



University Transportation Research Center - Region 2

Final Report



Real-Time Estimation of Transit OD Patterns and Delays Using Low Cost- Ubiquitous Advanced Technologies

Performing Organization: New York University



April 2017



Sponsor:
University Transportation Research Center - Region 2

University Transportation Research Center - Region 2

The Region 2 University Transportation Research Center (UTRC) is one of ten original University Transportation Centers established in 1987 by the U.S. Congress. These Centers were established with the recognition that transportation plays a key role in the nation's economy and the quality of life of its citizens. University faculty members provide a critical link in resolving our national and regional transportation problems while training the professionals who address our transportation systems and their customers on a daily basis.

The UTRC was established in order to support research, education and the transfer of technology in the field of transportation. The theme of the Center is "Planning and Managing Regional Transportation Systems in a Changing World." Presently, under the direction of Dr. Camille Kamga, the UTRC represents USDOT Region II, including New York, New Jersey, Puerto Rico and the U.S. Virgin Islands. Functioning as a consortium of twelve major Universities throughout the region, UTRC is located at the CUNY Institute for Transportation Systems at The City College of New York, the lead institution of the consortium. The Center, through its consortium, an Agency-Industry Council and its Director and Staff, supports research, education, and technology transfer under its theme. UTRC's three main goals are:

Research

The research program objectives are (1) to develop a theme based transportation research program that is responsive to the needs of regional transportation organizations and stakeholders, and (2) to conduct that program in cooperation with the partners. The program includes both studies that are identified with research partners of projects targeted to the theme, and targeted, short-term projects. The program develops competitive proposals, which are evaluated to insure the most responsive UTRC team conducts the work. The research program is responsive to the UTRC theme: "Planning and Managing Regional Transportation Systems in a Changing World." The complex transportation system of transit and infrastructure, and the rapidly changing environment impacts the nation's largest city and metropolitan area. The New York/New Jersey Metropolitan has over 19 million people, 600,000 businesses and 9 million workers. The Region's intermodal and multimodal systems must serve all customers and stakeholders within the region and globally. Under the current grant, the new research projects and the ongoing research projects concentrate the program efforts on the categories of Transportation Systems Performance and Information Infrastructure to provide needed services to the New Jersey Department of Transportation, New York City Department of Transportation, New York Metropolitan Transportation Council, New York State Department of Transportation, and the New York State Energy and Research Development Authority and others, all while enhancing the center's theme.

Education and Workforce Development

The modern professional must combine the technical skills of engineering and planning with knowledge of economics, environmental science, management, finance, and law as well as negotiation skills, psychology and sociology. And, she/he must be computer literate, wired to the web, and knowledgeable about advances in information technology. UTRC's education and training efforts provide a multidisciplinary program of course work and experiential learning to train students and provide advanced training or retraining of practitioners to plan and manage regional transportation systems. UTRC must meet the need to educate the undergraduate and graduate student with a foundation of transportation fundamentals that allows for solving complex problems in a world much more dynamic than even a decade ago. Simultaneously, the demand for continuing education is growing – either because of professional license requirements or because the workplace demands it – and provides the opportunity to combine State of Practice education with tailored ways of delivering content.

Technology Transfer

UTRC's Technology Transfer Program goes beyond what might be considered "traditional" technology transfer activities. Its main objectives are (1) to increase the awareness and level of information concerning transportation issues facing Region 2; (2) to improve the knowledge base and approach to problem solving of the region's transportation workforce, from those operating the systems to those at the most senior level of managing the system; and by doing so, to improve the overall professional capability of the transportation workforce; (3) to stimulate discussion and debate concerning the integration of new technologies into our culture, our work and our transportation systems; (4) to provide the more traditional but extremely important job of disseminating research and project reports, studies, analysis and use of tools to the education, research and practicing community both nationally and internationally; and (5) to provide unbiased information and testimony to decision-makers concerning regional transportation issues consistent with the UTRC theme.

Project No(s):

UTRC/RF Grant No: 49997-44-25

Project Date:

April 2017

Project Title: Real-Time Estimation of Transit Origin-Destination Patterns and Delays Using Low Cost-Ubiquitous Advanced Technologies

Project's Website:

<http://www.utrc2.org/research/projects/real-time-estimation-transit-origin>

Principal Investigator(s):

Kaan Ozbay, Ph.D.

Professor and Director

C2SMART Center (A Tier 1 USDOT UTC)

Department of Civil and Urban Engineering & Center for Urban Science & Progress (CUSP)

Tandon School of Engineering

New York University

Brooklyn, NY 11201

Tel: (646) 997-3691

Email: kaan.ozbay@nyu.edu

Co Principal Investigator(s):

Neveen Shlayan, Ph.D.

Assistant Professor

Department of Electrical Engineering

The Cooper Union for the Advancement of Science and Art

New York, NY 10003

Tel: (212) 353-4333

Email: nshlayan@cooper.edu

Hani Nassif, P.E., Ph.D.

Professor

Department of Civil & Environmental Engineering

Rutgers University

Piscataway, NJ 08854

Tel: (848) 445-4414

Email: nassif@soe.rutgers.edu

Performing Organization(s):

New York University

Sponsor(s):

University Transportation Research Center (UTRC)

To request a hard copy of our final reports, please send us an email at utrc@utrc2.org

Mailing Address:

University Transportation Research Center

The City College of New York

Marshak Hall, Suite 910

160 Convent Avenue

New York, NY 10031

Tel: 212-650-8051

Fax: 212-650-8374

Web: www.utrc2.org

Board of Directors

The UTRC Board of Directors consists of one or two members from each Consortium school (each school receives two votes regardless of the number of representatives on the board). The Center Director is an ex-officio member of the Board and The Center management team serves as staff to the Board.

City University of New York

Dr. Hongmian Gong - Geography/Hunter College
Dr. Neville A. Parker - Civil Engineering/CCNY

Clarkson University

Dr. Kerop D. Janoyan - Civil Engineering

Columbia University

Dr. Raimondo Betti - Civil Engineering
Dr. Elliott Sclar - Urban and Regional Planning

Cornell University

Dr. Huaizhu (Oliver) Gao - Civil Engineering

Hofstra University

Dr. Jean-Paul Rodrigue - Global Studies and Geography

Manhattan College

Dr. Anirban De - Civil & Environmental Engineering
Dr. Matthew Volovski - Civil & Environmental Engineering

New Jersey Institute of Technology

Dr. Steven I-Jy Chien - Civil Engineering
Dr. Joyoung Lee - Civil & Environmental Engineering

New York University

Dr. Mitchell L. Moss - Urban Policy and Planning
Dr. Rae Zimmerman - Planning and Public Administration

Polytechnic Institute of NYU

Dr. Kaan Ozbay - Civil Engineering
Dr. John C. Falcocchio - Civil Engineering
Dr. Elena Prassas - Civil Engineering

Rensselaer Polytechnic Institute

Dr. José Holguín-Veras - Civil Engineering
Dr. William "Al" Wallace - Systems Engineering

Rochester Institute of Technology

Dr. James Winebrake - Science, Technology and Society/Public Policy
Dr. J. Scott Hawker - Software Engineering

Rowan University

Dr. Yusuf Mehta - Civil Engineering
Dr. Beena Sukumaran - Civil Engineering

State University of New York

Michael M. Fancher - Nanoscience
Dr. Catherine T. Lawson - City & Regional Planning
Dr. Adel W. Sadek - Transportation Systems Engineering
Dr. Shmuel Yahalom - Economics

Stevens Institute of Technology

Dr. Sophia Hassiotis - Civil Engineering
Dr. Thomas H. Wakeman III - Civil Engineering

Syracuse University

Dr. Riyad S. Aboutaha - Civil Engineering
Dr. O. Sam Salem - Construction Engineering and Management

The College of New Jersey

Dr. Thomas M. Brennan Jr - Civil Engineering

University of Puerto Rico - Mayagüez

Dr. Ismael Pagán-Trinidad - Civil Engineering
Dr. Didier M. Valdés-Díaz - Civil Engineering

UTRC Consortium Universities

The following universities/colleges are members of the UTRC consortium.

City University of New York (CUNY)
Clarkson University (Clarkson)
Columbia University (Columbia)
Cornell University (Cornell)
Hofstra University (Hofstra)
Manhattan College (MC)
New Jersey Institute of Technology (NJIT)
New York Institute of Technology (NYIT)
New York University (NYU)
Rensselaer Polytechnic Institute (RPI)
Rochester Institute of Technology (RIT)
Rowan University (Rowan)
State University of New York (SUNY)
Stevens Institute of Technology (Stevens)
Syracuse University (SU)
The College of New Jersey (TCNJ)
University of Puerto Rico - Mayagüez (UPRM)

UTRC Key Staff

Dr. Camille Kamga: *Director, Assistant Professor of Civil Engineering*

Dr. Robert E. Paaswell: *Director Emeritus of UTRC and Distinguished Professor of Civil Engineering, The City College of New York*

Herbert Levinson: *UTRC Icon Mentor, Transportation Consultant and Professor Emeritus of Transportation*

Dr. Ellen Thorson: *Senior Research Fellow, University Transportation Research Center*

Penny Eickemeyer: *Associate Director for Research, UTRC*

Dr. Alison Conway: *Associate Director for Education*

Nadia Aslam: *Assistant Director for Technology Transfer*

Nathalie Martinez: *Research Associate/Budget Analyst*

Tierra Fisher: *Office Assistant*

Bahman Moghimi: *Research Assistant;
Ph.D. Student, Transportation Program*

Wei Hao: *Research Fellow*

Andriy Blagay: *Graphic Intern*

**REAL-TIME ESTIMATION OF TRANSIT OD PATTERNS AND DELAYS USING
LOW COST-UBIQUITOUS ADVANCED TECHNOLOGIES**

PRINCIPAL INVESTIGATOR:

Kaan Ozbay, Ph.D.
Professor & Director
C²SMART Center (A Tier 1 USDOT UTC)
Department of Civil and Urban Engineering &
Center for Urban Science & Progress (CUSP)
Tandon School of Engineering
New York University
Six MetroTech Center, Room 404
Brooklyn, NY, 11201
<http://engineering.nyu.edu/urbanmits/index.php>
Tel (NYU CUE): 646.997.3691
Email: kaan.ozbay@nyu.edu

CO-PRINCIPAL INVESTIGATORS:

Neveen Shlayan, Ph.D.
Assistant Professor
Department of Electrical Engineering
The Cooper Union for the Advancement of Science and Art
Tel: 212.353.4333
Email: nshlayan@cooper.edu

Hani Nassif, P.E., Ph.D., Professor, FACI
Dept. of Civil & Env. Engineering
Rutgers, The State Univ. of New Jersey
96 Frelinghuysen Road, Piscataway, NJ 08854
Contact Info:
Office: (848) 445-4414
RIME: (732) 445-5858
Fax: (732) 445-8268
Email: nassif@soe.rutgers.edu
RIME Group Website: <http://www.rime.rutgers.edu>
I-BEAM Laboratory Website: <http://ibeam.rutgers.edu/>

PROJECT DURATION:

DATE SUBMITTED: April 1st, 2017

**REAL-TIME ESTIMATION OF TRANSIT OD PATTERNS AND DELAYS USING
LOW COST-UBIQUITOUS ADVANCED TECHNOLOGIES**

Kaan Ozbay, Ph.D.

Professor & Director

Department of Civil and Urban Engineering & Center for Urban Science and Progress (CUSP)
C2SMART USDOT' Tier 1 University Transportation Center, Tandon School of Engineering
New York University

Abdullah Kurkcu, M.Eng.

Research Assistant

Department of Civil and Urban Engineering & Center for Urban Science and Progress (CUSP)
Tandon School of Engineering, New York University

Neveen Shlayan, Ph.D.

Assistant Professor

Department of Electrical Engineering, The Cooper Union for the Advancement of Science and Art

Sami Demireluk, Ph.D.

Research Associate

Department of Civil and Environmental Engineering, Rutgers, The State Univ. of New Jersey

Matthew D. Maggio

Software Programmer

Department of Civil and Environmental Engineering, Rutgers, The State Univ. of New Jersey

Hong Yang, Ph.D.

Assistant Professor

Department of Modeling, Simulation & Visualization Engineering, Old Dominion University

Hani Nassif, P.E., Ph.D.

Professor, FACI

Department of Civil & Environmental Engineering, Rutgers, The State Univ. of New Jersey

1. Report No.	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle REAL-TIME ESTIMATION OF TRANSIT OD PATTERNS AND DELAYS USING LOW COST-UBIQUITOUS ADVANCED TECHNOLOGIES		5. Report Date April 1, 2017	
		6. Performing Organization Code	
7. Author(s) Kaan Ozbay, PhD Neveen Shlayan, Ph.D Hani Nassif, P.E., Ph.D.,		8. Performing Organization Report No.	
9. Performing Organization Name and Address Center for Urban Science & Progress (CUSP) Tandon School of Engineering New York University Six MetroTech Center, Room 404 Brooklyn, NY, 11201		10. Work Unit No.	
		11. Contract or Grant No. 49997-44-25	
12. Sponsoring Agency Name and Address UTRC The City College of New York, Marshak 910 160 Convent Avenue New York, NY 10031		13. Type of Report and Period Covered Final, June 1, 2014 to April 2017	
		14. Sponsoring Agency Code	
15. Supplementary Notes			
16. Abstract <p>The main objective of this project is to develop and conduct limited testing of novel sensors using Bluetooth technology (BT) to estimate OD demands and station wait times for users of public transit stations. The NYU research team tested the feasibility of the utilization of sensors with Bluetooth technology to estimate Origin-Destination (OD) demands and station wait-times of users of transit systems with a focus on subway systems. For example, if the entrance and exit turnstiles at subway stations were equipped with this type of sensors, it is possible to capture OD information for some of the riders with activated devices.</p> <p>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 Bluetooth IDs uses inexpensive, small, 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.</p> <p>Following a literature review and device testing, a series of one-day pilot tests are conducted in coordination with the MTA to iron out all of the possible hardware and software issues. Following further consultation with the MTA, a full one day to one week indoor tests are conducted with continuous data collection and monitoring to assess the feasibility and usefulness of long-term data collection using the proposed sensor technology. Two software tools to post process the collected data and to perform self-diagnosis and remote data acquisition functions are developed as part of the overall research project. The results and recommendations are provided to the MTA and other interested transit agencies.</p>			
17. Key Words origin-destination, Bluetooth technology, station wait times, wireless detectors		18. Distribution Statement	
19. Security Classif. (of this report) Unclassified	20. Security Classif. (of this page) Unclassified	21. No of Pages 53	22. Price

Disclaimer

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. The contents do not necessarily reflect the official views or policies of the University Transportation Research Center. This report does not constitute a standard, specification or regulation. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government assumes no liability for the contents or use thereof.

Executive Summary

Monitoring non-motorized traffic is gaining increased attention in the context of transportation studies. Most of the traditional pedestrian monitoring technologies focus on counting pedestrians passing through a fixed location in the network. It is thus not possible to determine the movement of individuals or groups as they move outside of each particular sensor's range. Moreover, most transportation agencies do not have continuous pedestrian counts mainly because of technological limitations. However, wireless data collection technologies can capture crowd dynamics by scanning mobile devices. Data collection methods that take advantage of mobile devices have gained much interest in the transportation literature due to its low cost, ease of implementation and richness of captured data (Carpenter, Fowler, & Adler).

The main objective of this project is to develop and conduct limited testing of novel sensors using Bluetooth technology (BT) to estimate OD demands and station wait-times for users of public transit stations. The NYU research team tested the feasibility of the utilization of sensors with Bluetooth technology to estimate Origin-Destination (OD) demands and station wait-times of users of transit systems with a focus on subway systems. For example, if the entrance and exit turnstiles at subway stations were equipped with this type of sensors, it is possible to capture OD information for some of the riders with activated devices.

Estimation of daily and hourly OD demands and delays is important for transit agencies because it can help improve their operations and make time decisions to reduce delays and mitigate cost, among other benefits. The proposed method of tracking Bluetooth IDs uses inexpensive, small, 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.

Following a literature review and device testing, a series of one-day pilot tests are conducted in several locations with different pedestrian traffic characteristics to identify and address all of the possible hardware and software issues. A full one day to one week indoor tests are conducted with continuous data collection and monitoring to assess the feasibility and usefulness of long-term data collection using the proposed sensing technology. Two software tools to post process the collected data and to perform self-

diagnosis and remote data acquisition functions are developed as part of the overall research project.

Finally, A Markov Chain Model (MCM) is developed to model the general attributes of a pedestrian network such as density, dwell times, and OD flows. Markov Chains provide a mathematical structure appropriate to modeling discrete events, which we find to be suitable for understanding pedestrian behavior in our case study.

This project demonstrated the feasibility of building low-cost ubiquitous sensors that can be used to collect pedestrian data to estimate OD flows and waiting times of pedestrians. However, due to the complexity of pedestrian traffic, more work is needed to improve data quality of these sensors by developing advanced filtering methods that can be deployed in the field. The MCM model is tested for a toy network with limited data. It can also be improved to be used under real-world conditions for large size indoor and outdoor transportation networks with heavy pedestrian traffic.

Table of Contents

1	INTRODUCTION.....	5
1.1	Research Objective	5
1.2	Research Approach	7
2	LITERATURE REVIEW & BACKGROUND	8
2.1	Wireless Technology and Its Use in Transportation Studies	8
2.2	Privacy Issues.....	11
2.3	Commercial Solutions Using Bluetooth	12
3	HARDWARE, SOFTWARE, AND DATA ANALYTICS	14
3.1	Hardware.....	15
3.2	Software	16
3.3	Anonymization.....	18
3.4	Data Access and Remote Self-Diagnostics.....	19
4	EQUIPMENT AND SOFTWARE TESTING	21
4.1	Path Discovery of Pedestrians using Bluetooth Technology: Student Center.....	22
4.1.1	Introduction	22
4.1.2	Test Objectives and Experimental Setup.....	22
4.1.3	Summary of Results.....	23
4.2	Pedestrians in an Office Environment (Brooklyn IV)	26
4.2.1	All Data.....	26
4.2.2	Data by Sensor.....	28
4.2.3	Data by Unique BT ID.....	29
4.3	Controlled Experiments	32
4.3.1	Objectives	33
4.3.2	Experimental Setup and Results	33
4.3.3	Conclusion	36
4.4	The Subway Station Test	37
4.4.1	Introduction	37
4.4.2	Test Objectives and Parameters.....	38
4.4.3	Summary of Results.....	38
4.4.4	Individual Results for each Testing Period.....	40
5	PEDESTRIAN MOBILITY PREDICTION MODEL	45

*Real-time Estimation of Transit OD Patterns and Delays Using Low-Cost Ubiquitous
Advanced Technologies*

5.1	Introduction.....	45
5.1.1	Crowd Modeling and our Approach.....	45
5.2	Brooklyn IV Case Study	46
5.2.1	Data-based MCM	47
5.2.2	Analysis and Results.....	48
6	CONCLUSIONS	51
	DISCLAIMER	53
	References	54

List of Figures

Figure 1: MAC Address Architecture.....	15
Figure 2: Physical Features of the Tablet	16
Figure 3: Android App Interface.....	17
Figure 4: Traffic Tracker Scan Screen.....	17
Figure 5: Depiction of the Web-based Self-Diagnosis and Data Access Application	19
Figure 6: Sketch of Student Center.....	22
Figure 7: Paths of the Pedestrians Detected at All Locations on Day II.....	25
Figure 8: Detections Received from All Three Sensors: Sensor 1 (blue), Sensor 2 (green), and Sensor 3 (red).....	27
Figure 9: Visualization of BT IDs at Sensor Locations over Time	28
Figure 10: Elapsed Time of Transitions between Sensors that are less than 20 mins	31
Figure 11: Elapsed Time of Transitions between Sensors that are less than 1.5 mins	32
Figure 12: Trajectories of Two Unique BT IDs.....	33
Figure 13: Path Verification Experiment Plan.....	34
Figure 14: Ground Truth Data (Green) vs. Sensor Data.....	36
Figure 15: Locations of the Tablets in the Subway Station	37
Figure 16: Hourly Counts from each Device during Test #1.....	41
Figure 17: Hourly Average Wait-times from each Device during Test #1.....	42
Figure 18: Hourly Movement Counts - Tablet 1 vs. Others during Test #1	43
Figure 19: Hourly Movement Counts - Tablet 2 vs. Tablets during Test #1	44
Figure 20: Hourly Movement Counts - Tablet 3 vs. Others during Test #1	44
Figure 21: Markov Chain Model for the Brooklyn IV Case Study	47
Figure 22: Convergence of the n-step transition matrix as $n \rightarrow \infty$	49
Figure 23: The Density Distribution for all three Locations.....	50

List of Tables

Table 1: Information Structure Collected by the Bluetooth Reader	7
Table 2: Asus Nexus 7 Tablet Specifications	15
Table 3: The Accuracy of Detections using the Last N Digits	18
Table 4: Conducted Tests.....	21
Table 5: Detected Devices and Pedestrian Counts at Test Locations	23
Table 6: The Number of Detected Bluetooth Devices at Multiple Locations	23
Table 7: Summary Statistics for eT between Subsequent Data Points for each Sensor.	29
Table 8: Intersection of Unique BT IDs Between Sensors.	30
Table 9: The Weighted Aggregation of the Summary Statistics for the Elapsed Time.	31
Table 10: Period of Data Collection by Device	37
Table 11: Total Number of Detected Devices and Average Waiting Times	38
Table 12: Daily Average Counts and Waiting Times for Tablet 3	39
Table 13: The Number of Matches between Tablets per Testing Period	40
Table 14: Daily Counts and Average Wait Times for Peak and Off-peak Periods	40
Table 15: The Movement Codes between Tablets.....	42

1 INTRODUCTION

Estimation of daily and hourly Origin-Destination (OD) demands and wait-times at transit stations are important for transit agencies because it can help improve their operations and make timely decisions to reduce delays and cost, among other benefits. In the past, transit agencies such as New York City's MTA and NJ Transit have been mainly using surveys to estimate OD demands and wait-times. However, surveys are generally very expensive and time consuming for large systems, such as the NYC subway, and consequently limited in size and information they provide. Moreover, they do not provide information that can accurately capture time and frequency of OD demands between stations. The estimation of wait-times for transit riders using surveys or travel diaries is also problematic due to the difficulties of sample accuracy, bias, to name a few. It is well known that survey respondents can have difficulties in accurately remembering and recording their wait-times. Some other technology oriented methods for pedestrian mobility monitoring also include fixed pedestrian counters and vision based technologies but these cannot capture OD demands. There are also new ticketing technologies that can capture OD information but their current level of implementation in the NY-NJ-CT tri-State area, coupled with privacy issues, make acquiring this data several years away. In addition, the wait-time information cannot be not easily captured by any of the ticketing technologies and additional sensors will be needed. In the recent years, there have been several studies in the literature to automate pedestrian detection or counting to explore economical and reliable methods (Ozbay, Bartin, Yang, Walla, & Williams, 2010; Ozbay, Yang, & Bartin, 2010).

1.1 Research Objective

The main objective of this project is to develop and conduct limited testing of novel sensors using Bluetooth technology (BT) to estimate OD demands and station wait-times for users of public transit stations. If the entrance and exit turnstiles at subway stations are equipped with this type of sensors, it is possible to capture OD information for some percentage of the riders with visible Bluetooth devices. This information can be used anonymously to detect origin and destination of riders by matching data collected at entrances and exits from the system. Assuming that visible Bluetooth enabled devices are

uniformly distributed among the riders, it is possible to estimate a transit OD matrix for the entire system in terms of demand percentages. Moreover, other publicly available transaction data can be used in conjunction with the OD demand percentages generated from the wireless sensor observations to estimate number of actual OD trips. It is clearly valuable for any transit agency to estimate heavily used OD pairs for scheduling and operations, as well as agency costs. The wait time estimation utilizes the same sensors that will be deployed at specific locations by recording the time each individual BT enabled device remains at a given location before departing to their final destination. This information can help transit agencies to optimize their train and bus schedules in order to minimize wait times. More importantly, the optimization of schedules can be done in real or almost real-time due to the real-time availability of wait times from the sensors.

1.2 Research Approach

The research team has developed a Java based application that can be run on any Android device to detect and record the active Bluetooth devices within a certain perimeter. The developed app running on a tablet records, timestamp and location of detection of active mobile devices. This tablet running the developed app can be placed in an encasement that can be secured at a certain location in a pedestrian facility or transit station. Additional batteries can be added to prolong battery life for a single usage. Then, this encasement can be placed near the entrance and the exit points to determine OD flows. The device can store the OD and wait information and upload it to a central server such so that the information can be accessed remotely.

Table 1: Information Structure Collected by the Bluetooth Reader

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

Table 1 illustrates the structure of the information that is collected by the Bluetooth reader and stored in the database. By matching the IDs which are anonymized encrypted at the time of detection and unidentifiable to individual devices or persons, the entrance and exit stations of a rider can be known. The timestamp can also be used to identify the travel-time, including wait-time, of the rider. The aggregated OD flow values and time-dependent average wait times will be the only stored information when these types of sensors are actually deployed. Thus, no individual data point will need to be stored in the case of actual deployment.

2 LITERATURE REVIEW & BACKGROUND

2.1 Wireless Technology and Its Use in Transportation Studies

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. Bluetooth is a wireless technology for short-range communications from fixed or mobile devices. The technology was first designed to replace data cables while maintaining the high levels of security. The key features of Bluetooth technology are robustness, low power, and low cost (Jie, Na-na, Ji-lin, Yong-feng, & Zheng, 2008). Bluetooth 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 Bluetooth 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 Bluetooth devices are discoverable, but in general it has been reported that 5%-12% of devices are discoverable via Bluetooth (Brennan Jr et al., 2010). This sample size is adequate for most transportation studies.

The use of anonymous Bluetooth data is gaining popularity for data collection for pedestrian detection. The general idea is to track anonymous Bluetooth IDs using inexpensive, small and easy to deploy Bluetooth detectors that consist of low cost smart phones and/or tablets that are loaded with specialized software developed by the research team. This type of technology is now widely used by highway agencies and the most important technical and privacy issues have already been resolved (Puckett & Vickich, 2010). The work presented by Ahmed et. al. (2008) is among the first to utilize Bluetooth detection for vehicle monitoring. The contribution of this work is the deployment of a very low cost and low power device/software combination for transit related OD estimation applications for the first time.

Kostakos (2008) used Bluetooth devices to trace passenger journeys on public buses and derive passenger OD matrices. Bullock et al. (2010) deployed a Bluetooth 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. Hamed et al. (2009) investigated the quality of vehicle probe data using new traffic surveillance devices based on Bluetooth technology. Their results showed that the technology is a promising method for collecting high quality travel time data that can be used for evaluating other sources of travel time and intelligent transportation systems (ITS) applications. Haseman et al. (2009) also used Bluetooth probe data from multiple field collection sites to quantify delay and to assess diversion rates at a rural Interstate Highway work zone along I-65 in Northwestern Indiana.

Malinovskiy et. al. (2012) presented a study of pedestrian detection using Bluetooth at two separate sites, Montreal and Seattle. They investigated the feasibility of using Bluetooth for pedestrian studies and found out that it can provide useful information for pedestrian travel behavior. Barceló et. al. (2012) presented Ad Hoc procedures based on Kalman Filtering. Their approach used the explicit travel time measurements from Bluetooth detectors for estimating time dependent OD matrices. Results showed that the proposed approach to dynamic OD matrix estimation provides good estimates of target values. Although the time of detection is known precisely, it is really challenging to find the location of the device. Tuning the antenna features and power levels can reduce the detection radius. However, this may lead to another potential problem of not being able to detect devices that are not within the detection range due to random delays in the process of detection which can go up to 10 seconds. Most of these problems were addressed in Lees-Miller et.al.'s (2013) study. They tried to recover the path of a vehicle using only Bluetooth detection data and used Hidden Markov Models. The proposed approach was able to reconstruct vehicle trajectories outperforming a simple deterministic strategy by 30-50%.

Michau et. al. (2014) pointed out that the position of the detectors is of great importance and that the Bluetooth signals are easily weakened by physical conditions as well as weather. The detection process of a Bluetooth device can be described as a cycle during which the sniffer will transmit messages on different range of frequencies and waiting for devices to pick up that message. This requires some time to be completed. Therefore, Bluetooth devices have to be in discoverable mode approximately 10 seconds

within the detection zone in order to detect them (Michau et al., 2014). The filtering methodology of Bluetooth data plays a key role in estimating any kind of traffic state estimation or prediction in terms of research. Laharotte et. al. (2015) provided some insights on how BT data can be used for flow prediction. Their filtering algorithm reconstructs traffic states at a network scale using non-parametric pattern recognition techniques with a k-nearest-neighbors (kNN) procedure. Their prediction of the network traffic state with a kNN approach showed convincing results using 31 days of data.

Integration of Wi-Fi systems to Bluetooth sensors can be seen in the recent studies of real-time data collection and monitoring of pedestrian networks. However, Wi-Fi monitoring requires that devices are connected to a certain wireless network and that the network covers the entire study area. Lesani et. al. (2016) investigated the advantages and the feasibility of a Wi-Fi data collection system as an alternative and a supplement to BT technology. They found that the detection rate for BT is as low as 2.0% and the combination of Wi-Fi and BT systems showed promising results. Hourly travel time estimations errors were around 3.8%. The average and median prediction errors of pedestrian flows were 15% and 9% respectively. Weppner, Lukowicz, Blanke, and Troster (2014) used Bluetooth scanners to count the number of devices in a fixed region. Nicolai and Kenn (2007) presented a method to find out the relationship between detected Bluetooth devices and the ground truth data. Kalogianni et al. (2015) used passive Wi-Fi scanning method to sense the movements of students, employees and visitors in a university campus. They investigated what kind of patterns can be captured by WiFi monitoring and how people utilize the buildings at the campus. The results pointed out that passive Wi-Fi monitoring is an effective way to identify building usage and movement between buildings. Bonne, Barzan, Quax, and Lamotte (2013) developed a low-cost crowd counting system based on a single-board computer with the addition of a LED to provide a status indicator and an Android cell phone as an operator. 15 devices were deployed in a music festival and 4 in a university campus. They concluded that tracking visitors at mass events can be achieved by using Raspberry PI sensors at a very low cost. N Abedi, Bhaskar, and Chung (2013) used a commercial sensor with the capability of scanning both Bluetooth and Wi-Fi addresses simultaneously. They compared the standards for both technologies regarding architecture, discovery time,

signal strength and popularity of use. The results pointed out that Wi-Fi has shorter discovery time, the distance from the sensor can be estimated based on the signal strength, and Wi-Fi is accepted as the more appropriate standard compared to Bluetooth for pedestrian data collection. Naeim Abedi, Bhaskar, Chung, and Miska (2015) evaluated antenna characteristics and concluded that the bigger antenna gains capture more data, but they may not be useful for small scales of monitoring due to overlapping detection areas. Schauer, Werner, and Marcus (2014) used both Wi-Fi and Bluetooth sensors to estimate crowd densities and pedestrian flow at a major German airport. Additional studies are dealing with pattern mining in tourist attraction visits (Versichele et al., 2014), Bayesian approach to detect destinations (Danalet, Farooq, & Bierlaire, 2014), and location popularity and visit patterns (Vu, Nguyen, Nahrstedt, & Richerzhagen, 2015) can be found in the literature.

Most of the mentioned studies allude to the fact that Bluetooth 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 highly depend on the system at hand.

2.2 Privacy Issues

Privacy is an important issue that needs to be addressed in any data collection studies involving the public. The electronic identifier for Bluetooth devices (MAC ID) contains two parts: the first part is assigned by the manufacturer of the device, and the second part is assigned to the specific device. Data collection systems should not store personally identifiable data; therefore, the MAC addresses collected by the Bluetooth reader should not be associated with specific users. In a research study by Texas Transportation Institute (2010), a routine is added to Bluetooth data collection software to encrypt the MAC addresses collected. This is done to make sure that actual device addresses are not stored anywhere, but rather a random set of characters is used. For example, a device with the MAC address 00:24:9F:E1:FE:98 might be encrypted to MDA6MjM6RDc6REQ6MzI6QkM by their software (Puckett & Vickich, 2010). In this study, a similar encryption approach is proposed to ensure maximum privacy while

maintaining persistent records, which is explained in detail in the next section. The information that will be collected will be encrypted at the moment of data collection so that no one will have access to any personal information other than the authorized person who holds the secure encryption key.

The other studies in the literature utilized similar encryption approaches to achieve anonymous detection. The detected electronic identifier of the BT device is converted into an encrypted hash code and this hash code is stored on the device in some studies. The unique identifiers that are not matched by two sensors were deleted at the site in studies exploring pedestrian movements. (Malinovskiy et al., 2012; Michau et al., 2014) . In others, it has been stated that either privacy concerns for the end users are a non-issue when the data collected through Bluetooth or the data was kept anonymous without being directly tied to individuals (Barceló et al., 2012; Carpenter et al., 2012; Laharotte et al., 2015).

For the use cases identified in this study namely, estimation of OD flows and wait times, there is no need to store any individual data even if they are anonymized and encrypted. The OD flow and with time data will be in the form of aggregated total number of detections and average wait times for a given time period by obviating the need for storing any individual data points.

2.3 Commercial Solutions Using Bluetooth

There are some commercial solutions available for pedestrian tracking. Sensys Networks presents a traffic data collected via wireless sensors and analytical platform for managing corridors and intersections (SENSYS). They provide real-time bicycle counts, travel times, intersection delay, OD, volume, occupancy and traffic speed data through their sensors. Another company that provides traffic data using BT sensors is Clearview Intelligence (Clearview). Their sensor is designed to integrate with physical infrastructure. It can provide temporary or permanent installations for monitoring the flow of traffic. Although Tyco's BT unit is similar, it is designed to be mounted on a pole (Tyco). It identifies and time stamps detected agents for transmission to a server via mobile wireless or wired networks. Trafficast is another wireless device computing travel

times and vehicle behaviors using BT signals emanating from passing vehicles. The detection radius for most of these sensors stated as 50 meters. BlipTrack sensors from Denmark-based BLIP Systems measure and predict movements of people and vehicles (Blip). It allows decision makers in various situations, including airports, road traffic, train stations, ports, ski resorts, amusement parks and more, to reduce delays and queues, travel times, optimize resources and planning, and improve the traveler experience and retail.

3 HARDWARE, SOFTWARE, AND DATA ANALYTICS

It is possible to estimate the proximity of pedestrians to the sensors when their devices are actively looking for other devices. Majority of commuters carry either a mobile phone or a smart device that is basically a handheld minicomputer with multiple means of data transmission such as Bluetooth and Wi-Fi. A Bluetooth device can either operate in slave mode, meaning that the device is controlled by another device, or in master mode. Whenever a BT device is powered on, it may try to operate as one of the aforementioned modes. Once the master device and the slave device know each other's addresses, two devices synchronize over the frequency hopping sequence, in other words, piconets. A piconet is a network of connected devices via Bluetooth technology. There can be as many as 8 active devices at once in a piconet; however, only one device can be a master.

The Bluetooth protocol for establishing connections has a layered structure. These layers can be grouped into two main parts as a controller and a host stack. The controller stack is installed on the hardware. It contains the Bluetooth radio and a microprocessor. The host stack, on the other hand, is implemented as software and deals with high-level data. While the controller establishes the connection, the host stack controls the protocols for packet handling and modifies parameters for discoverability. During the connection, BT devices send message packets including their anonymous MAC IDs and Received Signal Strength Indicator (RSSI). MAC addresses are the most common unique identifiers in IEEE 802 network technologies. It consists of 6 bytes/48 bits which make it possible to generate 248 potential unique MAC addresses. The first three bytes contains an organizationally unique identifier (OUI), and the following three are assigned by the organization in any manner as long as it is unique. Figure 1 shows the structure of a MAC address.

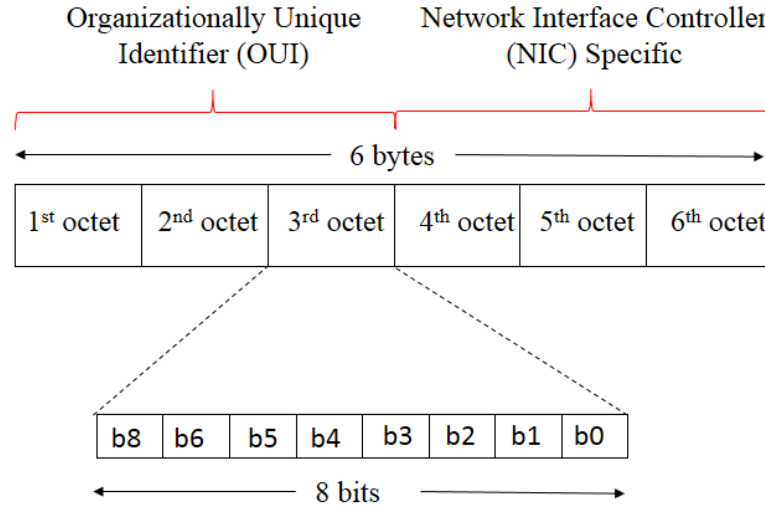


Figure 1: MAC Address Architecture

3.1 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 Bluetooth data. The specifications of the device are given in detail in Table 2. 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.

Table 2: Asus Nexus 7 Tablet Specifications

Operating System	Android 4.1
Display	7" WXGA (1280x800) Screen IPS Panel - 10 finger multi-touch support
CPU	NVIDIA® Tegra® 3 T30L Quad-Core @1.2Ghz
Memory	1GB
Storage	16GB*1
Wireless Data Network	WLAN 802.11 b/g/n@2.4GHz - Bluetooth
Camera	1.2 MP Front Camera
Interface	2-in-1 Audio Jack (head-out/MIC), 1x Docking PIN 1x micro-USB, 2x Digital microphone, 2x High Quality Speakers,
Sensor	G-Sensor, Light Sensor, Gyroscope, E-compass, GPS, NFC, Hall Sensor
Battery	9.5 hours, 4325mAh,*2 16Wh Li-polymer
Dimensions	198.5 x 120 x 10.45 mm
Weight	340 g

A sample Asus Nexus 7 tablet and its physical features can be seen from Figure 2 below. 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.

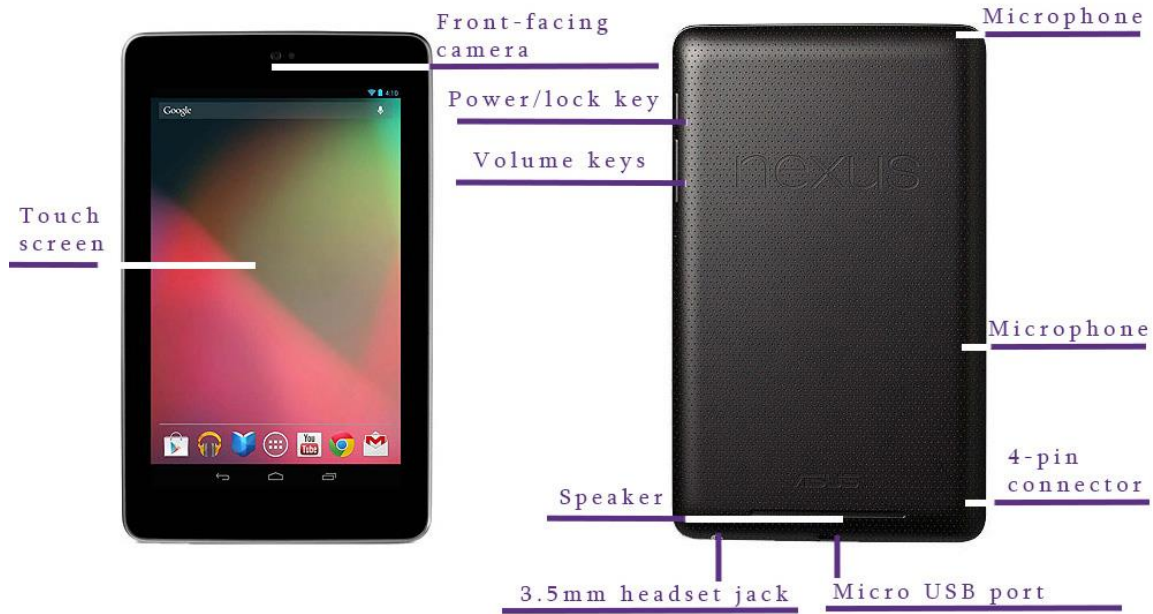


Figure 2: Physical Features of the Tablet

3.2 Software

The research team developed an application (app) called “Traffic Tracker” working on any Android device to detect BT devices. Traffic tracker scans the discoverable Bluetooth devices nearby and monitors messages in a way that their unique identifiers and signal strength information can be extracted and saved in tables as well as the detection times. Figure 3 shows the main screen when the app is initiated.

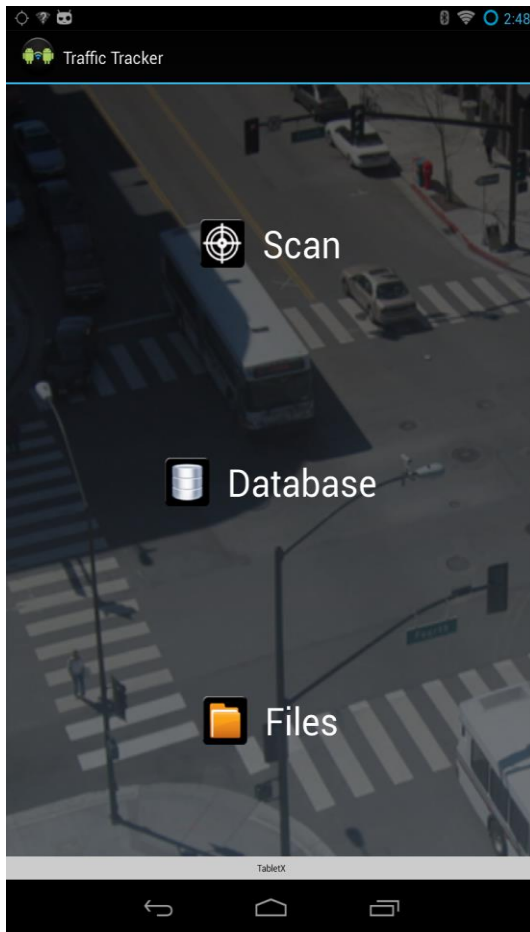


Figure 3: Android App Interface

During a scan, users can press the “Refresh” button to view the last 10 detected devices. Figure 4 shows the sample screen when users clicked on the “Refresh” button.

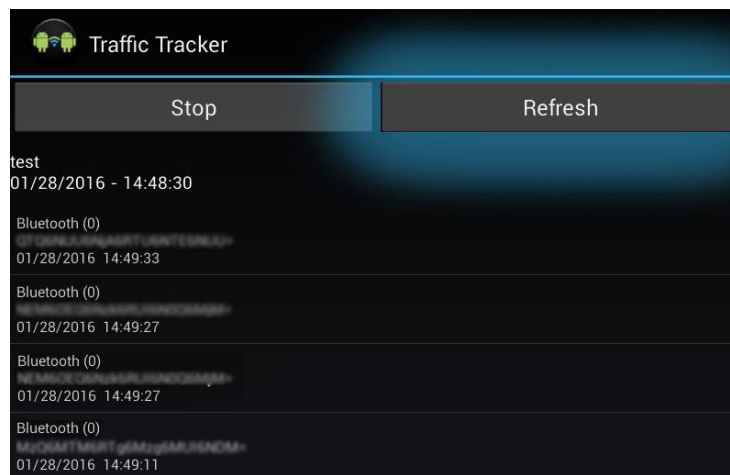


Figure 4: Traffic Tracker Scan Screen

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.

3.3 Anonymization

The MAC addresses of the detected devices consist of a unique identifier that can be tied to individuals. Therefore, this information has to be encrypted in a way such that it cannot be decrypted. 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)
- Aggregation (for OD and average wait time calculation studies)

In the encryption method, the MAC id is cut in half and the first seven digits of it are deleted. Then, the part containing the last five digits is encrypted using a key that is updated every 24 hours automatically in the app. To test the algorithm, the research team collected 8803 MAC addresses using the device and then analyzed the uniqueness of the ids using the last 6, 5 and 4 digits of the MAC address. The results can be seen in the table below.

Table 3: The Accuracy of Detections using the Last N Digits

# of Digits Used	# of actual Detections	# of Detections using N digits	Accuracy
6	8803	8802	99.98%
5	8803	8762	99.53%
4	8803	8252	93.74%

Since there are no major differences between using the last 6 digits and 5 digits in terms of accuracy, the research team decided to use the last 5 digits of the MAC address to make it even more challenging to retrieve the original address of the device. The last 5 digits of the address are then encrypted with the key generated every day before storing the data on the sensors. The encryption key is first randomly generated on a remote server. After the initial key is generated, it is then encrypted again before uploading to the devices on site.

In the aggregation method, only matching unique addresses between devices are stored in the dataset. These matching addresses are then replaced with a unique identifier that is derived from the order and the timestamp of the arrival. The original address is immediately then deleted. As mentioned previously, for the actual deployment, even

these anonymized data points will be deleted once OF flows and wait times for a predetermined time period is calculated.

3.4 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. There are many cloud-based file storage services available such as Dropbox, Google Drive, Microsoft Skydrive, etc. These services were investigated for suitability and security prior to the application development. 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.

One important question about the lack of internet access in most of the pedestrian facilities such as subway stations can be addressed by using one of the two approaches proposed by the research team: 1) for the filed study, choose stations with Wi-Fi and or internet access 2) for stations without internet and /or Wi-Fi access, use some of the deployed dPEDBT2 devices as Wi-Fi hubs that are connected to one device with internet access that is deployed at an area of the subway station, most likely at a location close to the exit where there is internet connection.

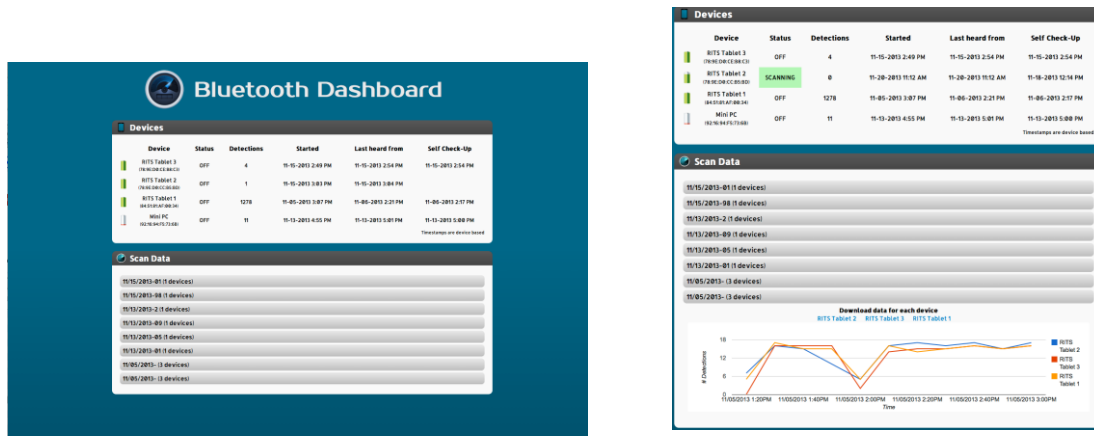


Figure 5: Depiction of the Web-based Self-Diagnosis and Data Access Application

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. The developed web page can be seen in in Figure 5. It enables authorized users to access the collected data and sensors. This is one of the most important improvements of the current set-up since the team will be able to identify any hardware and software related problems in near real-time. This allows a prompt treatment of the issues that may arise. This software-oriented task can be done in coordination with the equipment testing to ensure that the data is captured adequately.

4 EQUIPMENT AND SOFTWARE TESTING

Table 4: Conducted Tests

Study	Length	Data Points	Data Collection Locations	Power Usage	Coverage
Student Center	4 hours (over 2 days)	161	3 locations	2 hours	Passing pedestrians at 3 Exit/ Entrance of the building (additional details below)
Roadside (US 27)	2 hours	85	1 location	2 hours	Vehicles + Pedestrians
Roadside (I-287)	1 hours	39	1 location	2 hours	Vehicles
Brooklyn I	5 days	7792	3 locations	5 days (2 devices), 6 hours (1 device)	Pedestrians in a subway station
Brooklyn II	1 days	2159	2 locations	1 day, 7 hours	Pedestrians in a subway station
Brooklyn III	5 days	9755	2 locations	5 days, 38 hours	Pedestrians in a subway station
Brooklyn Bus	1.5 hours	11	2 locations	1.5 hours	Commuters inside a bus
University Building	4 days	104	4 locations	50 hours, 45 hours, 4 days, 4 days	4 floors of University building
Brooklyn IV	7 days		3 locations	Plugged in	Pedestrians in an office environment (additional details below)
Brooklyn V	1 day		3 locations	Plugged in	Pedestrians in an office environment
Student Center	1 hour		3 locations	3 hours	Passing pedestrians at 3 Exit/ Entrance of the building
Brooklyn VI	2 hours		4 locations	4 hours	Pedestrians at 2 Exit/Entrance and at 2 different floors of the building

Existing equipment is tested for various locations, situations, study lengths, and different purposes. Table 4 shows all the tests conducted. Developed BT detection and counting algorithms were improved and calibrated using the data generated from these field experiments. Fields tests with shorter durations were mostly conducted for calibration purposes. Longer tests provided more information about the data quality, battery life, and self-diagnostics. The most important field tests and their results will be explained in detail in the following sections.

4.1 Path Discovery of Pedestrians using Bluetooth Technology: Student Center

4.1.1 Introduction

For this study, three computers equipped with BT sensors were placed at the three exits of student center. Figure 6 shows the locations of Bluetooth receivers and attractions in the building. The student center experiences its busiest times during lunchtime. In addition, there is a bus stop in front of the center. Most students use the campus center, walk through the center, to reach the dorms or use the public transportation system on campus.

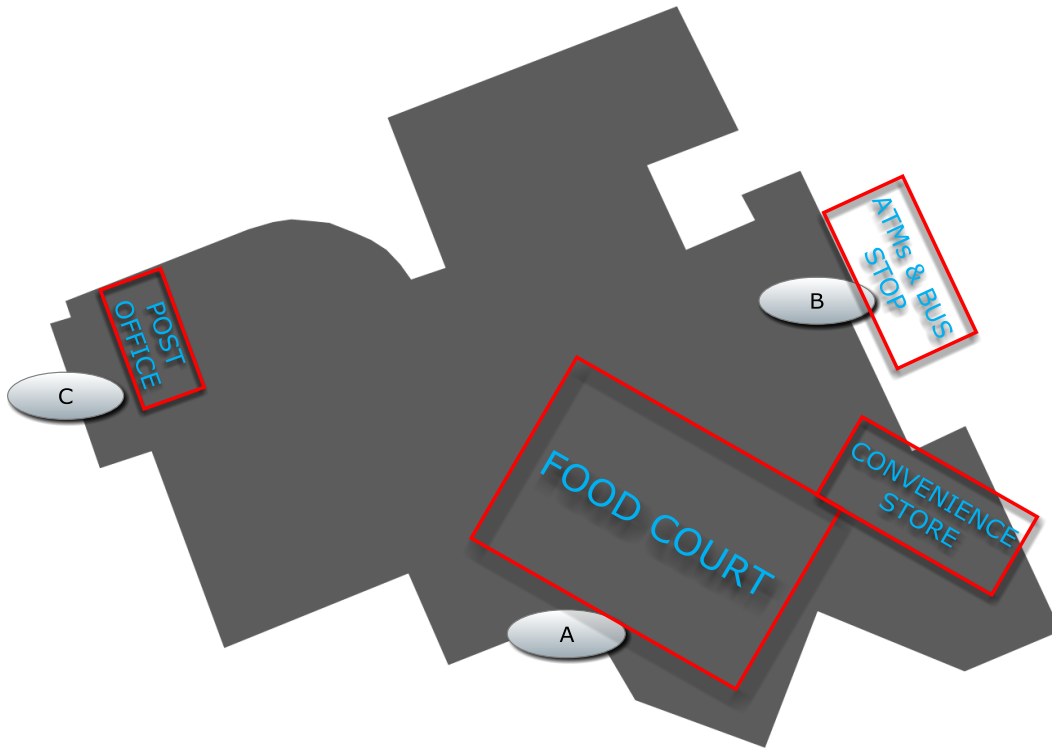


Figure 6: Sketch of Student Center

4.1.2 Test Objectives and Experimental Setup

The research team conducted an additional test using a different Bluetooth device to understand the maximum detection range of a BT sensor. At the control points, BT sensors, which have a capability to detect mobile devices up to 100 meters, were used. It is investigated that whether more devices than 5% of the population can be captured with

a higher detection range capability. For the experimental set up, the devices are connected to a laptop at locations A, B and C in Figure 6.

4.1.3 Summary of Results

At two different days, three graduate students were positioned close to exits of the center. The study was conducted between 12:00 PM to 2:00 PM since the student center was the most crowded at this time frame due to lunch. On Day I of the study, total of 75 Bluetooth devices were detected and on Day II, 86 devices were detected at the three locations. Table 5 shows the detected Bluetooth devices by day and by location with pedestrian counts. From the values in Table 5, the percentage of discoverable devices was found as 3.16%.

Table 5: Detected Devices and Pedestrian Counts at Test Locations

	# of Detected Bluetooth Devices			# of Pedestrians Counted		
	A	B	C	A	B	C
Day I	22	33	20	690	672	914
Day II	27	37	22	734	1022	1071

The purpose of this test was to observe the pedestrian behavior and track their paths using Bluetooth receivers; hence, the counts alone did not provide useful information. The paths of the pedestrians were found by matching the MAC addresses of Bluetooth devices at three exits. Table 6 shows the number of the devices seen on more than one location. On Day I, 22 devices were detected at more than one location and on Day II, 27 devices are detected at multiple locations.

Table 6: The Number of Detected Bluetooth Devices at Multiple Locations

	A-B	A-C	B-C	A-B-C
Day I	5	2	12	5
Day II	9	1	10	7

By matching the Bluetooth devices at multiple locations, the paths used by the pedestrian who was carrying the discovered Bluetooth device can be found. Moreover,

since the records also have timestamp, their activities while taking the path might be predicted to a degree. Figure 7 shows the paths of the pedestrians carrying Bluetooth devices, which are detected at all three locations on Day II.

From the Bluetooth data, the durations of the paths and the time spent close to a control point was calculated. Table 6 shows the durations for the paths. In this table, the time spent close to a control point is shown as “X-X”. From the results,

- Pedestrians, generally spend time close to point B. This is an expected result as the point B is located very close to food court, convenience store and ATMs.
- Pedestrians leave the building from the same point where they enter. This result is also plausible as on-campus employees are going from their offices to the student center and then returning to their offices.
- In general, pedestrians do not have complex paths. Only four pedestrians are observed to have 4 or 5 paths.

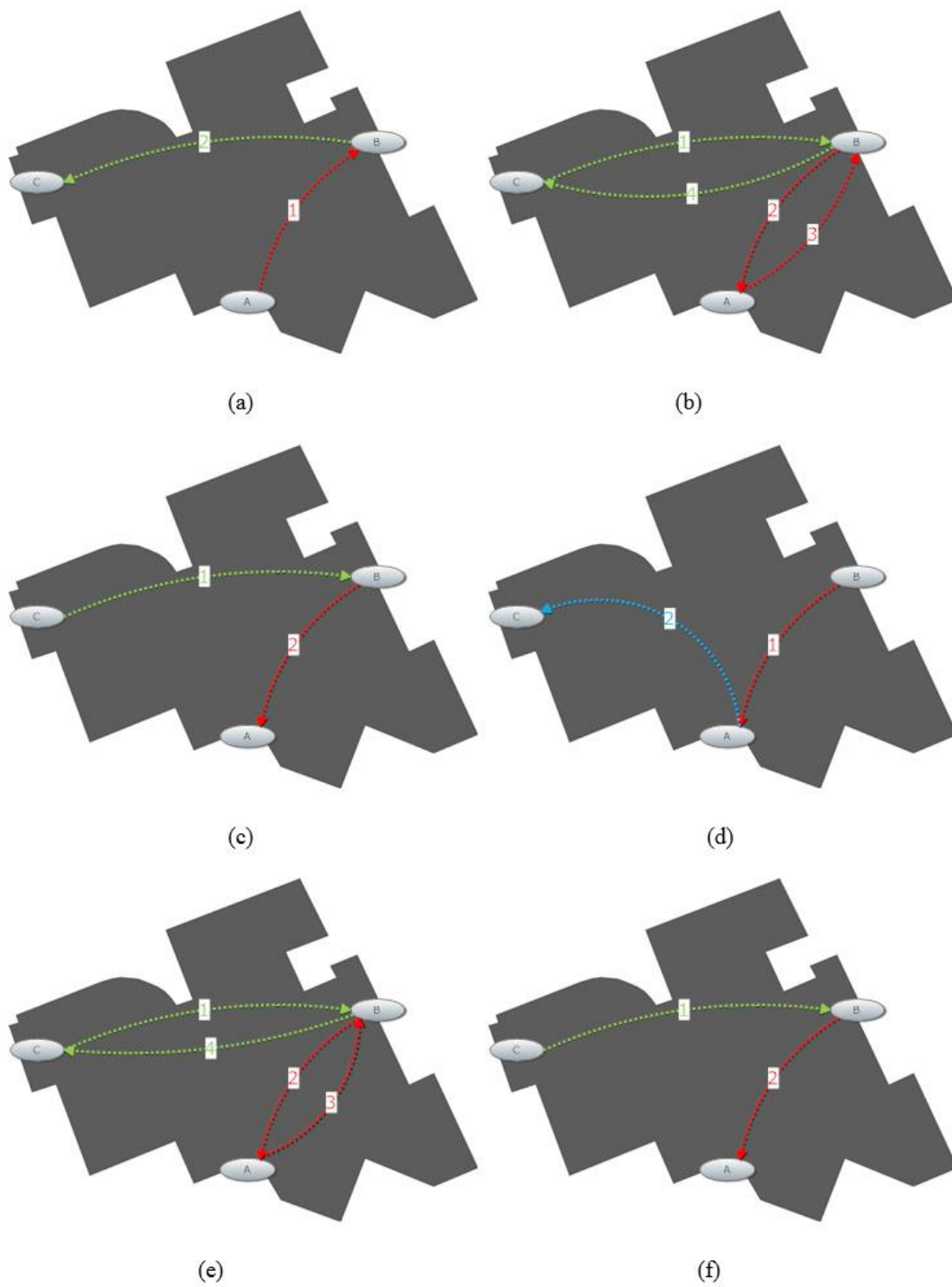


Figure 7: Paths of the Pedestrians Detected at All Locations on Day II

4.2 Pedestrians in an Office Environment (Brooklyn IV)

The data obtained from this pilot test conducted is processed to obtain basic visualizations and summary statistics of various parameters. The presented analysis is focused on mainly three quantities *travel time*, *wait time*, and *transition patterns*. Both, travel time and wait time branch out from the derived quantity of *elapsed time* (eT) and can be directly inferred from the data. Transition patterns, on the other hand, can be inferred from the sensor location data. Additional quantities may be obtained and will be briefly discussed. The data is processed at three aggregation levels all data, by sensor, and by unique BT ID.

4.2.1 All Data

Data from three different sensors, referred to as sensor 1 (blue), sensor 2 (green), and sensor 3 (red), is obtained. The data indicates 8161 total observations and 26 unique BT MAC addresses. Figure 8 shows the operation periods of each sensor; they are found to be somewhat consistent with the previously reported downtimes. However, there are unexplained gaps in the data particularly from sensors 1 and 2. Note that data recorded by sensor 3 consistently spans hours throughout the whole day, whereas data from sensor 2 is mainly obtained during the afternoon and night hours. The data recorded by sensor 1 is inconsistent, yet limited and, therefore, conclusions cannot be drawn. The reasons for such patterns are yet to be explored. Consistent early morning BT ID detection can occur due to a stationary device for instance. Other detection methods such as IR or video may assist in verifying whether the BT detections are actual pedestrians.

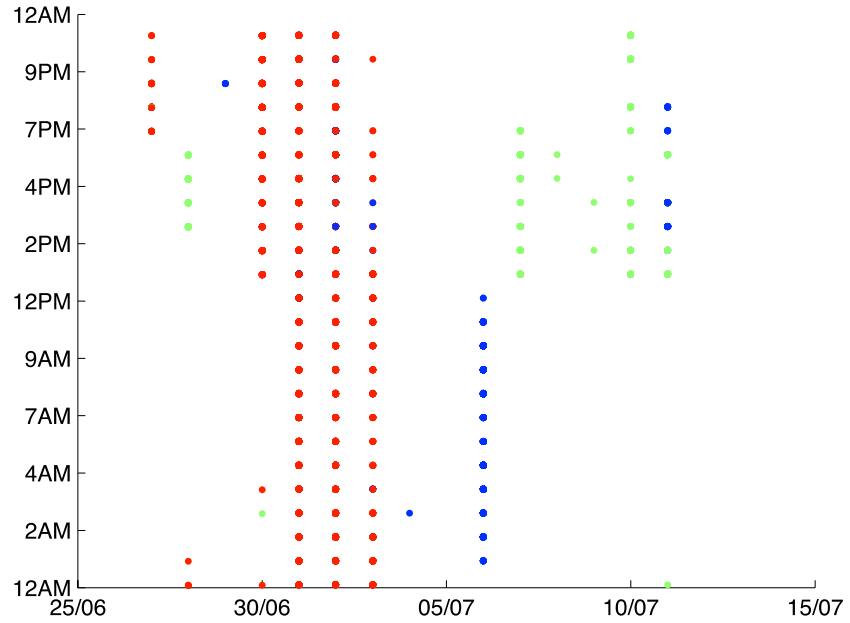


Figure 8: Detections Received from All Three Sensors: Sensor 1 (blue), Sensor 2 (green), and Sensor 3 (red)

To gain a better understanding of the data, we initially attempt to obtain a holistic visualization taking into consideration all attributes, namely, date and time, unique BT ID, and sensor location as depicted in Figure 9.

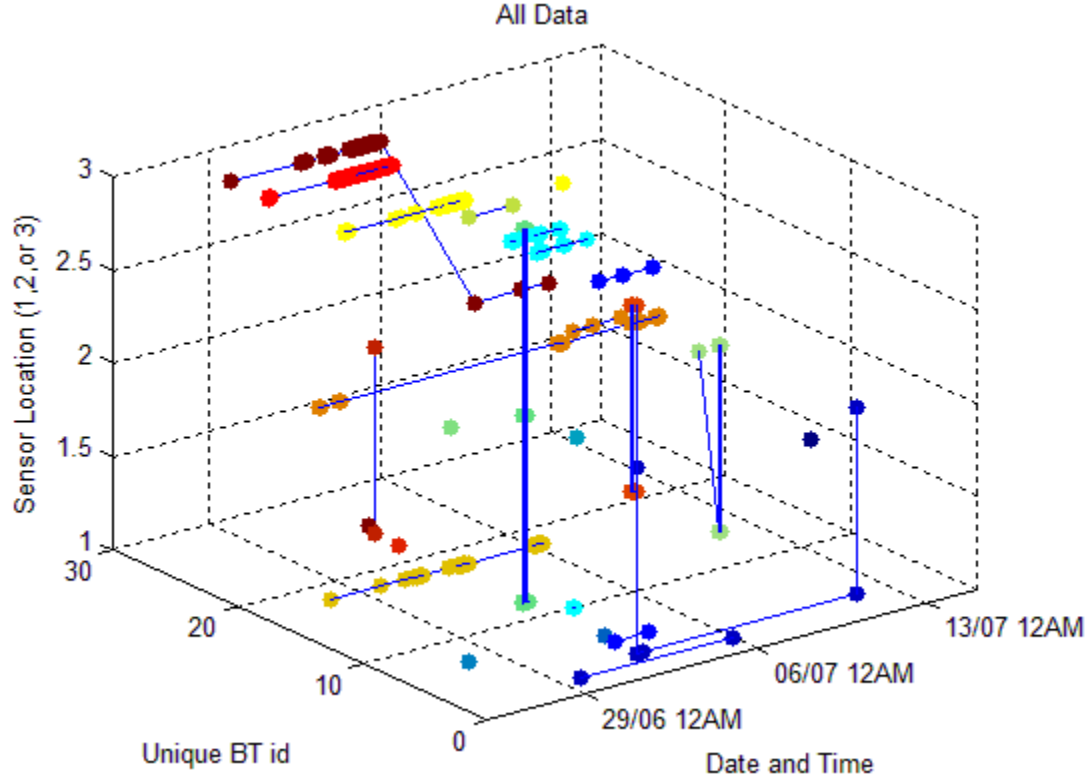


Figure 9: Visualization of BT IDs at Sensor Locations over Time

From this visualization, it can be seen that the sensors are functioning as intended for the most part, but an issue arises in that transitions between sensors can occur very rapidly, most likely attributable to the device being in detectable range of different sensors at the same time. This makes simple tracking of devices more difficult, but rejecting transitions that occur within a certain timeframe, perhaps 30 seconds to a minute, could allow for a better overall picture of traffic patterns.

4.2.2 Data by Sensor

Considering the fact that less than 10% of the data is detected using BT, it becomes necessary to develop statistical models for better estimation of the desired physical quantity. For Instance, estimation of pedestrian arrival using Poisson process or parameters estimation for the pedestrian flow dynamical model. In this section, descriptive statistics for the elapsed time (eT) between subsequent events categorized by sensor are presented. Herein, we highlight the potential use of eT to infer various

parameters such as travel time and wait time. We propose the extraction of eT quantities based on a specific range since it may differentiate between multiple pedestrian actions, for example:

- **1hr<eT<24hr**: may indicate **time spent** in a zone.
- **1min<eT<60min**: may indicate **travel time** between nodes and/or time spent.
- **1sec<eT<60sec**: may indicate **wait time** at a specific location.

Table 7 presents the summary statistics for the eT between subsequent data points for each sensor. On average, the elapsed time between 1hr and 24hr is approximately 4.8hr \pm 4.5hr, where the elapsed time in the range of 1min and 60min is, on average, 3.05min \pm 5.3min, and finally, the average elapsed time between 1sec and 60sec is 37.1sec \pm 10.8sec. The large standard deviation with respect to the mean indicates that the time elapsed data follows an exponential distribution which gives rise to Poisson distributed pedestrian arrivals. The upper and lower limits need further investigation in order to accurately infer the desired pedestrian actions. In addition to time spent, travel time, and wait time, the rate of pedestrian arrival at each sensor, which may represent rate of arrival or exit of pedestrians with active BT devices at a building's entrances.

Table 7: Summary Statistics for eT between Subsequent Data Points for each Sensor.

	Data Points	Elapsed time	Min	Max	Mean	Median	Mode	Std.	Var.
Sensor 1	1264	1hr<eT<24hr	1.1765	23.898	5.493	2.1819	1.1765	7.3321	53.759
		1min<eT<60min	1.0013	40.847	2.6104	1.43	1.2168	4.3665	19.067
		1sec<eT<60sec	1.484	59.692	39.769	38.666	30.339	9.488	90.021
Sensor 2	1103	1hr<eT<24hr	1.1406	22.629	8.2412	1.9726	1.1406	9.0601	82.085
		1min<eT<60min	1.0034	59.991	4.7862	1.7506	1.0034	8.8361	78.077
		1sec<eT<60sec	1.05	59.973	39.245	38.728	30.9	7.9232	62.777
Sensor 3	5794	1hr<eT<24hr	1.218	9.1951	4.0547	3.091	1.2183	3.0454	9.2744
		1min<eT<60min	1.008	58.2	2.8	1.36	1.02	4.88	23.79
		1sec<eT<60sec	1.023	60	36.1	36.92	38.45	11.67	136.14

4.2.3 Data by Unique BT ID

There are 26 unique BT IDs present in the data from all three sensors, thirteen of which detected by sensor 1, eleven by sensor 2, and nine unique BT IDs detected by sensor 3. In this section, we attempt to extract tracking information for each unique BT ID and obtain

summary statistics regarding the elapsed time and the transitions between sensors. This analysis provides us with information pertaining to the behavior of the individual, as oppose to the rates of detection (occurrence of an event) obtained in the analysis of the previous section.

The trajectories of every unique BT ID can be obtained. This can assist in exploring recurring patterns in individuals' trajectories such as hourly, daily, and weekly habits, as well as commonly used entrances and exits. In the current data some patterns can be observed at BT IDs 9, 11, 17, 18, 19, 23, and 26. The rest of the detections are somewhat occasional. It is also observed that transitions between sensors did not occur often. As Table 8 demonstrates, sensors 1 and 2 had only five IDs in common, whereas two IDs in common are found between sensor 2 and 3, and only 1 ID between sensors 1 and 2.

Table 8: Intersection of Unique BT IDs Between Sensors.

Sensors	Intersection of Unique BT IDs
1 and 2	3, 12, 14, 21, 24
1 and 3	12
2 and 3	12, 26

A summary statistics of the elapsed time for each unique BT ID was obtained. The number of total detections as well as the time between each data input is considered. These are divided into three time elapsed subintervals: between one second and one minute, between one minute and one hour, and between one hour and 24 hours. This allows for extraction of false alarms, e.g. unrealistic quick transitions between sensors, in order to gain a better understanding of true traffic patterns. However, further analysis can be done to allow for even better rejection of data. In addition to data cleaning, parameters such as time spent, travel time, and wait time can also be extracted as mentioned in the previous section. The weighted aggregation of the results is shown in Table 9. Note that the expected values of the elapsed time for the first and the second intervals are close in value to the standard deviation. Thus, Poisson distribution is appropriate to use in this case. Further investigation is needed for the limits, particularly for the third interval.

Table 9: The Weighted Aggregation of the Summary Statistics for the Elapsed Time.

Elapsed time	Min	Max	Mean	Median	Mode	Std.	Var.
1hr<eT<24hr	2.37	18.62	6.81	3.76	2.37	6.35	44.18
1min<eT<60min	1.07	41.29	3.74	2.04	1.11	5.54	34.52
1sec<eT<60sec	28.03	59.73	40.37	39.26	35.8	7.7	61.21

The elapsed time information for each unique ID is segregated based on the location transitions made by the pedestrian. Transitions between sensors are distinguished by the sensor location as well as the direction of movement giving rise to seven possible transitions 3 to 1, 3 to 2, 2 to 1, no transition, 1 to 2, 2 to 3, and 1 to 3. As expected, a Poisson process is apparent as demonstrated in Figure 10. In this plot, the elapsed time that is less than 20 min is provided with respect to the transition performed. Transitions between sensors 1 and 2 outnumber other transitions in both directions. Taking a closer look by zooming onto the data to display transitions that took less than 1.5min, depicted in Figure 11, it shows that some transitions took less than 30sec (the detection rate). This phenomenon may occur due to measurement noise or large detection range of adjacent detectors. This calls for the requirement of performing data validation since a more accurate representation of traffic would likely be realized.

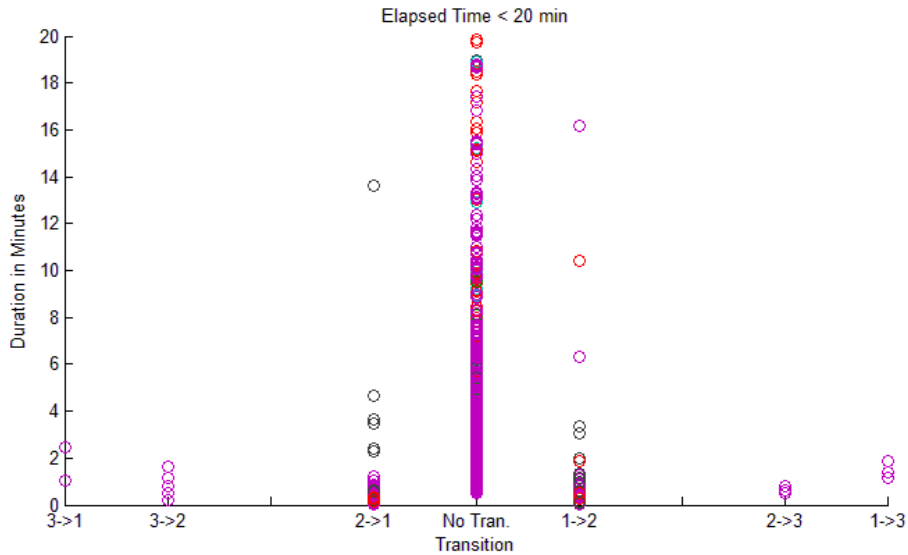


Figure 10: Elapsed Time of Transitions between Sensors that are less than 20 mins

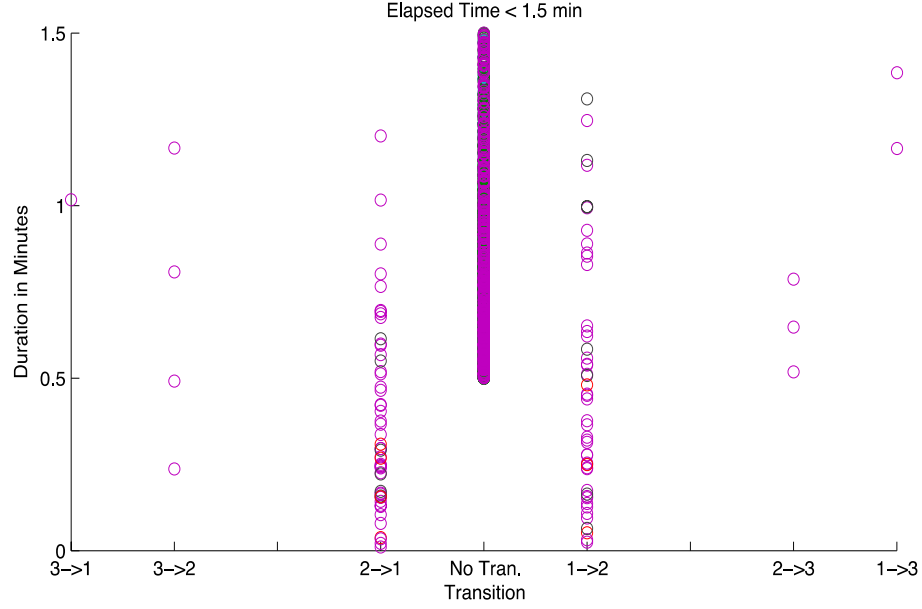


Figure 11: Elapsed Time of Transitions between Sensors that are less than 1.5 mins

4.3 Controlled Experiments

As mentioned previously, case specific data filtering techniques are necessary. However, the preliminary data analysis performed on the above experiments 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 12, a sample trajectory of unique BT ID selected to demonstrate the issues stated above. The depicted sample trajectories show the time and location of detection for each BT ID. The 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.

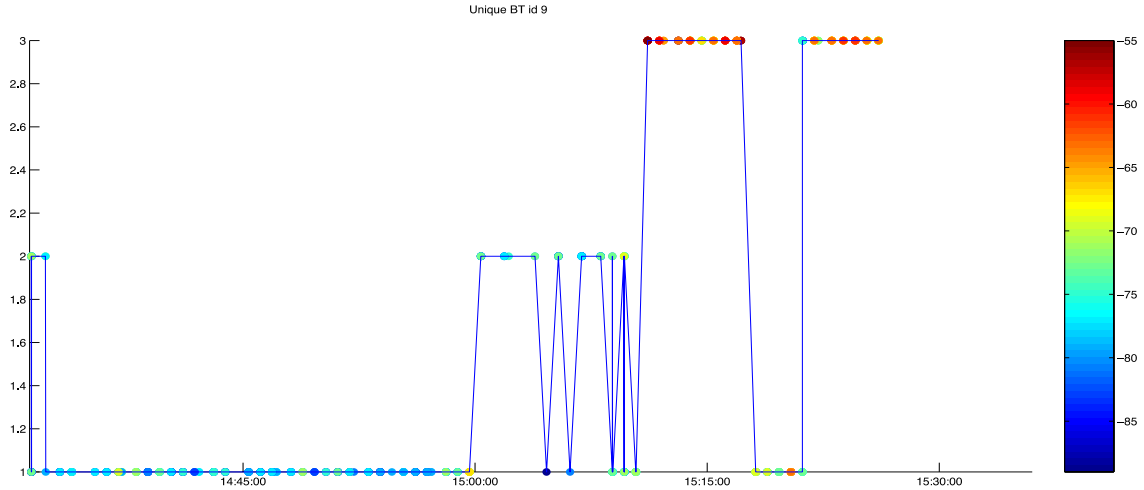


Figure 12: Trajectories of Two Unique BT IDs

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

4.3.1 Objectives

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.

4.3.2 Experimental Setup and Results

In this experiment, the goal is to verify how well we can identify trajectories made by users. Depicted in Figure 13 the plans for the path verification experiment. Three students

held BT active devices and walked the same path five times. Typical walking speed is approximately 3 miles/hour. The students used three different devices: Samsung tablet (Android), iPhone 6 (iOS), and windows phone.

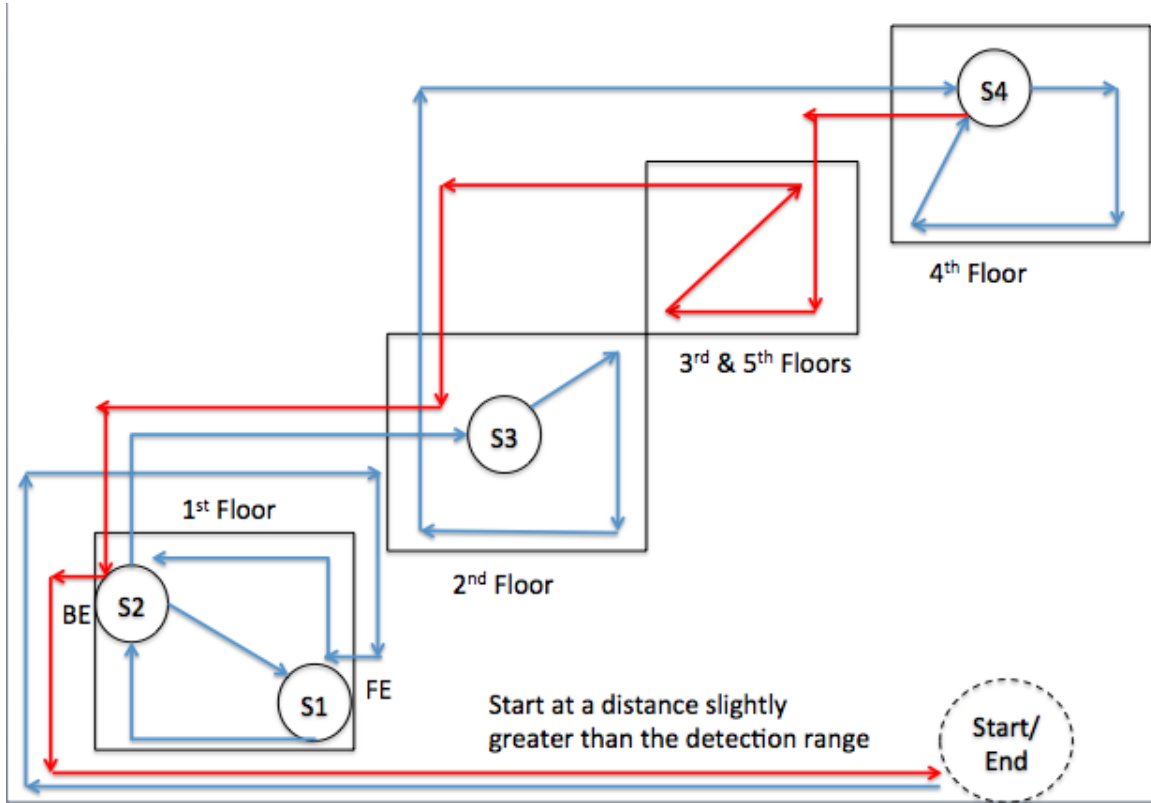


Figure 13: Path Verification Experiment Plan

The results of the path verification experiment can be depicted in figures 14.a, 14.b, and 14.c for the Samsung tablet, iPhone 6, and Windows phone, respectively. The ground truth data, the green line, is plotted in conjunction with the sensor data, the blue line.

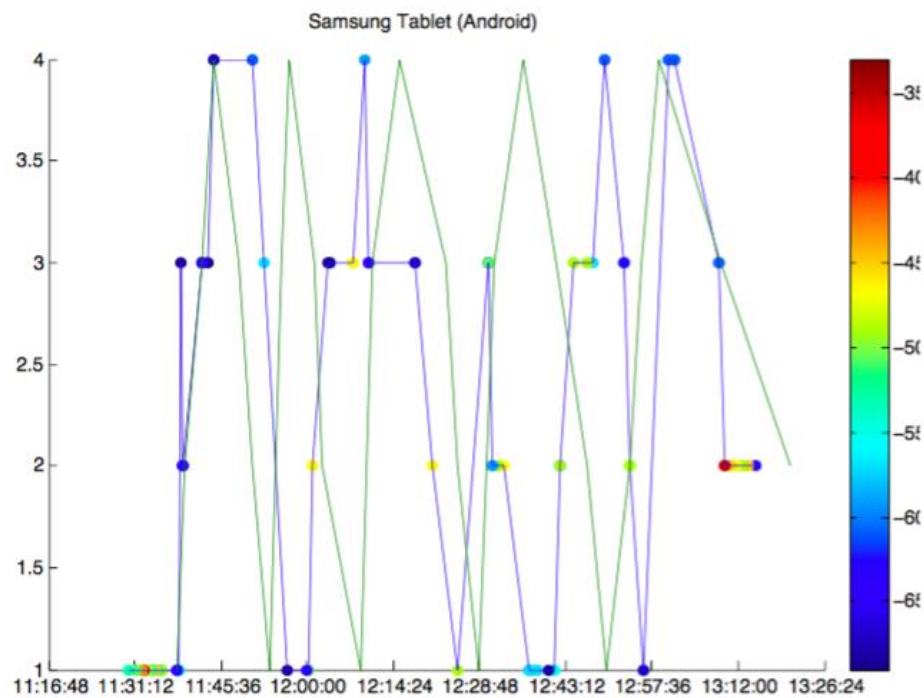


Figure 14.a: Samsung Tablet

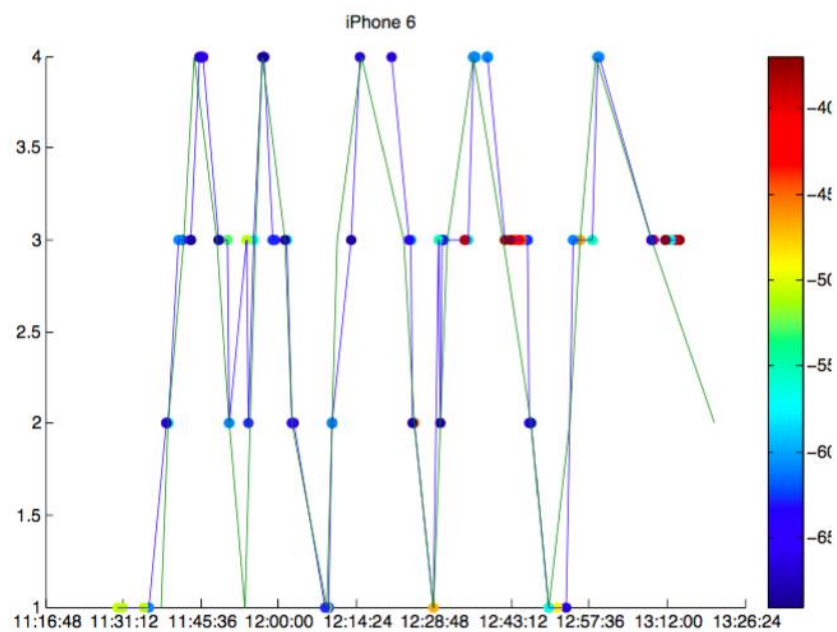


Figure 14.b: iPhone 6

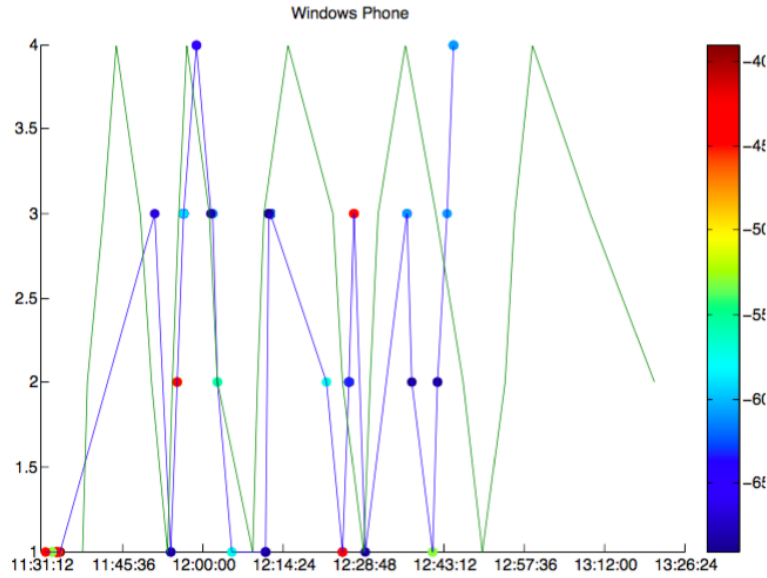


Figure 14.c: Windows Phone

Figure 14: Ground Truth Data (Green) vs. Sensor Data

4.3.3 Conclusion

The results demonstrate that there are some errors associated with the sensors accurately reporting the location of 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. For more indicative results, the subjects should have been walking together in order to guarantee that they have followed an identical path. Inaccuracy of time recording has also contributed to the errors. This reiterates the importance of the three main aspects: detection range, sensor placement, and filtering techniques.

4.4 The Subway Station Test

4.4.1 Introduction

The research team in collaboration with the transit authority conducted a pilot study at three different locations in a major subway station serving multiple subway lines. Figure 15 shows the locations of the Bluetooth based pedestrian (PEDBT2) sensors inside the subway station. In spite of the challenges such as software problems encountered during the pilot test, the outcome was 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 count and origin destination data and waiting times can be collected using a small number of devices.

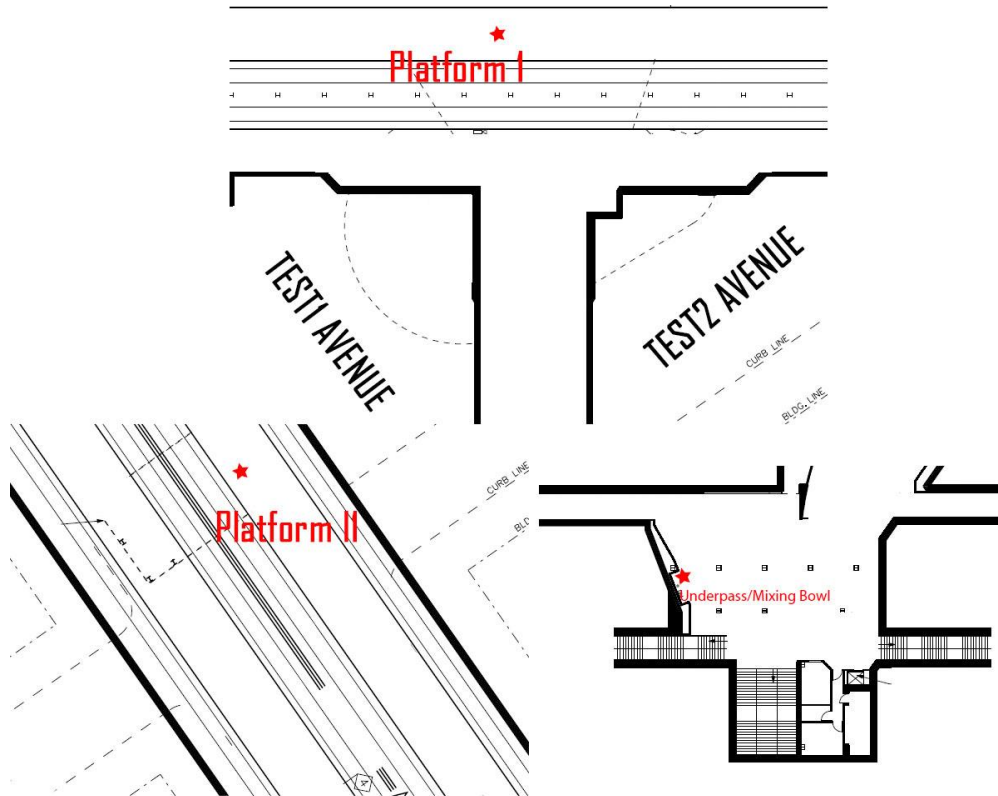


Figure 15: Locations of the Tablets in the Subway Station

Table 10: Period of Data Collection by Device

	Duration	Tablet 1 (Platform II)	Tablet 2 (Mixing Bowl)	Tablet 3 (Platform I)
Test #1	7/30-8/3	7/30 10AM - 8/3 9AM	7/30 10AM - 7/30 4PM	7/30 10AM - 8/3 10AM
Test #2	8/13-8/14	-	8/13 11AM - 8/14 11AM	8/13 11AM - 8/13 6PM
Test #3	8/14-8/18	-	8/14 11AM - 8/18 6PM	8/14 11AM - 8/16 1AM

As seen in Table 10, the length of periods during each test period is different. Moreover, there were a number of problems with the app that we have identified during the pilot test. The most important one was the quitting of the app unexpectedly. This was mainly observed for tablet 1. Since we did not have an online connection to the devices, we were not able to detect these problems in real-time. The other problem was the management of device memory and power since the incoming data can be quite large.

4.4.2 Test Objectives and Parameters

The primary objectives of the study are:

1. Conducting a test to observe the effectiveness of the BT sensing app running in tablets for tracking foot traffic, and
2. Feasibility of capturing the before/after travel pattern changes due to the closure of an alternative segment necessitating a transfer at the test station.

In the test, we focused on the counts from the tablets to observe the foot traffic at the selected locations. Moreover, 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 sensors which can be used to identify some travel patterns of riders as well as approximate OD demands. However, due to the problems with the tablets, it was not possible to collect data continuously and thus conduct a reliable before and after study. Thus, the research team focused on objective 1.

4.4.3 Summary of Results

Table 11 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.

Table 11: Total Number of Detected Devices and Average Waiting Times

	# of records by device			Average wait (min)		
	Tablet 1	Tablet 2	Tablet 3	Tablet 1	Tablet 2	Tablet 3
Test #1	2762	392	4638	1.34	1.13	2.7
Test #2		1221	938		1.36	2.33
Test #3		5133	4622		1.79	3.02

Some of the important observations based on the limited 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. The reason for varying number of samples shown in Table 11 are due to the time the app worked properly without failing, not due the change in percentages of devices that could be detected.

2. Average waiting times at the platform (Tablet 3) increase from 2.7 minutes to 3.02 minutes after August 14th. A slight increase in waiting times at the mixing bowl (Tablet 2) after August 14th is also observed.

3. We saw an increase in movement percentages at Platform II (Tablet 3) after August 14th.

4. Due to software problems with tablet 1, data from the tablet is not available for Test 2 and Test 3.

Table 12 shows the daily average counts and waiting times for Tablet 3. Note that, the weekend counts (8/16 and 8/17) are excluded from averaging for Test #2 since the data was collected during weekdays in Test #1 and Test #3. From Table 12, it can be observed that after the closure of Montague Tunnel, the number of people waiting in proximity of Tablet 3 (Platform II) and their waiting times increased.

Table 12: Daily Average Counts and Waiting Times for Tablet 3

	12AM-6AM		6AM-10AM		10AM-4PM		4PM-7PM		7PM-12AM	
	Cnt	Wait	Cnt	Wait	Cnt	Wait	Cnt	Wait	Cnt	Wait
Test #1	36	1.32	174	1.95	364	2.86	202	1.52	135	2.4
Test #2					477	2.93	314	1.26		
Test #3	52	3.95	616	1.86	626	3.44	449	2.72	333	3.84

Table 13 presents the counts of movement between tables for each testing period.

Table 13: The Number of Matches between Tablets per Testing Period

	1-1	1-2	1-3	2-1	2-2	2-3	3-1	3-2	3-3
Test #1	2628	2	132	144	45	321	598	30	4010
Test #2					316	433		53	885
Test #3					776	1810		161	4461

4.4.4 Individual Results for each Testing Period

Test #1

1) Counts and Average Wait Times

Table 14 shows the daily counts and average wait times for peak and off-peak periods for each tablet. During the first test, Tablet 3 is the one that detected the most BT devices almost every period of each day in the study. Average wait times for Tablet 3 are higher than the average waiting times recorded by other devices.

Table 14: Daily Counts and Average Wait Times for Peak and Off-peak Periods

	12AM-6AM		6AM-10AM		10AM-4PM		4PM-7PM		7PM-12AM	
	Cnt	Wait	Cnt	Wait	Cnt	Wait	Cnt	Wait	Cnt	Wait
<i>30-Jul</i>										
Tablet 1					323	1.44				
Tablet 2					370	1.86	165	0.79	125	1.18
Tablet 3					767	2.62	501	1.35	283	1.63
<i>31-Jul</i>										
Tablet 1	54	1.94	408	1.22	211	1.53	105	0.73	81	0.39
Tablet 3	72	1.36	235	2.18	235	3.36	115	1.14	89	3.12
<i>1-Aug</i>										
Tablet 1	23	1.41	179	0.72	91	1.17	26	1.56	19	1.3
Tablet 3	27	1.49	163	2.43	216	2.67	104	1.83	69	3.41
<i>2-Aug</i>										
Tablet 1	7	0.62	65	2.39	103	1.62	34	1.22	25	0.96
Tablet 3	21	2.51	124	2.11	239	3.26	89	2.61	99	3.24
<i>3-Aug</i>										
Tablet 1	17	2.93	39	1.89						
Tablet 3	24	5.3	77	1.94						

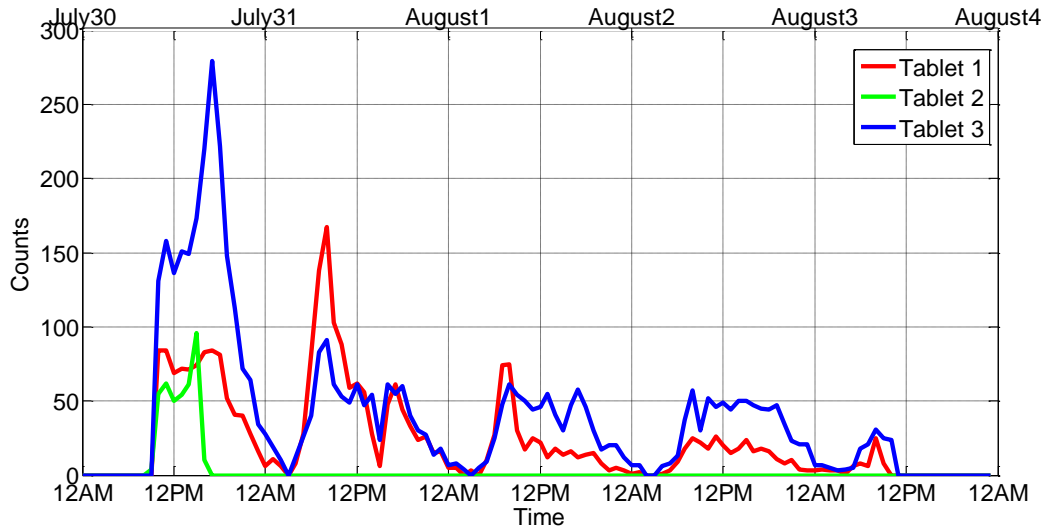


Figure 16: Hourly Counts from each Device during Test #1

Figure 16 shows the hourly counts from each device during test #1. It can be seen that Tablet 2 stopped working on July 30th and did not record any activity after approximately 4 PM. However, the other two devices kept working and recording BT detections properly. It can be interpreted that the locations of Tablet 3 and Tablet 2 are experiencing the busiest times around morning and evening peak hours. The results showed that this station is heavily used by commuters and it has almost no traffic around midnight.

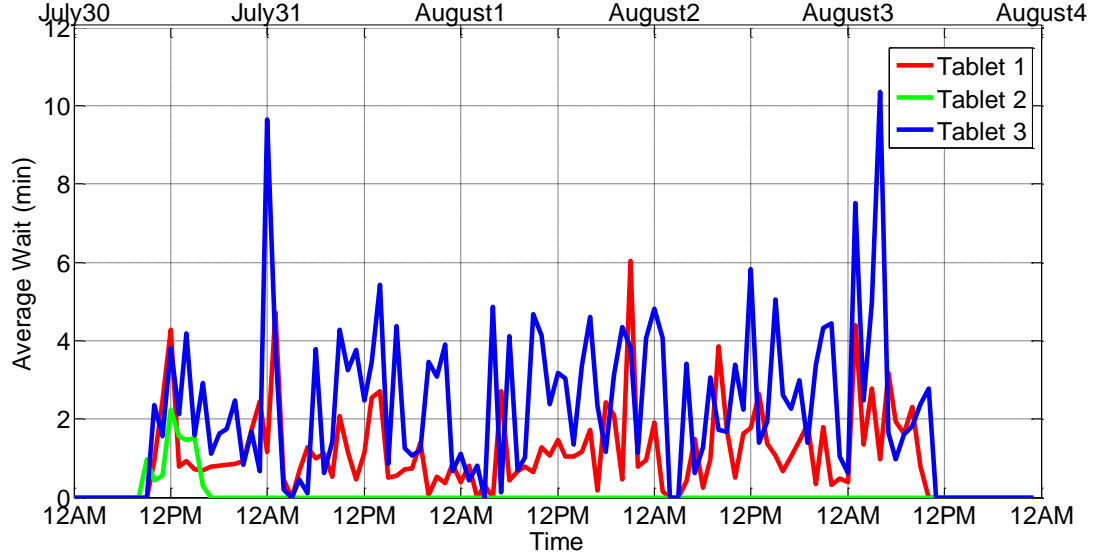


Figure 17: Hourly Average Wait-times from each Device during Test #1

Although the station's busiest periods are morning and evening peak hours, users experience the most waiting times during midday and midnight. This can be explained by frequent train schedules for morning and evening peak hours. We can assume that the train service during midday and midnight time periods is infrequent resulting in longer average wait-times.

2) Movement between Tablets

Table 15 exhibits the movement codes between tablets. Some users were detected only by a single sensor. In that case, repeated tablet number represents the movement code. For example, 1-1 corresponds to a user who is detected only by Tablet 1. In other cases, the first number represents the initial detection and the second number represents the movement direction of a user.

Table 15: The Movement Codes between Tablets

Movement Code	Explanation
1-1	Detected only by Tablet 1 (Platform I)
1-2	Movement from Tablet 1 to Tablet 2
1-3	Movement from Tablet 1 to Tablet 3
2-1	Movement from Tablet 2 to Tablet 1
2-2	Detected only by Tablet 2 (Mixing Ball)
2-3	Movement from Tablet 2 to Tablet 3

3-1	Movement from Tablet 3 to Tablet 1
3-2	Movement from Tablet 3 to Tablet 2
3-3	Detected only by Tablet 3 (Platform II)

Figure 18, Figure 19 and Figure 20 show the hourly movement counts from Tablet 1, 2 and 3 respectively. It should be noted that Tablet 2 stopped working after 4 PM. Therefore, its plot looks different than the other plots showing a shorter period of time. Figure 18 shows that the movements from Tablet 1 to Tablet 3 constituting approximately 10%. On the other hand, about 20% of the users moved from Tablet 3 to Tablet 1 which can be seen in Figure 12. This indicates that more people moved from the location of Tablet 3 to Tablet 2 since Tablet 3 detected more users than any other tablets in the study.

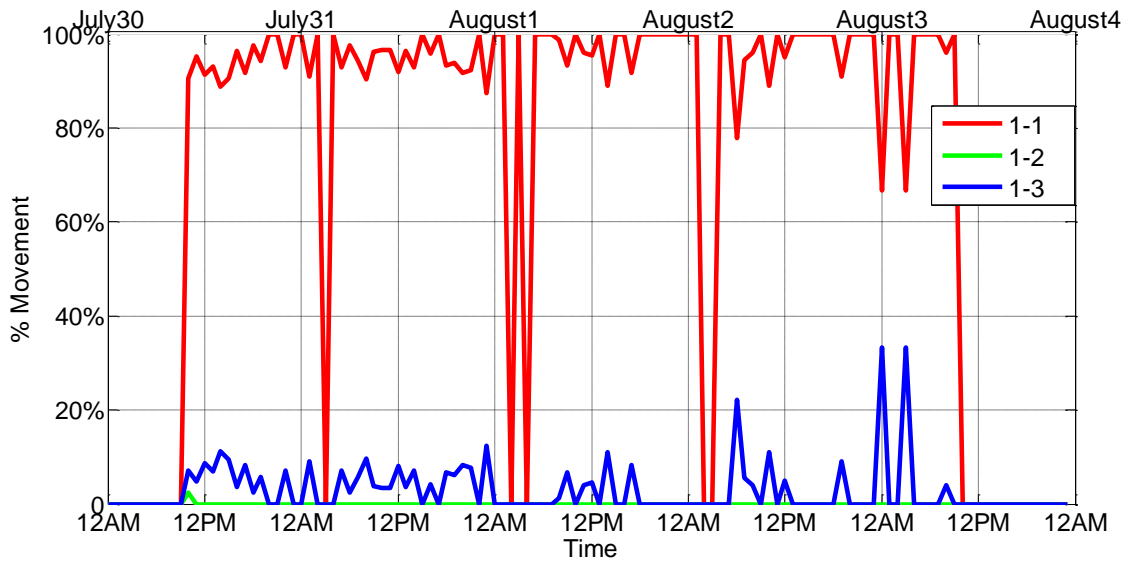


Figure 18: Hourly Movement Counts - Tablet 1 vs. Others during Test #1

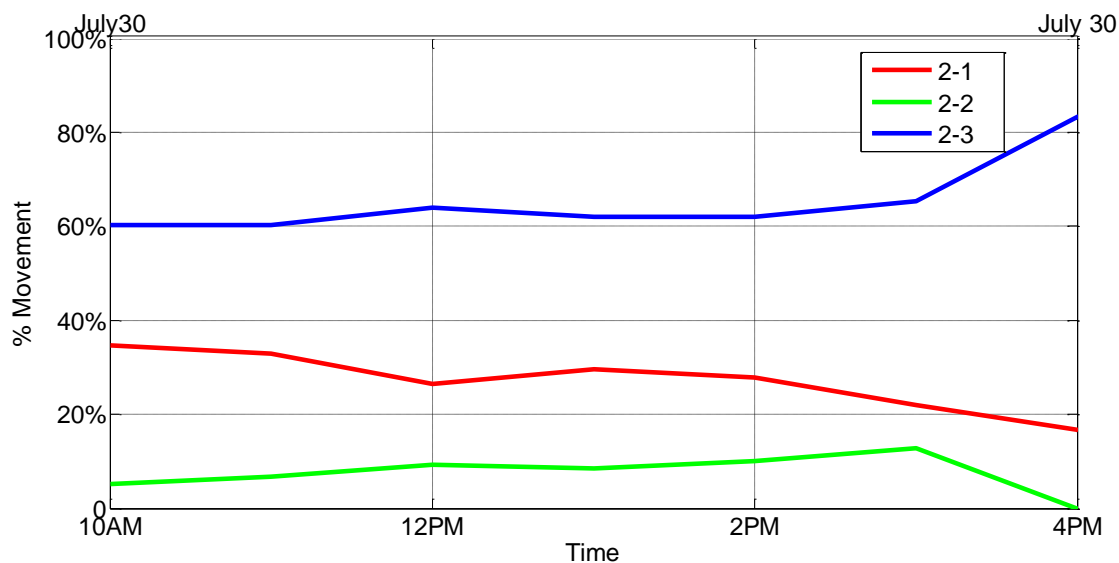


Figure 19: Hourly Movement Counts - Tablet 2 vs. Tablets during Test #1

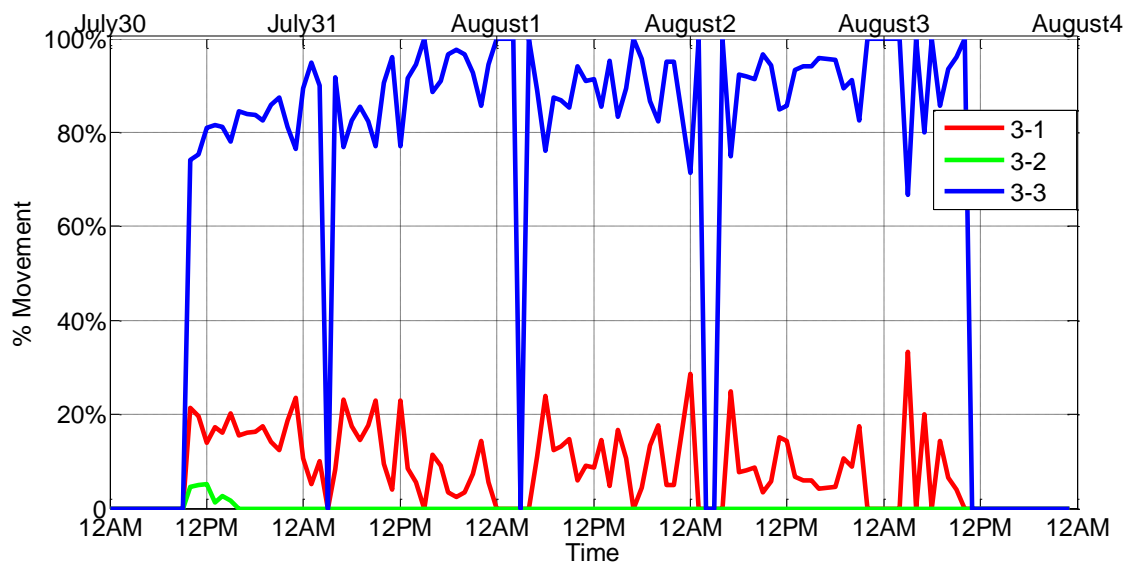


Figure 20: Hourly Movement Counts - Tablet 3 vs. Others during Test #1

5 PEDESTRIAN MOBILITY PREDICTION MODEL

5.1 Introduction

Traditionally, various versions of linear and nonlinear filters have been proposed for the dynamic OD matrix estimation problem for vehicles where flow data is obtained from classical flow detectors. Van Lint et al. and Treiber et al. discuss a variety of nonlinear filters (Lint 2009 & Treiber 2011). A refined version of the Kalman Filter (Kalman 1960), is proposed in Barcelo et al. 2010. Chang et al. 1995 proposed using time dependent traffic data along with traffic flow models to estimate travel times dynamically for every OD pair using an Extended Kalman Filter to treat nonlinearities. Lin et al. 2007 extended the work proposed by Change to include traffic dynamics and travel time data. Hu et al. 2001 also uses an Extended Kalman Filtering algorithm to estimate dynamic OD matrices. In recent years, as BT 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. Barcelo et al. 2012 proposed a recursive linear Kalman-Filter for state variable estimation that combines and modifies the earlier work of Chang and Wu 1995}, Hu et al. 2001, Choi et al. 2009, and Van Der Zijpp & Hamerslag 1994, adapting their models to take advantage of travel times and traffic counts collected by tracking BT equipped vehicles and conventional detection technologies. They proposed a linear formulation of the Kalman Filtering approach that uses deviations of OD path flows as state variables, as suggested by Ashok 2000, and calculated with respect to Historic OD path flows for detected vehicles without using an assignment matrix.

5.1.1 Crowd Modeling and our Approach

A wide range of models have been employed to describe the distribution of crowds, such as, dynamic cellular automaton model, lattice gas models, social force models, fluid-dynamic models, agent-based models, and game theoretic models (Alizadeh 2011 & Guo 2008). Some of these models capture the microscopic behavior and other capture the macroscopic behavior of crowd dynamics. Even though this study is concerned with pedestrian networks, we first define the network as a general graph, as demonstrated in Figure 21, in order to take advantage of existing graph theoretic properties and develop additional ones based on the unique data sets BT detectors provide. This allows us to

capture the topological attributes of the infrastructure as well as the flow dynamics (Xiaoping 2009). This also provides a flexible framework where the flow on each edge of the graph can potentially be governed by the desired dynamics using various pedestrian crowd dynamic models such as fluid-dynamics for macroscopic behavior, or lattice gas models for macroscopic behavior.

Defining the network as a graph lays the ground for the development of a Markov Chain Model (MCM) to capture the general attributes of a pedestrian network such as density, dwell times, and OD flows. Markov Chains provide a mathematical structure appropriate to modeling discrete events, which we find to be suitable in our case study or our understanding of pedestrian behavior. Pedestrian networks can be conceptualized as a general graph. A given structure of interest, such as a residential building, a shopping mall, an airport terminal, or a transit station, can be represented as a network by defining a couple $G(I, J)$ where I is a finite collection of N edges and J is a finite collection of M vertices that include all entrances and exits of the space of interest.

In this section, we demonstrate a possible modeling scheme using the data obtained from the Brooklyn IV case study. We then draw conclusions on how this model can be improved in order for it to be used for prediction.

5.2 Brooklyn IV Case Study

In the Brooklyn IV case study, three sensors with BT detection capabilities were placed in a building floor with two main entrances. The MC model for this scenario is given in Figure 21. A sensor was placed at each entrance; therefore, for this case study, we have three transient states representing the local areas surrounding each of the three sensors s_1 , s_2 , and s_3 and one absorbing state “ s_4 ” representing areas near the entrance/exit locations.

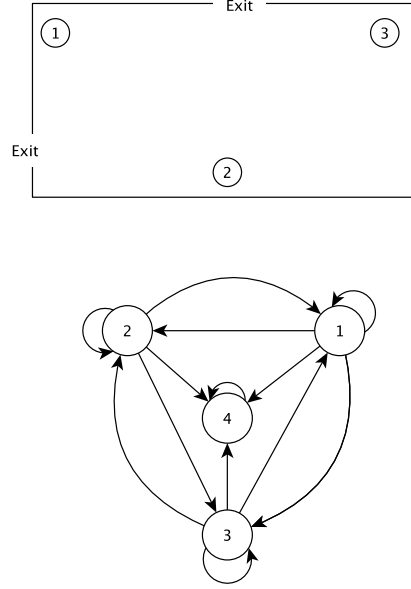


Figure 21: Markov Chain Model for the Brooklyn IV Case Study

5.2.1 Data-based MCM

Twenty-six unique BT devices were recorded during the case study. The data includes several attributes, namely, signal strength, timestamp and encrypted device IDs. This data is initially processed to estimate the transition matrix. This part involves two main processes: obtaining the flow matrix and maximum likelihood estimation.

The data was segregated by unique IDs where for each a list of time-stamps and corresponding states were provided. For each unique ID, the data was then segregated by day where the transitions at each time step for every unique ID were obtained. In general terms, let F be the $M \times M$ flow matrix where the entry f_{ij} is the count of occurrences during which a person transitioned to j from i . In this case study, the flow matrix from the data for the proposed model of the given network was found to be:

$$F_w = \begin{bmatrix} 3806 & 121 & 54 & 531 \\ 117 & 551 & 86 & 251 \\ 55 & 78 & 353 & 230 \\ 0 & 0 & 0 & 1012 \end{bmatrix}$$

Maximum Likelihood Estimation (MLE) is then used in order to estimate the transition matrix T from the flow matrix F_W . In general, each entry of T , a_{ij} is estimated as follows:

$$\hat{a}_{ij} = \frac{f_{ij}}{\sum_{k=1}^M f_{ik}} \quad (1)$$

This indicates the likelihood of a single step transitioning from state i to state j . In this case study, the transition matrix for the proposed model of the given network using the data was found to be:

$$\hat{T}_w = \begin{bmatrix} 0.84 & 0.03 & 0.01 & 0.12 \\ 0.12 & 0.55 & 0.09 & 0.25 \\ 0.08 & 0.11 & 0.49 & 0.32 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

5.2.2 Analysis and Results

Obtaining the transition matrix allows us to compute other important quantities, i.e. convergence, time to absorption, and state density distribution, which provide an insight to key pedestrians' behavioral properties.

The n-step transition matrix, T^n , clearly shows that the transition probability, as n becomes large, to any state other than the absorbing state will eventually become zero; whereas, the transition probability to the absorbing state from any state will eventually converge to one. As depicted in Figure 22, it takes 22 steps or less for a person to leave the test site in the case study under discussion.

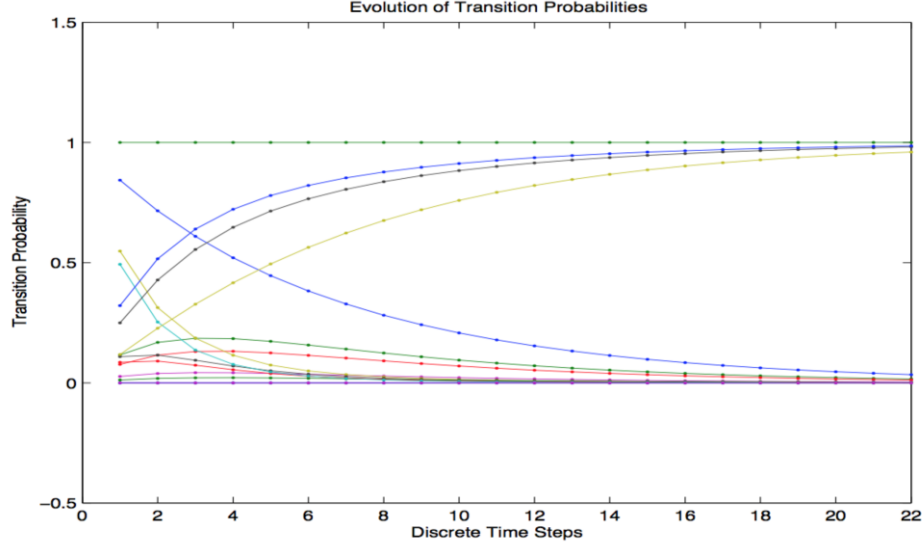


Figure 22: Convergence of the n -step transition matrix as $n \rightarrow \infty$

Time to absorption is given by, t_i , the expected number of steps needed before the process is absorbed when starting from the i^{th} state. The times to absorption for all three nonabsorbent states in our case study were found to be:

$$t_w = \begin{bmatrix} 7.56 \\ 4.95 \\ 4.18 \end{bmatrix}$$

We will now use the estimated MCM in order to predict the density distribution within the various states. In practical situations, this can also be used as demand prediction for resource allocation. Firstly, the initial density distribution, μ_0 , was obtained from the data for each state where each entry provides the probability of being in a state at initial time $n = 0$. In addition, d , the total number of pedestrians entering the building, and d_i , the number of pedestrians entering the building at time step i , were also obtained from the data. Then, we can show that the density distribution of pedestrians at the n^{th} time step at each state is given by:

$$D_n = \mu_0 \sum_{i=0}^n d_i T^{n-i} \quad (2)$$

Figure 23 illustrates how the relative density is distributed among the three tested locations in the aforementioned case study.

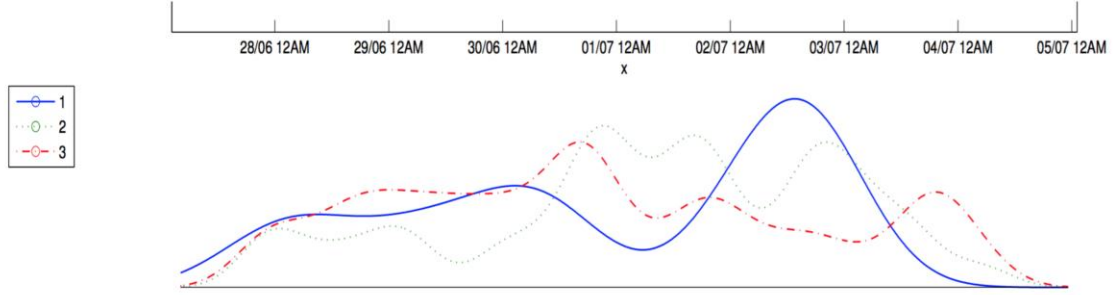


Figure 23: The Density Distribution for all three Locations

We are continuing with our efforts to further develop and validate the proposed model. Future work must encompass two fronts. The first is to develop a data collection plan that includes ground-truth data to be able to validate the predictive model and the second is to enhance the quality of the data by fusing it with data collected from other detection methods, for example infrared counts or vision.

6 CONCLUSIONS

Human movement behavior research has recently received increased attention particularly in the field of transportation planning and engineering. The traditional methods for pedestrian mobility monitoring include surveys, fixed pedestrian counters, and vision-based technologies. With the increase of smart devices, research has started focusing on tracking mobile phones to estimate pedestrian movements. The research team showed that if the detection system is equipped with Bluetooth receivers, it is possible to capture Origin-Destination (OD), travel time, wait time and flow information for some subset of the pedestrians with visible Bluetooth devices.

There are of course limitations of these procedures 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 sensors. However, robust algorithms can be developed to alleviate the inaccuracies originated from the outliers that are inherent in the collected wireless traces. Such algorithms can aim to remove low-quality detections, eliminate periodic/cyclic behavior, and improve detection and counting performance of devices.

The initial filtering and analysis of the data showed that it is probable to capture re-occurring patterns of the passengers in the terminal. The peak periods and busiest hours can also be detected at sensor locations. This information makes it easier to estimate passenger demand at a transit terminal. However, the initial results only represent the number of detected devices and should not be used as actual pedestrian counts. If the location and time specific penetration rate of Bluetooth devices is known for the study sites, the device counts can be used to estimate total pedestrian counts. In conclusion, we suggest that the wireless data should be used with great care and well-tested filters have to be used to clean the collected data.

The encryption technique used in the data collection for this research not also provides an extra layer of protection for sensitive information but also preserves the identifiers unique for approximately 96% of the cases. The proposed MC model has demonstrated to be a promising approach in describing pedestrian networks that generally

lack structure and consistency. The measures obtained, such as convergence, time to absorption, and state density distribution, can be further developed to provide us with better understanding of pedestrians behavioral patterns and estimating key states of the pedestrian network, such as density, dwell times, and OD flows. As a future study, the BT sensors will be developed using micro-computers such as Raspberry PI on open source operating systems and application platforms such as UNIX and Python. This will not only provide more robust environment to enhance data collection and filtering processes but also the opportunity to integrate various sensor technologies together.

DISCLAIMER

The contents of this report reflect the views of the authors, who are responsible for the facts and the accuracy of the information presented herein. This study is partially funded by UTRC at CUNY. The authors would like to thank MTA for their support in the early stages of this study. The contents do not necessarily reflect the official views or policies of the UTRC, NYU, MTA or the Federal Highway Administration. This report does not constitute a standard, specification or regulation. This document is disseminated under the sponsorship of the Department of Transportation, University Transportation Centers Program, in the interest of information exchange. The U.S. Government and NYU assume no liability for the contents or use thereof. The contents of this paper only reflect views of the authors who are responsible for the facts and accuracy of the data and results presented herein.

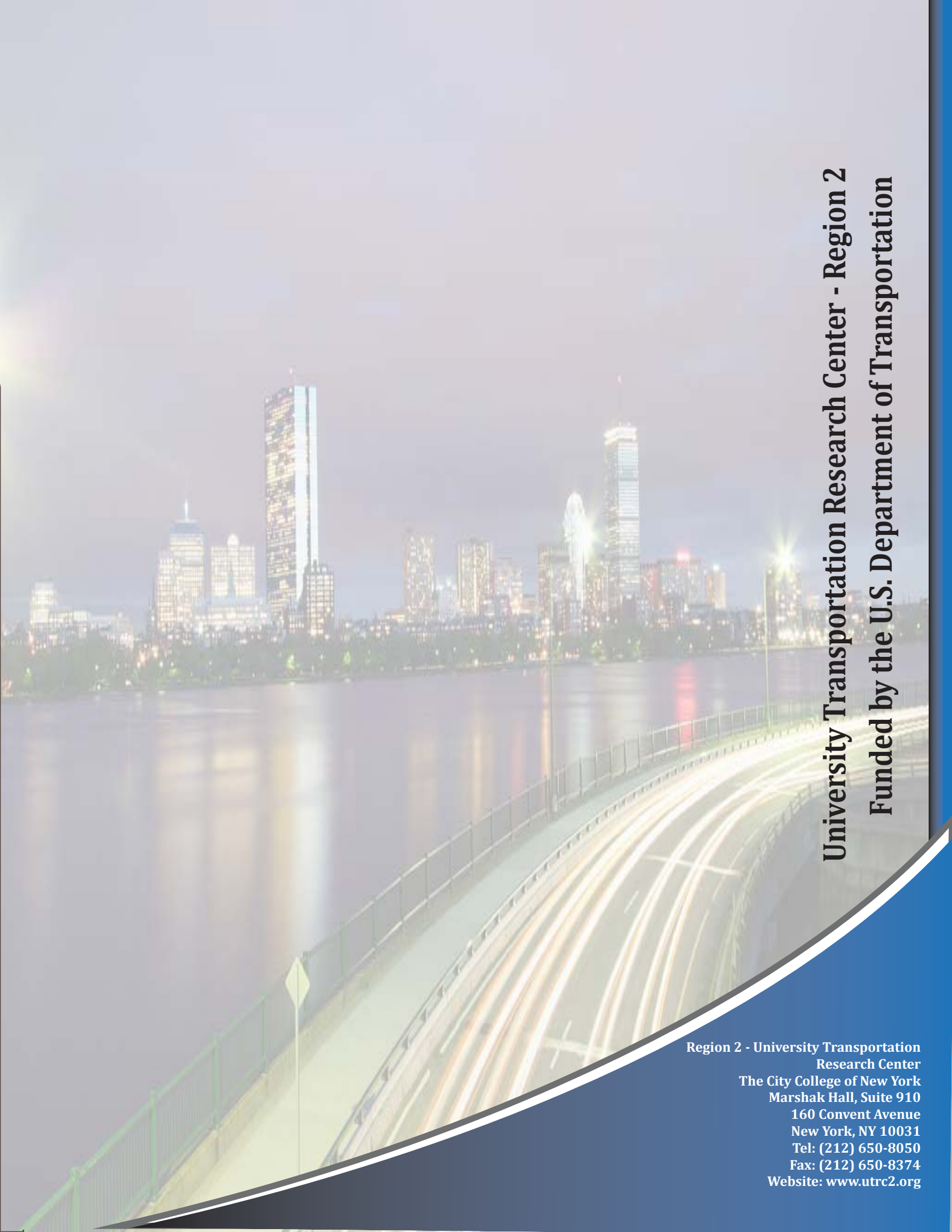
REFERENCES

- Abedi, N., Bhaskar, A., & Chung, E. (2013). *Bluetooth and Wi-Fi MAC address based crowd data collection and monitoring: benefits, challenges and enhancement*. Paper presented at the Australasian Transport Research Forum (ATRF), 36th, 2013, Brisbane, Queensland, Australia.
- Abedi, N., Bhaskar, A., Chung, E., & Miska, M. (2015). Assessment of antenna characteristic effects on pedestrian and cyclists travel-time estimation based on Bluetooth and WiFi MAC addresses. *Transportation Research Part C: Emerging Technologies*, 60, 124-141.
- Ahmed, H., EL-Dariebey, M., Abdulhai, B., & Morgan, Y. (2008). *Bluetooth-and Wi-Fi-based mesh network platform for traffic monitoring*. Paper presented at the Transportation Research Board 87th Annual Meeting.
- Alizadeh, R. (2011). *A dynamic cellular automaton model for evacuation process with obstacles*. *Safety Science*, 49(2):315–323.
- Ashok, K. & Ben-Akiva, M. (2000). *Alternative approaches for real-time estimation and prediction of time-dependent origin- destination flows*. *Transportation Science*, 34(1):21–36.
- Barceló, J., Montero, L., Bullejos, M., Serch, O., & Carmona, C. (2012). *A Kalman filter approach for the estimation of time dependent OD matrices exploiting Bluetooth traffic data collection*. Paper presented at the Transportation Research Board 91st Annual Meeting.
- Barcelo, J., Montero, L., Bullejos, M., Serch, O., & Carmona, C., (2012). *Dynamic od matrix estimation exploiting Bluetooth data in urban networks*. 14th International Conference on Automatic Control, Modeling and Simulation, Saint Malo, France.
- Barcelo, L. Montero, L. Marques, and C. Carmona (2010). *Travel time forecasting and dynamic of estimation in freeways based on Bluetooth traffic monitoring*. *Transportation Research Record*, 2179:19–27.
- Blip. Blip Systems. Retrieved from <http://blipsystems.com/jfk-displays-actual-wait-times-using-sensors-that-monitor-mobile-phones/>
- Bonne, B., Barzan, A., Quax, P., & Lamotte, W. (2013). *WiFiPi: Involuntary tracking of visitors at mass events*. Paper presented at the World of Wireless, Mobile and Multimedia Networks (WoWMoM), 2013 IEEE 14th International Symposium and Workshops on a.
- Brennan Jr, T. M., Ernst, J. M., Day, C. M., Bullock, D. M., Krogmeier, J. V., & Martchouk, M. (2010). Influence of vertical sensor placement on data collection efficiency from bluetooth MAC address collection devices. *Journal of Transportation Engineering*.
- Bullock, D., Haseman, R., Wasson, J., & Spitler, R. (2010). *Anonymous bluetooth probes for measuring airport security screening passage time: the indianapolis pilot deployment*. Paper presented at the Transportation Research Board 89th Annual Meeting. CD-ROM. Transportation Research Board, Washington DC.
- Carpenter, C., Fowler, M., & Adler, T. (2012). Generating Route-Specific Origin-Destination Tables Using Bluetooth Technology. *Transportation Research Record: Journal of the Transportation Research Board*(2308), 96-102.

- Chang, G. L. & Wu, J. (1994). *Recursive estimation of time-varying origin-destination flows from traffic counts in freeway corridors*. Transportation Research B, 28B(2):141–160.
- Choi, K., Jayakrishnan, R., Kim, H., Yang, I., & Lee, J. (2009). *Dynamic od estimation using dynamic traffic simulation model in an urban arterial corridor*. 88th TRB Meeting Washington D.C..
- Clearview. Clearview Intelligence. Retrieved from <http://www.clearview-intelligence.com/products/m830-bluetooth-traffic-monitoring-system>
- 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.
- Guo, R., & Huang, H. (2008). *A modified floor field cellular automata model for pedestrian evacuation simulation*. Journal of Physics A: Mathematical and Theoretical, 41(38).
- Hamed, M., Haghani, A., & Sadabadi, F. (2009). *Using Bluetooth technology for validating vehicle probe data*. Paper presented at the 16th ITS World Congress and Exhibition on Intelligent Transport Systems and Services.
- Haseman, R. J., Wasson, J. S., & Bullock, D. M. (2009). Real time measurement of work zone travel time delay and evaluation metrics using bluetooth probe tracking. *TRB Paper*, 10-1442.
- Hu, S., Madanat, S. M., Krogmeier, J., & Peeta, S. (2001). *Estimation of dynamic assignment matrices and OD demands using adaptive Kalman filtering*. Intelligent Transportation Systems Journal, 6:281–300.
- Jie, S., Na-na, L., Ji-lin, C., Yong-feng, D., & Zheng, Z. (2008). *Design and implementation of intelligent transportation system based on GPRS and Bluetooth hybrid model*. Paper presented at the Automation and Logistics, 2008. ICAL 2008. IEEE International Conference on.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problem. *Journal of Basic Engineering*, 2(82):35–45.
- Kalogianni, E., Sileryte, R., Lam, M., Zhou, K., Van der Ham, M., Van der Spek, S., & Verbree, E. (2015). *Passive WiFi monitoring of the rhythm of the campus*. Paper presented at the Proceedings of The 18th AGILE International Conference on Geographic Information Science; Geographics Information Science as an Enabler of Smarter Cities and Communities, Lisboa (Portugal), June 9-14, 2015; Authors version.
- Kostakos, V. (2008). Using Bluetooth to capture passenger trips on public transport buses. *arXiv preprint arXiv:0806.0874*.
- Laharotte, P.-A., Billot, R., Faouzi, E., & Rakha, H. A. (2015). *Network-Wide Traffic State Prediction Using Bluetooth Data*. Paper presented at the Transportation Research Board 94th Annual Meeting.
- Lees-Miller, J., Wilson, R. E., & Box, S. (2013). Hidden markov models for vehicle tracking with bluetooth.
- Lesani, A., & Miranda-Moreno, L. F. (2016). *Development and Testing of a Real-Time Wifi-Bluetooth System for Pedestrian Network Monitoring and Data Extrapolation*. Paper presented at the Transportation Research Board 95th Annual Meeting.

- Lin, P. and Chang, G. (2007). *A generalized model and solution algorithm for estimation of the dynamic freeway origin-destination matrix*. Transportation Research B, 41:554– 572.
- Malinovskiy, Y., Saunier, N., & Wang, Y. (2012). Analysis of pedestrian travel with static bluetooth sensors. *Transportation Research Record: Journal of the Transportation Research Board* (2299), 137-149.
- Michau, G., Nantes, A., Chung, E., Abry, P., & Borgnat, P. (2014). Retrieving dynamic origin-destination matrices from Bluetooth data.
- Nicolai, T., & Kenn, H. (2007). *About the relationship between people and discoverable Bluetooth devices in urban environments*. Paper presented at the Proceedings of the 4th international conference on mobile technology, applications, and systems and the 1st international symposium on Computer human interaction in mobile technology.
- Ozbay, K., Bartin, B., Yang, H., Walla, R., & Williams, R. (2010). *Automatic pedestrian counter*.
- Ozbay, K., Yang, H., & Bartin, B. O. (2010). *Calibration of an infrared-based automatic counting system for pedestrian traffic flow data collection*. Paper presented at the Transportation Research Board 89th Annual Meeting.
- Puckett, D. D., & Vickich, M. J. (2010). *Bluetooth®-based travel time/speed measuring systems development*.
- Schauer, L., Werner, M., & Marcus, P. (2014). *Estimating crowd densities and pedestrian flows using wi-fi and bluetooth*. Paper presented at the Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and Services.
- SENSYS. Retrieved from <http://www.sensysnetworks.com/products/re-id>
- Treiber, M., Kesting, A., & Wilson, R. E. (2011). *Reconstructing the traffic state by fusion of heterogeneous data*. Computer-Aided Civil and Infrastructure Engineering, (26):408–419, 2011.
- Tyco. Tyco Traffic & Transportation. Retrieved from <http://www.tyco-its.com/products-and-services/urban-traffic-control/bluetooth-travel-timespeed-measurement-system>
- Van Der Zijpp, N. J. & Hamerslag, R. (1994). *An improved Kalman filtering approach to estimate origin-destination matrices for freeway corridors*. Transportation Research Records, 1443:54–64.
- Van Lint, J.W.C., Hoogendoorn, S.P. (2009). *Robust and efficient method for fusing heterogeneous data from traffic sensors on freeways*. Computer-Aided Civil and Infrastructure Engineering, (24): 1-17.
- Versichele, M., De Groote, L., Bouuaert, M. C., Neutens, T., Moerman, I., & Van de Weghe, N. (2014). Pattern mining in tourist attraction visits through association rule learning on Bluetooth tracking data: A case study of Ghent, Belgium. *Tourism Management*, 44, 67-81.
- Vu, L., Nguyen, P., Nahrstedt, K., & Richerzhagen, B. (2015). Characterizing and modeling people movement from mobile phone sensing traces. *Pervasive and Mobile Computing*, 17, 220-235.
- Weppner, J., & Lukowicz, P. (2013). *Bluetooth based collaborative crowd density estimation with mobile phones*. Paper presented at the Pervasive computing and communications (PerCom), 2013 IEEE international conference.

- Weppner, J., Lukowicz, P., Blanke, U., & Troster, G. (2014). Participatory Bluetooth scans serving as urban crowd probes. *Sensors Journal, IEEE*, 14(12), 4196-4206.
- Xiaoping, Z., Tingkuan, Z., & Mengting, L. (2009). *Modeling crowd evacuation of a building based on seven methodological approaches*. Building and Environment, 44:437–445.

A long-exposure photograph of a highway bridge at night. The bridge has multiple lanes with light trails from cars. In the background, a city skyline is visible across a body of water, with several tall buildings illuminated. The sky is dark with some light clouds.

University Transportation Research Center - Region 2

Funded by the U.S. Department of Transportation

**Region 2 - University Transportation
Research Center**
The City College of New York
Marshak Hall, Suite 910
160 Convent Avenue
New York, NY 10031
Tel: (212) 650-8050
Fax: (212) 650-8374
Website: www.utrc2.org