Paper ID #

Exploring the Feasibility of Bluetooth and Wi-Fi Technologies for Measuring Transit Passengers Wait-Times and Origin-Destination Travel Times

John Whittier[1], Saerona Choi[1,2], B. Brian Park[1*], and Samy El-Tawab[3]
1. University of Virginia, USA
* PO Box 400742, Charlottesville VA 22904-4742, 424-924-6347, bpark@virginia.edu
2. DGIST, Republic of Korea
3. James Madison University, USA

Abstract
The goal of this study was to explore a feasibility of obtaining passengers' bus stop waiting times and passengers' origin and destination (OD) travel times using low cost devices (e.g., Bluetooth or WiFi Readers). We developed and demonstrated a low cost Wi-Fi reader based building block that can estimate passengers bus stop waiting times and OD travel times by collecting and analyzing WiFi MAC addresses data.

Our case study, conducted at the University of Virginia, focused on collecting passengers' waiting times and their origin-destination travel times from one bus stop to another. We found that the Wi-Fi reader was more effective than the Bluetooth reader at both producing a large sample size and matching time values to the samples. The Wi-Fi reader was able to extract an average wait time of approximately between 98 and 118 seconds at the Chemistry Building bus stop and around 248 seconds for origin-destination trips from the Chemistry Building to the Garrett Hall. We expect this building block can be scalable to cover large number of bus stops and be implemented in real-time.

KEYWORDS: WiFi Reader, Transit Passenger Waiting Time, Transit Operations

Introduction
As the volume of vehicles have increased on the roadways, traffic engineers have been exploring methods of effectively collecting traffic counts to help determine the best solutions when analyzing current traffic patterns. The problem is that current counting measures are either slow or unable to reflect the intended sample size. With new innovations in the technology field, engineers are wondering if there is a way to incorporate this technology into traffic counts. Bluetooth and Wi-Fi readers are two examples of technologies that have been used to test the viability of the technology in transit systems in terms of measuring transit passenger data. The purpose of this study is to determine if either Bluetooth or Wi-Fi technologies are useful data collection methods to analyze passenger waiting times and origin-destination (O-D) data. The study focused on passengers boarding buses on routes maintained by the University Transit Services (UTS) at the University of Virginia, collecting the Bluetooth and Wi-Fi information from personal cellular devices. From this data collection, we intended to gather information about the effectiveness of these technologies, along with a framework for launching larger data collection efforts around the University of Virginia, and beyond.

Literature Review
With the development of new technologies, such as Bluetooth and Wi-Fi readers, bus riders are beginning to expect instant gratification when using transit services. Some of these riders have started using bus tracking applications to monitor the progress of buses along routes and determine what time a bus will arrive at the nearest location to them (Thiagarajan, 2010). However, there is no way to guarantee that the buses will arrive at the times stated on the applications. That is where studying passenger waiting time and bus travel time studies with Bluetooth or WiFi Readers can be useful. Typical methods for determining origin-destination data include person survey, license plate matching, manual car following, and others, but Bluetooth or WiFi reading can also be a helpful tool for data collection. Bluetooth readers will be placed

along corridors and then be analyzed for matches along the routes (Blogg et al., 2010).

When the research project was initially developed, the primary focus was on Bluetooth data collection. The problem with focusing solely on Bluetooth for data collection is in the literature. In a study testing the validity of Bluetooth in traffic system data collection, it was found that around 18% of the devices were detected, while approximately 80% of devices in the study area had detectable Bluetooth devices (Kostakos et al., 2013). This threshold of study is not very comforting when trying to study a particular data collection method for its usefulness. Even though it could be assumed that this experiment could be an anomaly, we decided to include Wi-Fi as another method that can be discovered and implemented into collecting traffic information from WiFi devices.

One of the biggest obstacles with the inclusion of Wi-Fi as the second data collection method is the inability to locate where a Wi-Fi signal is being detected. Typically, commercial Wi-Fi sensors are able to detect signals within a wide range of distances. In order to alleviate this problem, the signals need to have Wi-Fi signal strength included in their measurements as well. Based on the signals strength information, signals are at excellent strength if the detected values are in a range of -1 decibels to -35 decibels (Crane, 2015). The signals are of good strength if they range between -36 decibels and -67 decibels (Crane, 2015). These values could lead to better determination of the locations of the users in proximity to the data collection box.

One of the bigger justifications for using technologies for data collection is the unreliability of people to collect traffic data. Often, people are susceptible to commit errors, either systematic or random in nature. Even when new technologies implement these data collection methods, sometimes the users do not select the correct data. An example of this trend was an experiment conducted by researchers at the Nanyang Technical University in Singapore. Their experiment used user interface to note when buses were arriving and leaving from stops, along with how long they were waiting at the particular stops (Zhou et al., 2012). The problem is that riders are in control of starting and stopping the data collection device. With user control still being a factor, the error has not been completely eliminated from the system. Even if it is better than following similar riders for longer periods of time to predict their ridership trends (Mishalani et al., 2006), the system does not benefit if people are required to participate in the data collection process. The best way to gather data is to create methods that automatically gather and store the data, which is in the design parameters of the experiment.

The research mentioned above is not a unique idea, but in fact an emerging topic in the field of transportation engineering. In an article released by *Metro Magazine* in February of 2016, the University of Washington has developed a system that uses Wi-Fi and Bluetooth signals to implement real time data about the college's bus services (Dunlap et al., 2016). The set up for this experiment is almost same as the experiments being conducted in this research. They were able to collect the data for a better understanding of service issues and timing of buses. The University of Washington is not the only research University conducting similar studies. The University of Illinois has developed a system based on GPS-location software that collects both media access control (MAC) addresses and geographical position system coordinates for traffic counts (Vu et al., 2010). Their experimental structure is similar to those of ours and the University of Washington (Langston, 2016). One major difference is that Dunlap et al. (2016) have used the WiFi Readers within the buses, while our research relies on WiFi Readers from both in the bus and at the bus stops.

Methodology

Because this study emphasized the needs of the University community, we sent out an initial survey to gather mode choice information and preferences from university students. These questions consist of general information, such as year in school, housing location, and bus route preferences. The survey was distributed by email, social media, and word of mouth. From this information, the researchers gained insight into developing the optimal test route to best service the needs of the students who use the system on a daily basis.

One of the goals of the survey was to determine what routes students at the University of Virginia (UVA) are choosing to ride the most. At UVA, there are 10 different bus loops; the Northline Connector, Inter University Loop, Outer University Loop, Central Grounds Shuttle, Colonnade Shuttle, Green Route, Hereford/IRC Shuttle, Nursing Clinical Shuttle, Stadium/Hospital Shuttle, and the Special Route. Based on the independent survey results, there are three buses that are best service students in traveling to, from, and around Central Grounds, which are the Northline route, the Inner Loop route, and Outer- Loop route. To avoid any differences in route behaviors, we selected to focus on the Northline bus route during the data collection process.

After careful consideration of student ridership trends, we decided to narrow down the study area to focus on the Northline route from the Chemistry Building to the Garrett Hall. This route was chosen because it is a popular route by the student body and has no direct stops between one point and another. The research also only focused one direction, from the Chemistry Building to the Garrett Hall.

The experiment was conducted in a few phases. The first phase consisted of analyzing the bus routes between the Chemistry Building and Garrett Hall bus stops of the Northline of the University of Virginia's University Transit Services (UTS) using both WiFi and Bluetooth Reader technologies simultaneously to compare sample output values for each. Figure 1 shows a study area of this study. The second phase analyzed the route using the Wi-Fi signal readers and personal validation of data points using pencil and paper. The third phase of the experiment looks into tracking a mobile researcher and matching the travel time and wait time information of that particular subject as a proof of concept for whether the technology is effective at detecting a known data point. Each phase was conducted using three Raspberry Pi data collectors, two stationary readers at the bus stop and one mobile device being operated by a researcher along the route. The readers were used to determine sample size output of the two data collector options, passenger wait time, and Origin-Destination (OD) time for passengers utilizing the UTS bus system. Similar study was conducted by our partners at James Madison University (El-Tawab et al., 2016).

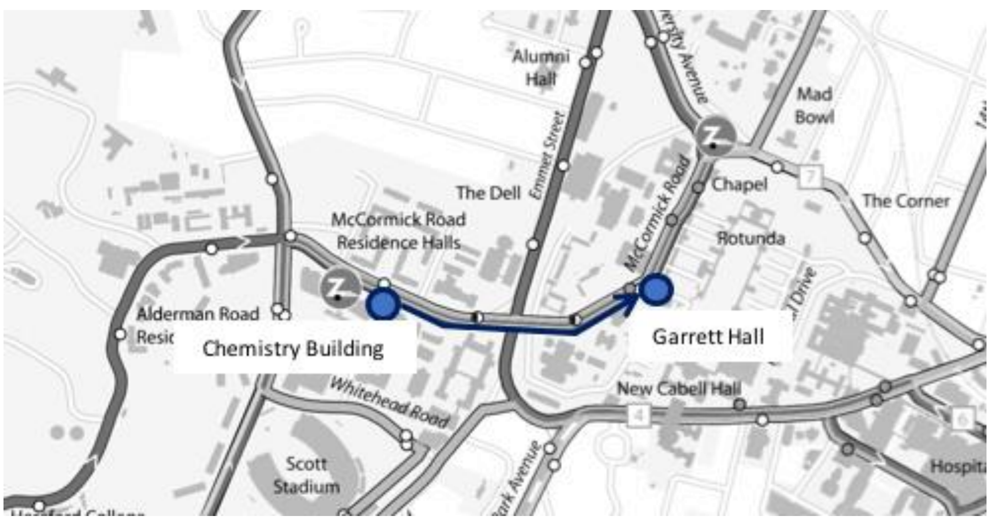

Figure 1. A study area of this study (source: www.virginia.edu/parking/uts/)

All three phases of the experiment were conducted on the same date in March of 2016. This was done to ensure that consistency in time and weather patterns. All the data from this day was analyzed using Microsoft software packages and sorted through to find and analyzed desired results.

Sample Output Data

The Bluetooth and Wi-Fi readers would collect and store data on the Raspberry Pi's (as shown in Figure 2) before being extracted onto computers for analysis purposes. This data was parsed to specific formats to help the user understand the output and analyze the data in the most effective manner possible.

For the Bluetooth reader, the reader scanned for data and stored it every 5 seconds. In the analysis process, we looked for how long a signal was emitted before not being recognized by the reader anymore (out of range). All the data was detected and stored in text files that were then converted over to excel for filtering and storage. The data was logically structured within the Excel document to allow for separating and filtering out of irrelevant information. The data included the python code in the first column, followed by the date, time, and devices detected during that moment in time. The final lines indicated each new MAC addresses that were obtained during the study period. Each new line in the table represents a new five second data collection period that either was unique or showed similar devices to trace waiting time and origin-destination travel time.

Figure 2. Bluetooth and WiFi Readers connected to Rapsberry Pi

The Wi-Fi reader stored information in a different manner entirely. The Wi-Fi reader operated by storing all devices detected during the data collection every five seconds, creating greater lists of data points as the collection continued. Rather than dividing detected devices separately, it printed out a history of these points to the file every new time interval, with updated times if a certain device was detected multiple times. The data was provided in a csv files that was then exported into Excel for data management. Each column in the dataset was setup to provide organization of the data in the system and allow for optimal filtering based on necessary data elements. The output data listed the unique device identifier (MAC Address), time paired, time unpaired, and signal strength. Instead of having to go list by list to determine the points, we needed to look at the final output time and extract the necessary points and timestamps that meet the standards of the experimental data.

To protect the privacy of individuals, we elected not to include any sample output data in this paper due to the MAC addresses being present in the data. It is noted that we have used these addresses solely for counting purposes. No individual MAC address would be traced back as a result of this research. From this data, it helps to understand how to evaluate the nature of bus ridership at the University of Virginia.

Phase 1: Bluetooth vs. WiFi Sample Collection

For the first phase of the experiment, the data was taken from both the Bluetooth and WiFi sample readers to determine which method should be used for the other phases of the experiment. The two readers were run simultaneously on all three Raspberry Pi readers to ensure that the Bluetooth and WiFi Readers were collecting data for the same period of time. The total sizes are shown in Table 1.

Table 1. Comparison of sample sizes for Bluetooth and WiFi Readers on March 21st, 2016

	Bluetooth	WiFi
Chemistry Building	15	1,789
Garrett Hall	14	1,696

From the data presented, it can be seen that the WiFi Reader was able to detect more samples than the Bluetooth readers. Based on the fact that there were far more people riding the buses during those times than were detected by the Bluetooth reader. Based on the desire to analyze which method is more appropriate, we elected to proceed with experimentation based on the WiFi data presented on March 21, 2016.

Phase 2: Origin-Destination Travel Time for Bus Passengers

After determining a better technology to test in the experiment, the next phase involved going through the data from March 21st to determine device matches between the three Raspberry Pi's put out in the field. Each Raspberry Pi was programmed with a specific number to identify that particular device. Table 2 labels the devices by those numbers. Device 65 corresponds to the Chemistry Building bus stop. Device 63 corresponds to the Garrett Hall location. Device 53, located within the bus, was used as a mobile device, tracking devices from one point to the other.

Over the one-hour time frame that data collection occurred, 205 devices were matched, all of which were detected by WiFi Readers. Looking at Table 2, it can be seen that the average trip time, which is defined as the origin-destination time for the individuals, was approximately 4 minutes and 8 seconds. This average time calculated does not take into account the third measurement in Table 2 as it is an outlier in the dataset. The 248 seconds origin-destination time is a reasonable expectation for such a short bus route with no stops between the two data collection points. Having a reasonable average helps to show the validity of using WiFi to collect passenger information on a transit network.

One thing to note is that the experimental data possesses a few outliers in the data set. First, some of the MAC addresses were not detected at only one or two of the devices, which then produced some data points at 0 minutes and 0 seconds. Second, another set of outliers deals with longer detection times that the average headway. There are a few samples in the dataset that well exceed the standard detection time expected for an average passenger riding a bus. This could have resulted from either lingering pedestrians or uneven bus headway due to traffic or other uncontrollable issues. Third, some devices showed that detection time at Garrett Hall was earlier than the Chemistry Building. It is likely that they traveled from the Garrett Hall to the Chemistry Building, which was not analyzed in this research.

Table 2. Passenger Origin-Destination Travel Times

Bus #	Chemistry Building (Device #65)			Garrett Hall (Device #63)		
	MAC Detected	Bus Arrival Time	# pax onboard	MAC Detected	Bus Departure Time	OD Travel Time (sec)
1	17:32:39	17:34	2	17:36:53	17:36	254
2	17:49:09	17:51	9	17:54:10	17:54	301
	17:28:18			17:53:14		1496*

	17:49:09			17:54:10		301
	17:51:11			17:53:53		162
	17:50:24			17:53:16		172
3	18:02:56	18:03	16	18:06:15	17:06	199
	17:59:43			18:06:12		389
	17:59:07			18:06:06		419
	18:02:56			18:06:15		199
4	18:25:17	18:25	4	18:26:47	17:26	90

* an outlier as this person did not board on the bus #1 at 5:34pm while started waiting at 17:28:18.

In addition, from Table 2, we could estimate percent of sample with respect to actual number of passengers. Using the number of passengers boarded and the number of valid MAC addresses detected at both bus stops, the sample size is between 25% to 50%. It should be noted that MAC addresses from the stop before the Chemistry building, in which not available in this experiment, can confirm that the MAC addresses detected at the Chemistry building were not the passengers already on the bus. Similarly, MAC addresses from the stop after the Garrett Hall could confirm that MAC addresses detected at the Garrett Hall were indeed exited the bus.

Phase 3: Tracking of Researcher's Mobile Device

In order to validate the process, we decided to track a known device along the route to see if the data was being properly recorded and plausible when compared to experimental data.

To help establish the validity of both the field detection and device detection, we used the device of Raspberry Pi #53 (i.e., a mobile WiFi Reader placed within the bus) to simulate if the technology can detect and find an individual person in the network. The results of tracking that individual sample can be seen in Table 3. From the data, it can be seen that the device matches the boarding and departure times within an acceptable threshold of 1 minute for the times when the mobile devices were detected.

Table 3. Tracking of Researcher's Wireless Device Compared to Field Verification

Bus #	Time on Bus – Field Verification	Time off Bus – Field Verification	Estimated Trip Duration (sec)	Time on Bus – Reader Verification	Time off Bus – Reader Verification	Duration of Trip (sec)
1	17:34	17:36	120-180	-	-	-
2	17:51	17:54	180-240	17:51:39	17:54:29	170
3	18:03	18:06	180-240	18:03:41	18:06:54	193
4	18:25	18:26	60-120	-	-	-
	Average		160	Average		181.5

Phase 4: Bus Stop Waiting Times

Table 4 includes field verification of bus stop waiting times along the route for the data collection period. Please note that the "Chemistry Building" and "Garrett Hall" columns indicate points detected at either 65 or 63, respectively, and it was verified by device 53 whether the passengers took the bus or not. It is noted that we could also verify this without using the device 53 placed within the bus by setting up a threshold value of bus travel time.

For these routes, it was found that passengers' waiting time varied between 14 seconds to 3 minutes and 31 seconds, after removing apparent outlier showing 22 minutes 53 seconds waiting time. About the very short waiting time such as 14 seconds, it is possible that the device was detected when the bus arrives. It is very likely that the device inside bus could have been detected. It is noted that this can be screened if we had installed WiFi Reader at a previous bus stop. It is noted that if a smartphone (e.g., iPhone) is on power-saving mode, it may take more time to detect the MAC address of this phone.

Table 4. Bus stop waiting times using field counts of passengers recorded

Bus #	Chemistry Building (#65)				Garrett Hall (#63)	
	MAC Detected	**Bus Arrival Time**	**MAC Disappeared**	**Bus Stop Waiting Time**	**MAC Detected**	**Bus Departure Time**
1	17:32:39	17:34	17:33:58	0:01:19	17:36:53	17:36
2	17:49:09	17:51	17:51:13	0:02:04	17:54:10	17:54
	17:28:18		17:51:11	0:22:53*	17:53:14	
	17:49:09		17:51:21	0:02:12	17:54:10	
	17:51:11		17:51:35	0:00:24	17:53:53	
	17:50:24		17:51:27	0:01:03	17:53:16	
3	18:02:56	18:03	18:03:14	0:00:18	18:06:15	18:06
	17:59:43		18:03:14	0:03:31	18:06:12	
	17:59:07		18:03:49	0:04:42	18:06:06	
	18:02:56		18:03:30	0:00:34	18:06:15	
4	18:25:17	18:25	18:25:31	0:00:14	18:26:47	18:26

* an outlier as this person did not board on the bus #1 at 5:34pm while started waiting at 17:28:18.

Conclusions

The experiment conducted in this research shows that WiFi Reader is an effective means of detecting passengers waiting and boarding buses in a transit network. The data shows that it can collect and validate meaningful amounts of data on passenger trips. Using the known individual device demonstrates the reliability of collecting data on a small portion of the large data collection network. Even though the data had some outliers, it could be verified by simple data analysis using data collected by multiple WiFi Readers. Using the proposed method, we expect that passengers' waiting times and origin destination travel time could be collected in real-time by collecting and analyzing their handheld devices' WiFi MAC addresses.

Future studies should look into expanding the network to include a larger network incorporating more stops to explore popular trends. The study could also test for peak travel times on the system to indicate where improvements would need to be made to best serve passengers' need. Finally, instead of having a portable reader, we should explore placing stationary reader on each bus stop along the selected bus routes and have the collected data being sent to the cloud-based database (e.g., Amazon Cloud Services) for real-time processing. A comparison between the data collected at University of Virginia and our partners at James Madison University can be very beneficial.

Acknowledgements

This project is in part supported by 4VA (http://4-va.org) grant and the GRL program though the NRF funded by the Ministry of Science, ICT & Future Planning (2013K1A1A2A02078326). We would like to thank Professor Joyoung Lee at NJIT for providing technical support on the implementation of WiFi Reader, and Seongah Hong and Huifu Jiang for their help with the field data collection on March 21, 2016.

References

1. Blogg, M., Semler, C., Hingorami, M., and Troutbeck, R. (2010). Time travel and origin-destination data collection using Bluetooth mac address readers. *Australasian Transport Research Forum:* 1-15.
2. Crane, J. (2015). Acceptable Wi-Fi signal strengths. *MetaGeek.* Retrieved 7 April, 2016, from https://support.metageek.com/hc/en-us/articles/201955754-Acceptable-Wi-Fi-Signal-Strengths.
3. Danalet, A., Farooq, B., & Bierlaire, M. (2014). A Bayesian approach to detect pedestrian destination-sequences from WiFi signatures. *Transportation Research Part C: Emerging Technologies, 44;* 146-170.
4. Dornbush, S. & Joshi, A. (2010). StreetSmart traffic: discovering and disseminating automobile networks using VANETs. *University of Maryland Baltimore County Department of Computer Science and Electrical Engineering:* 1-5.
5. Dunlap, M., Li, Z., Henrickson, K., and Wang, Y. (2016). Estimation of origin and destination information from Bluetooth and Wi-fi sensing for transit. *TRB 2016 Annual Meeting;* 1-14.
6. Kostakos, V., Camacho, T., & Mantero, C. (2013). Towards proximity-based passenger sensing on public transport buses. *Personal & Ubiquitous Computing, 17(8)*, 1807-1816.
7. Leccese, F., Cagnetti, M., & Trinca, D. (2014). A smart city application: a fully controlled street lighting isle based on raspberry-pi card, a ZigBee sensor network and WiMAX. *Sensors, 14,* 24408-24424.
8. Mishalani, R., Mccord, M., & Wirtz, J. (2006). Passenger wait time perceptions at bus stops: empirical results and impacts on evaluating real-time bus arrival information. *Journal of Public Transportation, 9*(2), 89-106.
9. Thiagarajan, A., Biagioni, J., Gerlich, T., & Eriksson, J. (2010). Cooperative transit tracking using smart-phones. *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems,* 85-98.
10. Vu, L., Nahrstedt, K., Retika, S., & Gupta, I. (2010). Joint Bluetooth/WiFi scanning framework for characterizing and leveraging people movement in university campus. *Department of Computer Science, University of Illinois;* 1-10.
11. Zhou, P., Zhang, Y., & Li, M. (2012). How long to wait?: predicting bus arrival time with mobile phone based participatory sensing. *Nanyang Technical University, Singapore*; 1-14
12. S. El-Tawab, R. Oram, M. Garcia, C. Johns and B. B. Park, "Poster: Monitoring transit systems using low cost WiFi technology," 2016 IEEE Vehicular Networking Conference (VNC), Columbus, OH, 2016, pp. 1-2.