

## ORIGIN-DESTINATION-BASED PUBLIC TRANSPORT SERVICE GAP

Minhyuck Lee<sup>1</sup>, Hyunwoo Nam<sup>1</sup>, Chulmin Jun<sup>2,\*</sup>

<sup>1</sup>Dept. of Geoinformatics, University of Seoul, Korea, Ph.D students – (lmhll123, hwnam)@uos.ac.kr

<sup>2</sup>Dept. of Geoinformatics, University of Seoul, Korea, Professor - cmjun@uos.ac.kr

**KEY WORDS:** Public transport, Service gap, Supply-Demand, Origin-Destination, Spatial disparity

### ABSTRACT:

There are some studies that analyze public transport service gap by region from comparing supply and demand. However, due to data limitations, previous studies were limited in subdividing region-based service gap by Origin-Destination (O-D). This study analyzes the gaps of public transport services based on O-D, a micro spatial unit. The data used in this study include timetable of public transport and smart card data stored with transportation used records of individual users. The supply index presented in this study is based on O-D travel time considering for temporal fluctuation. And the demand index is explained in terms of actual traffic of O-D. The proposed methodology is applied to Seoul metropolitan city and the analysis for identifying service gap is conducted along major time periods of a day. Visualization is performed on some O-D pairs that require improvements in supply relative to demand. The areas where disparities in service exist were identified

### 1. INTRODUCTION

Studies that analyzed the gaps in public transport services using demand and supply can be found mainly in the field of transport geography. These studies provide indices that explain the demand and the supply of public transport on a regional basis, and analyze the disparities between the two indices by region (Currie, 2004). In this way, they identify the spatial disparity in public transport service and the influence factors that lead to such inequalities.

Unlike studies about region-based public transport service gap, studies on service disparities based on Origin-Destination (O-D) level are very scarce. This is largely attributed to data limitations. Most studies that evaluate public transport services by comparing demand and supply rely on data that are spatially aggregated. In general, provision of transport is explained in terms of the frequency and capacity of public transport in the area, density of transport facilities and so on, while demand is explained by the population of the socially disadvantaged class. Thus, there is a limit to analyzing service gaps on the basis of O-D with data used in existing studies.

Therefore, this study aims to analyze gaps in public transport services based on O-D level using micro data. The micro data used in this study include smart card data, which contain the record of ride/alight of individual users, and GIS data such as public transportation and road networks. The supply index is based on the travel time of O-D, and the demand index is explained by the actual traffic of O-D. The service gap represents the relative difference between the two indices.

The proposed methodology was applied to public transportation in Seoul. The O-D-based service gap was measured and analyzed along major time periods of a day to consider the characteristics of public transport supply and demand expressed fluidly depending on flow of time.

### 2. RELATED WORK

#### 2.1 Measuring public transport supply

Currie (2010) analyzed the public transport gaps between the supply level and the social demand in Melbourne, Australia, by census district. For the supply index, the study considered the space occupancy of public transport facilities within a census district and the frequency of operation by means of public transportation. Space occupancy of a public transport facility refers to the ratio of the public transport catchment area within a given census district. In the study, different catchment area was applied according to the means of public transportation such as bus, tram and urban railway.

Jaramillo et al. (2012), who conducted a study in Santiago de Cali, Colombia, further considered the capacity of public transport for the supply indicators proposed by Currie (2010). For the 22 districts covered in the study, the supply index for each district reflected the number of stops, capacity, and service frequency by means of public transportation. In addition, the supply index was sorted into absolute supply index, which accounts for the supply level to unit area of a given district, and relative provision index, which is calculated by dividing the absolute supply index by the population.

A travel time-based provision index was proposed in the study by Fransen et al. (2015) that analyzed the public transit service gap in the Flanders area of Belgium. Fransen et al. (2015) used the General Transit Feed Specification (GTFS), which allows network configuration based on public transport timetables. Instead of using existing supply indices, which only consider the aspects of public transport infrastructure within a given area, the study proposed a provision index accounting for the accessibility to a destination (Kaza, 2015). The provision index for a traffic analysis zone (TAZ) was computed as the number of major facilities that can be reached within a given travel time. Here, the major facilities correspond to schools, work places, hospitals, and so on.

Fayyaz et al. (2017) explained a supply index in terms of gravity-based accessibility. The gravity-based accessibility of a TAZ is the weighted average travel time, with the travel time to each zone calculated by taking the opportunity in each destination zone as the weighted value. Here, opportunity refers to the number of jobs that take into account the wage level of the destination zone. In other words, Fayyaz et al. (2017) provided a provision index of TAZ through accessibility, reflecting travel time and employment requirements. In addition, the proposed provision index used GTFS to reflect the travel time of public transportation, which varies according to the departure time (Farber et al., 2014; Farber et al., 2017).

## 2.2 Measuring public transport demand

In most related studies, a demand index is estimated relying on the sociodemographic statistical data (Currie, 2010; Jaramillo et al., 2012; Fransen et al., 2015). Factors inducing public transportation needs such as the population with vehicle ownership, the population of minors and the elderly, the number of students and workers, income level, etc. are considered in calculating the demand index for a TAZ. An index of demand is determined by using the principal component analysis (Jolliffe, 2002) based on the statistical values of the mentioned factors.

The service gap is explained by the difference between supply and demand indices. In general, the service gap is calculated by converting the values through a standard score (Z score) and minimum-maximum normalization because supply and demand indices have different dimensions. Quantitative interpretation is difficult since the service gap computed based on normalized values implies a relative difference between supply and demand.

Instead of the method mentioned above, Fayyaz et al. (2017) proposed an index of transport demand that considers the income level in a TAZ using the weighted average number of workers. Then, the service gap was explained by the product of supply index multiplied by demand index. The supply index corresponds to the travel time, and the demand index to the number of workers. This study explained that the service gap of a zone is large where workers with low income levels are

concentrated and the travel time to move to another zone is long.

The supply indices presented in recent studies incorporate the GTFS-based public transport operation schedule, and hence, reflect the characteristics of public transportation supply that varies with destination and time. This method allows analysis based on O-D level. On the other hand, the demand indices are estimated based on statistical data related to the less privileged population in a region. Similar to the supply of public transport, the demand also varies depending on destination and time. However, related studies have failed to provide a demand index reflecting these characteristics due to data limitations. This is a fundamental reason for analyzing the gaps in public transport services from a macro perspective.

## 3. METHODOLOGY

This study analyses the public transport service gap in Seoul metropolitan city with O-D level. The data used for the analysis are smart card data, timetable of public transportations (bus and urban railway), public transport and road networks, and so on. The service gap is computed as the difference between supply and demand in a specific time period. In this study, the index of public transport supply is based on the travel time of O-D, and the index of public transport demand indicates the actual traffic.

### 3.1 Study area and data

Seoul is the capital city of the Republic of Korea, with an area of about 605km<sup>2</sup> and a population of approximately 9.8 million. The public transportation of Seoul shows a modal split of 65% (bus 28%, urban railway 37%), with the number of daily average users reaching 7.2 million and 4.4 million for urban railway and bus, respectively. There are 9 urban railway lines in operation and the total length of the entire system is approximately 330km. There are approximately 420 bus lines and close to 7,500 buses in operation. Figure 1 shows the distribution of public transport stops about 12,000 in Seoul, including urban railway and bus.

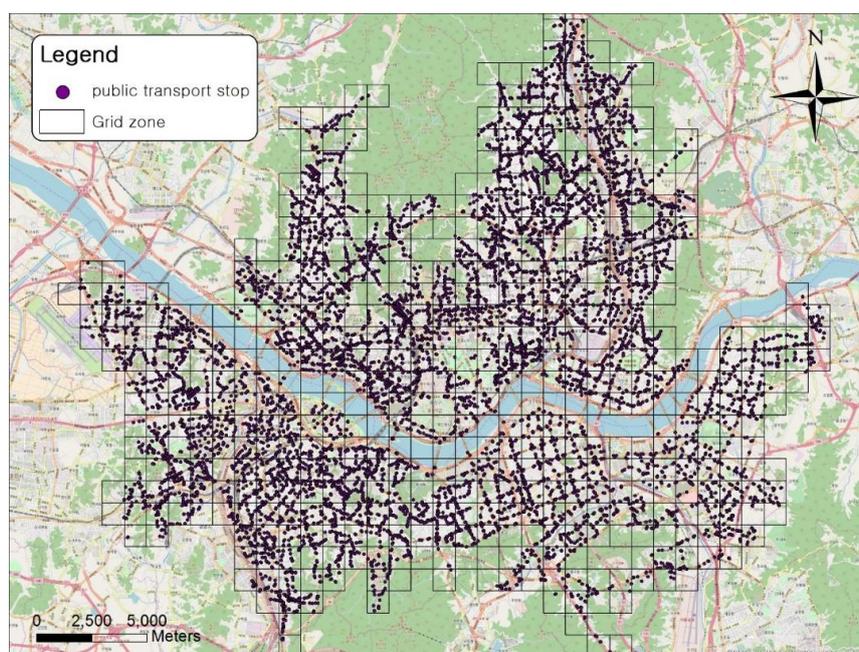


Figure 1. Public transport stops and grid unit traffic zones in Seoul

Data	Description	Source	Time range
Smart card data	Departure stop, arrival stop, departure time, arrival time, route taken, etc. by public transport users	Seoul TOPIS	2017.10.12. ~2017.10.18
Public transport network	Public transportation (urban railway, bus) stops (nodes) and routes (links)	Seoul TOPIS, Seoul Metropolitan Rapid Transit Corporation	2017.10
Public transport timetable	Departure/arrival Time by route, dispatch interval, etc.	Seoul TOPIS, Seoul Metropolitan Rapid Transit Corporation	2017.10.19
Road network	Nodes and links that make up public road network	ITS National Transport Information Center	2017.9

Table 1. Description and source of major data

As shown in Figure 1, the Seoul metropolitan area is divided with a 1km \* 1km grid in order to conduct a grid-based O-D analysis. When a stop level O-D analysis is adopted, the computation becomes overwhelming and the results too complicated, making it difficult to interpret the results. In addition, an analysis with O-D based on administrative district will yield rather macroscopic results in comparison to the detailed data used. Therefore, this study constructed grid-based transport zones in order to reduce the complication in computation and interpretation and enhance the usage of micro data. The size of the grid was set to 1km\*1km by applying the general 500m pedestrian mobility range to the centroid of the grid. The total number of grids is 548, and the number of possible O-D pairs is about 300,000.

The data used for calculating the travel time and traffic of O-D are shown in Table 1. To computing the travel time of O-D, this study used the RAPTOR (Round-bAsed Public Transit Optimized Router) algorithm which finds optimal route which has minimum travel time based on the timetables of public transport (Delling et al., 2014). Based on a user's time of arrival at the departure stop, minimum travel time is considered as the sum of waiting time, in-vehicle time, transfer time and transfer penalty. Here, transfer penalty is the time converted physiological burden due to making transfer.

Record such as the users' ride/alight stops and time, routes taken, and so on are stored in the smart cards (Pelletier et al., 2011). Data is generated when a user tags the card to a reader when ride/alight a public transportation. Because the smart card use rate in among public transportation users in Seoul is 99% and 100% for bus and urban railway, respectively, data for complete enumeration of public transport usage can be obtained from smart card data. This means that traffic information on a stop-to-stop level can be aggregated by minute time unit. Since the timetable used in computing the index of public transport supply was the weekday timetable, traffic also extracted from only weekday smart card.

### 3.2 Index of public transport supply (IPTS)

The IPTS between departure zone  $p$  and arrival zone  $q$  ( $S_{pq,T}$ ) in a specific time period ( $T$ ) is the difference between the average travel time of  $p-q$  ( $A_{pq,T}$ ) and the average travel of all O-D ( $\bar{A}_{X,T}$ ) that has similar distance difference with that of  $p-q$  ( $d(p,q)$ ). In other words, it is the deviation from the mean travel

time of a set (based on distance) to which  $p-q$  belongs. The distance of O-D is calculated as the shortest distance based on the road network.

In the time period of  $T$ , the average travel time of  $p-q$ ,  $A_{pq,T}$ , is defined as Eq. 1. Let's assume that  $n$  routes were derived when searching for the minimum travel time route of  $p-q$  at a certain time interval in  $T$ .  $t_{pq,k}$  is the  $k$ th travel time of  $p-q$  (in-vehicle time + transfer time + transfer penalty) in the found routes, and  $w_{pq,k}$  is the waiting time between the  $k$ -1th route and  $k$ th route. When  $k=1$ , the waiting time is the time between the start of  $T$  and first route. However, only half of the waiting time is applied in the computation since the user's waiting time cannot be determined. In  $T$ , the IPTS of  $p-q$  ( $S_{pq,T}$ ) is shown in Eq. 2. If  $X$  is the set inclusive of all  $i-j$  with shortest distance similar to  $d(p,q)$ ,  $\bar{A}_{X,T}$  is the average travel time of set  $X$  in  $T$ .

$$A_{pq,T} = \frac{1}{n} \sum_{k=1}^n \left( \frac{w_{pq,k}}{2} + t_{pq,k} \right) \quad (1)$$

$$S_{pq,T} = \bar{A}_{X,T} - A_{pq,T} \quad (2)$$

$$X = \{(i,j) | d(i,j) \approx d(p,q)\}$$

Figure 2 (a) shows the optimal routes of  $p-q$  (198-370) from 6 AM to 7 AM at 10 minutes intervals. Regarding the optimal routes by query time, the ride time at the departure stop (located in the departure zone  $p$ ) and the travel time after boarding are generated. (b) shows the computed waiting time until the  $k$ th vehicle using the ride time of  $k$ -1th vehicle (Prev\_Ride\_Time). In result, the average travel time of  $p-q$  between 6 AM to 7 AM ( $T$ ) is calculated as 32 minutes.

Figure 3 shows the distribution of all O-D pairs according to shortest distance and average travel time in the specific time period of  $T$ . The shortest distance was computed in km and, as shown in Eq. 3, the set of similar distance was classified at intervals of 1km. The average travel time was computed in minutes. One point indicates the number of O-D with that shortest distance and average travel time. The darker the point, the greater the distribution of O-D. Examining the O-D pairs with shortest distance of 20km, many O-D pairs are distributed around average travel time of about 80 minutes.

$$X_l = \{(i,j) | [d(i,j) + 0.5] = l, l \in \mathbb{Z} \text{ (integer)}\} \quad (3)$$

Assuming Figure 3 accounts for the time span from 6 AM to 7 AM and that  $p-q$  belongs to the 10km O-D set, then the average travel time of the 10km O-D set is approximately 47 minutes. Because the average travel time of  $p-q$  is 32 minutes and the

average travel time of the set is 47 minutes, the IPTS of  $p-q$  is +15 minutes. Accordingly, it can be interpreted that the level of public transport supply of  $p-q$  is 15 minutes faster compared to the average time required to travel the distance similar to  $p-q$ .

	Origin	Destination	Query_Time	Ride_Time	Travel_Time	Transfer_Count
1	198	370	2017-10-19 06:00:00,000	2017-10-19 06:06:20,000	25	1
2	198	370	2017-10-19 06:10:00,000	2017-10-19 06:18:20,000	24	1
3	198	370	2017-10-19 06:20:00,000	2017-10-19 06:26:00,000	36	1
4	198	370	2017-10-19 06:30:00,000	2017-10-19 06:36:20,000	26	1
5	198	370	2017-10-19 06:40:00,000	2017-10-19 06:47:20,000	30	1
6	198	370	2017-10-19 06:50:00,000	2017-10-19 06:57:20,000	27	1
7	198	370	2017-10-19 07:00:00,000	2017-10-19 07:07:20,000	24	1

(a)

	Origin	Destination	Prev_Ride_Time	Ride_Time	Waiting_Time	Travel_Time
1	198	370	NULL	2017-10-19 06:06:20,000	6	25
2	198	370	2017-10-19 06:06:20,000	2017-10-19 06:18:20,000	12	24
3	198	370	2017-10-19 06:18:20,000	2017-10-19 06:26:00,000	8	36
4	198	370	2017-10-19 06:26:00,000	2017-10-19 06:36:20,000	10	26
5	198	370	2017-10-19 06:36:20,000	2017-10-19 06:47:20,000	11	30
6	198	370	2017-10-19 06:47:20,000	2017-10-19 06:57:20,000	10	27
7	198	370	2017-10-19 06:57:20,000	2017-10-19 07:07:20,000	10	24

(b)

Figure 2. Example of minimum travel time route search result (a) and wait time calculation (b)

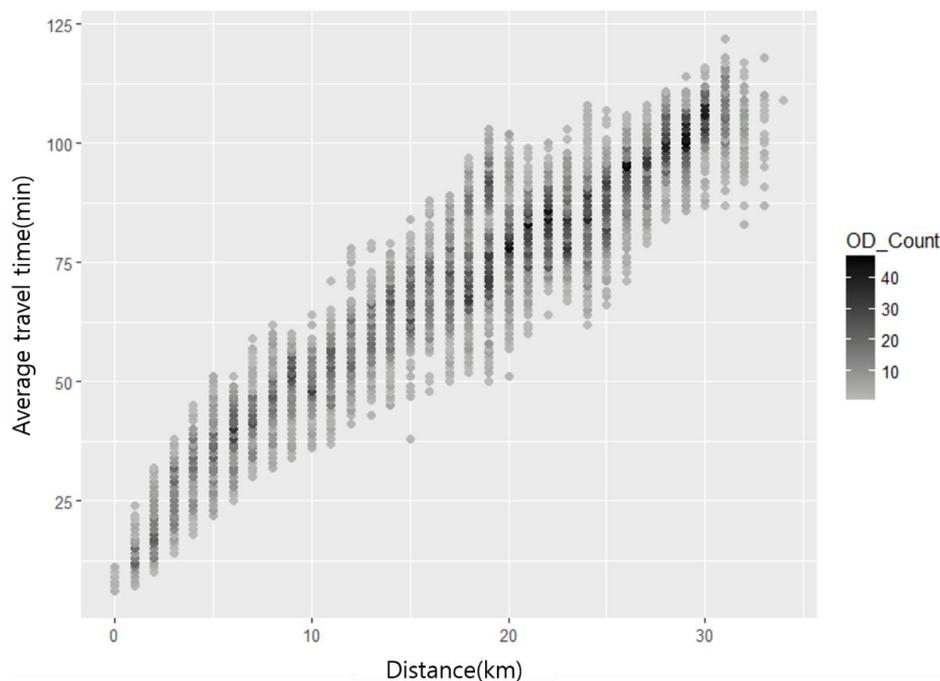


Figure 3. Distribution of all O-D pairs according to shortest distance and average travel time

### 3.3 Index of public transport demand (IPTD) and service gap (SG)

The IPTD of  $p-q$  ( $D_{p,q,T}$ ) is the total traffic that rided a vehicle at the  $p$  zone and alighted at the  $q$  zone. Alighting for transfer was not counted, only the cases where the initial boarding stop and the final alighting stop matched  $p-q$ .

The service gap of  $p-q$  ( $SG_{p,q,T}$ ) is defined as Eq. 4. Service gap is determined by the level of provision and needs. However, since the dimension of IPTS is time and the dimension of IPTD is traffic (number of people), absolute comparison of two indicators is impossible. Taking the stance that a relatively high level of provision needs to be arranged at O-D with high demand, this study computed the service gap by converting the

two indices into a standard score (Z score) that has a distribution with a mean of 0 and a standard deviation of 1.

$$SG_{pq,T} = Z(S_{pq,T}) - Z(D_{pq,T}) \quad (4)$$

#### 4. RESULT

O-D-based public transport service gap in Seoul was analyzed along three different time periods, AM and PM peak hours (07-09, 18-20) and off-peak hours (13-15). Analysis was conducted on 1,000 O-D pairs that show high traffic at each time period while excluding O-D pairs that are of walking distance (distance < 1.5km). The minimum travel time routes were investigated at 10 minutes interval and the set of similar distance was classified at 1km interval. As for O-D pairs that fall in the upper 2% of

insufficient provision (service gap < -2), visualization was conducted to examine the intuitive pattern.

#### 4.1 Service gap

Figure 4 shows the O-D-based IPTS and IPTD by time period. The IPTS forms a normal distribution since they are deviations from the average travel time of the set. The range of supply index by time period was -23 minutes to +17 minutes during morning peak hours, -15 minutes to +14 minutes during off-peak hours, and -17 minutes to +20 minutes during evening peak hours. The standard deviations of the IPTS by time period was 5.2 minutes for morning peak hours, 4.0 minutes for off-peak hours, and 4.6 minutes for evening peak hours. This means that the supply level difference between the O-Ds is relatively lower during off-peak hours than peak hours.

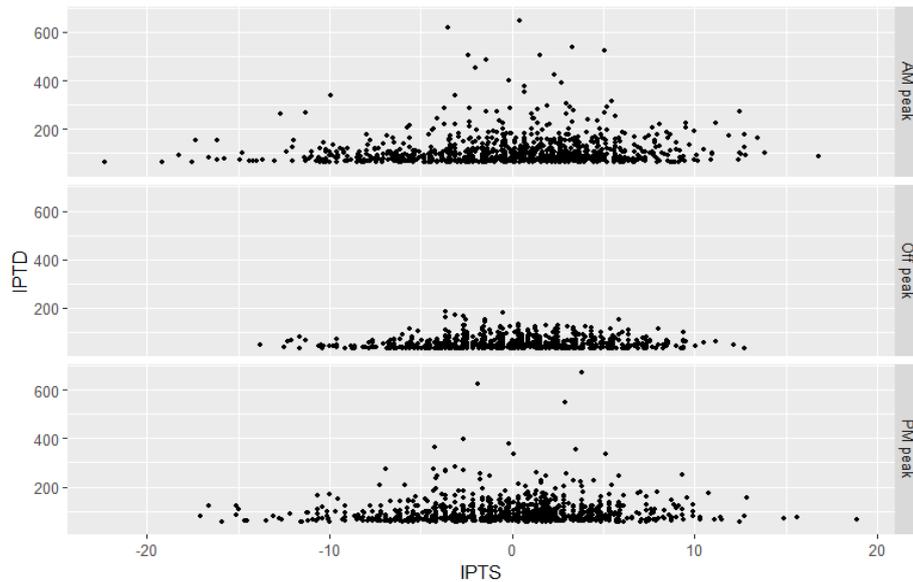


Figure 4. O-D-based IPTS and IPTD by time period

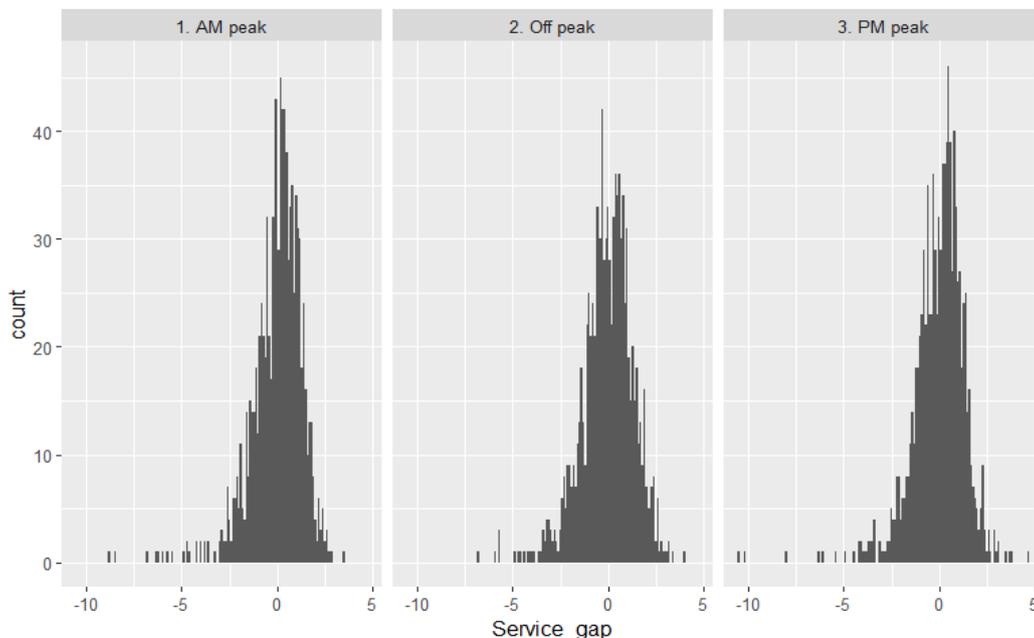


Figure 5. Histogram of service gap by time period

Sort		AM peak	Off-peak	PM peak
		# of O-Ds: 67	# of O-Ds: 75	# of O-Ds: 68
Z(IPTS)	Min	-4.26	-3.42	-3.71
	Average	-1.20	-0.93	-1.33
	Max	1.05	1.45	1.11
Z(IPTD)	Min	-0.72	-0.75	-0.77
	Average	2.12	2.07	1.95
	Max	8.57	5.86	11.05
Service gap	Min	8.84	-6.76	-10.50
	Average	-3.32	-3	-3.29
	Max	-2.04	-2.01	-2.03

Table 2. Descriptive statistics for O-D pairs with service gap in the lower 2%

Unlike the IPTS, the IPTD do not show a normal distribution since they are the traffic in O-D pairs. None of the O-D showed traffic above 200 during the off-peak hours, but there were O-D pairs that indicated traffic over 600 people during peak hours. This can be read as a result reflecting the commuting traffic.

Figure 5 shows the histogram of service gap by time period. All three time period presented O-D pairs with service gap below – 5, which can be read as the effect of demand index. The distribution of IPTD shows a positively skewed form with a large difference between the upper traffic and the mode. Therefore, for O-D pairs in the high traffic group, the service gap will yield low values because the demand are relatively higher even if the supply are high.

Table 2 shows the descriptive statistics for O-D pairs with service gap in the lower 2%. As discussed earlier, the lower 2% of the service gap corresponds to the upper 2% of insufficient supply to demand. In all time period, the average supply level of O-D pairs was at the lower 16%. Demand level was at the higher 2% on average, and a great gap was found between the average demand and the maximum demand during the peak

hours. Examining the minimum value of the service gap in each time period, it was found that their absolute values are very similar to the maximum demand values. This means that the service gap of the corresponding O-D is determined by a much higher demand regardless of the supply level.

#### 4.2 Spatial disparity

Figure 6 is a visualization of O-D pairs corresponding to the lower 2% of the service gap during the morning peak hours. During the morning peak hours, the O-D pairs heading for the major employment area indicate shortage of provision. The O-D pairs can be clustered into four groups (C1, C2, C3, C4) based on the destination/arrival area. Some of the O-D pairs heading for each destination area have urban railway, showing a supply level in the higher 16% (see Table 2). However, given the nature of morning peak hours, the demand is concentrated beyond the level of supply in these routes, resulting in service disparity. The O-D pairs from the western part of Seoul to C2 area in one of the most congested zones which has been covered in major domestic media.

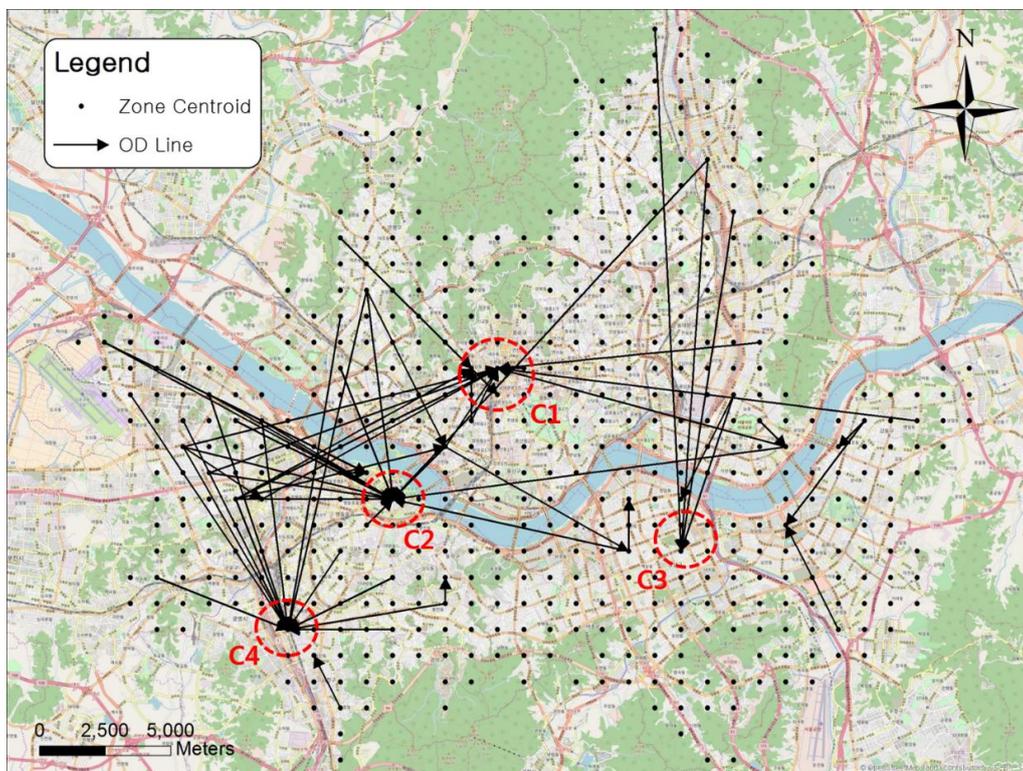


Figure 6. [AM peak] Service disparity O-D

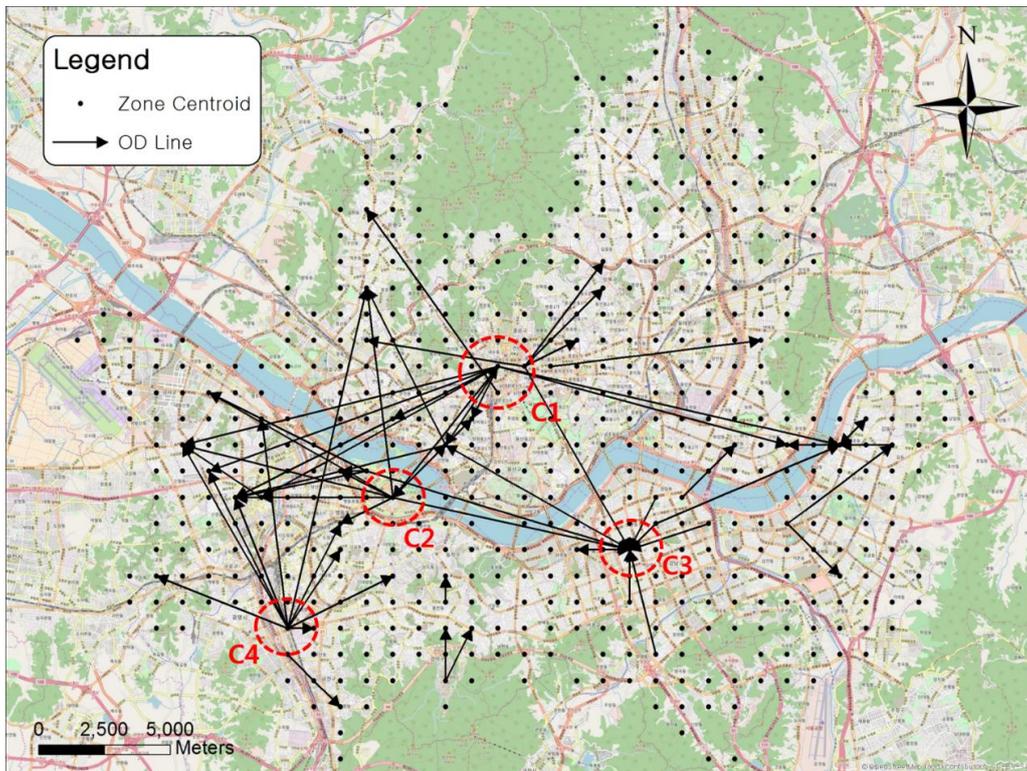


Figure 7. [PM peak] Service disparity O-D

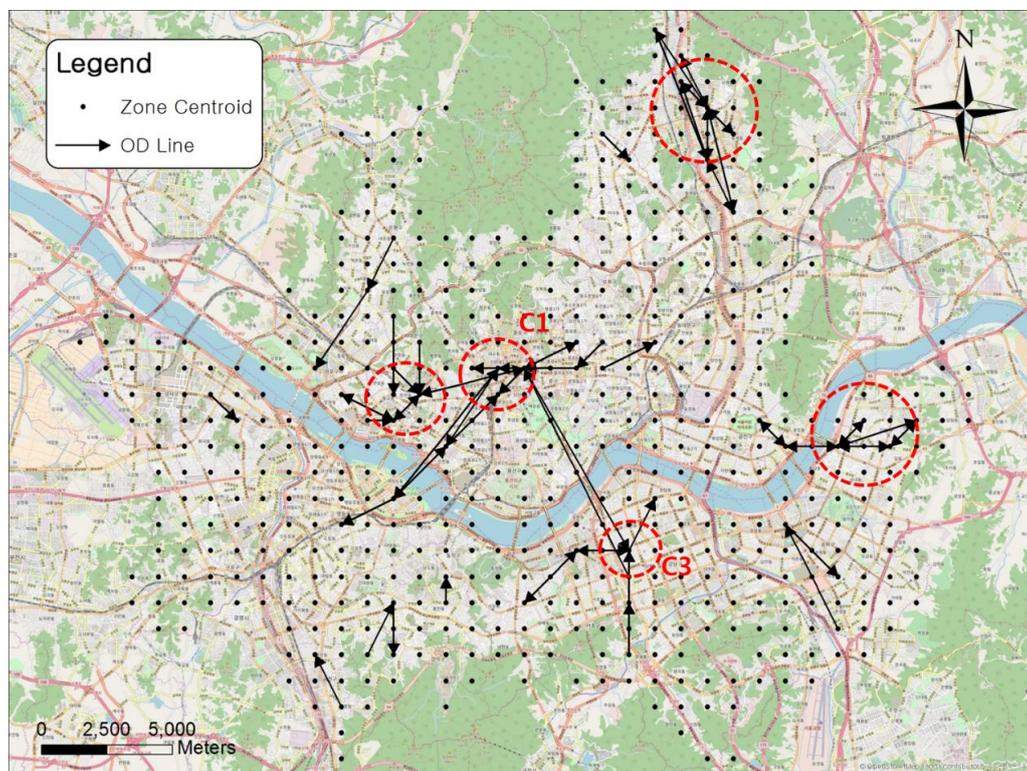


Figure 8. [Off-peak] Service disparity O-D

Figure 7 shows the O-D pairs corresponding to the lower 2% of the service gap during the evening peak hours. Contrary to the morning peak hours, the employment areas have become the departure areas with traffic dispersing away from them. Of the O-D pairs that showed insufficient supply during morning peak hours, 32 indicated service disparity even during evening peak

hours. The C3 area showed traffic concentration even during the evening peak hours. This area is Gangnam, which serves as the downtown of Seoul and is packed with amusement facilities and shopping centers. Therefore, it is understood that demand concentrated in this area as people enjoy their leisure activities after work.

Figure 8 shows the service disparity O-D pairs during the off-peak hours. It was found that O-D pairs that connect C1 and C3 areas, which have high volume of traffic all the time, need improvement in supply even during off-peak hours. In addition, disparity between supply and demand was found in some O-D pairs in college congested area near the C1 and the suburbs of Seoul.

## 5. CONCLUSIONS

In this study, we analyzed the gaps in public transport services on the basis of O-D from examining the disparity between supply and demand. As an index of public transport supply, we proposed relative travel time reflecting temporal fluctuation. As an index of public transport demand, we used the actual volume of traffic extracted from smart card data. Micro-grid space was applied to the transport zone composing O-D. Service gap in major time periods during a day was analyzed. And the findings were visualized to identify major areas where service disparity is occurring.

During peak hours, the O-D pairs found to be in the upper 2% of supply shortage to demand corresponded main to O-D pairs connecting the employment hubs and high density residential areas. In most cases, the direction of O-D in the morning and the direction of O-D in the evening were opposite, but the Gangnam area, which is the second downtown of Seoul, served as the arrival zone both in the morning and in the evening. It was confirmed that public transport supply needs to be improved for O-D connecting the Gangnam area even during off-peak hours. The public transport infrastructure is well-established in the Gangnam area. However, there is high demand for public transport services in this area since it is the central part of Seoul that fulfills various functions such as business, shopping, and residence, etc.

Analyzing public transport service gap with fine O-D level has its advantage in that specific O-D pairs that require improvements can be suggested. However, this study calculated the service gap through relative comparison rather than absolute comparison between supply and demand. Therefore, numerical interpretation of the calculated service gap is limited. In future studies, it is necessary to provide a supply index considering the capacity to quantitatively analyze the difference between the two indices.

## ACKNOWLEDGEMENTS

This research was supported by a grant (18CTAP-C133228-02) from Technology Advancement Research Program (TARP) funded by Ministry of Land, Infrastructure and Transport of Korean government

## REFERENCES

- Currie, G. (2004). Gap analysis of public transport needs: measuring spatial distribution of public transport needs and identifying gaps in the quality of public transport provision. *Transportation Research Record: Journal of the Transportation Research Board*, (1895), 137-146.
- Currie, G. (2010). Quantifying spatial gaps in public transport supply based on social needs. *Journal of Transport Geography*, 18(1), 31-41.
- Delling, D., Pajor, T., & Werneck, R. F. (2014). Round-based public transit routing. *Transportation Science*, 49(3), 591-604.
- Farber, S., Morang, M. Z., & Widener, M. J. (2014). Temporal variability in transit-based accessibility to supermarkets. *Applied Geography*, 53, 149-159.
- Farber, S., & Fu, L. (2017). Dynamic public transit accessibility using travel time cubes: Comparing the effects of infrastructure (dis) investments over time. *Computers, Environment and Urban Systems*, 62, 30-40.
- Fayyaz, S. K., Liu, X. C., & Porter, R. J. (2017). Dynamic transit accessibility and transit gap causality analysis. *Journal of Transport Geography*, 59, 27-39.
- Fransen, K., Neutens, T., Farber, S., De Maeyer, P., Deruyter, G., & Witlox, F. (2015). Identifying public transport gaps using time-dependent accessibility levels. *Journal of Transport Geography*, 48, 176-187.
- Jaramillo, C., Lizárraga, C., & Grindlay, A. L. (2012). Spatial disparity in transport social needs and public transport provision in Santiago de Cali (Colombia). *Journal of Transport Geography*, 24, 340-357.
- Jolliffe, I. T. (2002). *Principal Component Analysis*, Ed Springer. New York, US.
- Kaza, N. (2015). Time dependent accessibility. *Journal of Urban Management*, 4(1), 24-39.
- Pelletier, M. P., Trépanier, M., & Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557-568.

*Revised August 2018*