Origin-Destination Tracking Analysis of an Intelligent Transit Bus System using Internet of Things

Samy El-Tawab, Zachary Yorio, Ahmad Salman, and Raymond Oram
College of Integrated Science and Engineering
James Madison University, Virginia, USA
{eltawass, salmanaa}@jmu.edu,{yoriozp, oramrr}@dukes.jmu.edu

B. Brian Park
Link Lab
Engineering Systems and Environment Department
University of Virginia, Virginia, USA
bp6@virginia.edu

***Abstract*—There is a dramatic increase in the deployment of Internet of Things (IoT) devices in the last couple of years. In Intelligent Transportation Systems (ITS), we introduced a cyber-physical system that monitors the quality of service (QoS) for transit buses around a mid-size university city using Internet of Things (IoT). The sensing IoT devices detected the number of riders waiting for the bus system at each bus station. The experiments ran for six weeks continuously to monitor seven different bus stations around the university. The collected data reports the number of people waiting for the bus at each station, the wait time for a particular bus station at different time/day(s). In this paper, we analyze the collected data for the various bus stations and the origin/destination of some riders who used one of the seven stations as origin and destination stations. Also, several security measurements have been added to address privacy concerns that might occur with the collection, transmission, and storage of data in the Cloud (e.g., the privacy of the ridership Media Access Control (MAC) addresses and tracking of a particular bus rider in the system). We implement security measurements to emphasize the privacy protection of bus riders.**

***Index Terms*—IoT (Internet of Things); Security and Privacy; Intelligent Transportation Systems; Cyber-Physical System; Cloud Computing;**

I. Introduction

With the new era of sensing, Internet of Things and Edge nodes, suggested solution(s) are provided for several intelligent transportation systems applications (e.g., smart parking, traffic control, intelligent bus system, automatic incident detection) [1], [2]. The existence of massive number of wireless devices in various locations around our normal life, with a few computation steps, can help solve several other problems using the concept of Internet of Things (IoT) [3]. Besides, more benefit can occur with careful deployment of small wirelessly equipped devices in public places (e.g., university area) with little or no human interaction for a long period.

The US Department of Transportation uses Intelligent Transportation Systems (ITS) to improve the safety of its Transportation Systems and enhance their roads (e.g., highways or inner roads) by integrating sensing technology [4]. The US Department of Transportation crash-report for 2018 showed an estimate of $7,950$ fatalities in the first quarter of the year 2018. These numbers show a decrease by a 3.6% compared to 2017 [5]. The numbers of crashes have been decreasing in the last 10 years (e.g., there were 5.3 million crashes and 2.22 million injuries in 2011). Although all the effort has been made to improve public transportation, it is clear that there is still much effort needed to make the transportation system more safe and reliable. In this paper, we introduce the use of IoT data analysis to improve the bus system used in an educational campus (e.g., university). It encourages students/university staff members to use the bus system than other means of transportation (e.g., electric scooter). Also, we address some privacy and security concerns that were raised [6]. IoT devices collected data (e.g., Media Access Control (MAC) addresses) as a way to identify riders waiting for the bus around a specific bus station. We emphasize that we used security measurements to make sure no personal information has been collected. The uniqueness of the MAC addresses allows us to detect the number of riders waiting for the bus at a specific station. The rider who has been detected using their unique MAC address is afterward detected again at the destination station. With our data analysis, we can detect how many riders took the bus a specific "origin" station to another specific "destination" station. We can also conclude which stations are the most used at a certain hour. The paper is organized as follows: Section II highlights the background and related scholarly work. Section III presents detailed information about the deployment of the IoT devices and the proposed system. Section IV explains a complete analysis of the origin-destination tracking of the ridership of the proposed method. Then, section V discusses security and privacy concerns and measurements. Finally, we conclude the paper with section VI.

II. Background and Related Research

In the last ten years, several researchers have introduced the use of Wireless Communication to solve Transportation System problems. For example, Papadimitratos et al. [7], explored in a survey the use of sensing in vehicles and wireless

978-1-5386-9151-9/19/$31.00 ©2019 IEEE

communications as a part of what the authors called *vehicular communication systems*. Researchers introduced several applications in the field of Intelligent Transportation (e.g., Traffic Control Systems, and Smart Parking) [8], [9]. With the existence of the Cloud Computing world, more researchers studied the concept of Vehicular Cloud [10], also known as Internet of Vehicles [11]). With the increased use of the Internet of Things (IoT) devices, more researchers integrated solutions and application with the use of IoT small devices. With all this research, few researchers highlight the threat of wireless attacks [12].

Recently, different types of technologies (e.g., Bluetooth, WiFi, and Zigbee) have gained attention among researchers for more intelligent transportation application without depending on vehicle communication [9], [13], [14]. Dunlap et al. already introduced the idea of WiFi readers installed inside buses [15]. Having the WiFi readers inside the buses, don't give us accurate information about which rider may have gave up, or took another mean of transportation. The use of IoT devices for crowdsourcing has increased in the last couple of years [16].

Protecting the privacy and identity of data collected from passengers is important. De Montjoye et al. [17] demonstrate how looking at four random pieces of information from collected credit card metadata, can be used to uniquely identify individuals 90% of the time. In [18], the authors show that four spatial-temporal points gathered from a location tracking system where data is collected on hourly basis are enough to uniquely identify 95% of the individuals".

III. Overview of Cyber-Physical Tracking System:

We introduced a cyber-physical system (CPS) that monitors the quality of service (QoS) for transit buses around a mid-size university city using Internet of Things (IoT) [6], [19]. The system detects how many bus riders are waiting for the bus at any station. The smart node detects anyone was standing beside, or even walking near the station. The Intelligent Transportation CPS consists of several components: a monitor device for WiFi data (e.g., Raspberry Pi 3 connected to a chargeable battery). In this system, we refer to these devices as *Smart Nodes* (a package that can collect data, recharge itself using a solar panel, and runs for days with human interference). Smart node(s) use *sniffing* (also known as monitoring mode) to detect any wireless network traffic surrounding the bus station (with a radius of 7m) [20]. The location of the smart node around the bus station should be carefully considered as not to reach other "gathering areas" (e.g., another bus station or drop off area). While considering the privacy of each passenger (as described later in V, we use a network protocol analyzer (TShark [21]) to capture, record information needed. The system collected data including the MAC address of a device, arrival time, the strength of the WiFi signal, etc. from surrounding WiFi-enabled devices. The collected data is hashed and used to count the number of riders waiting for the bus with no personal data to be saved (encryption is added for stored data). The packets of data are sent to a cloud-based database, which is later parsed as shown in the Network Architecture Figure 1.

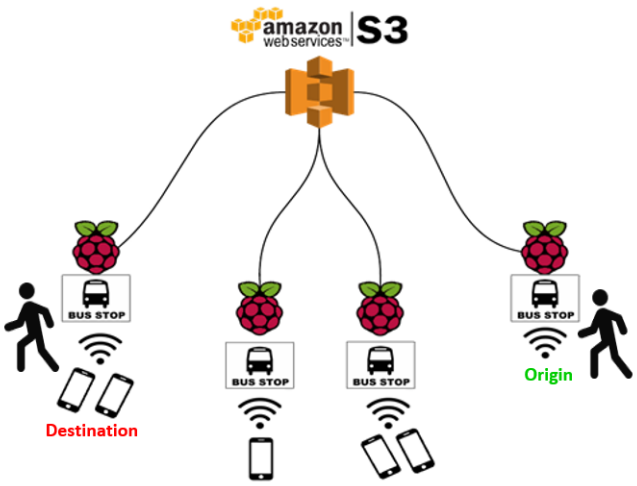

Fig. 1. Network Architecture Diagram describes the idea

We created and deployed seven nodes around seven different bus stations at James Madison University (JMU) as shown in Figure 2. James Madison University is a mid-size university located in Harrisonburg, Virginia, United States of America. It is worth mentioning that the bus system at JMU is considered as the primary transportation mode for students, staff and faculty members at the university. The collected Data in the cloud-based database consists of *MAC addresses, timestamps, and Received Signal Strength Indicator (RSSI(s))*. Using data analysis, the waiting times (maximum time appeared around the bus station "departure" - minimum time appeared around the same station "arrival") for a specific device in a bus station at a specific time frame is calculated. Further data analysis is used to eliminate false positives and false negatives. Since the system is installed in a public area, *False Positives* and *False Negatives* may occur. For example, WiFi signals of fixed or movable Wifi-enabled devices from the buildings surrounding the station and/or students walking beside the bus station. In the case of *False positives*, the system can eliminate it from the SSID (Service Set IDentifier) of the captured data (e.g., devices that appear for 30 seconds or less) and/or (devices that never left the bus station: such as someone sitting at the bus stations for longer than average). In the case of *False negatives*, such as students who do not have a smart cell-phone or their smartphone battery died, this case could be ignored with the big data collected of the riders at each station.

IV. Origin-Destination Tracking

In this section, we discuss the frequency of bus route used around the university, the average travel as calculated by the system, and a sample of the waiting time at one of the stations.

A. Frequency of Bus Route Use

From the database, we can search for MAC addresses initially recorded at one stop and again later on at subsequent

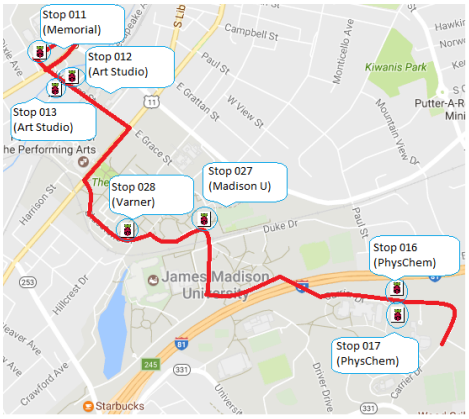

Fig. 2. Seven Smart Nodes locations around JMU

stops along the inner campus shuttle route. By including specific qualifiers in our analysis, we can also ensure that tracking in this manner only occurs in one way (e.g., PhysChem to Varner, but not Varner to PhysChem) and does not include those that indirectly traveled from one station to another (e.g., students walking or arriving at the specified destination later in the day). Filters such as these help further reduce the number of false positives in our results, improving the accuracy of our ridership numbers and estimated travel times.

Figure 3 shows the results of three different origin-destination analyses we conducted for the three most common bus stops students would depart at after boarding the bus at the PhysChem bus shelter: Madison Union, Varner, and Memorial. The three criteria were then repeated for one week. The results for Monday (M), Wednesday (W), and Friday (F) of that week are shown. The same analysis was also conducted for Tuesday (T) and Thursday (Th) of that week, as shown in Figure 4. We separate the M-W-F study from the T-Th study due to how classes at the university are scheduled differently on M-W-F to those on T-Th, requiring different bus route schedules.

As expected, students traveling from the PhysChem bus stop most commonly choose to depart the bus at Madison Union, with the second most popular being the Varner stop which enforces the notion that students need a fast means of transportation from the east side of campus to the main campus and quad area to arrive punctually for subsequent classes. The blue bar representing the bus trip from PhysChem Bus Station to Memorial Hall Bus Station is noticeably less used for the student body across all days. Possible explanations for this include how Education majors plan their class schedules to stay at Memorial Hall for the day and the use of alternative means of transportation that could be more effective for such students, such as a bike or personal vehicle.

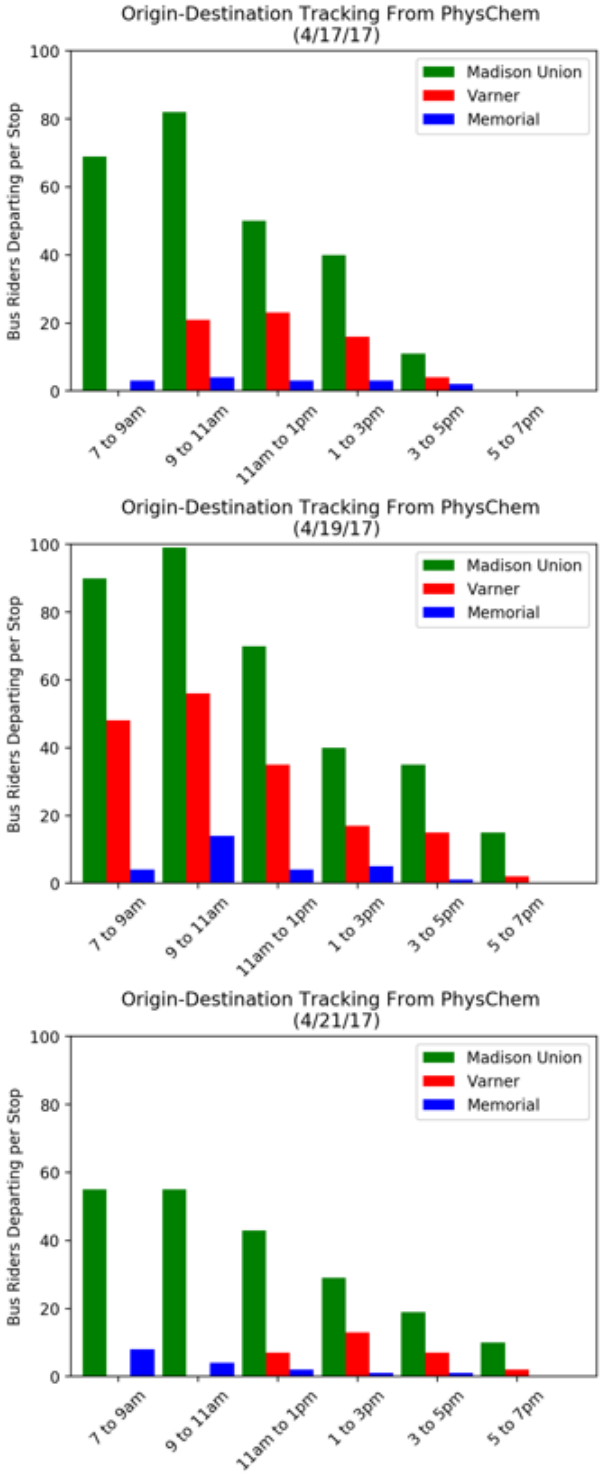

Fig. 3. Frequency of Routes from PhysChem Bus Station on MWF April 17^{th}, 19^{th} and 21^{st}, 2017

B. Average Travel Times

In addition to the frequency of which these routes are used, we also analyzed the time a student may expect each route to take when traveling by bus. To calculate this, we found the difference in minutes between the students' departures from PhysChem Bus station and their arrivals at the three destinations of interest. The times were then averaged per day of the week for which the information was queried and then an average of those daily averages is found for each route. Table I shows the average travel times on April 17^{th} through the 20^{th} for a student departing from PhysChem and traveling

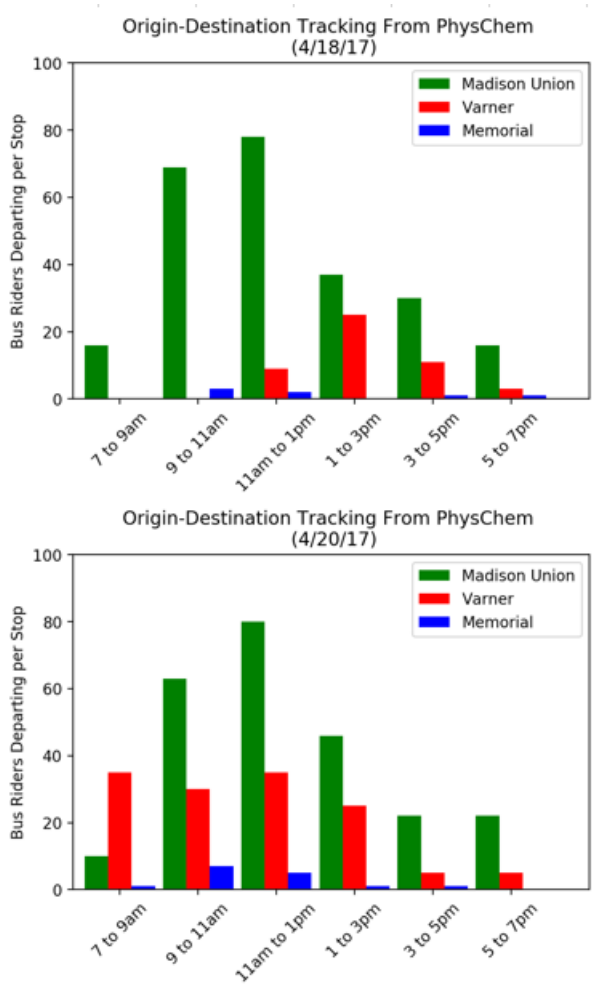

Fig. 4. Frequency of Routes from PhysChem Bus Station on TTh April 18^{th} and 20^{th}, 2017

to either Madison Union, Varner, or Memorial Hall. Due to insufficient data collection on Friday, April 21^{st}, travel time averages for that day were omitted from overall travel time averages for the week.

Generating realistic averages from a large number of trips (big data) in this manner can help inform the travel decisions of students, as well as help city Bus System schedule and deploy buses effectively, as well as craft more efficient routes in the future.

TABLE I
AVERAGES OF RIDERS' TRAVEL TIMES INCLUDING WAITING TIMES

PhysChem to:	Madison Union	Varner	Memorial
April 17^{th}	00:11:48	00:13:41	00:15:19
April 18^{th}	00:10:46	00:12:04	00:15:36
April 19^{th}	00:10:00	00:10:45	00:17:16
April 20^{th}	00:10:38	00:11:34	00:15:30
Average	00:10:48	00:12:01	00:15:55

C. Stop-Specific Wait Time Frequency

Another type of analysis performed on the cloud-based database was the calculation of wait times for each monitored bus stop. Figure 5 shows an example of the results found from these queries. The number of students waiting for buses is grouped into bins with ranges of two minutes, with wait times over eight minutes being of interest to us since inner campus shuttles are expected to arrive at all stops at least this often. Comparing similar days of the week, in this case - three Wednesdays in April, notable trends may be identified for route improvements.

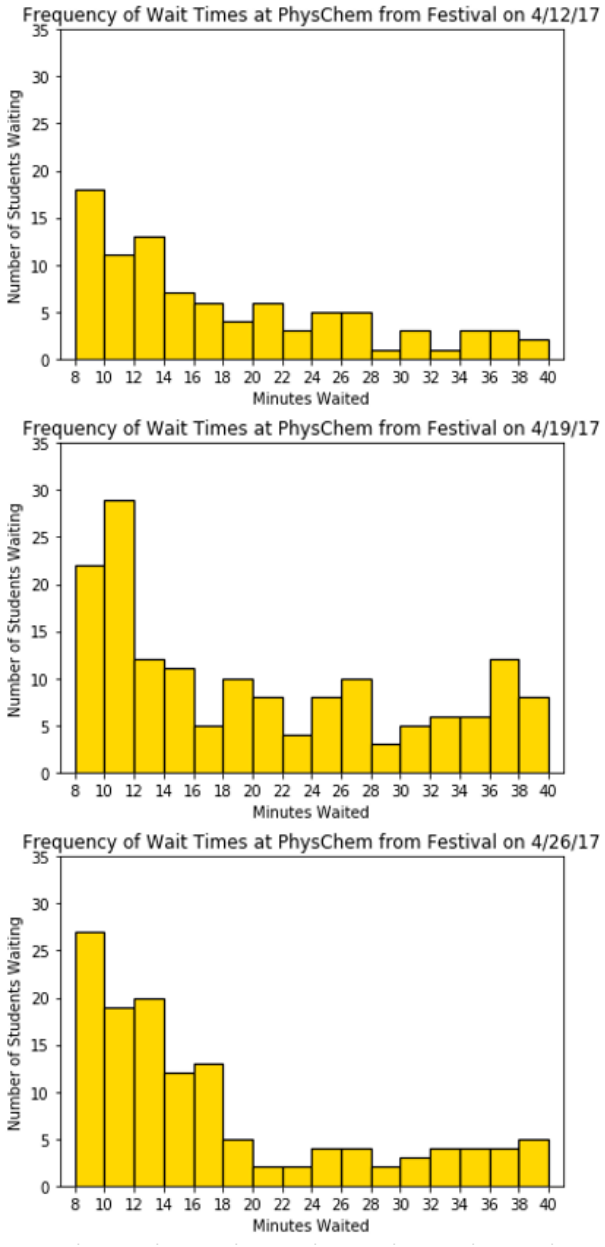

Fig. 5. Frequency of Wait Times at PhysChem from Festival on Wednesdays April 12^{th}, 19^{th} and 26^{th}, 2017

V. Security and Privacy Concerns

A. Protecting Privacy

As mentioned in the introduction section, bus riders at bus stations are identified through the collected MAC addresses from their WiFi enabled smart phones. The reason for collecting such information is to be able to uniquely identify passengers as MAC addresses are unique for each device. However, this process raises a concern regarding the invasion of one's privacy [22], [23]. To properly address the issue, we do not actually store any of the collected MAC addresses; Instead, we perform a hashing operation over the obtained MAC address to generate a unique hash value which we then store. Hashing is performed through a hash function which takes as an input a message of arbitrary length and produces an out put of fixed size, known as hash value, as shown in Fig. 6. An important property required by a hash function is that it's a one-way function; i.e. given a hash value of a message, $h(m)$, computing the original message m is an impossible task to perform. By making use of a hash function and its one-way property, storing the hash values of the MAC addresses will not violate the privacy of the passengers since a MAC address cannot be driven back from its stored hash value.

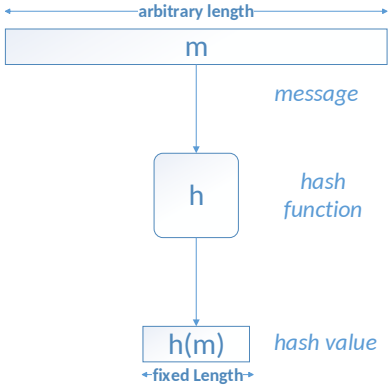

Fig. 6. Derivation of a hash value from a message

Our hash values are calculated using the Secure Hash Algorithm (SHA-256) [24] function. In addition to providing all the security requirements specified for hash functions as summarized in Table II, it has one of the fastest processing times in software implementations when compared to other hash functions [25]. This ensures that hashing of the sniffed MAC addresses would not require the use of large buffers.

TABLE II
Security Requirements of a Hash Function

Security Requirement	Given	Computationally Infeasible to Find
Preimage resistance (One-Way)	y	message x, such that $h(x) = y$
Second preimage resistance (weak collision resistance)	x, and $y = h(x)$	$x' \neq x$, such that $h(x') = h(x) = y$
Strong collision resistance		$x' \neq x$, such that $h(x') = h(x)$

The size of the input block of SHA-256 is 512-bits while a single MAC address is only 48-bits wide. For a message of the size 48-bits to be processed by SHA-2, a padding scheme should precede the hashing operation. The padding process for SHA-256 is as follows.

- Start with the original message x of size n-bit such that $n < 512$
- Concatenate x with a 1-bit of value '1'
- Append with m-bit of '0' such that m+n+1 = 448-bit
- The remaining of 512-bit block is the binary representation of n in 64-bit representation

In our case, the input message to the SHA-256 function is always a MAC address, which means that the padding size and value are fixed as shown in Fig. 7. This padding value is stored and appended to every MAC address detected then input to the hash function. The output hash value is then stored along with the WiFi signal strength and the time stamp in a separate entry to be transferred later to the database.

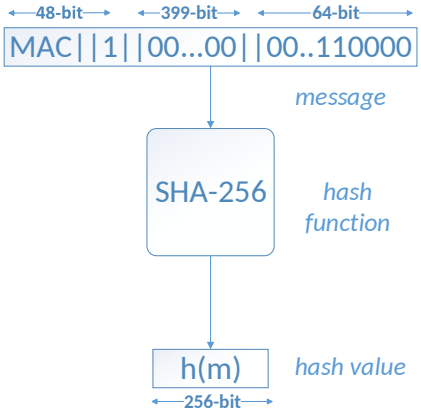

Fig. 7. The Padding Process of a MAC Address

Usually, information hashing operations, such as password hashing, require the underlying hash function to have special characteristics such as slow processing and having a salt as part of the input [26]. These requirements are to prevent attacks, such as dictionary attacks, aimed to retrieve the original password. However, we are not concerned about such attacks and for this reason we chose a hash function that would process the information fast to prevent bottlenecks when collecting the data from multiple stations at once.

B. MAC Address Randomization

MAC address randomization is a technique developed aiming to protect user privacy by preventing tracking through MAC addresses. The idea behind it is that instead of broadcasting MAC addresses, devices perform randomization to the MAC address and then broadcast that MAC address to other WiFi enabled devices and access points. Most of the smart phones running iOS or Android, perform MAC address randomization nowadays which affects our data collection methodology. However, in [27], the authors show that by sending certain control frame to client devices performing

MAC randomization, it was possible to reveal the global MAC address of these devices. The technique was applied to various iOS and Android devices with a success rate of 100%. This shows that MAC addresses can still be revealed after applying some minor modifications to our current methodology.

We should highlight that regardless of how we can obtain MAC addresses, we will always use the hashing technique proposed in the previous subsection and never store MAC addresses to protect user privacy.

VI. Conclusions and Future Directions

In this article, we discussed some of the use cases for origin-destination tracking using smart nodes. By collecting the MAC addresses of passengers' devices, along with the time stamps at which passengers arrive and depart from bus stops, multiple questions can be answered. We are able to discern the number of students traveling to and from bus stops of interest, the time required to make these trips, as well as the wait times leading up to these trips.

Future work in this endeavor will include meeting with the JMU and Harrisonburg bus management teams to better improve route efficiency. Redeployment of nodes will also be considered for the collection of more data. This will improve ridership and travel time averages for more informed decision making.

ACKNOWLEDGMENT

This work is supported by a 4-VA collaborative research grant between JMU and UVA: https://4-va.org/james-madison-university/ Fall 2017. The authors would like to thank James Madison University Public Safety, and transit bus manager (Mr. Lee Eshelman) for allowing us to conduct our experiments.

References

[1] A. Khanna and R. Anand, "IoT based smart parking system," in *Internet of Things and Applications (IOTA), International Conference on*. IEEE, 2016, pp. 266–270.

[2] O. Popescu, S. Sha-Mohammad, H. Abdel-Wahab, D. C. Popescu, and S. El-Tawab, "Automatic Incident Detection in Intelligent Transportation Systems Using Aggregation of Traffic Parameters Collected Through V2I Communications," *IEEE Intelligent Transportation Systems Magazine*, vol. 9, no. 2, pp. 64–75, Summer 2017.

[3] V. Karagiannis, P. Chatzimisios, F. Vazquez-Gallego, and J. Alonso-Zarate, "A survey on application layer protocols for the internet of things," *Transaction on IoT and Cloud Computing*, vol. 3, no. 1, pp. 11–17, 2015.

[4] U.S. Department of Transportation, "Strategic Plan for FY 2018-2022 ," https://www.transportation.gov/sites/dot.gov/files/docs/mission/administrations/office-policy/304866/dot-strategic-planfy2018-2022508.pdf, 2018.

[5] National Center for Statistics and Analysis (NCSA) Motor Vehicle Traffic Crash Data Resource Page, "Early Estimate of Motor Vehicle Traffic Fatalities for the First Quarter of 2018," National Highway Traffic Safety Administration, Tech. Rep., 2018.

[6] S. El-Tawab, R. Oram, M. Garcia, C. Johns, and B. B. Park, "Data Analysis of Transit Systems Using low-cost IoT Technology," in *First International Workshop on Mobile and Pervasive Internet of Things'17 - 2017 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*., Mar 2017.

[7] P. Papadimitratos, A. De La Fortelle, K. Evenssen, R. Brignolo, and S. Cosenza, "Vehicular communication systems: Enabling technologies, applications, and future outlook on intelligent transportation," *IEEE communications magazine*, vol. 47, no. 11, 2009.

[8] R. Sundar, S. Hebbar, and V. Golla, "Implementing intelligent traffic control system for congestion control, ambulance clearance, and stolen vehicle detection," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 1109–1113, 2015.

[9] M. Garcia, P. Rose, R. Sung, and S. El-Tawab, "Secure Smart Parking at James Madison University via the Cloud Environment (SPACE)," in *2016 IEEE Systems and Information Engineering Design Symposium (SIEDS)*, 2016, pp. pp–271.

[10] R. Florin, P. Ghazizadeh, A. G. Zadeh, S. El-Tawab, and S. Olariu, "Reasoning about job completion time in vehicular clouds," *IEEE Transactions on Intelligent Transportation Systems*, vol. PP, no. 99, pp. 1–10, 2016.

[11] M. Gerla, E.-K. Lee, G. Pau, and U. Lee, "Internet of vehicles: From intelligent grid to autonomous cars and vehicular clouds," in *Internet of Things (WF-IoT), 2014 IEEE World Forum on*. IEEE, 2014, pp. 241–246.

[12] J. Blum and A. Eskandarian, "The threat of intelligent collisions," *IT Professional*, vol. 6, no. 1, pp. 24–29, Jan 2004.

[13] M. Elhamshary, M. Youssef, A. Uchiyama, H. Yamaguchi, and T. Higashino, "Transitlabel: A crowd-sensing system for automatic labeling of transit stations semantics," in *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*. ACM, 2016, pp. 193–206.

[14] A. Salem, T. Nadeem, M. Cetin, and S. El-Tawab, "DriveBlue: Traffic Incident Prediction through Single Site Bluetooth," in *18th IEEE International Conference on Intelligent Transportation Systems, September 15-18, 2015*, 2015.

[15] M. Dunlap, Z. Li, K. Henrickson, and Y. Wang, "Estimation of Origin and Destination Information from Bluetooth and Wi-Fi Sensing for Transit," in *Transportation Research Board 95th Annual Meeting*, no. 16-6837, 2016.

[16] J. Liu, H. Shen, H. S. Narman, W. Chung, and Z. Lin, "A survey of mobile crowdsensing techniques: A critical component for the internet of things," *ACM Transactions on Cyber-Physical Systems*, vol. 2, no. 3, p. 18, 2018.

[17] Y.-A. de Montjoye, L. Radaelli, V. K. Singh, and A. ". Pentland, "Unique in the shopping mall: On the reidentifiability of credit card metadata," *Science*, vol. 347, no. 6221, pp. 536–539, 2015.

[18] Y.-A. De Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific reports*, vol. 3, p. 1376, 2013.

[19] Z. Yorio, R. Oram, S. El-Tawab, A. Salman, M. H. Heydari, and B. B. Park, "Data analysis and information security of an Internet of Things (IoT) intelligent transit system," in *2018 Systems and Information Engineering Design Symposium (SIEDS)*, April 2018, pp. 24–29.

[20] S. Dimatteo, P. Hui, B. Han, and V. O. K. Li, "Cellular traffic offloading through wifi networks," in *2011 IEEE Eighth International Conference on Mobile Ad-Hoc and Sensor Systems*, Oct 2011, pp. 192–201.

[21] B. Merino, *How-to Instant Traffic Analysis with Tshark*. Packt Publishing Ltd, 2013.

[22] A. Salman, A. Ferozpuri, E. Homsirikamol, P. Yalla, J. P. Kaps, and K. Gaj, "A scalable ecc processor implementation for high-speed and lightweight with side-channel countermeasures," in *2017 International Conference on ReConFigurable Computing and FPGAs (ReConFig)*, Dec 2017, pp. 1–8.

[23] K. Evers, R. Oram, S. El-Tawab, M. H. Heydari, and B. B. Park, "Security measurement on a cloud-based cyber-physical system used for Intelligent Transportation," in *2017 IEEE International Conference on Vehicular Electronics and Safety (ICVES)*, June 2017, pp. 97–102.

[24] National Institute of Standards and Technology, *FIPS PUB 180-4: Secure Hash Standard*, August 2015. [Online]. Available: https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.180-4.pdf

[25] E. Homsirikamol, M. Rogawski, and K. Gaj, "Comparing hardware performance of round 3 SHA-3 candidates using multiple hardware architectures in Xilinx and Altera FPGAs," ECRYPT II Hash Workshop 2011, May 2011.

[26] G. Hatzivasilis, I. Papaefstathiou, and C. Manifavas, "Password hashing competition - survey and benchmark," Cryptology ePrint Archive, Report 2015/265, 2015, https://eprint.iacr.org/2015/265.

[27] J. Martin, T. Mayberry, C. Donahue, L. Foppe, L. Brown, C. Riggins, E. C. Rye, and D. Brown, "A study of MAC address randomization in mobile devices and when it fails," *CoRR*, vol. abs/1703.02874, 2017. [Online]. Available: http://arxiv.org/abs/1703.02874