18	2	5640
111749426	20130310042030	8000880	7	24	8423	1333	18…23	1	587
111749426	20130310044202	7197	23	27	2067	373	23…24	2	328
111749426	20130310045614	72251	33	57	11015	2428	24…27	1	332
111749426	20130310050632	72255	39	56	7795	1608	18…23	1	0
111749430	20130310044655	72241	10	15	1655	464	10…15	1	464
111749430	20130310053704	72261	45	47	1824	308	45…47	1	308

buses. A graph that shows the total cost by the number of vehicles dispatched between 20:00…22:00 hours can be drawn in Figure 8. At the optimal dispatch interval, the total cost, which takes into account both passenger and operating costs, has the minimum value as shown in the graph.

Table 5. 20:00…22:00 hours cost per vehicle

| f	Passenger cost [won]	Operating cost [won]	Total cost [won]
1	2879.662	22181.6	15097212.2
2	780.666	44363.2	4131147.1
3	391.962	66544.9	2118468.3
4	255.916	88726.5	1428448.7
5	192.947	110908.1	1120982.9
6	158.741	133089.7	964096.8
7	138.116	155271.3	878306.3
8	124.729	177453.0	830409.8
9	115.551	199634.6	804546.1
10	108.987	221816.2	792361.2
11	104.130	243997.8	789115.43
12	100.435	266179.4	791957.5
13	97.560	288361.1	799088.34
14	95.279	310542.7	809327.7
15	93.439	332724.3	821874.9
16	91.932	354905.9	836171.4
17	90.684	377087.5	851818.0
18	89.638	399269.2	868523.3
19	88.753	421450.8	886070.3
20	87.997	443632.4	904294.9
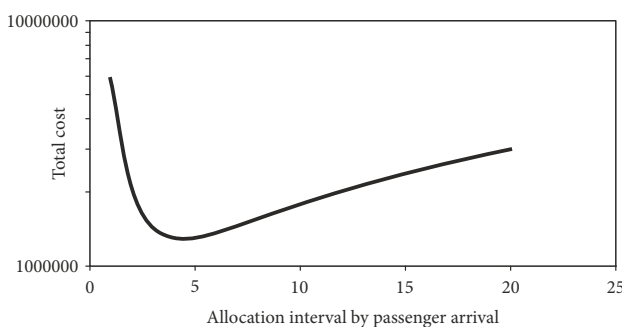
Figure 8. Total cost by allocation interval by passenger arrival

3.4. Model application result

We compare and analyse the traffic volume analysed based on the current bus traffic and the research model of this study. For each time zone, it was confirmed that the optimal number of dispatch vehicles was at the time when there were 7 vehicles in the time zone 4:00…6:00 hours, 11 vehicles in the time zone 14:00…16:00, and 11 vehicles in the time zone 20:00…22:00 hours (Figure 9).
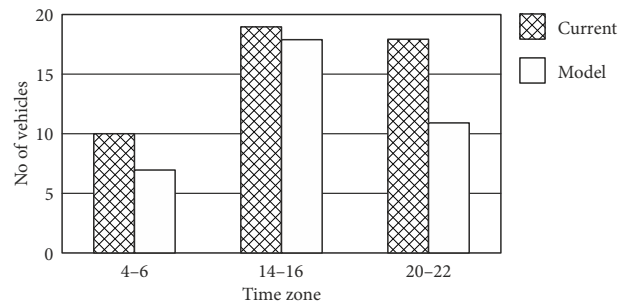

Figure 9. Comparison of current bus traffic and model applied traffic

In this way, we confirmed that we can determine the optimal dispatch interval for each time zone and day of the week.

The graph (Figure 9) compares the currently operating bus traffic by time zone and the optimal number of vehicles derived from the model application. Between 14:00…16:00 hours, it is shown that there is not any big difference from the result of applying 18 models of bus traffic at 19:00 o'clock. However, it is shown that bus traffic between 4:00…6:00 and 20:00…22:00 hours can be greatly reduced by applying the model. In particular, since the current bus traffic of the remaining time zones except for the early morning hours is set in accordance with the time zone having the maximum demand, more buses are operated than the necessary number of buses in the evening hours when the demand is reduced. Therefore, it is necessary to control flexible traffic through immediate demand reflection. By applying the model proposed in this study, bus traffic can be reduced drastically and the reduction of CHG emissions during bus operation and congestion of urban transportation network are expected to be reduced.

Conclusions

In this study, we designed a model, which can find the optimal bus dispatch interval based on Seoul bus fare cost model that is designed by Yu et al. (2010) and Park et al. (2008). The total cost consists of the Seoul bus operation cost and Seoul passenger time cost. Here, passenger time cost is divided into waiting time cost and travel time cost and the bus operation costs are divided into fixed costs and operating costs. In this model, the transaction data of Seoul Transportation Card is substituted to each cost item. As the number of buses increases, the time cost of the passengers decreases, but the cost of bus operation increases. Therefore, there is a dispatch interval with the total and minimum cost. It is concluded that the dispatch interval at this time is the optimal dispatch interval for each route in Seoul in each time slot. It is confirmed that the optimal number of buses is smaller than the current service and the cost reduction effect can be obtained with less traffic than that of the current.

As a result of applying the model, it is shown that the bus operation policy is inefficient in cost. Additionally, since more buses are currently being operated than

necessary, more CHGs are being emitted. If we increase the reality of the models presented in this study and assign actual usage details, more efficient bus operation can be achieved. In the case of Seoul, since the proportion of public transportation, especially buses, is significantly high, reduction level of air pollution of Seoul can be dramatic if buses are operated efficiently by applying the proposed model. In addition to the quantitative variables such as the amount of air pollution due to the bus operation, which is the amount of carbon emissions, it also can be utilized as a model, which considers passenger cost, bus operation cost and environmental cost into the total cost by adding the quantitative variable such as the amount of greenhouse emissions that's generated by the bus operation, which shows the level of air pollution due to the bus operation.

In addition, the research model can be applied not only to Seoul, but also to cities with similar problems in various countries around the world. Most countries in the world have similar structural problems with public transportation and have similar data. If we insert a variable that reflects the characteristics of each country in the proposed research model, you can make a more effective model.

Some limitations of this study are that the proposed model does not accurately reflect the current cost of bus operation in Seoul. Considering the changes in oil prices and other incidental costs and inflation rates, more realistic model designs can be possible. Additionally, given the data on CHG emissions and the cost of environmental pollution due to the bus distance, the effect of reducing environmental pollution can be expressed numerically. As the dispatch interval decreases, passengers benefit from their reduced time of use. However, the benefits of additional new passengers other than the existing passengers are not considered. If the additional passenger benefits are taken into account, the optimal system is expected to change. Therefore, in order to increase the reliability of the model constructed in this study, more attention should be paid to collecting the data on the variables that account for changes in passenger demand and explaining the degree of air pollution.

Acknowledgements

References

Ally, J.; Pryor, T. 2007. Life-cycle assessment of diesel, natural gas and hydrogen fuel cell bus transportation systems, *Journal of Power Sources* 170(2): 401–411.
https://doi.org/10.1016/j.jpowsour.2007.04.036

Bagchi, M.; White, P. R. 2005. The potential of public transport smart card data, *Transport Policy* 12(5): 464–474.
https://doi.org/10.1016/j.tranpol.2005.06.008

Ben-Akiva, M.; Lerman, S. R. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. The MIT Press. 384 p.

Bowman, L. A.; Turnquist, M. A. 1981. Service frequency, schedule reliability and passenger wait times at transit stops, *Transportation Research Part A: General* 15(6): 465–471.
https://doi.org/10.1016/0191-2607(81)90114-X

Ceder, A. 2002. Urban transit scheduling: framework, review and examples, *Journal of Urban Planning and Development* 128(4): 225–244.
https://doi.org/10.1061/(ASCE)0733-9488(2002)128:4(225)

Cesaroni, G.; Badaloni, C.; Porta, D.; Forastiere, F.; Perucci, C. A. 2008. Comparison between various indices of exposure to traffic-related air pollution and their impact on respiratory health in adults, *Occupational and Environmental Medicine* 65(10): 683–690. https://doi.org/10.1136/oem.2007.037846

Chester, M. V.; Horvath, A. 2009. Environmental assessment of passenger transportation should include infrastructure and supply chains, *Environmental Research Letters* 4(2): 1–8.
https://doi.org/10.1088/1748-9326/4/2/024008

Chester, M.; Pincetl, S.; Elizabeth, Z.; Eisenstein, W.; Matute, J. 2013. Infrastructure and automobile shifts: positioning transit to reduce life-cycle environmental impacts for urban sustainability goals, *Environmental Research Letters* 8(1): 1–10.
https://doi.org/10.1088/1748-9326/8/1/015041

Constantin, I.; Florian, M. 1995. Optimizing frequencies in a transit network: a nonlinear bi-level programming approach, *International Transactions in Operational Research* 2(2): 149–164. https://doi.org/10.1016/0969-6016(94)00023-M

Cooney, G.; Hawkins, T. R.; Marriott, J. 2013. Life cycle assessment of diesel and electric public transportation buses, *Journal of Industrial Ecology* 17(5): 689–699.
https://doi.org/10.1111/jiec.12024

Forbes, M. A.; Holt, J. N.; Watts, A. M. 1994. An exact algorithm for multiple depot bus scheduling, *European Journal of Operational Research* 72(1): 115–124.
https://doi.org/10.1016/0377-2217(94)90334-4

Herndon, S. C.; Shorter, J. H.; Zahniser, M. S.; Wormhoudt, J.; Nelson, D. D.; Demerjian, K. L.; Kolb, C. E. 2005. Real-time measurements of SO_2, H_2CO, and CH_4 emissions from in-use curbside passenger buses in New York City using a chase vehicle, *Environmental Science & Technology* 39(20): 7984–7990.
https://doi.org/10.1021/es0482942

Jansson, J. O. 1980. A simple bus line model for optimisation of service frequency and bus size, *Journal of Transport Economics and Policy* 14(1): 53–80.

Ko, J. H. 2009. *Strategies to Promote Green Car Supply in Seoul*. Korean Policy Report No 52. 22 p. (in Korean).

Manyika, J.; Chui, M.; Brown, B.; Bughin, J.; Dobbs, R.; Roxburgh, C.; Byers, A. H. 2011. *Big data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute. 156 p.

Messa, C. 2006. Comparison of emissions from light rail transit, electric commuter rail, and diesel multiple units, *Transportation Research Record: Journal of the Transportation Research Board* 1955: 26–33. https://doi.org/10.3141/1955-04

Ministry of Land, Infrastructure and Transport. 2011. *The 1st Sustainable National Transportation and Logistics Development Basic Plan*. Republic of Korea. Available from Internet: http://www.molit.go.kr (in Korean).

Mohring, H. 1972. Optimization and scale economies in urban bus transportation, *The American Economic Review* 62(4): 591–604.

Munizaga, M. A.; Palma, C. 2012. Estimation of a disaggregate multimodal public transport origin–destination matrix from

passive smartcard data from Santiago, Chile, *Transportation Research Part C: Emerging Technologies* 24: 9–18. https://doi.org/10.1016/j.trc.2012.01.007

Oldfield, R. H.; Bly, P. H. 1988. An analytic investigation of optimal bus size, *Transportation Research Part B: Methodological* 22(5): 319–337. https://doi.org/10.1016/0191-2615(88)90038-0

Park, J. S.; Lee, K. S. 2007. Analysis of traffic pattern exploration and traffic behavior in large traffic card transaction database, *Journal of Economic Geography* 10(1): 44–63. (in Korean).

Park, J.-S.; Oh, D.-K.; Kho, S.-Y. 2008. Estimation of operating cost and efficiency of the introduction of urban subway, *Journal of Korean Society of Transportation* 26(6): 113–122. (in Korean).

Puchalsky, C. 2005. Comparison of emissions from light rail transit and bus rapid transit, *Transportation Research Record: Journal of the Transportation Research Board* 1927: 31–37. https://doi.org/10.3141/1927-04

Salzborn, F. J. M. 1972. Optimum bus scheduling, *Transportation Science* 6(2): 137–148. https://doi.org/10.1287/trsc.6.2.137

Seoul Metro. 2017. *Website of the Seoul Metro*. Available from Internet: www.seoulmetro.co.kr/en

Tom, V. M.; Mohan, S. 2003. Transit route network design using frequency coded genetic algorithm, *Journal of Transportation Engineering* 129(2): 186–195.
https://doi.org/10.1061/(ASCE)0733-947X(2003)129:2(186)

Trépanier, M.; Morency, C.; Agard, B. 2009. Calculation of transit performance measures using smartcard data, *Journal of Public Transportation* 12(1): 79–96.
https://doi.org/10.5038/2375-0901.12.1.5

Trépanier, M.; Tranchant, N.; Chapleau, R. 2007. Individual trip destination estimation in a transit smart card automated fare collection system, *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations* 11(1): 1–14. https://doi.org/10.1080/15472450601122256

Utsunomiya, M.; Attanucci, J.; Wilson, N. 2006. Potential uses of transit smart card registration and transaction data to improve transit planning, *Transportation Research Record: Journal of the Transportation Research Board* 1971: 119–126. https://doi.org/10.3141/1971-16

Yu, B.; Yang, Z.; Yao, J. 2010. Genetic algorithm for bus frequency optimization, *Journal of Transportation Engineering* 136(6): 576–583.
https://doi.org/10.1061/(ASCE)TE.1943-5436.0000119

APPENDIX

Smart traffic card transaction data

pcard_no	run_depart_dtime	trans_id	transp_method_cd	traf_freq	bus_route_id	ride_transp_bizr_id	vehc_id	pcard_user_class_cd	fare_user_class_cd1	ride_dtime	ride_sta_id	alight_dtime	alight_sta_id	pasgr_num1	use_dist	use_seconds
1186678	2.0130E+13	18	115	0	11110005	111004400	111749424	1	1	1 2.013E+13	72239	2.013E+13	71613	1	3820	743
797763	2.0130E+13	30	115	0	11110005	111004400	111749420	7	7	1 2.013E+13	71613	2.013E+13	8952	1	4550	643
29860	2.0130E+13	13	115	0	11110005	111004400	111749426	1	1	1 2.013E+13	8000880	2.013E+13	71613	1	4446	709
152337	2.0130E+13	7	115	1	11110005	111004400	111749426	4	4	1 2.013E+13	8000880	2.013E+13	71198	1	8423	1333
535717	2.0130E+13	95	115	1	11110005	111004400	111749424	1	1	1 2.013E+13	71613	2.013E+13	72214	1	6062	806
829110	2.0130E+13	25	115	0	11110005	111004400	111749424	4	4	1 2.013E+13	72012	2.013E+13	72215	1	5206	784
189514	2.0130E+13	13	115	0	11110005	111004400	111749428	4	4	1 2.013E+13	8002702	2.013E+13	72217	1	13051	2343
842393	2.0130E+13	172	115	0	11110005	111004400	111749430	8	8	8 2.013E+13	8000880			1	0	0
46173	2.0130E+13	2	115	1	11110005	111004400	111749426	1	1	1 2.013E+13	7197	2.013E+13	72213	1	2067	373
406375	2.0130E+13	53	115	0	11110005	111004400	111749428	1	1	1 2.013E+13	72247	2.013E+13	72250	1	6093	1162
1191403	2.0130E+13	1	115	0	11110005	111004400	111749424	1	1	1 2.013E+13	72251	2.013E+13	7710	1	8778	1898
141709	2.0130E+13	7	115	0	11110005	111004400	111749430	1	1	1 2.013E+13	72241	2.013E+13	72246	1	1655	464
711282	2.0130E+13	59	115	1	11110005	111004400	111749424	6	6	1 2.013E+13	72208	2.013E+13	72268	1	9667	2153
1206511	2.0130E+13	75	115	0	11110005	111004400	111749428	1	1	1 2.013E+13	71196	2.013E+13	72214	1	3797	797
3840	2.0130E+13	13	115	1	11110005	111004400	111749426	4	4	1 2.013E+13	72251	2.013E+13	72270	1	11015	2428
868680	2.0130E+13	69	115	0	11110005	111004400	111749432	6	6	1 2.013E+13	72244	2.013E+13	8952	1	6690	1182
711494	2.0130E+13	70	115	1	11110005	111004400	111749424	6	6	1 2.013E+13	72259	2.013E+13	13199	1	9736	2228
842175	2.0130E+13	168	115	0	11110005	111004400	111749432	8	8	8 2.013E+13	72245			1	0	
1040098	2.0130E+13	2	115	0	11110005	111004400	111749432	7	7	7 2.013E+13	72245			1	0	
583282	2.0130E+13	59	115	1	11110005	111004400	111749426	1	1	1 2.013E+13	72255	2.013E+13	72269	1	7795	1608
67967	2.0130E+13	10	115	0	11110005	111004400	111749446	1	1	1 2.013E+13	13198	2.013E+13	72012	1	6951	1151
637803	2.0130E+13	66	115	1	11110005	111004400	111749428	1	1	1 2.013E+13	72208	2.013E+13	13199	1	13002	2419
287067	2.0130E+13	12	115	0	11110005	111004400	111749447	1	1	1 2.013E+13	72244	2.013E+13	71613	1	2140	334
377664	2.0130E+13	63	115	0	11110005	111004400	111749447	1	1	1 2.013E+13	72245	2.013E+13	72012	1	3029	523
810476	2.0130E+13	81	115	0	11110005	111004400	111749428	1	1	1 2.013E+13	71205	2.013E+13	8000884	1	6880	1248
508031	2.0130E+13	14	115	0	11110005	111004400	111749447	1	1	1 2.013E+13	72246	2.013E+13	71613	1	1422	210
896639	2.0130E+13	87	115	0	11110005	111004400	111749428	6	6	1 2.013E+13	72262	2.013E+13	72266	1	2006	383
289978	2.0130E+13	12	115	0	11110005	111004400	111749430	1	1	1 2.013E+13	72261	2.013E+13	72027	1	1824	308
544003	2.0130E+13	5	115	0	11110005	111004400	111749448	1	1	1 2.013E+13	72245	2.013E+13	72011	1	2348	405
191273	2.0130E+13	12	115	1	11110005	111004400	111749449	4	4	1 2.013E+13	72239	2.013E+13	72242	1	1238	242
295766	2.0130E+13	3	115	1	11110005	111004400	111749432	1	1	1 2.013E+13	72256	2.013E+13	72257	1	274	145
1158058	2.0130E+13	42	115	0	11110005	111004400	111749449	1	1	1 2.013E+13	72239	2.013E+13	72245	1	2101	490
824699	2.0130E+13	69	115	1	11110005	111004400	111749430	1	1	1 2.013E+13	72262	2.013E+13	72266	1	2006	412
639807	2.0130E+13	18	115	0	11110005	111004400	111749449	1	1	1 2.013E+13	72241	2.013E+13	72245	1	1358	277
73479	2.0130E+13	4	115	0	11110005	111004400	111749448	1	1	1 2.013E+13	74700	2.013E+13	72213	1	2565	466
972467	2.0130E+13	63	115	0	11110005	111004400	111749432	6	6	1 2.013E+13	72027	2.013E+13	72275	1	6765	1441