TIME TRAVEL ESTIMATIONS USING MAC ADDRESSES OF BUS, PASSENGERS: A POINT TO PATH-QGIS ANALYSIS

A. Hidayat, S. Terabe, H. Yaginuma

Urban and Transportation Planning Laboratory, Department of Civil Engineering

Abstract: Currently, the development of Wi-Fi is proliferating. Especially in the field of transportation and smart cities. At the same time, Wi-Fi is a low-cost technology, which offers a longer survey time and is able to support the big data era. This paper describes our study, which first uses a Wi-Fi scanner to capture media access control (MAC) address data of bus passengers. The MAC address is a unique ID for each device used such as mobile phones, smartphones, laptops, tablets, and other Wi-Fi-enabled equipment. The Wi-Fi scanner was placed inside the bus to capture all the MAC addresses inside and around the bus. The survey was conducted for one day (eight hours). The paper describes the procedure of the time travel estimation for each MAC address using the “point to path” analysis in QGIS open source software. This procedure, using point to path-GIS, produced 70,000-80,000 raw data points cleaned into 100-130 new data points. The procedure determined how many passengers traveled and explained which bus passengers used based on travel time.

1. INTRODUCTION

Technology development is growing every day. This development also applies to information technology systems such as Wi-Fi. Wi-Fi is a network connection system that is currently used almost universally. The Wi-Fi connection is built into devices such as smartphones, laptops, tablets, and other devices that receive Internet signals or data (Hidayat et al., 2017a). The development of Wi-Fi and information technology systems has penetrated the transportation engineering sector. The use of Wi-Fi in transportation is currently advancing, especially in terms of the development of travel data related to origin-destination (OD), speed estimation, travel time, and the estimation of passengers. This advancement pertains to transportation sectors such as a pedestrian, motor vehicle, and others (Abedi, 2014; Al-Husainy & Fadhil, 2013; Xia et al., 2014). Wi-Fi thrives because of its low cost, accessibility, energy efficiency, and mobile scanner capacity (non-static scanner). Furthermore, almost everyone uses Wi-Fi daily due to the easy data retrieval process.

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). It is a popular method to provide Internet access for wireless users (Xu et al., 2013). The more common Wi-Fi mode of operation is 802.11, called the infrastructure mode, where stations communicate with other wireless stations and wired networks (typically Ethernet) through an access point. The access point bridges traffic between wireless stations through the lookup of the destination address in the 802.11 frame (Sridhar, 2008). The infrastructure mode supports smartphones, tablets, routers, and laptop, among others. A smartphone can be identified by its unique ID such as its international mobile equipment identity (IMEI) number or the media access control (MAC) address. The IMEI is received when the mobile device is registered on a network, whereas the MAC address is on every data packet sent by the Wi-Fi-enabled mobile handset. 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). The Wi-Fi connection for smartphones is designed to periodically transmit a probe-request-frame to determine a known access point (Matte, 2017; Yaik et al., 2016). Probe requests are the active scans by the mobile device (Sun et al., 2017; Verbree et al., 2013). The probe request content includes the sender’s MAC address (Musa & Eriksson, 2012). The Wi-Fi scanner probe request can load all MAC address data into a single log file. This system accesses the MAC address without connecting to the Internet and is a passive scanning activity that collects data. The Wi-Fi scanner as a probe request mode was developed to collect MAC addresses included in the infrastructure mode. This study uses a Wi-Fi scanner on a bus to collect MAC address data from bus passengers and non-passengers.

Wi-Fi systems capture MAC address data from Wi-Fi device users (Abedi et al., 2015; Dunlap et al., 2016; Jackson et al., 2014). The MAC address is the same on each device that is Wi-Fi-enabled. MAC addresses are unique numbers and letters for each device and no device has two MAC address. In addition, one MAC address cannot be assigned to two devices (Asija, 2016; Hidayat et al., 2017a; Hidayat et al., 2018b; Sapiezynski et al., 2015; Shiravi, et al., 2016). In intersection estimation research today, there is a relationship between Wi-Fi data and travel time. Such intersection estimation research seeks to confirm the accuracy of the Bluetooth and Wi-Fi data on urban roads against reliable travel time results (Shiravi et al., 2016). The research that detects human movement uses high Wi-Fi frequencies, connected with GPS so that the position of the MAC addresses or access points can be identified (Sapiezynski et al., 2015). The use of Wi-Fi and Bluetooth in public terminal transportation has also been applied in a high and wide frequency to capture MAC addresses so the travel behavior of pedestrian patterns can be identified and understood in terms of seconds and minutes (Shlayan et al., 2016). This public terminal transportation research considers high-frequency Wi-Fi detected data compared with Bluetooth data. The reliability of travel time using Bluetooth has been investigated to identify the Bluetooth ability to detect MAC addresses (Araghi et al., 2015). In such studies, data processing was conducted by dividing the detection zone and detection time. Another empirical evaluation of Wi-Fi was conducted on road transport. The method used was to detect “exit to exit” with a procedure filtering the data with time as the main variable (Abbott-jard et al., 2013). “Exit to exit” is intended to be the “beginning” and the “end” of each MAC address identification time. Other travel research using static equipment has been conducted for bicycle users with time and speed filtering data processing to confirm the penetration rate of Wi-Fi data (Böhm et al., 2016; Ryeng et al., 2016). Research on the time travel estimation process is important in confirming the accuracy of the Wi-Fi data (Hidayat et al., 2018b).

GIS software explains spatial data distribution, in the case of transportation, this is through the use of open-source QGIS software. There are several studies related to the use of GIS for transportation and travel time. The distribution of Wi-Fi spatialized data can be identified if the data have specific XY coordinates captured through the Wi-Fi scanner and GPS (Feng & Liu, 2012; Odiyo, 2014). The GIS is also more efficient for wireless data deployment in the development of trade and services or urban planning (Aldasouqi & Salameh, 2014). Furthermore, research on travel diary data has been conducted based on travel time using GIS. In such research, the analyses use “starting” and “ending” person-trip data in combination with spatiotemporal data (Yu & Shaw, 2004). An analysis of “day-to-day” variations in travel time using GPS connected to a notebook PC can capture vehicle movement over multiple days. One study describes the tracking of vehicle movements using a GPS device based on travel time and travel speed (Ohmori et al., 2002). Path GIS research uses spatial trajectory analysis on all of the GIS data points obtained for vehicle movements (Zambrano et al., 2016). More specifically, examining the logical path of tourist movements using GPS data from tourism spots, the travel time data for each tourist (Meng-Lung Lin et al., 2009) or the fastest and shortest travel times can be classified using GIS modelling (Abousaeidi et al., 2016; Ilayaraja, 2013).

This research uses a Wi-Fi scanner to retrieve data from Wi-Fi users. For the study, the Wi-Fi scanner was placed on a bus to capture the MAC addresses of the bus users and non-bus users. The difference between this study and previous research is that this study uses non-static Wi-Fi placed on a moving vehicle. The data results are analyzed using a GIS procedure. An important variable in this research is “time,” which measures how long each MAC address is traveling. Travel time is essential for the identification and confirmation of the bus versus non-bus passengers. This study conducts a “point to path” procedure to
analyze the origin of destination or “beginning” and “end” of MAC address identification based on the travel time variable. The aim of a “point to path” analysis with a travel time variable is to estimate bus passenger usage.

2. DATA AND METHODS

2.1. Location
A survey was conducted in Obuse in the Nagano Prefecture in Japan (Figure 1). The Obuse bus is a tour bus that delivers passengers to a tourism spot in the town of Obuse. Obuse was selected for the Wi-Fi field test because Obuse is a tourism area with a hop-on-off bus system. This system is usually characterized as a per day, time pay system.

This field test was done by using a Wi-Fi scanner placed on the Obuse Bus from 09:50 to 17:10 (Hidayat et al., 2017a; Hidayat et al., 2017b; Hidayat at al., 2018a; Hidayat et al., 2018b; Terabe et al., 2017). The Obuse bus makes nine stops and the distance between each bus stop is about 500 meters or three minutes. The Obuse bus also makes seven loops, each called circulation time (CT), from bus stop one to bus stop nine and then starts back at one. Thus, the nine bus stops can categorize Obuse as a bus stop tourism spot.

![Figure 1. Obuse orientation map and bus route map](image)

2.2. Wi-Fi Scanner Equipment and Installation
As stated, the Wi-Fi scanner equipment captures MAC addresses of devices such as smartphones, laptops, tablets, computers, and other Wi-Fi-enabled devices. The MAC address only shows a unique identification for each device and it does not display personal data.

The Wi-Fi scanner equipment includes an antenna, GPS, and a mobile battery (Figure 2). This Wi-Fi scanner has an approximate range of about 200-300 meters and it detects bus passenger Wi-Fi-enabled devices, buildings, vehicles, and pedestrians (Hidayat et al., 2017a, Hidayat et al., 2017b, Hidayat et al., 2018b). The scanner uses a mini Raspberry Pi computer. This is a quad-core processor-powered single board computer running at 900MHz and the system has 1 GB RAM capacity. It also has a USB port, a pole stereo output, a video port, and an HDMI port, plus a micro SD port for loading the operating system and storing data. The scanner includes GPS Tracking BU-353 with high frequency. In terms of electricity, it is powered by a mobile battery 30,000 MAH, so it can be active up to 12 hours.

The Wi-Fi scanner was placed inside the bus near the driver. After the survey ended, the scanner was turned off and the MAC address recorded data was downloaded for further analysis. The data became raw data that would be confirmed for each travel time.
3. RESULTS AND DISCUSSION

This section describes the processing of the raw data into travel time data. This procedure is the cleaning of the Wi-Fi data. The Wi-Fi log data are in the form of time and MAC addresses, while the GPS log data are in the form of time, latitude, and longitude. There are seven processing steps for the Wi-Fi data, which start with the raw data and include combining the GPS and Wi-Fi data, converting the coordinates, inserting "point to path" QGIS analysis, analyzing travel time, confirming bus passengers based on circulation time, and validating or comparing driver data and Wi-Fi confirmed data. Figure 3 presents a chart of the Wi-Fi cleaning process converting the data into travel times. This procedure is performed using Python and QGIS open-source software. The Python software makes it easier to analyze the amount of Wi-Fi data. All analyses use open-source Anaconda Python 3.0 and QGIS 3.0 Girona applications.

3.1. The First Step Combines Data

The Wi-Fi scanner provides the Wi-Fi log data and the GPS provides the GPS log data. These data are still distinct, so they need to be merged so that each MAC address has a position or coordinate. The position and coordinate easily track the MAC address. To merge these data, the Pandas Python package and the concept command are used, with the MAC address as the merging "key." The screenshot in Figure 4 shows the structure of combining the Wi-Fi and GPS log data.

<table>
<thead>
<tr>
<th>Out[7]:</th>
<th>Time</th>
<th>Mac</th>
<th>Latitude</th>
<th>Longitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>17:01:04</td>
<td>91f6b77567567567564b2125d955a14</td>
<td>3641.9339</td>
<td>13818.7818</td>
</tr>
<tr>
<td>1</td>
<td>17:01:04</td>
<td>91f6b77567567567564b2125d955a14</td>
<td>3641.9339</td>
<td>13818.7818</td>
</tr>
<tr>
<td>2</td>
<td>17:01:04</td>
<td>91f6b77567567567564b2125d955a14</td>
<td>3641.9339</td>
<td>13818.7818</td>
</tr>
</tbody>
</table>

Figure 4. Screenshot Python - combining data
3.2. The Second Step Converts the Coordinate System

The data that have been merged need to be converted to UTM from decimal degrees. This change is done so that it is easier to calculate the distance the MAC addresses travel in the next analysis. The coordinate data transformation uses the Python’s Geo-Pandas module. The Figure 5 screenshot shows the structure of the data once they are converted. These data should appear in the QGIS interface. The data are entered as CSV-shaped so that the "add" data is a “delimited” file and the set of X and Y coordinates are the spatial positions. The UTM reference used is UTM WGS 84 Zone 54N. The MAC address data appear on the QGIS interface as data points (Figure 6).

<table>
<thead>
<tr>
<th>Out[13]:</th>
<th>Time</th>
<th>Mac</th>
<th>X</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>17:01:04</td>
<td>91f4be675d879d0b1b2125d9f93a14</td>
<td>259959.323048</td>
<td>4.064840e+06</td>
</tr>
<tr>
<td>1</td>
<td>17:01:04</td>
<td>91f4be675d879d0b1b2125d9f93a14</td>
<td>259959.323048</td>
<td>4.064840e+06</td>
</tr>
</tbody>
</table>

Figure 5. Screenshot python – UTM conversion

3.3. The Third Step is “Point to Path (PtP)” Analysis

Point to path (PtP) is one tool in QGIS that "connects the dots" based on a common attribute and a sequence field. The attribute field determines which points should be grouped together into a line (QGIS, 2011; Sherman, 2011). The sequence field determines the order in which the points will be connected. Before the analysis, PtP plugins must be installed first on the menu QGIS managed plugin. This tool has three variables: “group,” “begin,” and “end” (Sherman, 2011).

- Group - the name of the ID/MAC address taken from the field, we chose as the group field
- Begin - the time value of the first point order field used to create the path
- End - the time value of the last point order field used to create the path

In PtP analysis, two crucial factors must be considered when entering “group” and “time.” “Group” represents specific data based on the merged data and “time” is the time input. The “group” data capture the MAC addresses, while the “time” data capture the time-shaped column: “hour,” “minute,” and “second.” The results show the OD line of the movement of each MAC address. New attributes of the output data include the “begin” (time journey begins) and “end” times (time end of trip). “Begin” and “end” for each MAC address are used to analyze the travel time of each MAC address (Figure 7 and 8).
3.4. The Fourth Step is Time Travel Analysis

Time data were analyzed in Python with the equation “travel time = end – begin,” and show the number of seconds and minutes for each MAC address. Furthermore, the MAC addresses are divided by using the time classification as shown in Table 1. Once divided by time classification, the numbers of MAC addresses are shown that can be confirmed as passengers. This study uses three minutes because the distance between each bus stop is about three minutes and the distance for one loop (from bus stop one to nine and back again) is about 40 minutes (Figure 9a).

<table>
<thead>
<tr>
<th>No</th>
<th>Travel Time (minutes)</th>
<th>Classification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>&gt;40</td>
<td>Non-Passenger</td>
</tr>
<tr>
<td>2</td>
<td>&lt; 3</td>
<td>Non-Passenger</td>
</tr>
<tr>
<td>3</td>
<td>3-40</td>
<td>Passenger</td>
</tr>
</tbody>
</table>

3.5. The Fifth Step Confirms the Passengers

This stage confirms the passengers who use the bus that travels seven CTs from 9:50 to 17:10. In this process, the MAC address travel times are shown as well as the bus circulation times. The MAC addresses outside the CT can be removed. MAC addresses can be defined as passengers who are in the zone of CT. The bus circulation times are taken from the Obuse bus timetable or the bus stops (Table 2) (Figure 9b).

<table>
<thead>
<tr>
<th>Bus Circulation Time</th>
<th>CN Begin</th>
<th>CN End</th>
</tr>
</thead>
<tbody>
<tr>
<td>CT 1</td>
<td>9:50</td>
<td>10:40</td>
</tr>
<tr>
<td>CT 2</td>
<td>10:50</td>
<td>11:40</td>
</tr>
<tr>
<td>CT 3</td>
<td>11:50</td>
<td>12:40</td>
</tr>
<tr>
<td>CT 4</td>
<td>13:20</td>
<td>14:10</td>
</tr>
<tr>
<td>CT 5</td>
<td>14:20</td>
<td>15:10</td>
</tr>
<tr>
<td>CT 6</td>
<td>15:20</td>
<td>16:10</td>
</tr>
<tr>
<td>CT 7</td>
<td>16:20</td>
<td>17:10</td>
</tr>
</tbody>
</table>
3.6. Discussion

The raw travel time result shows the amount of travel time in the range of 0-400 minutes. This result still has to be reclassified based on the bus circulation time. There are 2,000 MAC addresses selected from about 75,000 raw data points for the CT analysis. The Figure 9 histogram shows the frequency of the number of MAC addresses with their travel times (Sridhar, 2008). After confirmation, based on the bus CT, the MAC addresses that are estimated as passengers equate to almost 120 MAC addresses with their frequency distribution as in Figure 9. From the histogram, we can see that more than 60% of the travel time is from 3-15 minutes, with the rest in the range of 15-40 minutes. A ground count and analysis test the data validity. Furthermore, the ground “truth” results are taken from driver data. The PtP result compared with driver data show that the difference trend is not too significant between the PtP data and ground data. Between 10:50 am and 11:40 am and 11:50 am and 12:40 pm passenger data tends to be high. This is because the morning before noon is a better time to travel. At other times, the circulation tends to decrease such as in early morning and late afternoon. This analysis is in line with some previous results that provides an illustration of how the PtP procedure can be used to calculate the number of passengers on a bus using Wi-Fi data (see Abousaeidi et al., 2016; Ilayaraja, 2013).

![Figure 9. Travel time histogram of differences between before and after the process](image1)

![Figure 10. Comparison between driver data with a PtP procedure](image2)
In Figure 10, the comparison between the driver field data and the Wi-Fi procedure results shows that CT1, CT2, CT5, CT5, and CT7 have insignificant value differences compared with CT3 and CT4. During CT3 and CT4, there is a significant number of visitors to Obuse using the bus or detected around the bus (Hidayat et al., 2017a, Hidayat et al., 2017b, Hidayat et al., 2018b). These time frames (circulation times) capture the highest number of passengers, which tends to be very significant from morning until noon. Passengers begin to decrease from noon into the afternoon. The difference in the number of passengers between the driver and Wi-Fi data is in the 5-10 passenger range.

4. CONCLUSION

The results imply that PtP analysis is advantageous (easy, instantaneous, intuitive) when used to process Wi-Fi scanner data, in particular, and Big Data in general. Travel time data can be categorized into two types: namely, passenger and non-passenger data. The analysis can be used as part of the development of smart city-based transportation and Big Data projects. Based on the comparison between driver data and Wi-Fi confirmed data, there was not a significant difference in the number value between these. The data cleaning process still needs to be developed with various additional analyses to get the confirmed Wi-Fi data closer to the field data. Further research can be fine-tuned to cover the non-passenger parts such as pedestrians, vehicles, and buildings and the making of an OD matrix for passenger data. The origin of the movement is still a straight line, which should be based on the route, so the distance calculations still include errors.

5. ACKNOWLEDGMENTS

The authors would like to thank all those who have contributed to this paper. Thank you to members of the Transportation Planning Laboratory, the Department of Civil Engineering, and the Tokyo University of Science, which provided surveys, data, and extensive assistance in support of this paper. Thank you to my workplace Universities Technology Sulawesi-Indonesia, my scholarship from the Ministry of Research, Technology and Higher Education and the Indonesia Endowment Fund for Education (LPDP) Ministry of Finance, The Republic of Indonesia. Thank you also to all the reviewers who provided corrections for this paper.

6. REFERENCES


Cunche, M. (2014). I know your MAC address: targeted tracking of individual using Wi-Fi. Journal of Computer Virology and Hacking Techniques, 10(4), 219–227. [Crossref]
Dunlap, M., Li, Z., Henrickson, K., & Wang, Y. (2016). Estimation of Origin and Destination Information from Bluetooth and Wi-Fi Sensing for Transit. Transportation Research Record: Journal of the Transportation Research Board, 2595(2595), 11–17.[Crossref]
Hidayat, A., Terabe, S., & Yaginuma, H. (2018a). Determine Non-Passenger Data from Wi-Fi Scanner Data (MAC Address), A Case Study: Romango Bus, Obuse, Nagano Prefecture, Japan. International Review for Spatial Planning and Sustainable Development, 6(3), 154–167. [Crossref]
Hidayat, A., Terabe, S., & Yaginuma, H. (2018b). Wi-Fi Scanner Technologies for Obtaining Travel Data about Circulator Bus Passengers: Case Study in Obuse, Nagano Prefecture, Japan. Transportation Research Record: Journal of the Transportation Research Board. [Crossref]
Ryeng, E. O., Haugen, T., Grønlund, H., & Overå, S. B. (2016). Evaluating Bluetooth and Wi-Fi Sensors as a Tool for Collecting Bicycle Speed at Varying Gradients. Transportation Research Procedia, 14(2352), 2289–2296. [Crossref]
Sapiezynski, P., Stopczynski, A., Gatej, R., & Lehmann, S. (2015). Tracking Human Mobility Using Wi-Fi Signals. PLOS ONE, 10(7), e0130824. [Crossref]


Xia, H., Qiao, Y., Jian, J., & Chang, Y. (2014). Using Smart Phone Sensors to Detect Transportation Modes. Sensors, 14(11), 20843–20865. [Crossref]


