TP0905

Investigating Transit Passenger Arrivals using Wi-Fi and Bluetooth Sensors

Abdullah Kurkcu[1*], Kaan Ozbay[1], Kuo Ren[1]

1. UrbanITS Center, Tandon School of Engineering, Center for Urban Science + Progress (CUSP), New York University (NYU), USA, ak4728@nyu.edu

Abstract

This paper developed a methodology to relate passenger data collected by Wi-Fi and Bluetooth sensors to scheduled bus departures for the purpose of studying passenger arrival behavior. One major advantage of using these sensors is their simplicity and cost. For stations at which AFC systems are not available, wireless sensors can easily be deployed. The developed Wi-Fi and Bluetooth sensors require are simple to operate, easily transported, and require minimal maintenance. The results pointed out that passenger arrival times at a transit stop are sensitive to the service frequency as it proposed in the literature. Furthermore, Wi-Fi and Bluetooth sensors can be a cost-effective alternative to understanding further and analyze the probabilistic distribution of passenger arrivals and wait times for stops and services where automated passenger data are not available to decision makers and researchers.

Keywords:

Passenger Arrivals, Wi-Fi & Bluetooth, Transit Planning

1. Background and Motivation

Passenger arrival rates at transit stations play a critical role in developing, planning, and coordinating public transportation schedules. The earlier efforts in the field of transit reliability assumed that passenger arrivals to the transit stops are independent of the vehicle departure process [1-3]. However, empirical evidence suggested that passengers may time their arrivals to match the scheduled bus departure under certain conditions [4]. The passenger arrival distribution has an effect on public transport stability. Large variations of passenger arrivals may cause schedule instability which will delay transit vehicles [5]. By investigating the arrival behavior, two main categories of passengers were established in the literature [4-8]. The first group of passengers is the ones who time their arrival to match the scheduled transit service. The other group comes randomly to a stop, thus, having to wait for a longer duration of time on average. However, the proportion of schedule-aware transit users declines as the headway or reliability decreases [6]. It is challenging to collect the data which can be used for analyzing passenger arrival behavior. Many scholars found out that the passengers' wait time at a station is a critical factor that affects passengers' decisions about the time of arrival.

Jolliffe and Hutchinson [7] identified three types of passengers based on their arrival patterns: The first type is passengers whose arrival time is causally coincidental with the bus. The second type is passengers who arrive at the optimal time, and the last one is passengers who arrive at totally random. Their results indicated that arrival times of buses and arrival times of passengers were strongly associated. Moreover, they observed that as headways increased, the ratio of non-random passenger arrivals increased. Assuming that passengers arrive at a bus stop randomly, Newell [9] and Mohring [10] proposed the model for average time that passengers have to wait before a bus comes:

$$E[W]=\frac{E[H^2]}{2E[H]}=\frac{E[H]}{2}(1+cv_H^2) \tag{1}$$

where W is the expected waiting time for a passenger arriving randomly, H is the service headways, $E[H]$ is the expected headway between buses, cv_H is the coefficient of variation defined as

$$cv_H=\frac{\sigma_H}{E[H]} \tag{2}$$

where σ_H is the standard deviation of H.

Bowman and Turnquist [4] developed a model to evaluate the sensitivity of passenger wait times with respect to service frequency and reliability. Their model contained a passenger decision-making process which represented a significant improvement in the explanation of observed wait times. The choice model indicated that passenger wait time is much more sensitive to schedule adherence improvements and less sensitive to services frequency.

Chang and Hsu [11] found out that the arrival behavior is different between intercity transit passengers and urban bus system passengers because intercity transit system usually comes with a long headway and fixed schedules. Their results indicated that the reliability of the bus system has a significant influence on passengers' waiting time. Fonzone, Schmöcker [12] suggested a model of passenger arrivals assuming that passengers tend to minimize their wait times and they may miss the bus because of service irregularity. Their research considers passengers' perceived departure time of buses. Passengers take the schedule and potential early or late departures into consideration to minimize wait time. Their main contribution is that non-uniform arrival patterns can cause severe bus bunching.

Avineri [13] introduced a decision-making model based on prospect theory that calculates cumulative prospect value of wait times. He found out that if passengers receive information concerning buses' headways, they could get confused by such information and are influenced to make a decision. Therefore, they concluded that passengers should not be provided with bus headway information. Neumann, Kaddoura [14] proposed a model to explore more realistic passenger arrival behavior. They mainly consider three different aspects concerning arrival behavior: passengers' degree of learning, service reliability, and bus headway. The result indicates that passengers consider the bus behavior based on their experience such as bus arrives late on purpose or it takes the time to handle other passengers. Therefore, passengers plan to add a buffer time and try to arrive at transit stops to minimize wait time. Their result also showed that different headways have different impacts on passengers: passengers arrive more randomly at a smaller headway. Similar to transit stops, Park and Ahn [15] mentioned that passenger arrival patterns at airports also depend on flight departure times. In addition, they investigated the effects of different types of aircraft and load factors on the passenger arrival behavior.

The relationship between passengers arriving behavior and bus departure time has also been studied since the 1970s. Although early studies in the literature used manually collected data sources [7, 16, 17], Automated Fare Collection (AFC) systems are used to assess the day-to-day consistency of passenger arrival behavior in the recent studies [6, 18-20]. AFC provides the disaggregate passenger journey data which allows researchers to analyze longitudinal aspects of arrival behavior conveniently. However, not every transit service is equipped with AFC systems. Moreover, bus passengers generally use their magnetic stripe cards when they get on board. Since the passenger arrival time cannot be logged, it may not be feasible to use AFC data for bus passenger studies. Therefore, the data collected by Wi-Fi and Bluetooth sensors can be used to analyze passenger behavior at bus stops. Wireless traces of passengers can provide information for passenger arrival times, wait times, and day-to-day variation of passenger arrival behavior.

The goal of this paper is to use data collected from low cost and ubiquitous customized sensors [21, 22] for understanding bus terminal operations and develop probabilistic distributions of passenger arrivals and wait times. Understanding passenger arrival behavior allows decision makers and researchers to execute transit plans that are based on realistic behavioral assumptions. The next section of this paper describes the developed sensors, the methodology to utilize and analyze wireless data collected by the sensors. Section 3 summarizes study results and describes a passenger arrival distribution model. The last section presents conclusions and discussion.

1. Methodology

This section proposes a new approach for analyzing passenger arrivals with respect to scheduled departure times by using wireless sensors data. Wi-Fi and Bluetooth sensors can detect the proximity of personal electronic devices when mobile devices are actively looking for other devices or networks. Wireless sensors can not only detect the mobile devices around them but also non-mobile devices, access points and other networks disseminating their presence. Therefore, MAC addresses, unique identifiers, only belonging to mobile devices are used to capture bus passengers in this paper. Sensors query the MAC address vendors online during data collection [23]. After retrieving the vendor name, the first seven digits of the MAC address are deleted, and the last five digits are encrypted and stored in the database as well as the vendor name. The following sections describe the technical aspects of the sensors, data collection, and analysis.

1.1. Wi-Fi and Bluetooth Sensors

Raspberry PI [24] is a low cost small sized computer which can run Linux [23] operating system and, use a standard keyboard and mouse. PI supports object-oriented programming languages such as Scratch and Python. In addition, PIs can be built to have the ability to interact with the environment. Raspberry PIs and similar mini PCs have been utilized as a sensor to detect motion [25], measure the noise [26] and air quality [27], for environmental monitoring [28] and various other smart city applications [29, 30] in the literature. The minimal installation of Raspbian, Linux operating system, is installed. Figure 1 shows the sensor in a custom-design case. Specifically, Raspberry PI 2 Model B, the second generation Raspberry PI, is used in the study. It has a 900MHz quad-core CPU, 1 GB RAM, 4 USB 2.0 ports, 40 GPIO pins, full HDMI port, Ethernet port, 3.5 mm audio jack, camera interface (CSI), a display interface (DSI), micro SD card slot, and VideoCore IV 3D graphics core. 16 GB SD card is installed to the sensor. The Raspberry PI can be powered by a +5.1V micro USB supply. Typically, the model B uses between 700-1000mA depending on what additional gadgets are connected. The maximum power it can use is 1 Amp. With the existing configuration, the sensor can run up to 12 hours on a 10,000mAh battery. One sensor with all the components costs approximately $75-100 depending on the configuration.

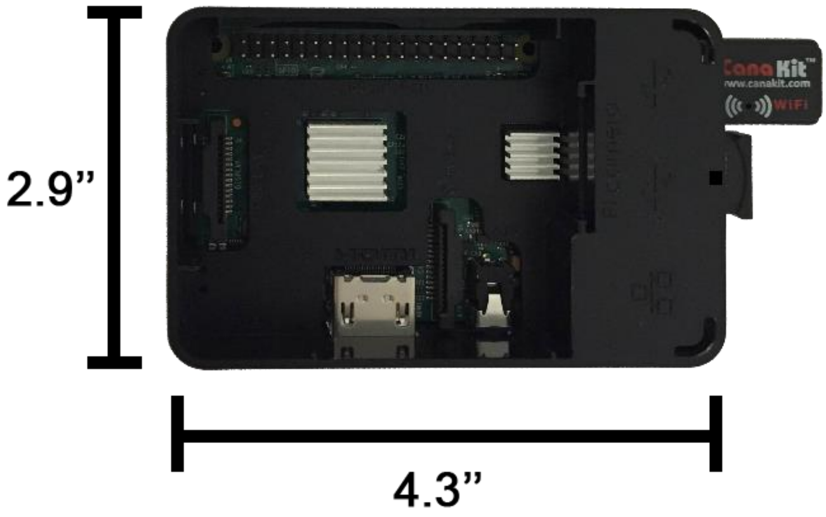

Figure 1 - Raspberry PI

1.2. Data

Six sensors collected data for ten weeks at a major bus terminal in New York City between May 2nd, 2016 and July 10th, 2016. 4 of the sensors are located at the bus gates, and the other two are located at the entrances of the terminal. The data from the busiest gate are selected for further analysis and filtered to have detections from mobile devices only. Figure 2 below shows the comparison of the smartphone brands in the market, reported by Parks Associates [31], and the brands identified by the sensors in the collected data at a bus stop. The same firm reported that almost 70% of U.S broadband households express interest in Wi-Fi based mobile service plans [32].

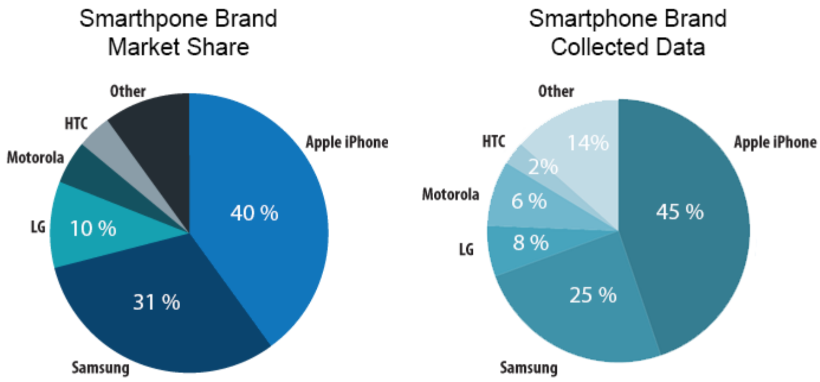

Figure 2 - Smartphone Brand Market Shares [31]

The signal strength (RSSI) of the devices can be used to create a detection zone. With tuning, this creates a circular detection circle that, when crossed, can trigger the detection system to store the information such as detection type (Wi-Fi or Bluetooth), encrypted MAC address, timestamp, and brand in SQLite tables. An illustration of the data that can be collected is shown in Table 1. There are more than 50 thousand detected unique mobile devices in the dataset for a 10-week period.

Table 1 - Sample Data

Type	Mac	RSSI	Timestamp	Brand
WiFi	...	-42	2016-05-16 16:25:22	Apple
WiFi	...	-61	2016-05-02 16:18:51	Samsung
Bluetooth	...	0	2016-05-02 16:19:25	Apple
WiFi	...	-64	2016-05-02 16:20:06	HTC
WiFi	...	-75	2016-05-02 16:20:44	Apple

1.3. Data Pre-processing

Instead of keeping all the digits of the MAC address, only the last three octets are stored to protect the embedded private information. Within the last three octets-6 digits, the last five digits are kept and then encrypted with an encryption key and stored on the instrument. This technique delivers an extra layer of protection. It also preserves the MAC addresses unique for approximately 96% of the cases. The encryption key is randomly generated on a remote server. After the initial key is generated, it is then encrypted again before uploading to the devices on site.

The data is first filtered for weekdays since the bus schedules, and routes change for weekends. The same device can be detected multiple times by the sensor depending on the frequency of the network requests sent by the mobile device. Such detections are grouped by the unique identifier, and the timestamp of the first detection of the device is used as the passenger arrival time. To create a detection zone that is approximately 20 ft. radius around the sensor, data are filtered for RSSI values of -80 or higher. While processing the raw data, how many different weekdays each unique identifier seen in the database is also stored. Having such information made it possible to investigate whether there are any differences between regular and irregular users of the same service.

The data are filtered for the periods between 6-9:00 AM, and 5-8:00 PM to compare the effects of different bus headways and time periods on the same transit service. While the average bus headway in the morning was 20.6 minutes, it was 6.2 minutes in the evening period.

2. Results

2.1. Wait Times

The average wait times at the selected gate are found to be 541 and 195 seconds for AM and PM peak periods respectively. More than 10 thousand data points are used in the wait time calculations. Figure 3 shows the density of wait times during AM peak period. Since the average headway in the morning period is 20 minutes, it is expected that the majority of passengers arrive right before the scheduled bus departure time. It can be seen from Figure 3 that most people wait shorter times favoring such expectation. However, there are also some passengers who wait 10-15 minutes for the next bus. Although this behavior is normally not expected for longer headways, it can be explained by the nature of the bus facility. It is one of the most important transportation hubs in the city, and most passengers travel to the bus facility using other means of transportation. When the headway is longer, it is more likely that passengers have missed their bus departure due to delays experienced at during their trips to the terminal from their origin. These delays may introduce some randomization of the passenger arrival times.

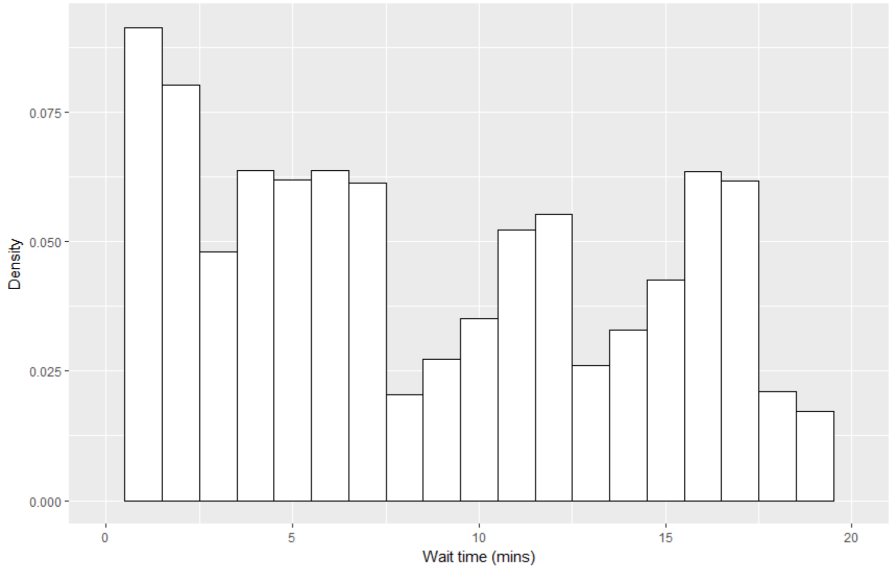

Figure 3 - The Density of Wait Times – AM Peak Period

On the other hand, increased random arrival behavior is expected for shorter headways. Figure 4 shows the density of wait times for PM peak period. Multiple normality tests showed that the probability distribution of the PM peak wait times does not follow a normal distribution (Kolmogorov-Smirnov test, D=0.996; $p<0.05$, Shapiro-Wilk normality test, W=0.9497, $p<0.05$).

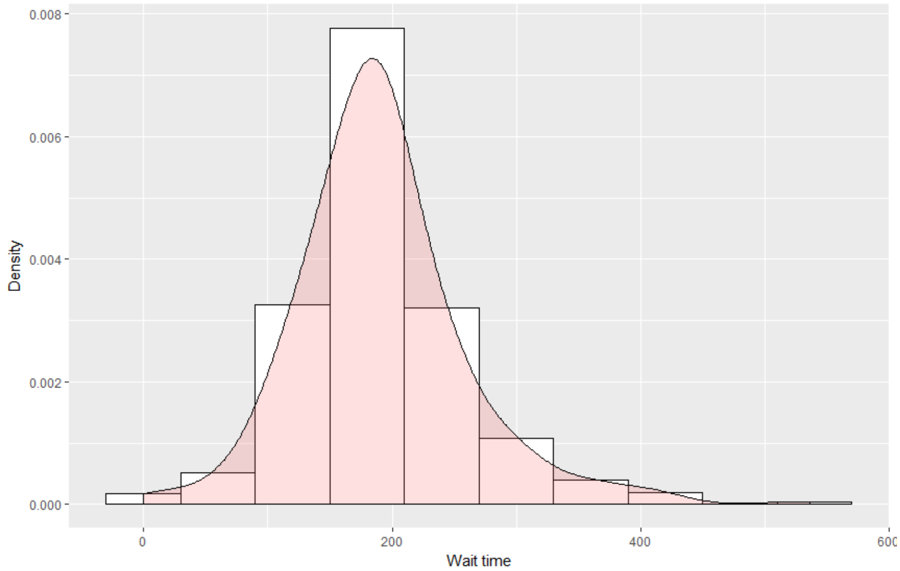

Figure 4 - The Density of Wait Times – PM Peak Period

Predicted wait times are also calculated for AM and PM peak periods using equation (1). The comparison of wait times can be seen in Table 2.

Table 2 - The Comparison of Observed and Estimated Wait Times

Period	Observed Average Wait (min.)	Predicted E[W] using Random Arrival Model (min.)
AM	9.0	10.7
PM	3.3	3.1

To check whether visiting the same bus stop multiple times impacts the experienced wait time, a new column added to the dataset showing the number visits for each unique identifier for different weekdays within the study period. The highest number of visits is recorded to be 35 out of 50 weekdays. It is worth to mention that sensors may not be able to capture the same passenger every day due to weak wireless signals, devices that are detectable at a low frequency, less frequent network requests, obstructions in the study area, and short living network addresses. Figure 5 illustrates AM peak period wait times for the passengers who used the gate from 1 to 28 different weekdays in 10 weeks. Although there is some variation in the calculated wait times, there is no significant trend which proves that familiar passengers have shorter wait times for the AM peak period.

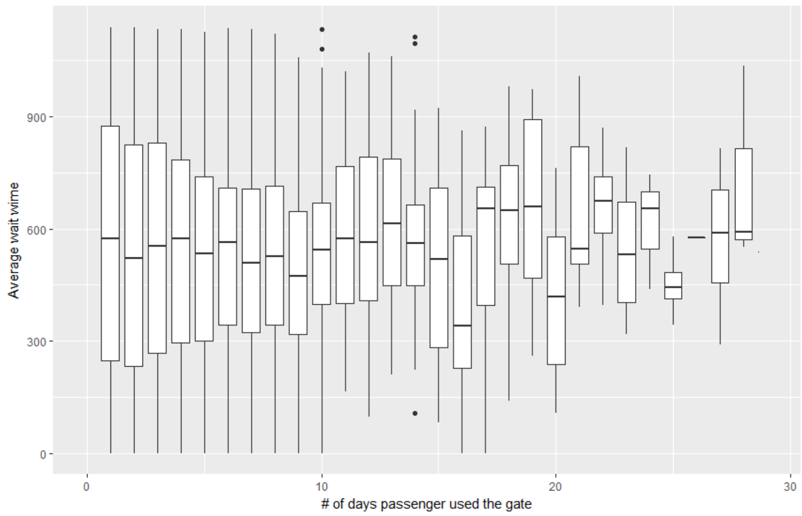

Figure 5 - AM Peak Wait Times

A similar figure is generated to reflect wait times for the passengers who used the same gate from 1 to 27 different weekdays in 10 weeks. It can be seen from Figure 6 that wait times are much closer to each other in the PM peak period and much less variation is observed.

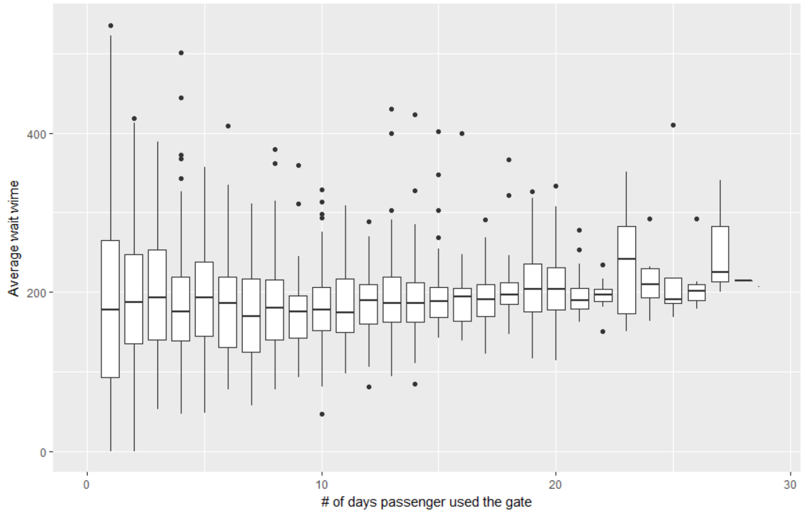

Figure 6 - PM Peak Wait Times

2.2. Probability of Arrival

The probability of passenger arrivals is calculated to understand the effects of shorter and longer headways. Figure 7 illustrates the distribution of passenger arrivals for a 20-minute service during the AM peak. It can be seen from the figure that passenger arrivals are more coordinated with bus departures rather than being random. Bowman and Turnquist [4] suggested that when services are reliable (standard deviation = 0.5 min), a timed arrival pattern would be more obvious. Therefore, the impact of reliability on wait times should be investigated at this stop during AM peak.

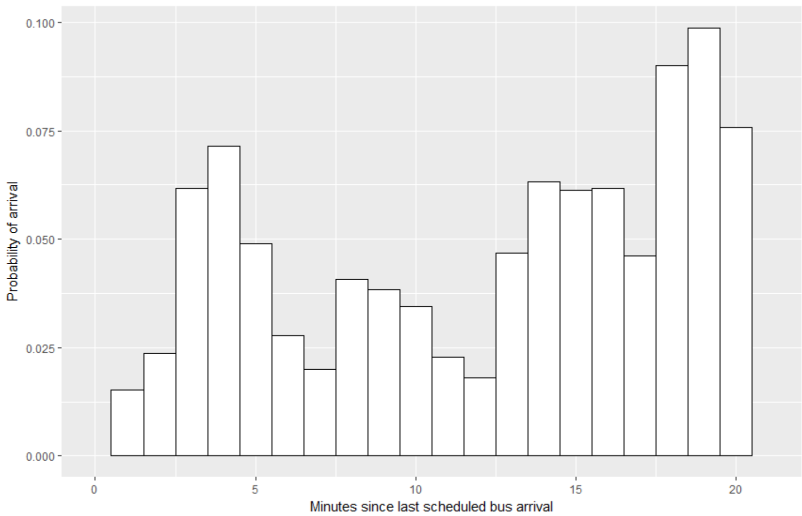

Figure 7 - Distribution of Passenger Arrivals (Headway = 20.6 mins)

Figure 8 shows that the distribution of passenger arrivals during the PM peak period which tends to be more random for a 6-minute service. The distribution is flatter in the evening peak period than the morning peak period which is also consistent with the literature [6, 7, 18]. It seems that with longer headways commuters in the morning peak period are more likely to have the knowledge of the schedule and the service. On the other hand, they also exhibit a more random arrival behavior in the evening peak period where headways are shorter.

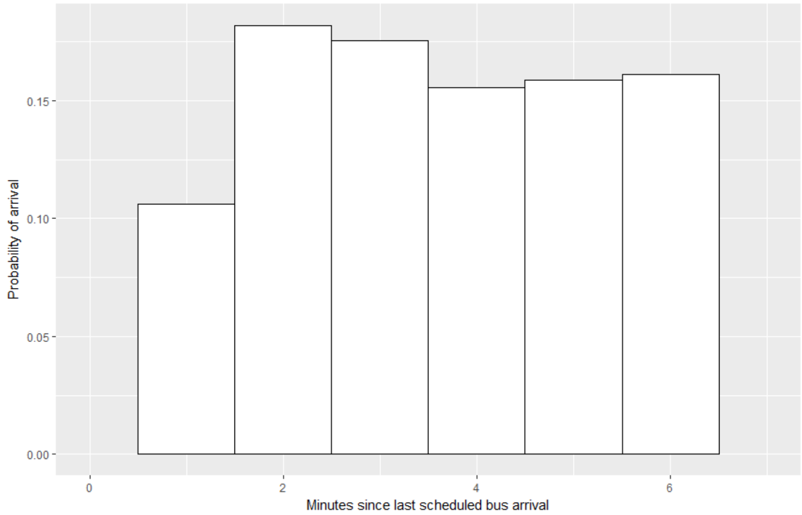

Figure 8 - Distribution of Passenger Arrivals (Headway = 6.2 mins)

3. Conclusion and Discussion

This paper developed a methodology to relate passenger data collected by Wi-Fi and Bluetooth sensors to scheduled bus departures for the purpose of studying passenger arrival behavior. The sensor with the most foot traffic at a bus gate is selected for further analysis. One major advantage of using these sensors is their simplicity and cost. For stations at which AFC systems are not available, wireless sensors can easily be deployed. The developed Wi-Fi and Bluetooth sensors require are simple to operate, easily transported, and require minimal maintenance. Sensors can run up to 3 months when directly connected to the power source. The encryption technique used in the data collection not also provides an extra layer of protection for sensitive information but also preserves the identifiers unique for approximately 96% of the cases. It is also possible to deploy the sensors on transit services which can provide data about bus stop utilization. Wi-Fi and Bluetooth sensor data collection system providing real-time and offline information about passenger arrival behavior will improve the quality, reliability and attractiveness of public transportation systems. The combination of such collected data with other applications such as real time bus locations will help passengers to optimize their travel and waiting times. There are of course limitations of the collected data that deserve mention. Short living network addresses, non-mobile devices that transmit intermittent probe requests and devices that are detectable at a low frequency can reduce the accuracy of the calculated arrival and wait times. However, the analyses showed that passenger arrival times at a transit stop are sensitive to the service frequency as it proposed in the literature. Furthermore, Wi-Fi and Bluetooth sensors can be a cost-effective alternative to understand further and analyze the probabilistic distribution of passenger arrivals and wait times for stops and services where automated passenger data are not available to decision makers and researchers. Another limitation is that the proposed method relies on the schedule-based assignment, for which there are issues in the presence of bus delays or overcrowded buses. In conclusion, passenger arrival behavior analyzed with the data coming from Wi-Fi and Bluetooth sensors varies across times of day, and the differences reflect different headways. These findings are consistent with the findings in the literature. In a future study, the results based on the bus schedules should be supplemented with real-time bus schedules and additional long-term data to ensure that they represent seasonal variations. The correlation with subway arrivals to the terminal can also be investigated to understand passenger arrival patterns to bus gates.

Acknowledgments

This research was funded in part by the Tandon School of Engineering and the Center for Urban Science and Progress of New York University. The authors thank Todd Goldman for sharing data and technical assistance.

References

1. Barnett, A., *On controlling randomness in transit operations.* Transportation Science, 1974. **8**(2): p. 102-116.
2. Jackson, W.H., *Bus Passenger Arrival Times*. 1977.
3. De Pirey, Y., *Simulation of a bus line as a means to analyze and improve bus transit reliability*. 1971, Master's thesis, Northwestern University, Department of Civil Engineering, Evanston, Illinois.
4. Bowman, L.A. and M.A. Turnquist, *Service frequency, schedule reliability and passenger wait times at transit stops.* Transportation Research Part A: General, 1981. **15**(6): p. 465-471.
5. Luethi, M., U. Weidmann, and A. Nash, *Passenger arrival rates at public transport stations.* Institute for transport planning and systems, ETH Zurich, October, 2006.
6. Csikos, D. and G. Currie, *Investigating consistency in transit passenger arrivals: insights from longitudinal automated fare collection data.* Transportation Research Record: Journal of the Transportation Research Board, 2008(2042): p. 12-19.
7. Jolliffe, J. and T. Hutchinson, *A behavioural explanation of the association between bus and passenger arrivals at a bus stop.* Transportation Science, 1975. **9**(3): p. 248-282.
8. Dessouky, M., et al., *Bus dispatching at timed transfer transit stations using bus tracking technology.* Transportation Research Part C: Emerging Technologies, 1999. **7**(4): p. 187-208.
9. Newell, G.F., *Dispatching policies for a transportation route.* Transportation Science, 1971. **5**(1): p. 91-105.
10. Mohring, H., *Optimization and scale economies in urban bus transportation.* The American Economic Review, 1972. **62**(4): p. 591-604.
11. Chang, S. and C.-L. Hsu, *Modeling passenger waiting time for intermodal transit stations.* Transportation Research Record: Journal of the Transportation Research Board, 2001(1753): p. 69-75.
12. Fonzone, A., J.-D. Schmöcker, and R. Liu, *A model of bus bunching under reliability-based passenger arrival patterns.* Transportation Research Part C: Emerging Technologies, 2015. **59**: p. 164-182.
13. Avineri, E. *A cumulative prospect theory approach to passengers behavior modeling: waiting time paradox revisited*. in *Intelligent Transportation Systems*. 2004. Taylor & Francis.
14. Neumann, A., I. Kaddoura, and K. Nagel. *Mind the gap–Passenger arrival patterns in multi-agent simulations*. in *Conference on Agent-Based Modeling in Transportation Planning and Operations*. 2013.
15. Park, Y. and S.B. Ahn, *Optimal assignment for check-in counters based on passenger arrival behaviour at an airport.* Transportation Planning and Technology, 2003. **26**(5): p. 397-416.
16. O'FLAHERTY, C. and D. Mancan, *Bus passenger waiting times in central areas.* Traffic Engineering & Control, 1900.
17. Seddon, P. and M. Day, *Bus passenger waiting times in Greater Manchester.* Traffic

Engineering and Control, 1974. **15**(9): p. 442-445.

18. Frumin, M. and J. Zhao, *Analyzing passenger incidence behavior in heterogeneous transit services using smartcard data and schedule-based assignment.* Transportation Research Record: Journal of the Transportation Research Board, 2012(2274): p. 52-60.
19. Zhao, J., et al., *Unified estimator for excess journey time under heterogeneous passenger incidence behavior using smartcard data.* Transportation Research Part C: Emerging Technologies, 2013. **34**: p. 70-88.
20. Chakirov, A. and A. Erath, *Use of public transport smart card fare payment data for travel behaviour analysis in Singapore.* 2011.
21. Shlayan, N., A. Kurkcu, and K. Ozbay. *Exploring pedestrian Bluetooth and WiFi detection at public transportation terminals*. in *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. 2016. IEEE.
22. Kurkcu, A. and K. Ozbay. *Estimating Pedestrian Densities, Wait Times, and Flows using Wi-Fi and Bluetooth Sensors*. in *Transportation Research Board 96th Annual Meeting*. 2017. Washington, DC.
23. [cited 2017 January 1]; Available from: http://www.linux.org/.
24. *Rasberry Pi*. [cited 2016 July 11]; Available from: https://www.raspberrypi.org/.
25. Prasad, S., et al., *Smart Surveillance Monitoring System Using Raspberry PI and PIR Sensor.* Int. J. Comput. Sci. Inf. Technol, 2014. **5**: p. 7107-7109.
26. Mydlarz, C., J. Salamon, and J.P. Bello, *The Implementation of Low-cost Urban Acoustic Monitoring Devices.* arXiv preprint arXiv:1605.08450, 2016.
27. Cheng, Y., et al. *AirCloud: a cloud-based air-quality monitoring system for everyone*. in *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*. 2014. ACM.
28. Abraham, K. and S. Pandian. *A Low-Cost Mobile Urban Environmental Monitoring System*. in *2013 4th International Conference on Intelligent Systems, Modelling and Simulation*. 2013. IEEE.
29. Leccese, F., M. Cagnetti, and D. Trinca, *A smart city application: A fully controlled street lighting isle based on Raspberry-Pi card, a ZigBee sensor network and WiMAX.* Sensors, 2014. **14**(12): p. 24408-24424.
30. Catania, V. and D. Ventura. *An approch for monitoring and smart planning of urban solid waste management using smart-M3 platform*. in *Open Innovations Association FRUCT, Proceedings of 15th Conference of*. 2014. IEEE.
31. Associates, P. *Apple iPhones accounted for 40% of U.S. smartphone market in 2015*. February 2016 [cited 2017 January 1]; Available from: http://www.parksassociates.com/blog/article/pr-02102016-mwc.
32. Associates, P. *Over 30% of U.S. broadband households are not interested in Wi-Fi services with their mobile services*. February, 2016 [cited 2017 January 1]; Available from: http://www.parksassociates.com/blog/article/pr-02222016.