Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2018) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

World Conference on Transport Research - WCTR 2019 Mumbai 26-31 May 2019

Bus Passenger Volume and Origin-Destination Based on Field Surveys Using a Wi-Fi Scanner

Arief Hidayat[a,b,*], Shintaro Terabe[a], Hideki Yaginuma[a]

[a]*Urban and Transportation Planning Laboratory, Department of Civil Engineering, Tokyo University of Science, 2641 Yamazaki, Noda 278-8510, Japan*
[b]*Universitas Teknologi Sulawesi, Jalan Talasalapang No. 51 Makassar 90221, Makassar, Indonesia*

Abstract

Currently, transport survey methods are very diverse. Transportation data retrieval using information technology, such as Bluetooth, Wi-Fi, and smartcards, is prominent. This study aims to obtain media access control (MAC) addresses of individual bus passengers by using a Wi-Fi scanner. This Wi-Fi scanner is capable of engaging a probe request mode to capture MAC addresses from mobile devices or other Wi-Fi-enabled modalities without connecting to the internet. This study also describes a new data cleaning procedure that is used to characterize bus passenger volume and travel trends using a combination of MAC address and GPS data. The approach developed in the proposed study is capable of producing outputs, such as an origin-destination (OD) matrix and passenger volume for a bus route section. A comparison between passenger volumes obtained from the Wi-Fi data processing procedure and the data obtained using the ground truth procedure indicates the number of passengers determined using the Wi-Fi data acquisition and processing procedure is less than the number of passengers determined using the ground truth procedure.

Keywords: Wi-Fi Scanner; Passenger Volume; Origin-Destination; Bus Stop; Procedure; MAC address

1. Introduction

Nowadays, the development of transportation in the metropolitan city is very rapid. Owing to the recent developments in technology for metropolitan cities, the use of technology for transportation surveys is also advancing. Currently, smartcards, loop detectors, phone signals, wireless systems, and image processing techniques are used in origin-destination (OD) surveys. However, technologies used for transportation surveys all have distinct shortcomings

* Corresponding author. Tel.: +81-4-7124-1501(Ext: 4058); fax: +81-4-7123-9766.
E-mail address: ariefhidayat06@hotmail.com

and advantages. Loop detectors are widely used in developed countries to survey large volumes of vehicles. However, loop detectors cannot quantify the number of passengers in a vehicle and surveying technologies that are capable of quantifying the number of passengers in a vehicle is expensive. Smartcards, sensor systems, image processing techniques, and wireless devices are useful for surveying vehicles and passengers but similar to other tools. This equipment must also operate in a static state.

Otherwise, in sub-urban cities or small cities for origin-destination (OD) data retrieval collection is still entirely conventional. Conventional transportation surveys still use a manual surveying method, such as a vehicle, passenger, pedestrian, or cyclist. A manual survey involves manual counting by an individual within a specified period. Manual surveys require many surveyors, expensive, and not suitable compared to surveying developments in the big data era such as questioners, interview and manual counting. The small size of the city and adequate funding are the reason to still carry out conventional methods in terms of OD data collection. Therefore, it is needed a technology that is cheaper, easier, and provide a long time.

In recent years, Wi-Fi and Bluetooth technology have developed significantly everywhere in developed and developing countries. These are useful not only in computer science but also for transportation. The development of these technologies for data acquisition and processing is rapid. Wi-Fi and Bluetooth are cheaper and more convenient for the long term period of the survey. Wi-Fi and Bluetooth obtain MAC addresses with a high and low-frequency from Wi-Fi or Bluetooth-enabled devices. A media access control (MAC) address is a unique device identifier that consists of particular numbers and letters specific to a device with Wi-Fi or Bluetooth capability (Cunche, 2014; Dunlap et al., 2016; Musa and Eriksson, 2012; Vanhoef et al., 2016). Additionally, people at some places or regions do not use Bluetooth only for the headset on smartphone or printer devices. Surveying using Bluetooth requires a fixed survey tool in a static position for data acquisition. Therefore, Wi-Fi is more useful than Bluetooth and it is ubiquitous, whether it is a smartphone, tablet, laptop, vehicle, building, or other Wi-Fi-enabled devices. Wi-Fi-enabled devices can operate in either a static or dynamic states to obtain MAC address data. Wi-Fi technology is one of a solution to retrieve transportation data for sub-urban cities or small cities.

This study attempts to use Wi-Fi equipment to obtain OD data from passengers using bus services. The authors used Wi-Fi equipment in probe request mode to capture and collect the media access control (MAC) addresses of passenger devices while the bus service was traveling its route for a day. Wi-Fi offers two functional modes, an infrastructure mode, and a probe request mode. These terms will be explained in the main of “general Wi-Fi as an infrastructure mode and probe request mode section.” The MAC address is captured in a log file by Wi-Fi equipment. Along with the process of retrieving MAC addresses, the survey captures ground truth data for real bus passenger OD data.

This paper investigation employs a Wi-Fi scanner installed on a passenger bus in a small town called Obuse, Japan. Obuse bus is a small bus which operates around in Obuse city. This experiment to obtain the MAC address data from bus passengers’ Wi-Fi-enabled devices as well as Wi-Fi-enabled devices within the vicinity of the bus. A data cleaning procedure was used on the raw data to determine the number of bus passengers dynamically. The Wi-Fi scanner equipment is compact, inexpensive, energy efficient, and was manufactured by the Urban and Transportation Laboratory at the Tokyo University of Science (Hidayat et al., 2018a, 2018b, 2017a, 2017b; Terabe et al., 2017). The approach developed in this current investigation is capable of producing outputs such as the origin-destination (OD) matrix and passenger volume for a bus route section. A comparison between passenger volumes obtained from the Wi-Fi data processing procedure and the data obtained using the ground truth procedure indicates the number of passengers determined using the Wi-Fi data and processing procedure is less than the number of passengers determined using the ground truth procedure.

2. General Wi-Fi as an Infrastructure Mode and Probe Request Mode

Before collecting data, it better to understand this research equipment. Wi-Fi technology is based on IEEE 802.11 standards (including 802.11a, 802.11b, 802.11g, and 802.11n) (Cisco, 2008; Najafi et al., 2014). Wi-Fi is a popular modality for providing users with wireless internet access (Xu et al., 2013). The most common mode of operation for 802.11 is called *infrastructure mode*, where mobile devices or wireless access points communicate with other wireless access points and wired networks (typically Ethernet). Access points bridge traffic between wireless access points by referencing a destination address in the 802.11 frames (Sridhar, 2008). Most wireless devices, such as smartphones, tablets, routers, laptops, etc., are configured for infrastructure mode. A smartphone can be identified by its international mobile equipment identity (IMEI) number or MAC address. An IMEI number is sent when a mobile phone registers with a network, whereas a MAC address is attached to every data packet sent by a Wi-Fi enabled device. MAC addresses are designed to be persistent and globally unique (Martin et al., 2017). A MAC address is a 48-bit number used to identify a network interface (Cunche, 2014). Smartphone Wi-Fi is designed to periodically transmit probe requests to identify known access points (Matte, 2017; Yaik et al., 2016). Probe requests are active scans by mobile devices (Sun et al., 2017; Verbree et al., 2013), and the probe request content includes the sender's MAC address (Musa and Eriksson, 2012). A Wi-Fi scanner can load all MAC address data into a single log file. This tool can access MAC addresses without connecting to the internet and is capable of passively scanning devices to collect data. A Wi-Fi scanner, operating in *probe request mode*, can also be used to obtain MAC addresses that are operating in infrastructure mode. This survey used a Wi-Fi scanner as a probe request, installed on a bus, to collect MAC address data from bus passengers' Wi-Fi-enabled devices or Wi-Fi-enabled devices within the vicinity of the bus.

3. Related Works

To understand Wi-Fi in the context of transportation surveys, Bluetooth, and Wi-Fi technologies are reviewed below. The previous use of these technologies and their reported results are reviewed to understand the purpose of Wi-Fi for OD surveys and to build on these technologies to further advance transportation survey capabilities. There are several studies about Bluetooth and Wi-Fi for the pedestrian. Bluetooth is used to measure pedestrian activity with a detection rate of 2% is lower than the detection rate of Wi-Fi. Wi-Fi data is noisier due to the higher detection rate, but can also be more informative (Lesani and Moreno, 2016). Previous research on pedestrian activity demonstrated that Bluetooth could be used to determine pedestrian flow and density, but is less accurate compared to ground truth data (Schauer et al., 2014). However, the results were shown to vary depending on the location and time of the survey. Filtering data to isolate pedestrian activity uses the calculated speed to discriminate between pedestrians, cyclists, and cars (Abedi et al., 2015). Another investigation included a waiting time filter as one of the Bluetooth data filters that is capable of determining pedestrian flow more accurately (Kurkcu and Ozbay, 2017).

Bluetooth applications for determining vehicle travel time can be used to detect abnormal traffic. The index characteristic for travel time used a “min,” “max,” “average,” and “medium” scale (Namaki Araghi et al., 2015). Other research focuses on OD in urban areas. This approach uses MAC address data and calculates the travel time and speed of vehicles based on the “start” and “end” of a recorded MAC address (Khliefat and Shatnawi, 2017). The use of Bluetooth can also be applied to road surveys, and previous investigations have demonstrated the importance of the MAC address data cleaning process for the investigation of transport activity on roads using “entrance” and “exit” specifications (Abbott-jard et al., 2013). This process records the position of a MAC address at the beginning (entrance) and end (exit) of a probe by a Bluetooth scanner tool. This method also used a timestamp parameter to calculate the travel time of each vehicle either in a public terminal space or on the highway (Shlayan et al., 2016).

Bluetooth and Wi-Fi research is strongly associated with time and speed variables for static and mobile scanners. The filtering process, based on time variables, is vital for determining whether or not Bluetooth is useful for a transportation survey (Abedi et al., 2015; Erkan and Hastemoglu, 2016; Filgueiras et al., 2014; Young, 2012). Time analysis is used to determine the OD matrix based on travel time (J.Barceló et al., 2010). Other investigations (Namaki Araghi et al., 2015; Porter et al., 2013; Pourhassan, 2016) described the calculated travel time between multiple Bluetooth and Wi-Fi antenna sensors. Furthermore, Araghi and colleagues used a combination of methods to analyze travel time (Araghi et al., 2015). A previous investigation conducted by Romancyshyn and colleagues describes the use of Bluetooth to determine travel time, and delays in urban areas demonstrated that weather significantly affects travel time, and demonstrated that Bluetooth could be used to monitor vehicle activity accurately (Romancyshyn, 2016). The difference between Bluetooth and Wi-Fi has also been examined and determined that Wi-Fi has a more extensive operating range of 10 to 100 meters (Ferro and Potorti, 2005). Wi-Fi can operate using two frequency bands (2.5Ghz and 5Ghz), while Bluetooth can only operate at 2.5Ghz.

4. Methodology

4.1. Field Survey

This case study was conducted in Obuse, Nagano Prefecture, Japan, in October 2017. Obuse is one of a small city in Japan. Obuse is tourism destination such as temple, food, hot spring, and traditional tourism spot.

Obuse has a hop-on-off tour bus, named the Obuse Romango Bus (Figure 1). This bus is used for tourism purposes and can be used for a full day with only one payment. The data was collected twice such as Wi-Fi data and ground truth data to observe the different trends of data in same, bus and same day.

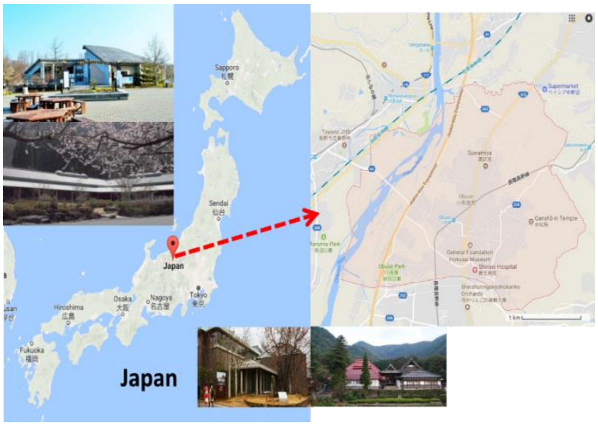


Fig 1. Obuse orientation map

4.2. Bus Specification

The Romango bus is a medium-sized bus with a maximum number of 25 passengers. The bus has nine bus stops along its route (Figure 2) and seven route circulations based on the timetable. This route circulation is referred to as circulation number (CN). A complete bus circulation is defined as a trip from bus stop no.1 to bus stop no.9 and a return to bus stop No.1. The bus route operates between 9:50 AM and 5:10 PM and the busiest bus stop is no.2, due to the Obuse train station.

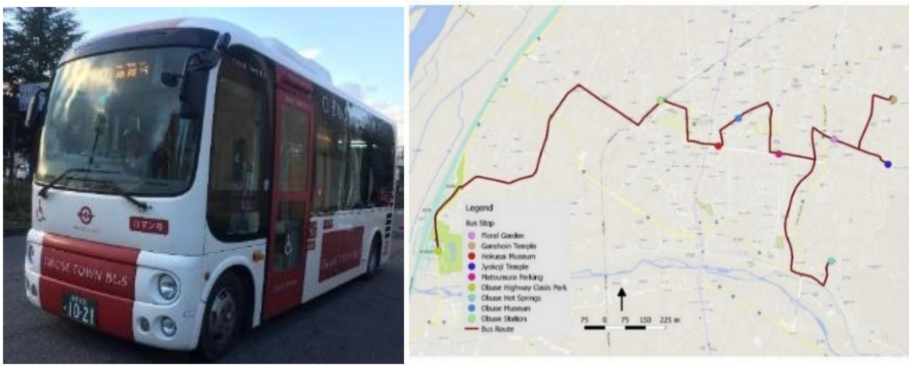


Fig 2. Obuse Romango bus and bus route map

4.3. Wi-Fi Scanner and Installation

Wi-Fi scanner equipment captures MAC addresses from devices such as smartphones, laptops, tablets, computers, and other Wi-Fi-enabled devices. A MAC address is a unique code, specific to a Wi-Fi-enabled device, and does not contain any personal information. The Wi-Fi scanner equipment includes an antenna, GPS, and a mobile battery (Figure 4). This scanner uses a Raspberry Pi mini-computer as a controlling device. The Raspberry Pi has a quad-core processor and a single-board CPU, running at 900MHz, powered by a 30000 MAH portable battery, with up to 12-hour battery life. This system has 1 GB RAM capacity, a USB port, pole stereo output, video port, and an HDMI input. The Raspberry Pi also includes a high-frequency, GPS tracking device and a Micro SD port, which is used for loading the operating system and storing data.

The Wi-Fi scanner was placed in the bus and positioned in proximity to the bus driver. The scanner has an approximate range of 200–300 meters (Figure 3) and detects bus passengers' and surrounding pedestrians' Wi-Fi devices as well as Wi-Fi devices located within buildings and vehicles. Upon concluding the field survey, the Wi-Fi scanner was turned off, and the MAC addresses were obtained for further analysis. The MAC addresses are referred to as raw data that will be used to estimate passenger.

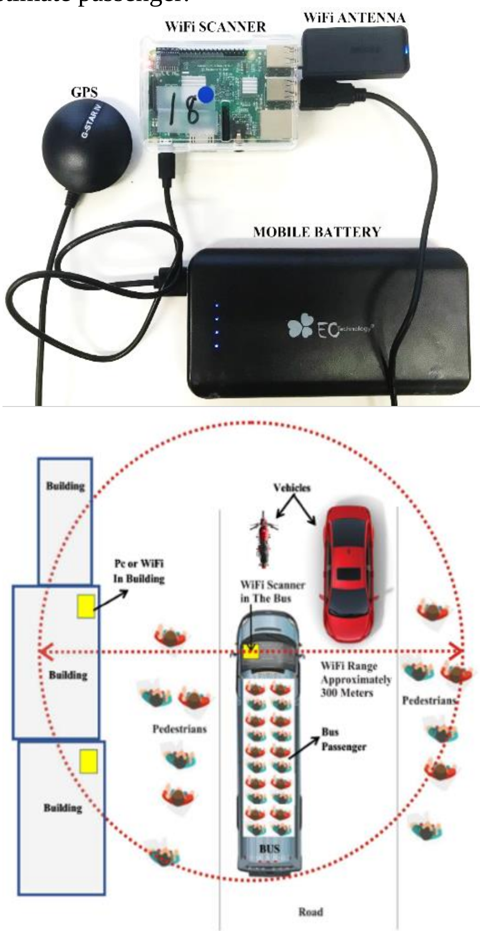


Fig 3. Wi-Fi scanner equipment and Wi-Fi scanner approximate range

5. Data Cleaning and Processing

The results of the survey provide both Wi-Fi and GPS log data. The Wi-Fi log contains the time and MAC address data, while the GPS log contains the time and coordinate (position) data. The Wi-Fi and GPS logs obtained from the

Wi-Fi scanner were distinct. The data processing procedure combines Wi-Fi and GPS logs, converts the coordinate system to UTM (universe transverse Mercator), calculates speed, and calculates passenger volume and determines an OD matrix for each route section (Figure 4).

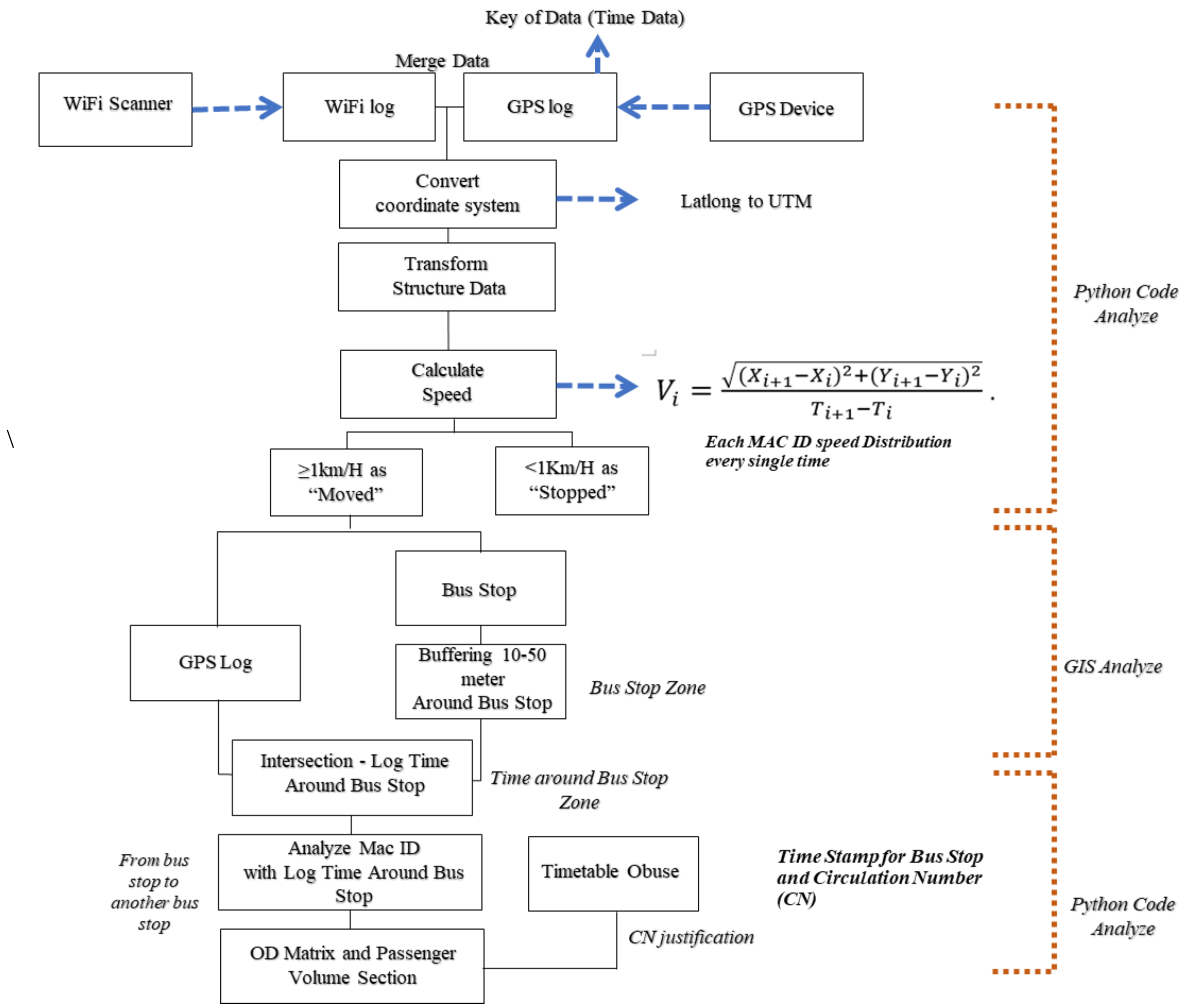


Fig 4. Wi-Fi procedure flowchart

5.1. Merging Wi-Fi and GPS Logs

This step of the data processing procedure combines Wi-Fi and GPS log data. Merging these two logs allows for the determination of the coordinates of each MAC address that is obtained by the Wi-Fi scanner. This merging technique uses Jupyther Notebook software's concate pandas module, and the time column data in the GPS log was used as a guide for merging the two datasets. Merged data was displayed on the GPS log because time is reported in units of data per second. This merging procedure focuses on time data that does not have a MAC address, but can still be used for further analysis, and also aims to determine how much GPS log time has elapsed between the detection of a MAC address (ping). Time data on the GPS log provides information regarding the amount of time elapsed between individual MAC address pings throughout a passenger's bus trip (Figure 6). This data is also used to calculate the total travel time for each MAC address. Data merging is essential for the next step of data processing: data cleaning.

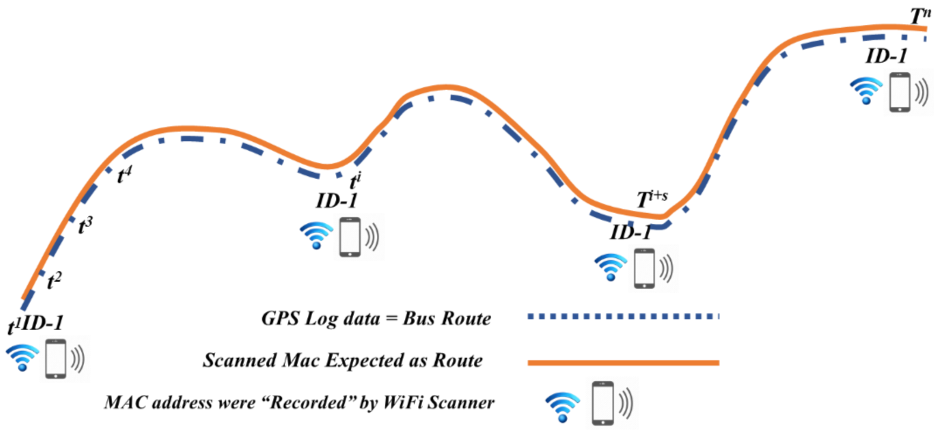


Fig 5. Schematic of combined Wi-Fi and GPS logs.

In Figure 5, the procedure for analyzing Wi-Fi data based on a bus route is shown and is referred to as route speed. Route speed is based on a MAC address' calculated speed based on the bus route, which is recorded in the GPS log. The GPS log significantly affects the calculation of the MAC address speed. Among the MAC address data, there is also time data that is not included with the MAC address that could be used to analyze speed, which tracks MAC addresses based on the route as well as the time.

5.2. Converting GPS Coordinates to UTM

This data processing procedure converts decimal degree coordinates to UTM (meter units). Converting decimal degree coordinates to UTM coordinate form enables the calculation of distance and speed for each MAC address. This data processing procedure is accomplished using Python software's geopandas module. After converting the coordinate data, the data is input into QGIS software. On a QGIS map, the Wi-Fi datapoints can be observed, and are in alignment with the bus route (Figure 6).

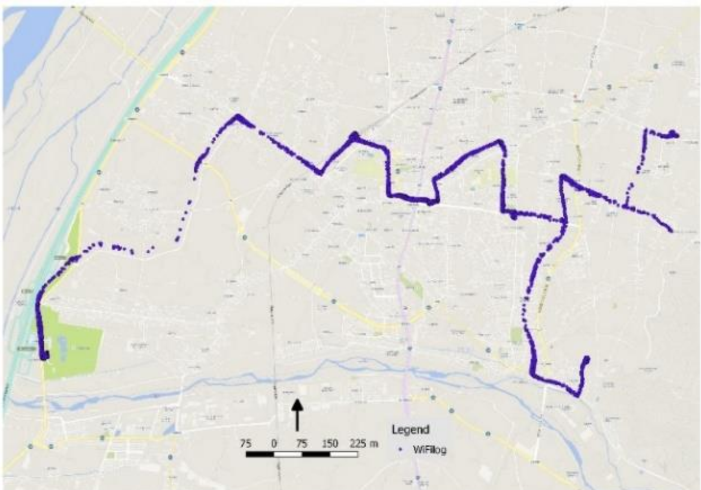
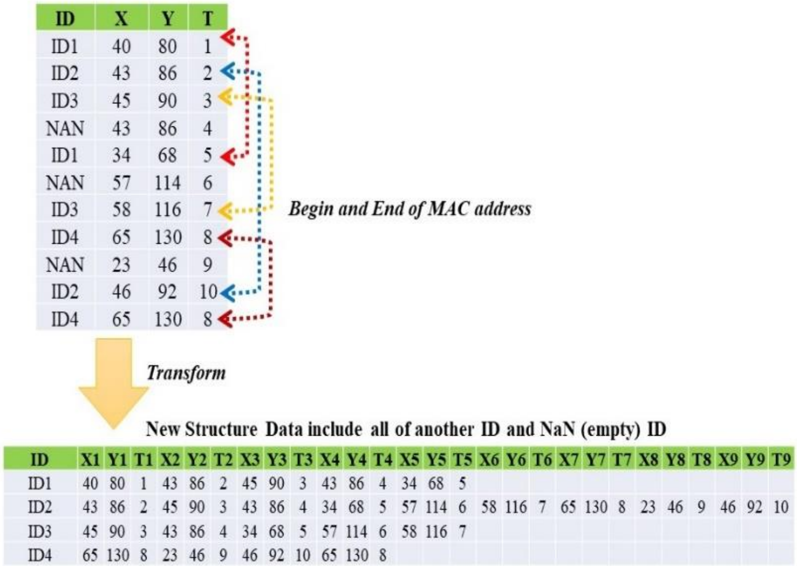


Fig 6. MAC address point map

5.3. Data Transformation

Data transformation is an advanced analytical step for Wi-Fi data cleaning and is accomplished using Python software's Iloc pandas data frame module. This data transformation procedure aims to identify the occurrence of related data between the beginning and end of MAC address detection (Figure 7). This analysis includes time and coordinate data that do not have MAC address data (NaN/Not a Number data).



ID	X	Y	T
ID1	40	80	1
ID2	43	86	2
ID3	45	90	3
NAN	43	86	4
ID1	34	68	5
NAN	57	114	6
ID3	58	116	7
ID4	65	130	8
NAN	23	46	9
ID2	46	92	10
ID4	65	130	8

ID	X1	Y1	T1	X2	Y2	T2	X3	Y3	T3	X4	Y4	T4	X5	Y5	T5	X6	Y6	T6	X7	Y7	T7	X8	Y8	T8	X9	Y9	T9
ID1	40	80	1	43	86	2	45	90	3	43	86	4	34	68	5												
ID2	43	86	2	45	90	3	43	86	4	34	68	5	57	114	6	58	116	7	65	130	8	23	46	9	46	92	10
ID3	45	90	3	43	86	4	34	68	5	57	114	6	58	116	7												
ID4	65	130	8	23	46	9	46	92	10	65	130	8															

Fig 7. Example of transformation data

Figure 7 shows a revised data structure for MAC address speed route calculations. The ID is a MAC address. X and Y are longitude and longitude coordinates and T is Time. Revised data structures have been combined by inserting the GPS data between the occurrences of a MAC address. The transformed data is interpolated between the beginning and end of a MAC ID.

5.4. Speed Calculation

The speed of each MAC address is intended for cleaning data based on the average amount of speed of each MAC address. Analysis of MAC address speed is essential for equating MAC address speed to bus speed. The results of this analysis reveal the distribution of speed for each time point that a MAC address is detected. This speed distribution also calculates speed count, variance, and average. The speed count represents the number of speed measurements calculated for each MAC address. Variance is the average of the squared differences from the mean of the speed. The average is the mean value of the overall speed for each MAC address. After this data processing step, passenger volume can be determined for each bus route section. The equation for calculating MAC address speed is shown below (1).

$$V_i = \frac{\sqrt{(X_{i+1}-X_i)^2+(Y_{i+1}-Y_i)^2}}{T_{i+1}-T_i}. \quad (1)$$

i = Timeframe of data acquisition
Vi = Speed at the i_th time frame
Xi = Longitude for the i_th MAC address
Yi = Latitude for i_th MAC address
Ti = Time for i_th MAC address

5.5. Moved Categorization

This data processing step categorizes each MAC address measurement as “moved” or “stopped.” The threshold for categorizing a MAC address as “moved” is 1km/h. Therefore, if a MAC address has moved faster than 1km/h, then the MAC address is categorized as “moved,” and if the speed of a MAC address is measured below the 1km/h threshold, the MAC address is categorized as “stopped.” The 1km/h threshold is based on the average of MAC address speed and bus speed. This research was conducted by trial and error. If it sets below than 1km/h, it will over-estimate, and if above than 2km/h, it is too underestimate. It refers to figure 8. Data below the threshold could be removed or not used in the following processing step. This paper used > 1 km/h for the threshold to justify "moved" and "stopped." Some thresholds are used to view the trend of MAC address data changing. The graph below (Figure 8) shows a different result by the thresholds value. The high number of the MAC address is the raw data (included “moved” and “stopped” MAC address data), and another amount the MAC address as a “moved” used for passenger justify analysis. The threshold is applied 1 km/h because it considers the bus speed. The 1 km/h threshold is moderate enough to be used for further analysis because the Obuse Romango Bus is hop on-off bus and the route quietly narrow. These make Obuse Romango Bus slower than the common bus. If the authors set without the threshold, 3931 unique MAC addresses are considered as passenger and non-passenger (moved and stopped classification). If the authors set the threshold as 1km/h that moved classification ($0 < Vi < 1$km/h → stopped and 1km/h $\leq Vi$ → moved).

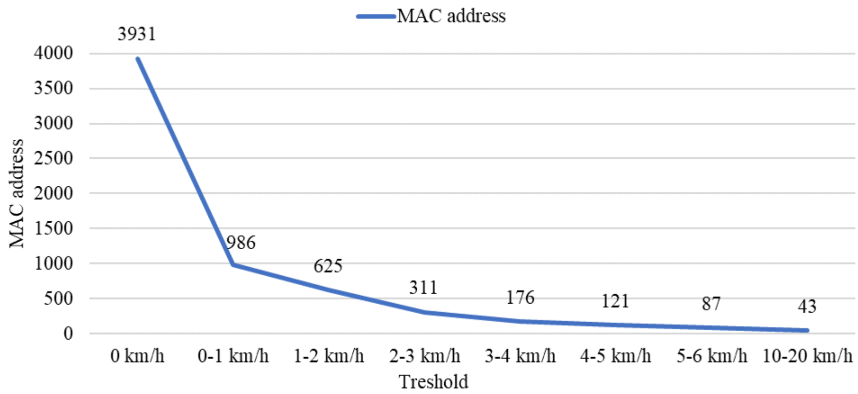


Fig 8. Threshold setting difference

This graph above shows how many MAC address data this research use. 3931 MAC addresses are raw data. Raw data without speed threshold will have result overestimated than ground truth data. This study uses 986 MAC addresses for the result further (1km/h threshold). This research results underestimated than ground truth data. After a trial and error experiment, therefore the threshold was used above than 2km/h the result too high overestimated (big gap) than ground truth data.

5.6. Bus Stop Zoning, Time Spent at Intersection, and Analysis of MAC Addresses in the Vicinity of Bus Stops

In this data processing step, each bus stop is buffering data using QGIS. The buffering zone radius surrounding a bus stop is approximately 10 meters. This part of determining how far queue or people standing around the bus stop; however, if too significant number, it will provide people not only around the bus stop also outside the bus stop. If too small, it will be inappropriate for data around the bus stop. This zone, surrounding the bus stop, is used for intersection analysis in conjunction with the time data on the GPS log. This zoning procedure analyzes the time of bus movement recorded at the bus stop or around the bus stop. The intersection is used to determine the existence of a MAC address at the bus stop. The GPS time log data and the intersection around the bus stop are used as the basis for classifying the time data of a MAC address which is either at the bus stop or outside of the bus stop. This processing step is used to determine the OD and passenger volume for each MAC address. This analysis uses Python’s classify function by classifying the time within the buffer zone.

5.7. Classification Based on Route Circulation

In this final processing step, the data from the previous processing step was re-analyzed based on route circulation. The Romango bus has seven route circulations in one day, and MAC addresses that are detected in circulation can be interpreted as traveling passengers. MAC addresses that are not in circulation can be eliminated from this dataset. This processing step uses Python's time classified program analysis.

6. Results and Comparisons

6.1. OD

The OD matrix data (Figure 9) is the result of the previously described data analysis and processing procedures conducted on raw data to obtain OD data. The data cleaning procedure was able to produce OD matrix data for the field surveys (Figure 9). From Figure 9, the highest number of estimate passenger origin-destination (OD) is on BS1-BS2, BS1-BS9, BS3-BS1, BS3-BS5, BS4-BS6, BS5-BS4, and BS7-BS9. For the average number of the estimate passenger OD is on BS3-BS2, BS5-BS6, BS6-BS7, and BS8-BS3. For others are of a small number of estimate passengers. This passenger OD movement is entirely dominated by the bus stop in the city center leading in-out of the city center. This OD is also based on the type of tourism spot that is adjacent to the bus stop. In the city center, obuse park, obuse station and museum are a high number of tourists. Therefore, these bus stops are a high number of estimate passengers OD. Outside of the city center is hot spring and temple.

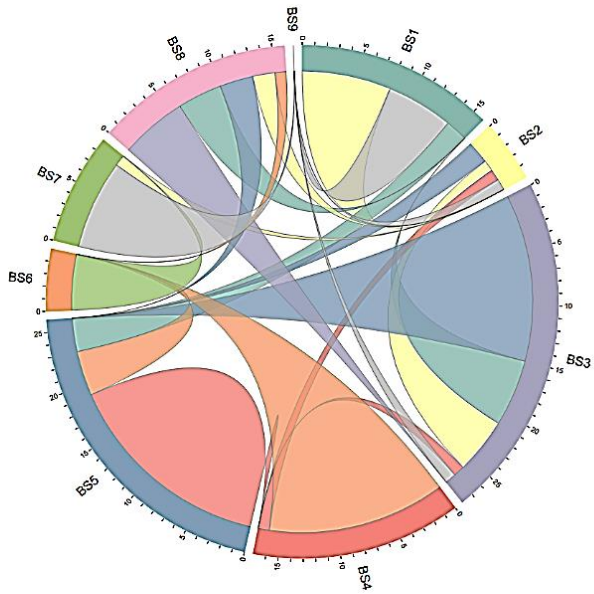


Fig 9. OD matrix results

6.2. Passenger Volume for a Section

Passenger volume for a section accounts for every passenger on-board the bus along the route at each bus stop between an individual passenger's origin and destination, even if a passenger's MAC address is not detected at one of the buses stops along the route traveled between their corresponding origin and destination (Figure 10). For example, if two people board the bus at bus stop No. 1, and travel to bus stop No. 3, they will still be included in the count at the bus stop No. 2, regardless of their MAC address being detected at bus stop No. 2 because the bus implicitly passed or stopped at the bus stop No. 2. Figure 10 shows the difference between OD and passenger volume for each section.

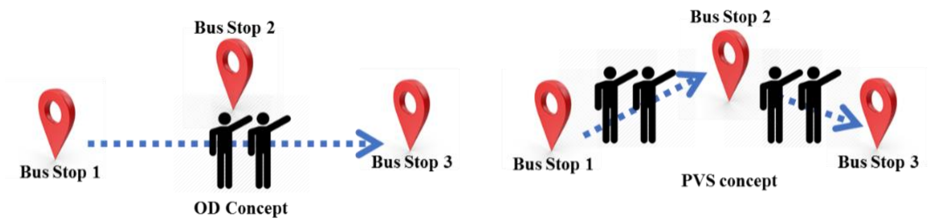


Fig 10. Difference OD and passenger volume for the section

Figure 11 demonstrates that data had different travel trends. The highest passenger volumes were observed in sections M2-3, M3-4, M4-5, M5-6, M6-7, M9-7, M7-5, M5-4, and M4-3, while average passenger volumes were observed in sections M1-2, M6-7, M7-8, M8-9, and M2-1. Additionally, the route circulations with the highest volume of passengers were CN1, CN4, and CN5. Considering the time, CN2, CN4, and CN6 is the high number of passengers volume because these CNs are 11:00 AM, 13:00 PM and 16:00 PM which is the best time to travel from one spot to another spot.

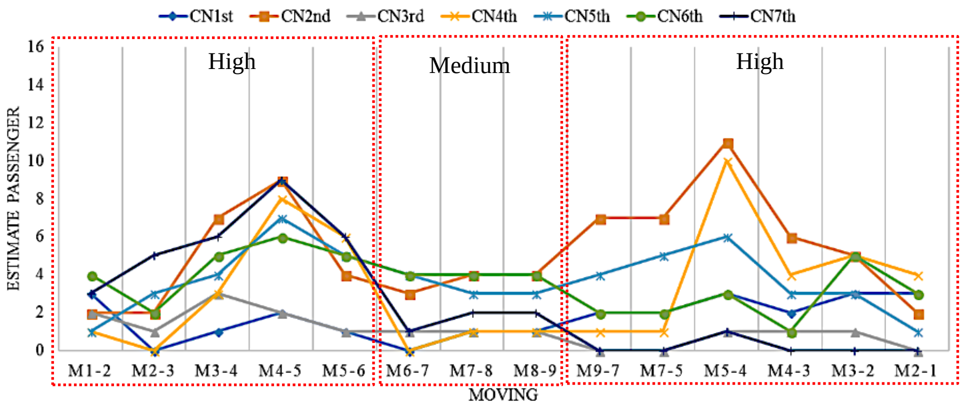
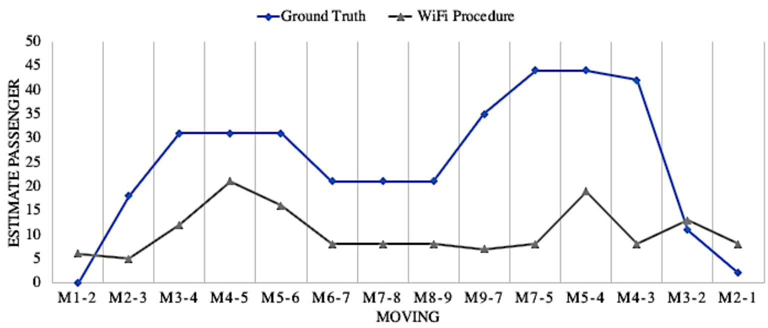


Fig 11. Passenger volume for sections

6.3. Ground Truth Comparison

Figure 12 demonstrates that the Wi-Fi procedure calculates fewer passengers than the ground truth method. The Wi-Fi procedure presumably calculated a lower number of passengers compared to the ground truth method because not all passengers may have a Wi-Fi-enabled device. Ratio Wi-Fi (RW) values obtained contain both positive and negative values (Figures 12). The number of passengers in sections M2-3 until M4-3 have negative values because the number of passengers determined using the Wi-Fi procedure is less than the number of passengers determined using the ground truth procedure. The number of passengers determined using the Wi-Fi method shares a similar trend to the number of passengers. The passenger volume determined using the Wi-Fi method is higher than the ground truth data (GRD) in sections M1-2, M3-2, and M2-1, while the passenger volume determined using the Wi-Fi method is less than the GRD in sections M2-3 and M4-3. There is still a bias between the direct and route speed procedure with ground truth data. The result must be underestimated because not all passengers using Wi-Fi devices or turn on the Wi-Fi. The trend from figure 12 reveals that the route speed procedure result tendency is quietly similar to ground truth data.



M	GRD	NWD	RW	Annotation
M1-2	0	6	6.0	Over Estimate
M2-3	18	5	-13.0	Lower Estimate
M3-4	31	12	-19.0	Lower Estimate
M4-5	31	21	-10.0	Lower Estimate
M5-6	31	16	-15.0	Lower Estimate
M6-7	21	8	-13.0	Lower Estimate
M7-8	21	8	-13.0	Lower Estimate
M8-9	21	8	-13.0	Lower Estimate
M9-7	35	7	-28.0	Lower Estimate
M7-5	44	8	-36.0	Lower Estimate
M5-4	44	19	-25.0	Lower Estimate
M4-3	42	8	-34.0	Lower Estimate
M3-2	11	13	2.0	Over Estimate
M2-1	2	8	6.0	Over Estimate

M : Moving/ passenger volume for the section
CN : Circulation Number
NWD : NewWi-Fi Data
RW : Ratio Wi-Fi
GRD : Ground Truth Data

Fig 12. Comparing ground truth, and Wi-Fi procedure

7. Conclusion

The analysis and interpretation of the Wi-Fi data yielded information regarding passenger volume as well as information regarding passengers' travel between destinations. This investigation describes the application of a new method for collecting and processing data to determine OD matrix data and passenger volumes for bus route sections. The results of this investigation can be summarized. A new processing procedure was developed to transform data into OD and passenger volume for each section, which determines passenger OD information, the number of passengers boarding a bus at a specific bus stop, and the number of passengers waiting at a bus stop. The results of the comparison between data obtained using the Wi-Fi procedure and observer data indicate similar travel patterns, and the number of passengers determined using the Wi-Fi procedure is generally less than the number of passengers determined using either the ground truth method.

The Wi-Fi scanner could be used for several other purposes: It could assist small bus operators in conducting their surveys. The device and the data processing procedure is appropriate for long-term data collection to observe daily variations in passenger bus transport. There is no need to communicate with passengers while collecting data. Movement demand (origin and destination) can be analyzed between bus stops, which can assist bus operators in determining hourly and daily demands posed by differing volumes of bus passengers and can determine bus capacity between bus stops. For example, bus demand sometimes increases during specific periods of the year, such as during the high autumn season and for festivals. The Wi-Fi scanner data can be used to dynamically determine passenger capacity during individual seasons, allowing bus operators to add or reduce the number of operating buses to accommodate the changing demand. Bus operators can also make timetable changes to adjust to high demand at certain times of the day. The Wi-Fi scanner data can also provide information regarding busy areas, such as bus stops, shops, parking lots, bus terminals, and others. If the Wi-Fi scanner detects an area that has a high volume of potential bus passengers, this data could inform operators to increase passenger capacity. For local governments, the Wi-Fi scanner and the data obtained by the Wi-Fi data processing procedure could be used to improve the function of facilities, provide input for urban planning related to transportation facilities and public transportation, and enhance marketing for tourism support facilities.

However, some limitation addressed to this research is the number of Wi-Fi device held by passengers. Authors assumed that each passenger carries only one Wi-Fi device. Accordingly, each detected device can be uniquely associated with one onboard petepete passenger. However, some passengers may bring more than one Wi-Fi devices. Activated/deactivated Wi-Fi device is another limitation to this research. This creates another source of error when attempting to estimate the number of passengers. MAC address changing is the one of limitation of this research because some device will change the MAC address if the software device was upgraded. Another limitation is the MAC address that is not in the right position (X and Y). Because sometime the MAC address will appear on another location. Wi-Fi has a long detection range, so the sensor covers a large area. Wi-Fi also adds a significant amount of spatial uncertainty to the data (noise data). This problem might be solved by using a directional Wi-Fi antenna or by reducing the Wi-Fi coverage. It would also be useful to take greater care in the data processing steps.

Acknowledgments

The authors would like to thank all those who have contributed to this paper especially to Urban and Transportation Planning Laboratory, Tokyo University of Science. Thanks to Universitas Teknologi Sulawesi-Indonesia, my scholarship from the Ministry of Research, Technology and Higher Education and Indonesia Endowment Fund for Education (LPDP) Ministry of Finance, The Republic of Indonesia. Thanks also to the anonymous reviewers who have provided corrections in this paper.

References

Abbott-jard, M., Shah, H., Bhaskar, A., 2013. Empirical evaluation of Bluetooth and Wi-Fi scanning for road transport. Proc. Australas. Transp. Res. Forum 1–14.

Abedi, N., Bhaskar, A., Chung, E., Miska, M., 2015. Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and Wi-Fi MAC addresses. Transp. Res. Part C Emerg. Technol. 60, 124–141. https://doi.org/10.1016/j.trc.2015.08.010

Araghi, B.N., Hammershøj Olesen, J., Krishnan, R., Tørholm Christensen, L., Lahrmann, H., 2015. Reliability of Bluetooth Technology for Travel Time Estimation. J. Intell. Transp. Syst. 19, 240–255. https://doi.org/10.1080/15472450.2013.856727

Cisco, 2008. 802.11 Security Summary. Wirel. Netw. Secur. Integr. Solut. Des. Guid. Chapter 3, 1–20.

Cunche, M., 2014. I know your MAC address: targeted tracking of individual using Wi-Fi. J. Comput. Virol. Hacking Tech. 10, 219–227. https://doi.org/10.1007/s11416-013-0196-1

Dunlap, M., Li, Z., Henrickson, K., Wang, Y., 2016. Estimation of Origin and Destination Information from Bluetooth and Wi-Fi Sensing for Transit. Transp. Res. Rec. J. Transp. Res. Board 2595, 11–17. https://doi.org/10.3141/2595-02

Erkan, İ., Hastemoglu, H., 2016. Bluetooth as a traffic sensor for stream travel time estimation under Bogazici Bosporus conditions in Turkey. J. Mod. Transp. 24, 207–214. https://doi.org/10.1007/s40534-016-0101-y

Ferro, E., Potorti, F., 2005. Bluetooth and wi-fi wireless protocols: a survey and a comparison. IEEE Wirel. Commun. 12, 12–26. https://doi.org/10.1109/MWC.2005.1404569

Filgueiras, J., Rossetti, R.J.F., Kokkinogenis, Z., Ferreira, M., Olaverri-Monreal, C., Paiva, M., Tavares, J.M.R.S., Gabriel, J., 2014. Sensing Bluetooth Mobility Data: Potentials and Applications. Adv. Intell. Syst. Comput. 262, 419–431. https://doi.org/10.1007/978-3-319-04630-3_31

Hidayat, A., Terabe, S., Yaginuma, H., 2018a. Wi-Fi Scanner Technologies for Obtaining Travel Data about Circulator Bus Passengers: Case Study in Obuse, Nagano Prefecture, Japan. Transp. Res. Rec. J. Transp. Res. Board. https://doi.org/10.1177/0361198118776153

Hidayat, A., Terabe, S., Yaginuma, H., 2018b. Determine Non-Passenger Data from Wi-Fi Scanner Data (MAC Address), A Case Study: Romango Bus, Obuse, Nagano Prefecture, Japan. Int. Rev. Spat. Plan. Sustain. Dev. 6, 154–167. https://doi.org/10.14246/irspsda.6.3_154

Hidayat, A., Terabe, S., Yaginuma, H., 2017a. Mapping of MAC Address with Moving Wi-Fi Scanner. Int. J. Artif. Intell. Res. 1, 34–29. https://doi.org/10.29099/ijair.v1i2.27

Hidayat, A., Terabe, S., Yaginuma, H., 2017b. Wi-Fi Scanner for Obtaining Pedestrian Data. Plano Madani 6, 128–136.

J.Barceló, L.Montero, Marqués, L., Carmona, C., 2010. Travel Time Forecasting and Dynamic OD Estimation in Freeways Based on Bluetooth Traffic Monitoring. 89th Transp. Res. Board 2010 Annu. Meet. 18.

Khliefat, I., Shatnawi, I., 2017. An Optimization of Bluetooth Sensor Locations for Origin-Destination in an Urban Network. J. Transp. Technol. 07, 367–375. https://doi.org/10.4236/jtts.2017.74024

Kurkcu, A., Ozbay, K., 2017. Estimating Pedestrian Densities, Wait Times, and Flows with Wi-Fi and Bluetooth Sensors. Transp. Res. Rec. J. Transp. Res. Board 2644, 72–82. https://doi.org/10.3141/2644-09

Lesani, A., Moreno, L.F., 2016. Development and Testing of a Real-Time Wi-Fi-Bluetooth System for Pedestrian

Network Monitoring and Data Extrapolation. Transp. Res. Board, 95th Annu. Meet. 1–13.
Martin, J., Mayberry, T., Donahue, C., Foppe, L., Brown, L., Riggins, C., Rye, E.C., Brown, D., 2017. A Study of MAC Address Randomization in Mobile Devices and When it Fails. Proc. Priv. Enhancing Technol. 2017, 268–286. https://doi.org/10.1515/popets-2017-0054
Matte, C., 2017. Wi-Fi Tracking: Fingerprinting Attacks and Counter-Measures. Universite de Lyon.
Musa, A.B.M., Eriksson, J., 2012. Tracking unmodified smartphones using wi-fi monitors. Proc. 10th ACM Conf. Embed. Netw. Sens. Syst. - SenSys '12 281. https://doi.org/10.1145/2426656.2426685
Najafi, P., Georgiou, A., Shachneva, D., Vlavianos, I., 2014. Privacy Leaks from Wi-Fi Probing. London.
Namaki Araghi, B., Skoven Pedersen, K., Tørholm Christensen, L., Krishnan, R., Lahrmann, H., 2015. Accuracy of Travel Time Estimation Using Bluetooth Technology: Case Study Limfjord Tunnel Aalborg. Int. J. Intell. Transp. Syst. Res. 13, 166–191. https://doi.org/10.1007/s13177-014-0094-z
Porter, J.D., Kim, D.S., Magaña, M.E., Poocharoen, P., Arriaga, C.A.G., 2013. Antenna Characterization for Bluetooth-Based Travel Time Data Collection. J. Intell. Transp. Syst. 17, 142–151. https://doi.org/10.1080/15472450.2012.696452
Pourhassan, K., 2016. How Reliable Is the Bluetooth-Based Origin-Destination Data ? Conf. Aust. Institutes Transp. Res. 1–12.
Romancyshyn, T., 2016. The Use of Bluetooth and Smartphone GPS Data to Investigate Factors Associated to Variations of Travel Times and Delays. McGill University.
Schauer, L., Werner, M., Marcus, P., 2014. Estimating Crowd Densities and Pedestrian Flows Using Wi-Fi and Bluetooth. Proc. 11th Int. Conf. Mob. Ubiquitous Syst. Comput. Netw. Serv. 171–177. https://doi.org/10.4108/icst.mobiquitous.2014.257870
Shlayan, N., Kurkcu, A., Ozbay, K., 2016. Exploring pedestrian Bluetooth and Wi-Fi detection at public transportation terminals. 2016 IEEE 19th Int. Conf. Intell. Transp. Syst. 229–234. https://doi.org/10.1109/ITSC.2016.7795559
Sridhar, T., 2008. Wi-Fi, Bluetooth and WiMax - Technology and Implementation. Internet Protoc. J. 11, 2–17.
Sun, L., Chen, S., Zheng, Z., Xu, L., 2017. Mobile Device Passive Localization Based on IEEE 802.11 Probe Request Frames. Mob. Inf. Syst. 2017, 1–10. https://doi.org/10.1155/2017/7821585
Terabe, S., Hidayat, A., Yaginuma, H., 2017. Wi-Fi Sensing Technologies for Affordable Travel Data Collection from Bus Passengers. 11th Int. Conf. Transp. Surv. Methods 1–11.
Vanhoef, M., Matte, C., Cunche, M., Cardoso, L.S., Piessens, F., 2016. Why MAC Address Randomization is not Enough. Proc. 11th ACM Asia Conf. Comput. Commun. Secur. - ASIA CCS '16 413–424. https://doi.org/10.1145/2897845.2897883
Verbree, E., Zlatanova, S., van Winden, K.B.A., van der Laan, E.B., Makri, A., Taizhou, L., Haojun, A., 2013. To localise or to be localised with Wi-Fi in the Hubei museum? ISPRS - Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. XL-4/W4, 31–35. https://doi.org/10.5194/isprsarchives-XL-4-W4-31-2013
Xu, Z., Sandrasegaran, K., Kong, X., Zhu, X., Zhao, J., Hu, B., Chung Lin, C.-, 2013. Pedestrain Monitoring System using Wi-Fi Technology And RSSI Based Localization. Int. J. Wirel. Mob. Networks 5, 17–34. https://doi.org/10.5121/ijwmn.2013.5402
Yaik, O.B., Wai, K.Z., K.T.Tan, I., Sheng, O.B., 2016. Measuring the Accuracy of Crowd Counting using Wi-Fi Probe-Request-Frame Counting Technique. J. Telecommun. Electron. Comput. Eng. 8, 79–81.
Young, S.E., 2012. Bluetooth Traffic Detectors for Use as Permanently Installed Travel Time Instrument.

AUTHORS RESPONSE

Reviewer 1

This paper proposes a method to estimate passenger volumes and OD from Wi-Fi data in a public transport network. This topic has been seen in many papers, so the literature review should emphasize on the problems with latter work and tell how this paper improves things. The paper is lacking many explanations in general and contains some assertions that should be backed by adequate citations.

Comments:
- I understand that Wi-Fi is "better' than Bluetooth, but the choice of BT is often related to the shorter range of transmission. A shorter range avoids having parasite data (ie. people not in the bus). This should be discussed.
Answer: Thank you for the comments and revise our paper. Authors tries to fix based on this comment. The reviewer can find an introduction section, the 3rd paragraph

- Section 3.1: 2% less, is this coming from the Lesani and Moreno study? In what conditions?
Answer: This is for pedestrian survey condition

- Section 3.2. Why are smart cards explained here? Is it used in the study? In addition, there is rarely "Real-time" smart card data. Data is collected and then available later in the company's servers.
Answer: for smartcards section, it was deleted in this paper

- Section 4.1. Passenger data was collected "twice". When? What day? Twice at the same time? Two different times?
Answer: Twice, it is means Wi-Fi survey data and ground truth data at the same time and same bus. The reviewer can find in field survey section, the 2nd paragraph

- Figure 7, I do not understand the transformation as is in the figure. Why store the XY in the second table?

Answer: Before calculating the speed, the data had to be transformed. This analysis included time and coordinate data that did not have MAC address data (NaN/Not-a-Number data). This transformation considered the GPS log data between the appearance of MAC addresses. The data transformation in Figure aimed at identifying and sorting the data between the beginning and end of MAC address detection. This transformation was useful before calculating the MAC address speed because MAC addresses were sorted by X, Y, and T (between of the first and the last appearance of the MAC address)

- Section 5.6. How was the buffer zone size chosen? Any sensitivity analysis on the parameters?
Answer: Authors tries to fix based on this comment. The reviewer can find in "Moved Categorization" section with threshold setting. 1 km was chosen because the authors tried "trial and error" for the threshold. If below the 1km/h it will include noise data or non-passenger data that is overestimated result. If above than 1km/h, this result is too underestimate, its means passenger data will remove.

- Section 6.2. There are many limitations here. Not all passengers use Wi-Fi devices. As a matter of fact,

they may not represent the whole spectrum of the population (especially older people). In addition, many users will have more than one device (computer, iPhone, Wi-Fi, watch). Knowing that, how can we use the results? Any correction procedure?
Answer: No correction procedure or ask the passenger directly about Wi-Fi turn on-off. This limitation included in the "conclusion" section, the 3rd paragraph.

- Figure 11. What is the TRR paper? Is this work published in TRR? A statistical test should be conducted here to see if the results are satisfying.
Answer: Syntax mistake, fixed it. The statistical test added in "ground truth comparison' section the 2nd paragraph.

- The conclusion should discuss about the numerous limitations of such kind of study.
Answer: *Authors tries to fix based on this comment. The reviewer can find in the "conclusion" section, the 3rd paragraph.*

- Please correct typos and English syntax
Answer: Authors tries to fix based on this comment.

Reviewer 2
It's a well written paper that, regrettably, demonstrates that the outcome of the proposed method depends on the unknown parameter
– fraction of travellers with the activated Wi-Fi. The authors report this discouraging result honestly and straightforwardly and I am strongly in favour of publication this paper. Wi-Fi scanning looks attractive to many researchers and the results of these first field experiments, no matter encouraging or discouraging are important for the current transportation research.
The investigated dataset seems yet below the threshold that is necessary for concluding on the usefulness of a technology. That is why I recommend publication in the conference proceedings only. However, if larger datasets became available after the paper was submitted, I would consider the paper, after the revision, for publication in a journal special issue.
Minor remarks by sections.

Answer: Thank you for the comments and revise our paper.

5.2 – Can be easily skipped.
Answer: Authors tries to explain more this part based on this comment.

5.3 – Report the fractions of errors
Answer: authors tried to revise the paper based on this comment. The errors were explained in the section "moved categorization" page 9 and conclusion page 12.

5.5 – Are the results robust to the chosen threshold of 1 km/h?
Answer: In the section "moved categorization" page 9. The threshold for categorizing a MAC address as "moved" is 1km/h. Therefore, if a MAC address has moved faster than 1km/h, then the MAC address is categorized as "moved," and if the speed of a MAC address is measured below the 1km/h threshold, the MAC address is categorized as "stopped." The 1km/h threshold is based on the average of MAC address speed and bus speed. This research was conducted by trial and error. If it sets below than 1km/h, it will

over-estimate, and if above than 2km/h, it is too underestimate.

6.3 – Do you have external estimates of the fraction of activated/ deactivated Wi-Fi? Should we ban the method based on your results?
Answer: No, we do not have an external estimate of the fraction of activated/ deactivated Wi-Fi. Authors tried to present a new procedure or method on how to distinguish MAC addresses data into passengers. However, there are still much lacks of errors or still need more detail in the analysis/procedure and we always try to improve the analysis.