Exploring Pedestrian Bluetooth and WiFi Detection at Public Transportation Terminals

Neveen Shlayan, Member, IEEE, Abdullah Kurkcu, Kaan Ozbay

Abstract—This is an on-going study that explores the potential benefits of using pedestrian data for evaluation and enhancement of public transportation. The research team proposes the utilization of Bluetooth (BT) and WiFi technologies to estimate time-dependent origin-destination (OD) demands and station wait-times of transit bus and subway users. The detection approach and a complete system design developed by the research team is discussed in this paper. Preliminary results from multiple pilot field studies, that were conducted at some of the major New York City (NYC) public transportation facilities, are also presented. The main objective of this study is to inquire into the various ways this extensive transit rider data can be used and to establish a general framework through data-driven pedestrian modeling within transit stations that renders estimation of key parameters and strategic control of public transportation services possible.

I. INTRODUCTION

It is highly important to understand pedestrians behavioral patterns and estimate key states of the pedestrian network, such as volume, dwell times, and OD flows. Obtaining this information is vital to a wide range of applications, for instance transportation demand forecasting, pedestrian evacuation, community wellness, sustainability, and traffic safety among others. However, the lack of sensor infrastructure and the unrestricted movement of pedestrians present a challenge as it pertains to obtaining the necessary data. Annual travel surveys are generally used to obtain pedestrian data which are not sufficient for most of the mentioned applications.

Thus, automated data collection methods are needed. Video-based pedestrian detection technologies are widely used yet these systems are quite costly and relatively complex to deploy, maintain, and operate.

Fan et al. proposed a novel approach for multi-person video-based tracking-by-detection using deformable part models in a Kalman Filtering framework for pedestrian detection and tracking. Tracking-by-detections suffers from high amount of false positives and missing detections. GPS-based pedestrian detection has also been used; however, this requires the user to register their device. Therefore, GPS may not be a feasible solution for applications with frequent data collection needs. Detection by Bluetooth (BT) and WiFi, significantly reduces false alarms leading to a more reliable and relatively cost effective dataset. This type of ubiquitous detection has become increasingly popular in recent years particularly among the motorized transportation community. Research is still very limited when considering BT and WiFi data collection for pedestrians and bicyclists.

The research team proposes utilizing BT and WiFi technology to estimate OD demands and station wait-times of users of transit facilities such as subway or bus stations in major metropolitan areas. More specifically, if the entrance and exit turnstiles at subway stations are equipped with BT/WiFi detectors, it is possible to capture OD information for some percentage of the riders with discoverable BT/WiFi enabled devices. Most electronic devices such as cell phones, iPods, and computers carry unique BT and WiFi MAC addresses. This information can be scrambled and used anonymously to detect the origin and destination of riders by matching data collected at entrances and exits from the system. Assuming that visible BT/WiFi devices are uniformly distributed among the riders, it is possible to estimate a transit OD matrix for the entire system not only on a daily basis but also over a time period allowing the agency to analyze time-dependent OD demand for different station pairs. Moreover the same wireless sensors proposed by the research teams will capture wait-times of the same sample of transit riders at fixed locations at each station. This information will then be converted to average hourly, daily, weekly delays that can be used in conjunction with OD matrices. Estimation of daily and hourly Origin-Destination (OD) demands and delays is important for transit agencies because it can help improve their operations, reduce delays, and mitigate cost, among other benefits. The proposed method of tracking BT and WiFi IDs uses inexpensive, portable, and easy to deploy wireless detectors/ readers with specialized software developed by the research team. This is a low-cost and viable alternative to traditionally used surveys or other advanced technologies. The paper presents the developmental aspects of the the hardware, software, and architectural total system design of the BT and WiFi detection technology.

The rest of the manuscript is organized as follows. After the background and literature review, a description of the developed technology is presented followed by an overview of some of the pilot tests implemented in NYC. The main results of the studies are then discussed which constitute the main motive of the future work planned. Finally conclusions

N. Shlayan is an Assistant Professor, Department of Electrical Engineering, State University of New York, Maritime (e-mail: nshlayan@sunymaritime.edu). Corresponding author.
A. Kurkcu is a Ph.D. Student, Civil and Urban Engineering and Center for Urban Science and Progress (CUSP), Tandon School of Engineering, New York University, 6 Metrotech, Brooklyn, NY, 11201, USA (e-mail: ak4728@nyu.edu).
K. Ozbay is a Professor, Civil and Urban Engineering and Center for Urban Science and Progress (CUSP), Tandon School of Engineering, New York University, 6 Metrotech, Brooklyn, NY, 11201, USA (e-mail: kaan.ozbay@nyu.edu).
II. BACKGROUND AND LITERATURE REVIEW

The literature is rich in studies and results related to BT and WiFi detection based estimation of vehicular traffic on transportation networks from which we can learn when applying similar methodologies to pedestrian networks.

In the study conducted by Fan et. al, video data was used in which they propose a multi-person tracking using deformable part models in a Kalman Filter framework. This method, though effective for a small number of pedestrians, its scalability to modeling large and complex pedestrian networks may present some issues. In Fardi et. al’s work, a multi sensor system, employing an infrared camera, a laser scanning device, and ego motion sensors, was discussed. The data was then analyzed using Kalman filter-based data fusion techniques that ultimately provided precise and detailed description of shape and motion of pedestrians. This work is also limited to individual pedestrians and expanding it to larger pedestrian networks may present a challenge. Pedestrians detection in vehicles using a laser scanner and a stereovision system along with estimation using velocity and GPS information is discussed in Garcia et. al in order to estimate danger. This work is very useful when taking accident preventive measures; however, solely it may not be sufficient for inferring pedestrian behavior at a larger scale in networks. LIDAR and a single camera pedestrian detection is presented in Preembrida et. al. Centralized and decentralized sensing architectures are described which can be greatly improved if combined with WiFi and BT data.

The emergence of the new Information and Communication Technologies (ICT), makes it possible to gather new types of traffic data with higher quality and accuracy. BT specification defines a uniform structure for a wide range of devices such as cell phones, GPS devices, mp3 players, and hands free devices to connect and communicate with each other. Since every BT and WiFi device has a unique MAC address, if this information is captured at a single or multiple locations, it is possible to use it in transportation studies. Although not all BT devices are discoverable, in general it has been reported that 5%-12% of devices are discoverable via BT. This sample size is adequate for most transportation studies. This type of technology is now widely used by highway agencies and the most important technical and privacy issues have already been resolved. In recent years, as BT and WiFi data became more common, researchers adapted previous approaches to fuse the new data with the classical flow data in order to solve the estimation problem. O’Neill et al. found that approximately 7% of pedestrians were carrying BT devices. Malinovskiy et al. conducted studies to estimate dwell times and travel times in two sites, Montreal, Quebec, and Seattle, Washington, to investigate the feasibility of pedestrian detection using BT and WiFi and found that high-level trend analysis can provide insights into pedestrian travel behavior provided sufficient populations. Liebig et. al discuss optimal dynamic placement of a limited number of BT sensors during various phases of a soccer match. It was reported that 14% of the population was detected. They also proposed a nonparametric Bayesian method, Gaussian Processes (GP) with a random-walk based trajectory kernel to estimate traffic volumes at unmeasured locations; however, dwell time and travel time estimation was not attempted. Kostakos used BT devices to trace passenger journeys on public buses and derive passenger OD matrices. Bullock et. al. deployed a BT tracking system at the new Indianapolis International Airport to measure the time for passengers to transit from the non-sterile side of the airport (pre-security), clear the security screening checkpoint, and enter the walkway to the sterile side.

Most of the mentioned studies allude to the fact that BT and WiFi detection technologies are revolutionary compared with traditional sensing and surveying methods as it pertains to the quality and richness of the data and relatively low cost and simplicity of the technology. Filtering, sensor placement, and sensor features are inevitably common themes in most of these studies and tweaking them highly depend on the system at hand. The type of system, for instance a train station, a shopping center, or a bus terminal, along with the desired parameters to be estimated, such as OD flows, travel-times, or wait-time, shape the development of techniques and algorithms used. In this paper, we conduct pilot studies to determine some of the issues that might be location dependent and unique to the NYC public transit.

III. TECHNOLOGY DESCRIPTION

It is possible to detect the proximity of a personal electronic device of users utilizing BT and WiFi probe requests when their devices are actively looking for other devices. Nowadays, most people carry either a mobile phone or a smart device with multiple means of data transmission such as BT and WiFi.

A. Hardware

The hardware used for testing is an Android tablet manufactured by ASUS called Nexus 7. It is a thin, light, portable and affordable 7” tablet that comes with Android 4.1. It has a 1.2GHz CPU, 1GB memory, and 16GB storage, which are sufficient for collecting and processing BT and WiFi data. It has a Li-polymer battery that can run up to 9.5 hours on its own and additional 6-7 hours by hooking up external 10kmAh batteries. It comes with a micro USB cable and a charging unit in a box. The device has double speakers, a micro-USB connector, 3.5 mm headphone jack, 2 microphones and a 4-pin connector. Although, it takes about 35 seconds to boot, applications load rapidly and respond briskly.

B. Software

The research team developed an application (app) called “Traffic Tracker” working on any Android device to detect BT and WiFi devices. Traffic tracker scans discoverable BT and WiFi enabled devices nearby in a way that their unique MAC IDs and signal strength information can be anonymously extracted and saved in tables including the
detection times. Figure 1 shows the main screen when the app is initiated which indicates the three main functions, Scan, Database, and Files.

Fig. 1. Android App Interface.

- **Scan**: This function allows users to start a new scan. The user has to name the new scan such as “Floor 2”. The scan name does not have to be unique and duplicate names can be differentiated by the timestamp of a scan. It is possible to get location updates providing GPS locations of a device when there is an internet connection.

- **Database**: After the scan is stopped, the app automatically creates a final table under the “Database” section. It shows the total number of records, scan name, duration, and occurrences of the same devices. These tables are saved in a relational database and can be imported to a text file.

- **Files**: This function allows users to view imported text files. The log file of the app is also stored under the “Files” section.

C. **Anonymization**

The MAC addresses of the detected devices consist of a unique identifier. This information is double encrypted by deleting part of the MAC address and then encrypting the remaining part using a state-of-the-art encryption software where double encryption is done. The encryption method can be chosen at the beginning of the scan depending on the purpose of the study. There are two main anonymization techniques used in the app: Encryption for counting studies, where no individual ID is saved after filtering and counting is completed, and Aggregation for OD matching studies.

D. **Data Access and Remote Self-Diagnostics**

Prior to deployment, a cloud-based server structure was implemented to enable tracking and accessing the data in real time, to observe the data collection, and also perform self-diagnostics. The research team implemented a cloud-based server that connects to all active devices and ensures data transfer between the device and the server. Then, these files can be accessible from any computer connected to the Internet to track the devices. An industry standard encryption method was integrated into the software app developed by the research team to guarantee maximum level of privacy. The developed web page can be seen in Figure 2. It enables authorized users to access the collected data and sensors. Moreover, a series of simple yet useful self-diagnostics web-enabled functions such as the current reporting status of each device, power levels of battery powered devices, and possible data errors are included. This software-oriented task will be done in coordination with the equipment testing to ensure that data is captured adequately. Figure 3 below depicts the system architecture that will be adopted for this purpose.

IV. **Data Validation and Filtering**

Before deployment, the team conducted multiple tests in order to validate the quality of the collected data. As mentioned previously, case specific data filtering techniques are necessary. However, the preliminary data analysis performed on the initial tests consistently demonstrated two main issues with the collected data. The first issue is that unique BT IDs can be detected at more than one sensor simultaneously; the second issue is that some transitions durations are less than a few seconds, which may not be possible considering the distance between the sensors. Depicted in Figure 4, a sample trajectory of a unique WiFi ID selected to demonstrate the issues stated above. The depicted sample trajectories show the time and location of detection for each WiFi ID. The Received Signal Strength Indication (RSSI) levels are also displayed by the heat map colors, where the highest value (in the negative sense) corresponds to the deepest red indicating closeness and the lowest value corresponds to the darkest blue.

The above results motivated the research team to conduct controlled experiments in order to verify the effectiveness of the data.

A. **Controlled Experiments**

The experiments were designed with the following objectives in mind:

- To understand how RSSI values are related to various distance/speed/devices
- To understand how to deal with a device that is detected by two or more sensors at the same time or within a very small amount of time that is less than the estimated transition time
- To determine whether we are able to distinguish between the devices detected outside or inside the building and those that are detected at other floors

Four sensors were used for the controlled experiments and plans were made to collect ground truth data in order to compare with sensor data. Two experiments were performed: path verification experiment and counting verification experiment.
1) **Summary of Results:** The results of the path verification experiment demonstrated promising results. All paths were accurately detected with a very minimal error rate. The results for one of the tested devices (iPhone 6) can be depicted in Figure 5. The ground truth data, the green line, is plotted in conjunction with the sensor data, the blue line.

At a first glance, the results of the count verification experiment demonstrated a significantly large variation as compared with the ground truth data. This issue may arise due to the presence of devices that are not necessarily associated with pedestrians within the detection perimeter. Another possible issue that contributes to this inaccuracy is due to sensor limitation. More specifically, the detection range cannot be directed and is radial in nature. This may lead to detection of pedestrians that are outside of the desired region. The research team developed a location specific filter was used to eliminate data entries that may have originated in the mentioned scenarios. Figure 6 shows the counting results for ground truth and sensor data with and without filter.

The results demonstrate that there are some errors associated with the sensors accurately reporting the location of the devices. This might be due to the detection range the sensors and their placement. Additional studies must be conducted to quantify this error. It is noted that the range of the RSSI the sensors picked up are somewhat consistent for all three devices.

V. PILOT TESTS

The initial objectives were to test the devices and to examine the data keeping in mind the estimation of OD flows and wait times for public transit. Multiple pilot tests were conducted in NYC; however, in this paper, we discuss only two studies:

- The Atlantic Avenue Subway Station Test, and
The Port Authority Transit facility in Manhattan

A. Atlantic Avenue Subway Station

In collaboration with the Metropolitan Transportation Authority (MTA), a pilot study was conducted at three different high-volume locations at Atlantic Avenue subway station. In spite of the challenges encountered during the pilot test, the outcome was very encouraging in terms of the quantity and detail of the collected data considering the investment needed for deployment. We were able to demonstrate that large amounts of time-dependent counts, OD, and waiting time data can be collected using a small number of devices.

Three different tests were conducted at this location encompassing various periods as shown in Table I.

<table>
<thead>
<tr>
<th>PERIOD OF DATA COLLECTION BY DEVICE.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tablet 1</td>
</tr>
<tr>
<td>7/30 - 9/3</td>
</tr>
<tr>
<td>10AM - 4PM</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>-</td>
</tr>
<tr>
<td>-</td>
</tr>
</tbody>
</table>

1) Summary of Results: In this test, we focused on the counts from the tablets to observe the foot traffic at the selected locations. The average wait-times are also of main interest as they might indicate the wait-time of commuters for the trains at the station. Another useful indicator is the counts of movement between tablets. By tracking the movement of the people in the station, we can identify some travel patterns of riders as well as approximate the demand and OD flows. Table II presents a summary of the data collected for three test periods. Due to difference in the length of testing periods, it is useful to focus on average values.

We can also obtain the hourly movement counts between all tablets. Figure 7 demonstrates the hourly movement counts from Tablet 3 to itself and other tablets.

Some of the key observations based on the data we were able to collect during the pilot test are as follows.
1) A relatively large number of BT enabled devices were detected by all three tablets.
2) Average waiting times at platform 3 (Tablet 3) increase from 2.7 minutes to 3.02 minutes after August 14th. A slight increase in waiting times at platform 2 (Tablet 2) after August 14th is also observed.
3) We saw an increase in movement percentages at platform 3 (Table 3) after August 14th.

B. Port Authority Transit Facility

In this test, we explore the pedestrian flows and device detections in a Port Authority transit facility with high level of pedestrian traffic (May 2016) based on the BT and WiFi data collected by the CitySMART Lab at NYU. The sensors are placed in such a way that focuses on the movements from two entrances, E1 and E2, to four different gates, G1, G2, G3, and G4, and attempts to find correlations or clear patterns. It also investigates the potential data discrepancies and sensor malfunctions.

1) Summary of Results: Table III summarizes the counts of the collected data at the E1 and E2 entrances. The average number of detected unique BT devices in an hour is 18.5 and the average number of detected unique WiFi devices in an hour is 444.1 at the E1 entrance. The number of detections within the study period is always higher at the E2 entrance. The percentage of detected devices that uses BT technology is 3.9% at the E1 and 0.9% at the E2 entrance.

The E1 entrance is selected for further investigations and the following figures are only generated for the E1 entrance. All of these analyses can be conducted for each sensor location in the future. Figure 8 below illustrates the number of detected unique devices for the first week. It is possible to see a repetitive pattern among different days indicating the
TABLE III
COUNTS SUMMARY OF E1 AND E2 ENTRANCES.

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>E1-WiFi</td>
<td>34</td>
<td>1803</td>
<td>444.1</td>
</tr>
<tr>
<td>E1-BT</td>
<td>1</td>
<td>77</td>
<td>18.5</td>
</tr>
<tr>
<td>E2-WiFi</td>
<td>48</td>
<td>5212</td>
<td>1935</td>
</tr>
<tr>
<td>E2-BT</td>
<td>1</td>
<td>107</td>
<td>22.35</td>
</tr>
</tbody>
</table>

busiest hours of the terminal. It is seen that 9:00 AM and 6 PM are the peak hours for this entrance.

The movements from both entrances to the gates for weekdays and weekends were examined and the results are presented in Table IV. Each quantity in the table represents the daily average number of detected unique devices moving from an entrance to a gate.

TABLE IV
MOVEMENTS BETWEEN SENSORS.

<table>
<thead>
<tr>
<th>From/To</th>
<th>G1</th>
<th>G2</th>
<th>G3</th>
<th>G4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>E1</td>
<td>3445</td>
<td>3383</td>
<td>1759</td>
<td>2253</td>
<td>Weekdays</td>
</tr>
<tr>
<td>E2</td>
<td>4704</td>
<td>3734</td>
<td>2878</td>
<td>2597</td>
<td>Weekdays</td>
</tr>
<tr>
<td>E1</td>
<td>1896</td>
<td>1929</td>
<td>1011</td>
<td>1099</td>
<td>Weekends</td>
</tr>
<tr>
<td>E2</td>
<td>2483</td>
<td>2209</td>
<td>1619</td>
<td>648</td>
<td>Weekends</td>
</tr>
</tbody>
</table>

From analyzing the data, it is evident that calculating accurate waiting times requires extensive filtering since it is possible to have some recurring MAC addresses within an hour for some individuals such as Port Authority workers. Mapping the regular users of the bus line might help reduce some errors. In addition, relaxing the signal strength filter, which detects only the individuals who are really close to the sensor at the moment, might help to accurately find the initial arrival time to the gate.

VI. IN PROGRESS AND FUTURE WORK
As demonstrated, the preliminary examination of the BT and WiFi data shows a great promise in its usefulness for evaluation and enhancement of public transportation. We are currently looking at stochastic models that are specific to pedestrians in order to be able to estimate and predict OD flows and wait times based on demand. We have established, through examining the data collected from our pilot tests, that the nature of pedestrian movement varies with respect to the public transportation system at hand (i.e., a bus terminal versus a subway station). Therefore, the case-specific models have to distinguish between such system variations. We are currently developing Markovian-based models that will enable accurate depiction of a transit user process. We are able to extract key parameters from the data such as the flow and transition matrices to feed the Markovian process. This allows the extraction of key indicative properties and quantities of the system such as convergence, time to absorption, absorption probabilities, and state density distributions. Ultimately, the models will evolve to be able to scale and handle systems of high complexity. However, location and system specific development and design of filtering techniques, sensor placement algorithms, and sensor features have to be developed and in place for more reliable modeling.

VII. CONCLUSION
In this paper, we presented an in-house total system design of a ubiquitous BT and WiFi detection technology. We proposed the use of anonymous and encrypted BT and WiFi data obtained from the users of a pedestrian network to make real-time decisions. The proposed system will have many implications on understanding pedestrians behavioral patterns and estimating key states of the public transit networks, such as density, dwell times, and OD flows. We conducted pilot tests in two high-volume public transit agencies in NYC and were able to demonstrate the usefulness of the data. However, all the conducted studies highlight the importance of location and system specific development and design of filtering techniques, sensor placement algorithms, and sensor features. System-specific stochastic modeling and estimation techniques are currently being developed.

ACKNOWLEDGMENT
The work is partially funded by UTRC at CUNY and New York University City Safety and Mobility Analysis and Resilient Transportation System (CitySMART) laboratory of NYU’s UrbanITS Center. The authors would like to thank MTA and PA for their support for field tests. The contents of this paper reflect views of the authors who are responsible for the facts and accuracy of the data presented herein. The contents of the paper do not necessarily reflect the official views or policies of the agencies.

REFERENCES


