AN ABSTRACT OF THE THESIS OF

Nadia Bathaee for the degree of Master of Science in Industrial Engineering presented on April 23, 2014.
Title: Differentiating Transportation Modes Using Bluetooth Sensor Data

Abstract approved

---

J. David Porter

State departments of transportation, transportation consultants, and other transportation agencies are always in need of data that can be used to better understand how different modes of transportation use the road and highway systems. A variety of automatic data collection technologies have been used to gather these data including video cameras, inductive loop detectors, license plate recognition, as well as wireless-based technologies such as infrared. These technologies have different capabilities with respect to the amount of information that can be derived from the collected data. Regardless of the richness of the collected data, the majority of the available technologies focus on collecting vehicle-based data because they either do not have the capability to collect data from other travel modes (e.g., bicycles and pedestrians), or may need to be deployed differently to support this capability (e.g., video technology). One type of

wireless-based data collection system that has been deployed recently is based on Bluetooth technology. A key feature of Bluetooth-based data collection systems that makes travel mode identification feasible is that the Bluetooth-enabled devices within vehicles are also present on bicyclists and pedestrians.

The main objective of this research was to explore the feasibility of utilizing the information contained in data collected by Bluetooth-based data collection units (DCU) to automatically identify three different modes of transportation (i.e., motor vehicles, bicyclists, and pedestrians) travelling through an intersection. To accomplish this objective, a methodology was developed that included three controlled data collection experiments and one uncontrolled data collection experiment where data were gathered from Bluetooth-enabled devices using several Bluetooth DCUs. The main performance metric utilized was the *duration of travel* which was calculated from the time-stamped MAC address data collected by the Bluetooth DCUs. The clustering methods k-Means, Fuzzy c-Means, and Partitioning Around Medoids were applied to the overall duration of travel data to distinguish vehicles, bicycles and pedestrians. The results obtained in this research prove that the Bluetooth-based data collection system can be a viable approach for distinguishing different modes of transportation travelling through intersections controlled by either a stop sign or traffic light.

# Differentiating Transportation Modes Using Bluetooth Sensor Data

by
Nadia Bathaee

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented April 23, 2014

Commencement June 2014

Master of Science thesis of Nadia Bathaee presented on April 23, 2014.

APPROVED:

Major Professor, representing Industrial Engineering

Head of the School of Mechanical, Industrial and Manufacturing Engineering

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Nadia Bathaee, Author

ACKNOWLEDGEMENTS

This thesis is dedicated to my parents Sima Ganjimanesh and Masoud Bathaei for their kind support and patience as I pursued this degree.

This work would not have been possible without the encouragement and mentoring from my major professor Dr. J. David Porter. Also, I am deeply grateful to Dr. David S. Kim for his support and invaluable advice during my graduate work. I would also like to thank Dr. Charlotte Wickham and Dr. John Sessions for their willingness to serve in my graduate committee and for their valuable comments and feedback.

Finally, thanks to Alireza, Sejoon, Amirali, Behnam, Elnaz, and Vahid who helped me in conducting my data collection experiments.

TABLE OF CONTENTS

Page

1.0 INTRODUCTION ..... 1

1.1 RESEARCH MOTIVATION ..... 2

1.2 RESEARCH OBJECTIVE ..... 3

1.3 RESEARCH CONTRIBUTIONS ..... 4

2.0 LITERATURE REVIEW ..... 5

2.1 BLUETOOTH TECHNOLOGY ..... 6

2.1.1 The Bluetooth Inquiry Process ..... 7

2.2 BLUETOOTH-BASED DATA COLLECTION FROM VEHICLES ..... 8

2.3 PEDESTRIAN TRAFFIC DATA COLLECTION ..... 12

2.4 BICYCLE TRAFFIC DATA COLLECTION ..... 16

2.5 MULTIPLE TRAFFIC MODE DATA COLLECTION AND CLASSIFICATION ..... 18

2.6 CLUSTERING METHODS ..... 20

2.7 LITERATURE REVIEW SUMMARY ..... 22

3.0 RESEARCH METHODOLOGY ..... 24

3.1 BLUETOOTH DATA COLLECTION UNIT ..... 26

3.2 BLUETOOTH-ENABLED DEVICES ..... 28

3.3 DATA COLLECTION EXPERIMENTS ..... 30

3.3.1 First Data Collection Experiment ..... 31

3.3.1.1 General Description of the Experiment ..... 31

TABLE OF CONTENTS (Continued)

Page


3.3.1.2 Experiment Hardware and Personnel ............................. 32

3.3.1.3 Experimental Factors and Response Variable ................ 33

3.3.1.4 Detailed Test Procedure ................................................. 37

3.3.2 Second Data Collection Experiment ............................................ 39

3.3.2.1 General Description of the Experiment .......................... 39

3.3.2.2 Experiment Hardware and Personnel ............................. 40

3.3.2.3 Experimental Factors and Response Variable ................ 41

3.3.2.4 Detailed Test Procedure ................................................. 45

3.3.3 Third Data Collection Experiment ............................................... 47

3.3.3.1 General Description of the Experiment .......................... 47

3.3.3.2 Experiment Hardware and Personnel ............................. 49

3.3.3.3 Experimental Factors and Response Variable ................ 49

3.3.3.4 Detailed Test Procedure ................................................. 53

3.3.4 Fourth Data Collection Experiment ............................................. 56

3.3.4.1 General Description of the Experiment .......................... 56

3.3.4.2 Experiment Hardware and Personnel ............................. 59

3.4 STATISTICS USED TO IDENTIFY MODES .................................................. 60

TABLE OF CONTENTS (Continued)

Page

3.5 DATA ANALYSIS METHODS ........ 61

3.5.1 Cluster Analysis ........ 61

3.5.1.1 K-Means Clustering Method ........ 62

3.5.1.2 Fuzzy c-Means Clustering Method ........ 65

3.5.1.3 Partitioning Around Medoids (PAM) ........ 68

3.5.2 Majority Voting ........ 74

4.0 RESULTS ........ 75

4.1 DATA PROCESSING AND ANALYSIS PROCEDURES FOR THE FIRST THREE DATA COLLECTION EXPERIMENTS ........ 76

4.1.1 Results from the First Data Collection Experiment ........ 79

4.1.2 Results from the Second Data Collection Experiment ........ 84

4.1.3 Results from the Third Data Collection Experiment ........ 89

4.1.4 Percentage Error of Data Analysis Methods ........ 100

4.2 DATA PROCESSING AND ANALYSIS PROCEDURES FOR THE FOURTH DATA COLLECTION EXPERIMENT ........ 101

4.2.1 Result from the Fourth Data Collection Experiment ........ 103

5.0 CONCLUSIONS AND OPORTUNITES FOR FUTURE WORK ........ 107

5.1 RESEARCH CONCLUSIONS ........ 108

5.2 OPPORTUNITIES FOR FUTURE WORK ........ 110

BIBLIOGRAPHY ........ 113

TABLE OF CONTENTS (Continued)

Page

APPENDIX .......... 119

LIST OF FIGURES


Figure Page

Figure 1: Categories of Clustering Methods (V. Filkov, 2006) ..............................22

Figure 2: Overview of the Research Methodology Chapter ...................................25

Figure 3: Bluetooth Data Collection Unit ..............................................................27

Figure 4: Structure of a Record Collected by a Bluetooth DCU.............................28

Figure 5: "Google Nexus One" Cell Phone (Google Nexus, 2010).........................29

Figure 6: "LG Shine CU720" Cell Phone (LG, 2008) ...........................................29

Figure 7: Bluetooth Data Collection System for the First Data Collection Experiment.......................................................................................... 32

Figure 8: Bluetooth Data Collection System for the Second Data Collection Experiment.......................................................................................... 40

Figure 9: Bluetooth Data Collection System for the Third Data Collection Experiment.......................................................................................... 48

Figure 10: Bluetooth Data Collection System for the Fourth Data Collection Experiment ......................................................................................... 57

Figure 11: Snapshot from the Fourth Data Collection Experiment .........................59

Figure 12: Example of k-Means Clustering Method...............................................64

Figure 13: Example of FCM Clustering Method ....................................................67

Figure 14: Example of PAM Clustering Method ....................................................70

Figure 15: Olimex iMX233-OLinuXino-MINI (Olimex, 2013)...........................120

Figure 16: SENA Parani UD100 Bluetooth USB Adapter (SENA, 2011) ...........121

Figure 17: MaxStreamXBee Series 2 OEM RF Modules (MaxStream, 2006)......123

Figure 18: L-com 2.4 GHz 8 dBi Round Patch Antenna ......................................125

LIST OF FIGURES (Continued)

Figure Page

Figure 19: L-com 2.4 GHz 8 dBi Round Patch Antenna Gain Patterns (L-com, 2005) ........ 126

Figure 20: L-com 2.4 GHz 5 dBi Rubber Duck Antenna (L-com, 2012) ........ 127

Figure 21: L-com 2.4 GHz 5 dBi Rubber Duck Antenna Gain Patterns (L-com, 2012) ........ 128

Figure 22: Energizer XP18000 Universal AC adapter (Energizer, 2010) ........ 129

Figure 23: GlobalSat BU-353 USB GPS Receiver (GlobalSat, 2009) ........ 130

LIST OF TABLES

Table Page

Table 1: Main Experimental Factors and Levels for the First Data Collection Experiment ........ 34

Table 2: Randomized Complete Block Design for First Data Collection Experiment ........ 35

Table 3: Order of Randomized Runs for First Experiment ........ 36

Table 4: Approximate Speeds (MPH) of Transportation Modes ........ 38

Table 5: Main Experimental Factors and Levels for Second Data Collection Experiment ........ 41

Table 6: Randomized Complete Block Design for Second Data Collection Experiment ........ 43

Table 7: Order of Randomized Runs for Second Data Collection Experiment ........ 44

Table 8: Main Experimental Factors and Levels for the Third Data Collection Experiment ........ 50

Table 9: Completely Randomized Design for Third Data Collection Experiment ........ 51

Table 10: Fully Randomized Experimental Runs for Third Data Collection Experiment ........ 52

Table 11: Bluetooth DCUs Used in Each Section for the Fourth Data Collection Experiment ........ 57

Table 12: Cluster Analysis Notation (Tan et al., 2006) ........ 62

Table 13: Main Characteristics of the k-Means Clustering Method ........ 71

Table 14: Main Characteristics of the FCM Clustering Method ........ 72

Table 15: Main Characteristics of the PAM Clustering Method ........ 73

Table 16: Example of Data Saved in an Excel File ........ 76

LIST OF TABLES (Continued)

Table Page

Table 17: Example of Group of Data Separated for Each Run ............................ 77

Table 18: k-Means and Majority Voting Results for the First Data Collection Experiment ........................................................................ 81

Table 19: FCM and Majority Voting Results for the First Data Collection Experiment ........................................................................ 82

Table 20: PAM and Majority Voting Results for the First Data Collection Experiment ........................................................................ 83

Table 21: k-Means and Majority Voting Results for the Second Data Collection Experiment ........................................................................ 86

Table 22: FCM and Majority Voting Results for the Second Data Collection Experiment ........................................................................ 87

Table 23: PAM and Majority Voting for the Second Data Collection Experiment ........................................................................ 88

Table 24: k-Means and Majority Voting Results for the Third Data Collection Experiment ........................................................................ 91

Table 25: FCM and Majority Voting for the Third Data Collection Experiment ........................................................................ 94

Table 26: PAM and Majority Voting for the Third Data Collection Experiment ........................................................................ 97

Table 27: Percentage Error for Different Analyses Methods............................. 100

Table 28: Ground Truth Data Collected from Video ......................................... 104

Table 29: Total Number of Runs in each Section Collected by Bluetooth DCUs ........................................................................ 104

Table 30: Clustering Results for Fourth Data Collection Experiment ................ 105

Table 31: Percentage of Detections for Fourth Data Collection Experiment ..... 106

LIST OF TABLES (Continued)

Table Page

Table 32: iMX233-OLinuXino-MINI features and specifications ..................... 121

Table 33: Parani UD100 Bluetooth USB Adapter Specifications ...................... 122

Table 34: XBee Series 2 OEM RF Modules Specifications ............................... 124

Table 35: 2.4 GHz 8 dBi Round Patch Antenna Electrical Specifications ......... 125

Table 36: 2.4 GHz 8 dBi Round Patch Antenna Mechanical Specifications...... 126

Table 37: 2.4 GHz 5 dBi Rubber Duck Antenna Electrical Specifications ........ 127

Table 38: 2.4 GHz 5 dBi Rubber Duck Antenna Mechanical Specifications..... 128

Table 39: Energizer XP18000 Universal AC Adapter Specifications ................ 129

Table 40: BU-353 USB GPS Receiver Specifications........................................ 130

# 1.0 INTRODUCTION

State departments of transportation, transportation consultants, and other transportation agencies are always in need of data that can be used to better understand how different modes of transportation use the road and highway systems. A variety of automatic data collection technologies have been used to gather these data including video cameras, inductive loop detectors, license plate recognition, as well as wireless-based technologies such as infrared. All these automatic data collection technologies have different capabilities with respect to the amount of information that can be derived from the data they collect. Also, their implementation, operation (i.e., data storage and processing), and maintenance cost increases as the richness of the dataset they collect also increases.

One type of wireless-based data collection system that has been deployed recently is based on Bluetooth technology. Since Bluetooth-based systems can identify and re-identify vehicles at different locations, they have primarily been deployed for vehicle travel time data collection and to conduct origin-destination studies. However, recent research has shown that the data collected by a Bluetooth-based systems contain more information than simply the identification of a vehicle at a specific time (Kim et. al, 2012; Saeedi et. al, 2012).

In the past, data collection efforts have largely focused on passenger and freight vehicles and a variety of solutions for monitoring traffic and travel behavior have been available for motorized modes for many years (Griffin et. al, 2014).

However, bicyclists and pedestrians play an integral role in the transportation system and must be accommodated to make the system efficient and safe.

Efforts to improve conditions for bicyclists and pedestrians and forecast future demand require accurate travel data. However, limitations in existing sources of data often interfere with these efforts. Therefore, transportation agencies have identified data relating to the counting and classification of bicycles and pedestrians as a high priority and have stressed the need for extensive research to identify technologies and methods (other than manual counting) to successfully obtain these data (Noyce et al., 2006).

## 1.1 RESEARCH MOTIVATION

The technologies used for the automatic collection of road and highway system data have different capabilities with respect to the amount of information that can be derived from the collected data. Also, the cost of implementation, operation (i.e., data storage and processing) and maintenance of a technology increases as the richness of the dataset it collects also increases. For example, a large amount of information can be obtained from systems based on video technology, but the systems are costly and often require manual data processing.

Regardless of the richness of the collected data, the majority of the available technologies focus on collecting vehicle-based data because they either do not have the capability to collect data from other travel modes (e.g., bicycles and pedestrians), or may need to be deployed differently to support this capability (e.g., video technology).

A key feature of Bluetooth-based data collection systems that makes travel mode identification feasible is that the Bluetooth-enabled devices within vehicles are also present on bicyclists and pedestrians. Also, Bluetooth-based data collection systems are less expensive and easier to deploy than the majority of competing technologies used to collect road and highway system data. These advantages were the main motivation for exploring the feasibility of using Bluetooth technology to collect data to differentiate vehicles, bicycles and pedestrians travelling through an intersection.

## 1.2 RESEARCH OBJECTIVE

The main objective of this research was to explore the feasibility of utilizing the information contained in data collected by Bluetooth-based data collection units (DCU) to automatically identify three different modes of transportation (i.e., motor vehicles, bicyclists, and pedestrians) travelling through an intersection.

To accomplish this objective, a methodology was developed that included three controlled data collection experiments and one uncontrolled data collection experiment where data were collected from Bluetooth-enabled devices using several Bluetooth DCUs. The main performance metric utilized was the *duration of travel* which was calculated from the time-stamped MAC address data collected by the Bluetooth DCUs.

## 1.3 RESEARCH CONTRIBUTIONS

The results obtained in this research prove that a Bluetooth-based data collection system is a viable approach for distinguishing different modes of transportation travelling through an intersection with high degree of accuracy.

More specifically, the development of a methodology that uses Bluetooth technology to collect data simultaneously from three modes of transportation and the application of the clustering methods k-Means, Fuzzy c-Means, and Partitioning Around Medoids to *the overall duration of travel* data to distinguish vehicles, bicycles and pedestrians are considered the main contributions of this research.

# 2.0 LITERATURE REVIEW

The objective of this research was to develop a methodology to differentiate three modes of transportation (i.e., vehicles, bicycles and pedestrians) by applying different clustering methods to data collected from Bluetooth-enabled devices. To accomplish this objective, the most relevant literature in the areas of Bluetooth technology, Bluetooth-based data collection, transportation mode differentiation using a variety of technologies, and data clustering methods was reviewed and the findings are synthesized in this chapter.

The rest of the chapter is organized as follows. Section 2.1 presents a broad description of the Bluetooth inquiry process and several studies about this subject. Sections 2.2, 2.3 and 2.4 present prior research work related to the use of different technologies to collect traffic data from vehicles, pedestrians, and bicyclists, respectively. Section 2.5 presents a summary of different studies on the topic of distinguishing multiple traffic modes. Section 2.6 presents a brief introduction to the topic of clustering methods. Finally, section 2.7 presents a summary of the literature reviewed in this chapter.

## 2.1 BLUETOOTH TECHNOLOGY

Bluetooth is a short-range, low-power, open standard wireless telecommunication technology that operates in the unlicensed 2.4-GHz industrial, scientific, and medical (ISM) frequency band. Bluetooth technology uses frequency hopping to achieve reliable communication even in noisy environments (Ringwald & Romer, 2007). Bluetooth technology is mainly used in the implementation of wireless personal area networks (WPAN) and is now a common feature available in a number of devices such as laptops, desktop computers, printers, mobile phones, and smart phones.

The applications of Bluetooth technology have grown significantly since it was first introduced in 1994 as a potential replacement to wired media (i.e., cables) to connect computers to peripherals such as printers. One of the recent applications of Bluetooth technology is in the collection of data that can be used to estimate transportation-related performance measures. Bluetooth technology offers several advantages when compared to other data collection technologies including:

- The ability to establish ad-hoc connections.
- The ability to withstand interference from other sources in open band.
- Its negligible power consumption.
- The use of an open interface standard.

### 2.1.1 The Bluetooth Inquiry Process

Bluetooth devices can form small networks called *piconets*. A piconet may include only one master and up to seven slave devices. The frequency hopping sequences to enable the synchronization between the master and the slave devices are controlled by the master device.

The process of forming a piconet consists of two steps: the *inquiry* process and the *page* process. After the inquiry process is complete, more formal connections can be established between the master and the slave devices via the page process. Transportation applications are only concerned with the data that can be collected during the inquiry process due to the fact that vehicles typically go past the discovery area of the Bluetooth device's antenna at high speeds.

In the inquiry process, the inquiring (i.e., master) device discovers neighboring slave devices and exchanges information about slave clock times for the purpose of synchronization. An inquiring device broadcasts inquiry packets on the sequence of available frequencies to detect potential slaves in the neighborhood and scans for replies. The inquiring device has a 28-bit clock, which ticks every 312.5μs. The inquiry process continues until all replies are received. On two consecutive time slots, the inquiring device transmits inquiry packets on two sequential frequencies and, during the next two time slots, it scans for a reply on these same two frequencies (Bisdikian, 2001).

Bluetooth scanning devices (i.e., slave devices) that want to be discovered enter the inquiry scan state and scan for inquiry packets on the same available frequencies that the inquiring device is using for transmission. The frequency of each scanning device cycles through the available frequencies in order and changes every 1.28 seconds. The hopping rate of scanning devices is much slower than that of the inquiring device to make sure that the frequencies used eventually coincide and that messages are successfully received. If the scanning device successfully hears a message by listening on the right frequency at the right time, it will switch to the inquiry response state. In this state, the scanning device waits two time slots and then sends a reply on the same frequency (Peterson et al., 2004).

## 2.2 Bluetooth-Based Data Collection from Vehicles

Several researchers have utilized Bluetooth technology for data collection in order to estimate transportation-related performance measures such as average speed and travel time. The units used to collect these data, referred to as data collection units (DCUs), usually include a board on which a Bluetooth module can be mounted directly or can also use a Bluetooth USB adapter. DCUs may use a battery as the power source, but can also draw AC power from the existing infrastructure on freeways or arterials.

Several studies have been conducted where data from Bluetooth-based devices were collected in a variety of environments. These environments included freeways experiencing free flow traffic conditions, as well as arterial roads with several intersections where traffic was being controlled with stop signs or traffic lights. In these studies, the data collected with Bluetooth technology varied depending on the purpose of the research and the target transportation system. However, a data element that is always collected is the media access control (MAC) address. A MAC address is a 48-bit, 12 alpha-numeric character unique identifier that is assigned to each Bluetooth device at the time of manufacturing. The date and time (i.e., the time stamp) at which the MAC address was detected is another data element that is commonly recorded. In more recent studies, a measure of the strength of the radio frequency (RF) signal known as the received signal strength indicator (RSSI) has also been collected. The rest of this section presents the most representative research efforts in this area.

Martchouk and Mannering (2009) collected time-stamped MAC addresses from Bluetooth-enabled devices travelling in vehicles on interstates in the state of Indiana. The data collection period lasted two weeks. The main objectives of this study were (1) to use Bluetooth technology to collect data on a freeways to calculate travel times, and (2) to observe potential variations in the calculated travel times induced by factors such as weather conditions, time of day, and type of vehicle and to find statistical models to predict this variability. The results of the study showed that indeed the average and the standard deviation of travel times varied significantly due to factors such as time of the day, the different speeds in adjacent lanes, and the behavior of the drivers. It was found that the highest travel times and lowest speeds were observed between 16:00 and 19:00. Thus, the average and the standard deviation of travel times increased during peak hours. Another factor observed to have a significant effect on the results was weather conditions. The analysis showed that both the average and standard deviation of travel times were significantly different under adverse versus normal weather conditions. The researchers indicated that two major benefits of using Bluetooth technology for data collection process are (1) no complex algorithms are required to calculate travel time, and (2) the technology is relatively inexpensive to implement.

Haghani et al. (2010) introduced Bluetooth as a new and effective way for data collection on freeways. Portable Bluetooth DCUs were developed at the Center for Advanced Transportation Technology at the University of Maryland and were positioned on a segment of Interstate I-95 between Washington, D.C., and Baltimore, Maryland. Time-stamped MAC addresses from Bluetooth-enabled devices were recorded by the DCUs. After the data collection process, a four-step offline filtering algorithm was designed to extract data from Bluetooth observations. The first two steps of the offline filtering algorithm were designed to identify and discard outliers in each time interval. The purpose of the third and fourth steps was to exclude time intervals during which they either did not have enough observations or there were large variations among individual observations within the time interval. Data from inductive loop detectors were also used to approximate the average sampling rate of Bluetooth sensors. On average, the Bluetooth DCUs sampled only between 2% and 3.4% of the vehicles in the traffic stream. However, the results indicated that Bluetooth technology is capable of providing reliable, high-volume travel time data on highways when the DCUs are not located very close to each other and there are no facilities such as gas stations or rest areas in between the two consecutive DCUs, which may result in the overestimation of the travel times.

Haseman et al. (2009) utilized Bluetooth technology in a study that resulted in proposed metrics to evaluate work zone mobility. Time-stamped MAC addresses from Bluetooth-enabled devices were collected over a 12-week period from a work zone segment of I-65 interstate highway in northwestern Indiana. It was indicated that approximately 8% of the passing vehicles were detected. The collected data were used to calculate travel delay times. The results showed that using Bluetooth technology provides information in construction zones that can be displayed on message boards to provide motorists with improved trip planning.

Malinovskiy et al. (2010) attempted to compare the travel times estimated with data collected with Bluetooth technology with the travel times obtained with automatic license plate recognition (ALPR) technology on a short corridor. The main objective was to better characterize errors with Bluetooth data collection by formulating a relationship between error, antenna type, and different configuration of the data collection units. Three types of antennae (i.e., 7dBi omnidirectional, 9dBi omnidirectional, and 12dBi directional) and three different sensor arrangements (i.e., one sensor, two sensors mounted on the same side of the street, and two sensors mounted on the opposite sides) were tested to determine the effects of these variables on travel time error. The authors concluded that from all the configurations attempted, combinations with omnidirectional antennae and large detection zones provided the best results, with low absolute error.

Saeedi et al. (2012) studied the use of Bluetooth-based data collection systems on arterial roads. The objective of this research was to develop a methodology to collect accurate and precise travel time data between signalized intersections using a Bluetooth-based data collection system. The methodology utilized RSSI data to significantly improve the accuracy of the travel times. A probe vehicle study was conducted that consisted of five DCUs permanently installed at consecutive signalized intersections, approximately one mile apart from each other, along a high-volume signalized arterial in Tigard, Oregon. The differences between travel times were calculated using time-stamped MAC addresses associated with the first detection, last detection, and average of the first and last detections in a group, and RSSI-based travel times. For ground truth, a laptop was used to record the exact time when the vehicle passed the DCUs. The average travel times generated using the RSSI-based method had an average error of 1.35 seconds compared to the ground truth travel times which was significantly smaller than other methods.

## 2.3 PEDESTRIAN TRAFFIC DATA COLLECTION

Several studies have been conducted to collect pedestrian traffic data. Individual studies have used a variety of technologies such as video, infrared, and more recently Bluetooth. The main purpose of these studies was to produce a count of the number of pedestrians and to study their behavior in a specific location. The main challenge of such studies, regardless of the data collection technology used, was to distinguish pedestrians from each other particularly in crowded places. Several studies have used image processing techniques on recorded video footage for this purpose. Bluetooth technology has been used recently in these types of studies. The data usually collected were the time-stamped MAC addresses of Bluetooth-enabled devices. The rest of this section presents the most representative research efforts in this area.

Rourke et al. (1994) captured information on pedestrian behavior using image processing techniques. In order to determine the presence of a pedestrian, two video frames were captured in a short period of time. Also, two algorithms were developed to measure pedestrian density in a populated area and to find the walking direction of each pedestrian. To overcome the variability associated with unpredictable changes in speed and direction of pedestrians, some assumptions were made. First, it was assumed that all pedestrians moved normally and, those who were not moving, would not stay this way for a significant amount of time. Second, it was assumed that the only moving objects in the scene were pedestrians. The results showed that walking direction was not identified correctly most of the times. The error can be caused by closely spaced pedestrians in a populated area or by moving arms and legs during walking. Identifying a pedestrian in a single-pedestrian scene worked without any error. However, specifying the individuals when there were more than one pedestrian in a crowded scene close to each other failed to work properly in a few cases.

Fujisawa et al. (2013) introduced a method of clustering optical flows in video frames to improve the counting accuracy of the pedestrians mostly in cases where occlusion occurs and many pedestrians are present at the same time. The main disadvantage of video technology, which is mentioned in several studies, is its inability to distinguish pedestrians in crowded areas. The accuracy of the proposed method was evaluated by using several video sequences, focusing on the effect of parameters for optical flow clustering. The optical flow clustering utilized the lengths, angles, and source locations of optical flows. The results showed an improvement in the counting accuracy by up to 25% as compared to a non-clustering method.

Dai et al. (2007) used infrared imagery technology to detect pedestrians. An algorithm was developed to separate the stationary and moving objects by checking the infrared images and using layering representation. The stationary objects were assigned to the background layer, whereas the moving objects were assigned to the foreground layer. Pedestrians were then separated from non-pedestrian moving objects in the foreground layer based on the shape of their image. The algorithm was applied to samples from the Ohio State University (OSU) thermal database, which covers different environmental conditions, and to the West Virginia University (WVU) database. At OSU, unlike WVU, the camera was kept still all the time and the distance between pedestrians and the infrared camera was large.

The proposed algorithm takes three to five seconds to process 30 frames and it performs well regardless of camera's motion and distance. It should be noted that highly occluded pedestrians were not counted in this study.

Wang (2011) used infrared monocular imagery to detect and track pedestrians with a stationary infrared camera. A shape-based pedestrian detection approach algorithm was developed to classify captured images into either a human or a non-human class and to improve the performance of pedestrian tracking. The images in this study were samples taken from two databases: Terravic Motion IR Database and OSU Color-Thermal Database. The algorithm had low computational complexity and managed to detect pedestrians most of the times, except for the times when a pedestrian was sitting or when pedestrians were very close to each other. Three different tracking scenarios were considered to evaluate the performance of the proposed infrared pedestrian tracking approach. First, tracking under good imaging conditions, which means that a human body image is light enough to be distinguished from the background. Second, tracking under image intensity changes in which a pedestrian walks from far to near toward the camera and the intensity and shape of the image changes. The third scenario involved tracking under heavy disturbances (i.e., two pedestrians walk very close together in front of the camera). In all three cases, even in the heavily disturbed situations, the proposed method improved the tracking performance compared to the previous studies in this area.

Malinovskiy et al. (2012) investigated the feasibility of using static Bluetooth DCUs for pedestrian data collection specifically to monitor their movements. The study was performed at two separate sites (i.e., Montreal, Canada, and Seattle, WA). At each site, two Bluetooth DCUs equipped with a 7dBi omnidirectional antenna were set 100m apart and mounted at a height of about 3 meters. Overall, 2,520 unique devices were seen in Montreal, whereas only 534 were seen in Seattle. In general, the sensors were able to capture roughly 5% of the population at the Montreal location and 2.25% at the Seattle location. However, the method that was used as ground truth was not mentioned in the study. Travel time and dwell time, which is the continuous presence of an individual (more than 60 seconds), were calculated from the data. The results indicated that even with a very low sample size, this approach can provide insights into pedestrian travel behavior.

## 2.4 BICYCLE TRAFFIC DATA COLLECTION

Several technologies have been used to collect bicycle traffic data and each has its advantages and disadvantages. This section synthesizes a number of the studies conducted on bicycle traffic data collection using either video or Bluetooth technology. The main purpose of these studies was to examine the behavior of the cyclists and to improve the traffic signal timing by calculating average speeds and travel times. As in other studies, time-stamped MAC addresses and RSSI values were the main data elements captured from Bluetooth-enabled devices.

Shladover et al. (2009) used video recordings taken at intersections to collect bicycle traffic data. The purpose was to improve the traffic signal timing and to determine how to specify the minimum green signal intervals throughout the state of California to give bicyclists sufficient time to cross wide arterials. If bicyclists could be distinguished from motor vehicles, it would be possible to modify the signal timing so that the bicyclists would receive longer green times than motor vehicles. Two video cameras were used in the experiment; one mounted at a height of 6 to 7 meters on top of a trailer parked near the test intersection with a view of the bicyclists' path; and a second camera mounted at a height of 2 to 3 meters, with a view of the traffic signal. After recording the observations, the video images were processed to extract the trajectories of bicyclists. These trajectories were synchronized with video images of the traffic signals so that the timing of the bicyclists could be determined relative to the signal phases. Testing was done for two days in Palo Alto, CA, and three days in Berkeley, CA. In Palo Alto, among 310 observed bicyclist crossing, 255 produced usable data. Out the 225 useful data points, 180 included a stop at the intersection and 75 were through-passes. In Berkeley, out of 439 usable crossings, 279 had stops and 160 were through-passes. Overall, however, there were more unusable data.

Mei et al. (2011) estimated bicycle travel time on a short corridor in China with data collected using Bluetooth sensors with directional antennas. Two video cameras were used as a ground truth in the experiments. On average, the Bluetooth sensors sampled between 2% and 3% of the bicycles in the bicycle traffic flow stream. In the paper, different filtering approaches were introduced to avoid errors and filter out outliers. A t-test was performed to compare the travel times calculated from Bluetooth sensors and video cameras. The statistical tests showed that the bicycle travel times estimated with the data collected with the Bluetooth sensors were not significantly different from the actual travel times. Therefore, Bluetooth sensor travel time data can be a good representative of actual travel times with less effort and lower cost compare to the video cameras.

## 2.5 MULTIPLE TRAFFIC MODE DATA COLLECTION AND CLASSIFICATION

This section presents the findings from studies that have attempted to collect data simultaneously from multiple traffic modes. Several of these studies have used infrared or video technology to collect data from pedestrians and bicyclists in an attempt to distinguish these two modes of transportation.

Noyce et al. (2006) used active-infrared overhead vehicle-imaging sensor technology to detect and classify pedestrians and bicycles. This technology creates an overhead, three-dimensional image of the passing objects by scanning the roadway with two laser beams and classifies them based on their shape and size. The effectiveness of their algorithm on pedestrians and bicycles was first tested on an overhead bridge located in the campus of the University of Massachusetts-Amherst. The active-infrared devices were mounted at a lower height than recommended, which led to a decrease in the distance between the laser beams. Different scenarios were tested including single pedestrian; single bicycle; different combinations of interaction between these two such as a combination of two pedestrians or bicycles moving closely to each other but one behind another; and a combination of two or three pedestrians and bicycles moving in opposite direction of travel. A total of 307 bicycles and 426 pedestrians were observed. The results showed that all 733 travelers were successfully detected and that approximately 92% of all the detections were correctly classified. Also, the results were very accurate when multiple pedestrians or multiple bicycles traveled together.

Schraml et al. (2010) presented a compact event-based 3D vision system including a spatio-temporal clustering method for real-time classification of pedestrians and bicyclists. The preliminary results on real-scenarios showed that the system can discriminate (in real-time) between pedestrians, riding bicyclists, and walking bicyclists in more than 92% of the cases using three criteria: length, width and time.

Very few studies have been conducted on the subject of distinguishing different transportation modes (i.e. vehicles, pedestrians, and bicyclists). Namaki Araghi et al. (2012) used Bluetooth technology to collect data from vehicles and bicyclists. The Bluetooth data collection system consisted of two Bluetooth DCUs positioned 550 meters apart from each other along a free flow road segment. For each detection captured by the Bluetooth data collection system, the time-stamped MAC address, the RSSI level, and the class of device (CoD) were recorded. To identify the type of mode, the researchers applied different clustering methods including hierarchical clustering, k-Means clustering, and two-step clustering techniques to the travel time data. The data clustering process was based on two underlying hypotheses, i.e., the travel time for vehicles is significantly lower than the bicycles and some class of devices would only be used by a specific mode. For instance, GPS-based navigation systems, headsets and speakers were clustered in the group of devices with short travel time (i.e., vehicles). For the experiment, a free-flow section of a road in the city of Aalborg, Denmark, was selected. This road did not have any access road connection and had a minimum number of pedestrian traffic, so all the passes for both traffic modes were considered through-passes without any stopping. In order to validate the collected data, real traffic was recorded by a video camera. Three-directional antennas were used which had a coverage range between 100 and 300 meters. The detection rate was about 20% of the traffic flow. Although the results of the three clustering methods were very close, the k-Means clustering method gave the most accurate estimate in most cases. Also, the results showed that in non-congested traffic situations, there was a significant difference between the average travel times of various traffic modes.

## 2.6 Clustering Methods

Clustering is the process of organizing objects into groups whose members are more similar to each other. A cluster is therefore a collection of objects which are similar within each cluster and are dissimilar to the objects belonging to other clusters. So, the goal of clustering is to determine the intrinsic grouping in a set of unlabeled data.

As Figure 1 shows, clustering methods are classified as “parametric” or “non-parametric” (also referred to as “hierarchical”). Non-parametric clustering methods produce a classification that seeks to build a hierarchy of clusters and are further subdivided into either “agglomerative” or “divisive” classifications algorithms. Agglomerative algorithms (also called “bottom-up algorithms”) treat each data point as a singleton cluster and then successively merge pairs of clusters until all clusters have been merged into a single cluster that contains all the data points. Divisive algorithms (also called “top-down algorithms”) proceed by splitting clusters recursively until individual data points are reached (Manning, 2008).

Parametric clustering methods generate a classification by partitioning a dataset creating a set of exclusive or overlapping groups having no hierarchical relationships among them. Parametric clustering methods are further subdivided into “reconstructive”, “graph models” and “generative.” Reconstructive algorithms are generally based on a cost function, which in some way incorporates the loss of information incurred by clustering methods when trying to reconstruct the original data from cluster representatives. Examples of this type of clustering are k-Means and partitioning around medoids (PAM). Graph models are used when the data can be presented as graphs. Generative clustering generates models to describe different clusters and use these models to classify data points. Fuzzy c-Means (FCM) is an example of this type of clustering (Guan, 2006).

Clustering methods can be applied in many different fields, such as marketing, biology, insurance, earthquake studies, etc. Clustering methods have also been used in data collected for transportation-related research (Namaki Araghi et al., 2012).

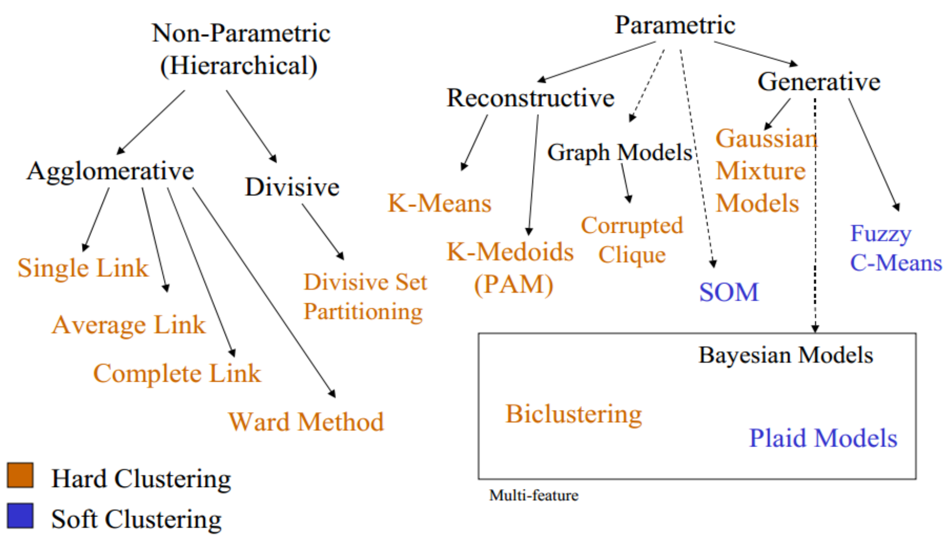


**Figure 1: Categories of Clustering Methods (V. Filkov, 2006)**

## 2.7 LITERATURE REVIEW SUMMARY

The previous sections in this chapter have synthesized a number of prior research efforts that focused on data collection from vehicles, bicyclists, and pedestrians using different technologies including infrared and video technology. The advantage of these technologies over Bluetooth is that they capture the whole population while Bluetooth technology is only capable of collecting a sample of the population. However, Bluetooth technology offers several advantages that have drawn the attention of researchers in this area including ease of use and low implementation cost. Another major advantage of using Bluetooth technology over infrared or video is that it works better in high density areas where, for example, pedestrians are walking very close to each other.

Despite the increasing research being conducted on data collection with Bluetooth technology, there is still an opportunity to use data collected from Bluetooth-enabled devices to distinguish different modes of transportation. Among the few studies that have been conducted, none of them considered complicated road segments such as intersections with stops signs or traffic lights and none could obtain a conclusive result with very few errors.

# 3.0 RESEARCH METHODOLOGY

The methodology followed in this research consisted of several phases, as depicted in Figure 2. Three controlled and one uncontrolled data collection experiments were conducted to examine the feasibility of utilizing Bluetooth-based data collection units (DCUs) to identify vehicles, bicycles, and pedestrians traveling on a road segment. The controlled data collection experiments were conducted in a parking lot where a simulated intersection controlled by a stop sign was set up. The uncontrolled experiment was conducted at a four-way intersection controlled by traffic lights. The clustering methods k-Means, Fuzzy c-Means (FCM), and partitioning around medoids (PAM), were applied to the collected data to identify the modes of transportation.

The rest of this chapter is organized as follows. Section 3.1 describes the Bluetooth DCUs used in this research. Section 3.2 describes the Bluetooth-enabled devices used in the first three data collection experiments. Section 3.3 describes in detail the procedures followed in the four data collection experiments conducted in this research, including the hardware and software used, experimental factors and response variables. Finally, section 3.4 and section 3.5 describe the statistics and data analysis methods used to identify different modes of transportation (i.e., vehicles, bicyclists, and pedestrians).

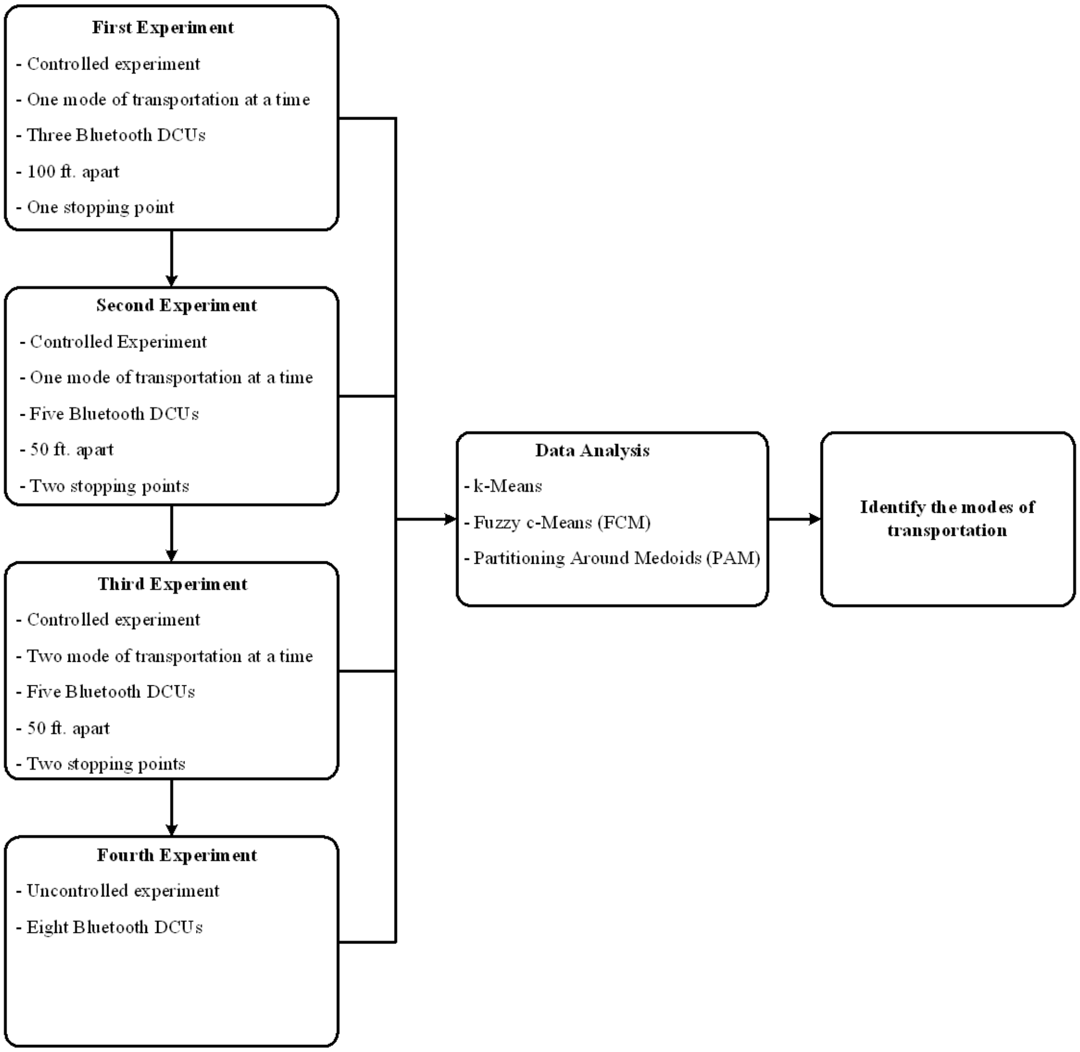


**Figure 2: Overview of the Research Methodology Chapter**

## 3.1 BLUETOOTH DATA COLLECTION UNIT

A system composed of several Bluetooth DCUs was used in this research to collect time-stamped data from Bluetooth-enabled devices. Each Bluetooth DCU was assembled using the following commercial-off-the-shelf (COTS) components:

- Olimex iMX233-OLinuXino-MINI Single Board Computer
- SENA Parani UD100 Class 1 Bluetooth USB adapter
- MaxStream XBee Series 2 OEM ZigBee module
- L-com 2.4 GHz 8dBi round patch antenna
- L-com 2.4 GHz 5dBi rubber duck antenna
- GlobalSat BU-353 USB GPS Receiver
- Energizer Lithium XP18000 Battery.

Figure 3 depicts how the COTS components were assembled to form a complete Bluetooth DCU.

**Figure 3: Bluetooth Data Collection Unit**

A script written in Python 2.7 was installed in the Bluetooth DCUs to enable the collection of data from Bluetooth-enabled devices. Each record collected by a Bluetooth DCU included the following fields:

- The media access control (MAC) address of the Bluetooth-enabled device
- A time stamp (i.e., date and time)
- An RSSI value
- The MAC address of the Bluetooth DCU.

Figure 4 shows the structure of a record collected by a Bluetooth DCU from Bluetooth-enabled devices during the data collection experiments conducted in this research.

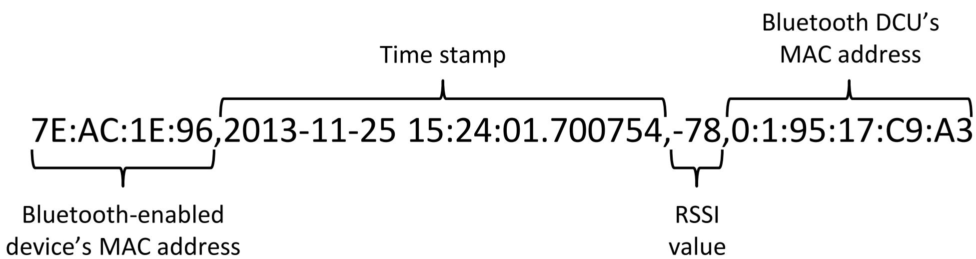


**Figure 4: Structure of a Record Collected by a Bluetooth DCU**

It is important to note that the first four characters (out of 12) from each MAC address collected from Bluetooth-enabled devices were removed to protect users' privacy.

### 3.2 BLUETOOTH-ENABLED DEVICES

Two types of Bluetooth-enable devices (i.e., cell phones) were used in the controlled experiments to collect data. One of the cell phones, depicted in Figure 5, was a "Google Nexus One" smartphone. The other cell phone, depicted in Figure 6, was an "LG Shine CU720."


**Figure 5: "Google Nexus One" Cell Phone (Google Nexus, 2010)**


**Figure 6: "LG Shine CU720" Cell Phone (LG, 2008)**

The "Google Nexus One" and the "LG Shine CU720" cell phones were used in the first experiment. After analyzing the data collected from the "LG Shine CU720," it was determined that the strength of the signal produced by its radio was too weak. Therefore, two “Google Nexus One” cell phones were used in the second and third data collection experiments.

For simplicity, the “Google Nexus One” and “LG Shine CU720” cell phones will be referred to simply as “Nexus” and “LG” for the rest of this chapter.

### 3.3 DATA COLLECTION EXPERIMENTS

A series of data collection experiments were conducted to examine the feasibility of utilizing Bluetooth DCUs to identify vehicles, bicycles, and pedestrians traveling on a road segment which included either a simulated or a real signalized intersection. To evaluate the feasibility of wirelessly identifying different travel modes, the following items were explored in the data collection experiments and data analysis:

- The number of Bluetooth DCUs along the road
- Spacing between consecutive Bluetooth DCUs
- Statistics utilized to identify modes
- Data analysis methods.

Three controlled data collection experiments were first conducted in a parking lot where the transportation modes could travel through their paths in a safe manner and without any interference. A simulated road segment was setup that allowed the different transportation modes to travel in both directions and also included a stop sign to control traffic flow. A fourth uncontrolled data collection experiment was conducted at a four-way intersection controlled by traffic lights.

Sections 3.3.1, 3.3.2, and 3.3.3 describe in more detail each of the three controlled data collection experiments. Section 3.3.4 describes the details of the uncontrolled data collection experiment.

### 3.3.1 First Data Collection Experiment

The objective of the first data collection experiment was to begin to explore the feasibility of using Bluetooth-based data collection and cluster analysis to identify different modes of transportation (i.e., vehicles, bicyclists, and pedestrians) in a simulated, two-way road segment with an intersection controlled by a stop sign. In the first data collection experiment, modes of transportation were only allowed to stop *before* the simulated intersection.

#### *3.3.1.1 General Description of the Experiment*

In this experiment, a Bluetooth-based data collection system composed of three Bluetooth DCUs was utilized. The Bluetooth DCUs were located 100 feet apart from each other along a straight line on a simulated road segment, as depicted in Figure 7. All three Bluetooth DCUs were time-synchronized using a laptop computer prior to running the experiment. To do this, the laptop computer was connected directly to each Bluetooth DCU to synchronize its time with the time reported by the laptop computer.

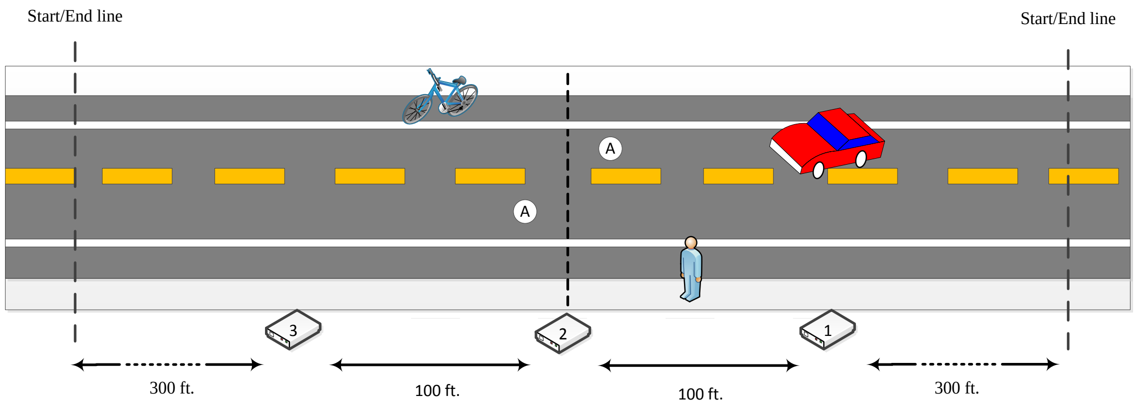


**Figure 7: Bluetooth Data Collection System for the First Data Collection Experiment**

The label "A" shown in Figure 7 identifies the point on the simulated road segment where a stop sign was located. In this experiment, not all experimental runs required a transportation mode to make a complete stop at point “A.” In some instances, the transportation mode traveled through the Bluetooth data collection system without stopping.

### *3.3.1.2 Experiment Hardware and Personnel*

The following hardware was utilized to run the first data collection experiment:

- Two cell phones (i.e., one “Nexus” and one “LG”)
- One car
- One bicycle
- Three Bluetooth DCUs.

Two operators coordinated and executed all the experimental runs. One operator was responsible for driving the vehicle, riding the bicycle and acting as a pedestrian. The other operator performed the time-synchronization of the Bluetooth DCUs at the beginning of the experiment.

#### *3.3.1.3 Experimental Factors and Response Variable*

The main experimental factors (and their levels) used in the first data collection experiment are shown in Table 1. The response variable was *total duration of travel* (in seconds) of each pass for each transportation mode. The total duration of travel was calculated by subtracting the time a transportation mode was first seen by any of the three Bluetooth DCUs from the time the transportation mode was last detected by any of the three Bluetooth DCUs. This does not necessarily mean that the transportation mode was seen first by the first Bluetooth DCU and last seen by the last Bluetooth DCU. For example, as a car traveled from the first Bluetooth DCU towards the third Bluetooth DCU, it might have been detected first by the third Bluetooth DCU and then detected last by the second Bluetooth DCU.

**Table 1: Main Experimental Factors and Levels for the First Data Collection Experiment**

| FACTOR | LEVELS |
|---|---|
| **Mode of Transportation** | 1. Car<br>2. Bicycle<br>3. Pedestrian |
| **Run Type** | 1. Through pass<br>2. Stop before the middle Bluetooth DCU (A) |
| **Cell Phone** | 1. LG<br>2. Nexus |

The experimental design chosen was a randomized complete block design with two replications. The blocking was based on the modes of transportation. A randomized complete block design was chosen because it is time-efficient. Considering the three main experimental factors and their levels, a total of 24 experimental runs were needed. Table 2 shows all the experimental runs and the coding used to identify them. For example, the experimental run coded as "P1s(L)" identifies the first replicate of a pedestrian run with a stop at “A” using the “LG” cell phone as the Bluetooth-enabled device. Table 3 shows the fully randomized experimental runs in each block in the order they were executed in the first data collection experiment.

**Table 2: Randomized Complete Block Design for First Data Collection Experiment**

| | **Pedestrian** | | | | **Bicyclist** | | | | **Car** | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **Stop** | | **Through** | | **Stop** | | **Through** | | **Stop** | | **Through** | |
| **LG** | P1s(L) | P2s(L) | P1t(L) | P2t(L) | B1s(L) | B2s(L) | B1t(L) | B2t(L) | C1s(L) | C2s(L) | C1t(L) | C2t(L) |
| **Nexus** | P1s(N) | P2s(N) | P1t(N) | P2t(N) | B1s(N) | B2s(N) | B1t(N) | B2t(N) | C1s(N) | C2s(N) | C1t(N) | C2t(N) |

**P: Pedestrian** **s: Stop at A** **L: LG** **1: 1st replicate**

**B: Bicyclist** **t: Through pass** **N: Nexus** **2: 2nd replicate**

**C: Car**

**Table 3: Order of Randomized Runs for First Experiment**

| Block | Pass # & Run Type | Block | Pass # & Run Type | Block | Pass # & Run Type |
|---|---|---|---|---|---|
| **Pedestrian** | 1. P2s(L) | **Bicyclist** | 9. B2t(L) | **Car** | 17. C1s(N) |
| | 2. P1t(L) | | 10. B2s(N) | | 18. C1t(N) |
| | 3. P1s(N) | | 11. B1t(N) | | 19. C2t(L) |
| | 4. P2s(N) | | 12. B1s(N) | | 20. C2t(N) |
| | 5. P1t(N) | | 13. B2t(N) | | 21. C1s(L) |
| | 6. P2t(N) | | 14. B2s(L) | | 22. C2s(L) |
| | 7. P2t(L) | | 15. B1s(L) | | 23. C2s(N) |
| | 8. P1s(L) | | 16. B1t(L) | | 24. C1t(L) |

#### *3.3.1.4 Detailed Test Procedure*

Before the first data collection experiment was conducted, three Bluetooth DCUs were placed in the test area exactly as depicted in Figure 7. Next, the following steps were completed before the execution of the randomized runs:

1. The batteries of the Bluetooth DCUs were checked to ensure they had enough charge (i.e., at least three bars).
2. It was verified that the L-com 2.4 GHz 8dBi round patch antenna and the SENA Parani UD100 Class 1 Bluetooth USB adapter were properly connected to each Bluetooth DCU.
3. Each individual Bluetooth DCU was connected to a laptop computer for time synchronization.

The following steps were executed for all the runs of the experiment:

1. It was verified that the person operating the transportation mode had only one Bluetooth-enabled cell phone and that all Bluetooth-enabled cell phones were turned OFF.
2. The operator positioned himself at the starting point of the run, i.e., 300 feet before the location of DCU1 (or DCU3). The 300 feet prior to the first DCU are referred to as the *acceleration zone* for the remainder of this chapter.
3. The operator turned ON the Bluetooth-enabled cell phone and placed it inside his pants' pocket.

4. Table 4 shows the average speed at which the different transportation modes traveled within the detection range of the Bluetooth DCUs. These speeds were measured during prior calibration experiments.

**Table 4: Approximate Speeds (MPH) of Transportation Modes**

| Mode of Transportation | Average Speed (MPH) |
|---|---|
| **Pedestrian** | 3 |
| **Bicycle*** | 15 |
| **Car** | 20 |

*** Typical speeds for bicycles are 10 to 20 mph. A default value of 15 mph may be used as the average bicycle running speed in the absence of local data.**

5. For the experimental runs with a through pass, the mode of transportation passed all three DCUs without stopping at the intersection.
6. For the experimental runs with a stop, the transportation mode remained at the stopping point “A” located at the imaginary intersection (see Figure 7) for 10 seconds. Once the 10 seconds elapsed, the transportation mode started accelerating until it reached its average speed (see Table 4) and left the detection zone.
7. After passing all three DCUs, the transportation modes ended their run approximately 300 feet past the last DCU in the system. The 300 feet after the last DCU are referred to as the *deceleration zone* for the remainder of this chapter.

8. Once the transportation modes reached the end of the deceleration zone, all Bluetooth-enabled cell phones were turned OFF.

### 3.3.2 Second Data Collection Experiment

The objective of the second data collection experiment was to compare and validate the results observed in the first data collection experiment, but now using slightly different experimental conditions. In the second data collection experiment, a simulated, two-way road segment with an intersection controlled by a stop sign was used. However, the modes of transportation were allowed to stop both *before* and *after* the simulated intersection.

#### *3.3.2.1 General Description of the Experiment*

In this experiment, a Bluetooth-based data collection system composed of five Bluetooth DCUs was utilized. The Bluetooth DCUs were located 50 feet apart from each other along a straight line on a simulated road segment, as depicted in Figure 8. All five Bluetooth DCUs were time-synchronized using a laptop computer prior to running the experiment using the same procedure as in the first data collection experiment.

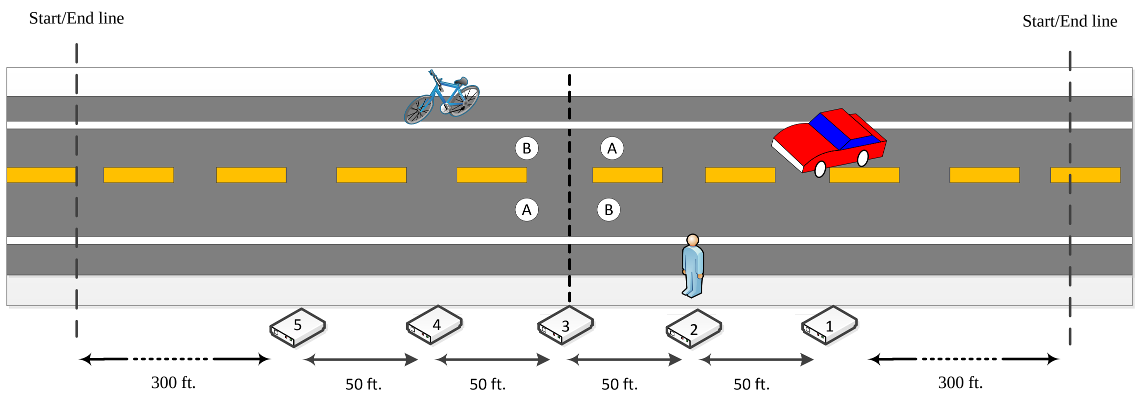


**Figure 8: Bluetooth Data Collection System for the Second Data Collection Experiment**

The labels "A" and "B" shown in Figure 8 identify the points on the simulated road segment where modes of transportation could stop. In this experiment, not all experimental runs required a transportation mode to make a complete stop at point “A” or “B.” In some instances, the transportation mode traveled through the Bluetooth data collection system without stopping.

### *3.3.2.2 Experiment Hardware and Personnel*

The following hardware was utilized to run the second data collection experiment:

- Two cell phones (i.e., both “Nexus”)
- One car
- One bicycle
- Five Bluetooth DCUs.

Two operators coordinated and executed all the experimental runs. One operator was responsible for driving the vehicle, riding the bicycle and acting as a pedestrian. The other operator performed the time-synchronization of the Bluetooth DCUs at the beginning of the data collection experiment.

#### *3.3.2.3 Experimental Factors and Response Variable*

The main experimental factors (and their levels) used in the second data collection experiment are shown in Table 5. The response variable was *total duration of travel* (in seconds) of each pass for each transportation mode. The total duration of travel was calculated by subtracting the time a transportation mode was first seen by any of the five Bluetooth DCUs from the time the transportation mode was last detected by any of the five Bluetooth DCUs.

**Table 5: Main Experimental Factors and Levels for Second Data Collection Experiment**

| FACTOR | LEVELS |
|---|---|
| **Mode of Transportation** | 1. Car<br>2. Bicycle<br>3. Pedestrian |
| **Run Type** | 1. Through pass<br>2. Stop before the middle Bluetooth DCU (at A)<br>3. Stop after the middle Bluetooth DCU (at B) |
| **Cell Phone** | 1. Nexus 1<br>2. Nexus 2 |

Due to resource and time limitations, the experimental design chosen was a randomized complete block design with a single replication. The blocking was done based on the modes of transportation. Considering the three main experimental factors and their levels, a total of 18 experimental runs were needed. Table 6 shows all the experimental runs and the coding used to identify them. For example, the experimental run coded as "bA2" identifies a bicyclist run with a stop at “A” using the “Nexus2” cell phone as the Bluetooth-enabled device. Table 7 shows the fully randomized experimental runs in each block in the order that they were executed in the second data collection experiment.

**Table 6: Randomized Complete Block Design for Second Data Collection Experiment**

| | **Pedestrian** | | | **Bicyclist** | | | **Car** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Through** | **Stop at A** | **Stop at B** | **Through** | **Stop at A** | **Stop at B** | **Through** | **Stop at A** | **Stop at B** |
| **Nexus1** | pT1 | pA1 | pB1 | bT1 | bA1 | bB1 | cT1 | cA1 | cB1 |
| **Nexus2** | pT2 | pA2 | pB2 | bT2 | bA2 | bB2 | cT2 | cA2 | cB2 |

**p: Pedestrian** **T: Through pass** **1: Nexus1**

**b: Bicyclist** **A: Stop at A** **2: Nexus2**

**c: Car** **B: Stop at B**

**Table 7: Order of Randomized Runs for Second Data Collection Experiment**

| Block | Pass # & Run Type | Block | Pass # & Run Type | Block | Pass # & Run Type |
|---|---|---|---|---|---|
| Pedestrian | 1. pT1 | Bicyclist | 7. bT1 | Car | 13. cA2 |
| | 2. pA2 | | 8. bA2 | | 14. cA1 |
| | 3. pB2 | | 9. bA1 | | 15. cB2 |
| | 4. pA1 | | 10. bB2 | | 16. cT1 |
| | 5. pT2 | | 11. bB1 | | 17. cT2 |
| | 6. pB1 | | 12. bT2 | | 18. cB1 |

*3.3.2.4 Detailed Test Procedure*

Before the second data collection experiment was conducted, five DCUs were placed in the test area exactly as depicted in Figure 8. Next, the following steps were completed before the execution of the randomized runs:

1. The batteries of the Bluetooth DCUs were checked to ensure they had enough charge (i.e., at least three bars).
2. It was verified that the L-com 2.4 GHz 8dBi round patch antenna and the SENA Parani UD100 Class 1 Bluetooth USB adapter were properly connected to each Bluetooth DCU.
3. Each individual Bluetooth DCU was connected to a laptop computer for time synchronization.

The following steps were executed for all the runs of the experiment:

1. It was verified that the person operating the transportation mode had only one Bluetooth-enabled cell phone and that all Bluetooth-enabled cell phones were turned OFF.
2. The operator positioned himself at the beginning of the acceleration zone (i.e., 300 feet before the location of DCU1 or DCU5).
3. The operator turned ON the Bluetooth-enabled cell phone and placed it inside his pants' pocket.
4. For the experimental runs with through pass, the mode of transportation passed all five DCUs without stopping at the intersection.

5. For the passes with a stop before the middle DCU, the transportation mode remained at the stopping point “A” located at the imaginary intersection (see Figure 8) for 10 seconds. Once the 10 seconds elapsed, the transportation mode started accelerating until it reached its average speed (see Table 4) and left the detection zone.
6. For the passes with a stop after the middle DCU, the transportation mode also remained at the stopping point “B” located at the imaginary intersection (see Figure 8) for 10 seconds. Once the 10 seconds elapsed, the transportation mode started accelerating until it reached its average speed (see Table 4) and left the detection zone.
7. After passing all five DCUs, the transportation modes ended their run after the deceleration zone (i.e., 300 feet after the location of DCU1 or DCU5).
8. Once the transportation modes were at the end of the deceleration zone, all Bluetooth-enabled cell phones were turned OFF.

### 3.3.3 Third Data Collection Experiment

In the third data collection experiment, a more complex environment was used to collect data from Bluetooth-enabled devices. Once again, a simulated, two-way road segment with an intersection controlled by a stop sign was used. However, two different modes of transportation could be present at the same time in the discovery range of the Bluetooth DCUs. In this experiment, the modes of transportation were allowed to stop *before* and *after* the simulated intersection.

#### *3.3.3.1 General Description of the Experiment*

In this experiment, a Bluetooth-based data collection system composed of five Bluetooth DCUs was utilized. The Bluetooth DCUs were located 50 feet apart from each other along a straight line on a simulated road segment, as depicted in Figure 9. All five Bluetooth DCUs were time-synchronized prior to running the experiment. A GlobalSat BU-353 USB GPS receiver was used to set the current time on Bluetooth DCU#3 (see Figure 9). Bluetooth DCU#3, referred to as the *coordinator DCU*, was placed in the middle to be within the reasonable distance from all the other Bluetooth DCUs (referred to as *router DCUs*). Once the time was set on the coordinator DCU, this unit started sending a time reference to the router DCUs via a ZigBee wireless communications protocol. At the beginning of the time synchronization process, the router DCUs wait for the time reference from the *coordinator DCU*. However, once the router DCUs receive the time reference from the coordinator DCU, they start the Bluetooth inquiry process (i.e., the data collection process) immediately.

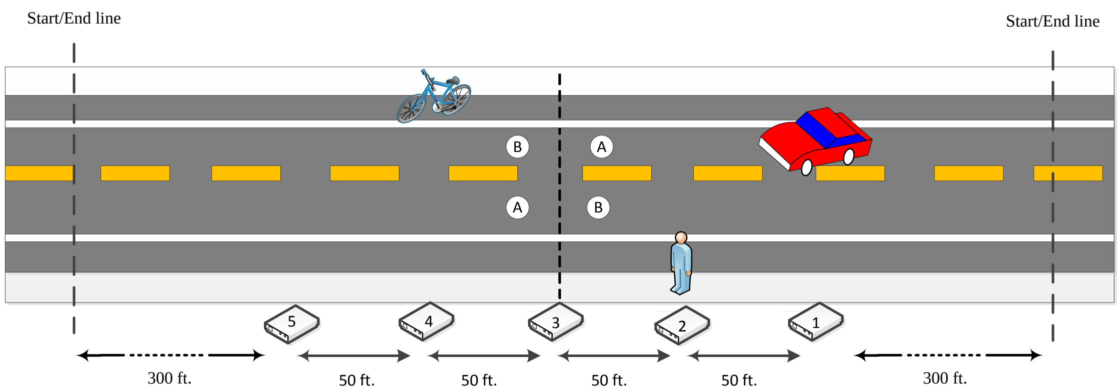


**Figure 9: Bluetooth Data Collection System for the Third Data Collection Experiment**

The labels "A" and "B" shown in Figure 9 identify the points on the simulated road segment where modes of transportation could stop. In this experiment, two different modes of transportation traveled through the Bluetooth data collection system simultaneously, either in the same direction or in opposite directions. In some experimental runs, the modes of transportation could both stop at the same location (i.e., either “A” or “B”). In other experimental runs, one transportation mode could stop at either “A” or “B” while the other mode could continue its travel through the Bluetooth data collection system without stopping.

#### *3.3.3.2 Experiment Hardware and Personnel*

The following hardware was utilized to run the third data collection experiment:

- Two cell phones (i.e., both “Nexus”)
- One car
- One bicycle
- Five Bluetooth DCUs
- One GlobalSat BU-353 USB GPS receiver.

Three operators coordinated and executed all the experimental runs. Two operators were responsible for driving vehicles, riding bicycles and acting as pedestrians. The other operator was responsible for time-synchronization of the Bluetooth DCUs at the beginning of the data collection experiment and managing the times at which each mode of transportation began its run.

#### *3.3.3.3 Experimental Factors and Response Variable*

The main experimental factors (and their levels) used in the third data collection experiment are shown in Table 8. The response variable was *total duration of travel* (in seconds) of each pass for each transportation mode. The total duration of travel was calculated by subtracting the time a transportation mode was first seen by any of the five Bluetooth DCUs from the time the transportation mode was last detected by any of the five Bluetooth DCUs.

**Table 8: Main Experimental Factors and Levels for the Third Data Collection Experiment**

| FACTOR | LEVELS |
|---|---|
| **Mode of Transportation** | 1. Car<br>2. Bicycle<br>3. Pedestrian |
| **Run Type** | 1. Both modes through pass<br>2. Both modes stop at A<br>3. Both modes stop at B<br>4. One mode stops at A; the other mode through pass<br>5. One mode stops at B; the other mode through pass |
| **Direction of Travel** | 1. Same direction<br>2. Opposite direction |

Due to resource and time limitations, the experimental design chosen was a completely randomized design with a single replication. Considering the three main experimental factors and their levels, a total of 30 experimental runs were needed. Table 9 shows all the experimental runs and the coding used to identify them. For example, the experimental run coded as "O: cT, bB" identifies a car and a bicycle traveling in opposite directions; the car executed a through pass, whereas the bicycle stopped at location B. Table 10 shows the fully randomized experimental runs in the order they were executed in the third data collection experiment.

**Table 9: Completely Randomized Design for Third Data Collection Experiment**

| | Run Type for Transportation Modes | Opposite direction of travel | Same direction of travel |
|---|---|---|---|
| **Pedestrian and Bicycle** | **Both through pass** | O: pT, bT | S: pT, bT |
| | **Both stop at A** | O: pA, bA | S: pA, bA |
| | **Both stop at B** | O: pB, bB | S: pB, bB |
| | **One stop, one through pass** | O: pA, bT | S: pA, bT |
| | | O: pT, bB | S: pT, bB |
| **Car and Bicycle** | **Both through pass** | O: cT, bT | S: cT, bT |
| | **Both stop at A** | O: cA, bA | S: cA, bA |
| | **Both stop at B** | O: cB, bB | S: cB, bB |
| | **One stop, one through pass** | O: cA, bT | S: cA, bT |
| | | O: cT, bB | S: cT, bB |
| **Car and Pedestrian** | **Both through pass** | O: cT, pT | S: cT, pT |
| | **Both stop at A** | O: cA, pA | S: cA, pA |
| | **Both stop at B** | O: cB, pB | S: cB, pB |
| | **One stop, one through pass** | O: cA, pT | S: cA, pT |
| | | O: cT, pB | S: cT, pB |

**Table 10: Fully Randomized Experimental Runs for Third Data Collection Experiment**

| Pass # | Run Type |
|---|---|
| 1 | O: cT, pT |
| 2 | S: cT, pB |
| 3 | O: cT, pB |
| 4 | S: cA, pA |
| 5 | O: cA, bT |
| 6 | O: pB, bB |
| 7 | S: pA, bA |
| 8 | S: cA, bA |
| 9 | S: pT, bB |
| 10 | O: pT, bB |
| 11 | O: cT, bT |
| 12 | S: cB, bB |
| 13 | S: pT, bT |
| 14 | S: cA, bT |
| 15 | O: pA, bA |

| Pass # | Run Type |
|---|---|
| 16 | O: pA, bT |
| 17 | O: cB, pB |
| 18 | S: cT, pT |
| 19 | S: cA, pT |
| 20 | O: cA, pA |
| 21 | O: cA, bA |
| 22 | O: pT, bT |
| 23 | S: pB, bB |
| 24 | S: cT, bT |
| 25 | S: pA, bT |
| 26 | O: cB, bB |
| 27 | O: cT, bB |
| 28 | S: cB, pB |
| 29 | O: cA, pT |
| 30 | S: cT, bB |

### *3.3.3.4 Detailed Test Procedure*

Before the third data collection experiment was conducted, five DCUs were placed in the test area exactly as depicted in Figure 9. Next, the following steps were completed before the execution of the randomized runs:

1. The batteries of the Bluetooth DCUs were checked to ensure they had enough charge (i.e., at least three bars).
2. It was verified that the L-com 2.4 GHz 8dBi round patch antenna and the SENA Parani UD100 Class 1 Bluetooth USB adapter were properly connected to each Bluetooth DCU.
3. A GlobalSat BU-353 USB GPS receiver was connected to the coordinator DCU to set the current time. The operator waited approximately 2 minutes and then unplugged the GPS receiver.

The following steps were executed for all the runs of the experiment:

1. It was verified that the persons operating the transportation modes had only one Bluetooth-enabled cell phone and that all Bluetooth-enabled cell phones were turned OFF.
2. The operators positioned themselves at the beginning of the acceleration zone (i.e., 300 feet before the location of DCU1 or DCU5).
3. The people operating both transportation modes turned ON their Bluetooth enabled cell phones and placed them inside their pants' pocket.

4. One operator managed the times at which each mode of transportation began its run. This was necessary due to the differences in speed among transportation modes. The main objective was to ensure that both modes of transportation appeared within the discovery range of the DCUs at approximately the same time. The process to accomplish this was as follows:

    a. Table 4 shows the average speed at which the different transportation modes traveled within the detection range of the DCUs.

    b. The slower mode of transportation started its run before the faster mode, i.e., pedestrians started before bicyclists; bicyclists started before cars; and pedestrians started before cars. Markings were made within the acceleration zone to identify a point that had to be reached by the slower transportation mode **before** the faster transportation mode could begin its run. ***The operators executed several preliminary runs before conducting any final runs to identify the location of these markings***. A marking could be a chalk line applied directly to the pavement or a small orange traffic cone.

    c. The operator responsible for managing the timing of the runs used a flag to signal the time at which the faster transportation mode was allowed to start its run using the markings identified in the previous step.

5. For the experimental runs with a through pass, the mode of transportation passed all five DCUs without stopping at the intersection.
6. For experimental runs with a stop before the middle DCU, the transportation mode remained at the stopping point “A” located at the imaginary intersection (see Figure 9) for 10 seconds. Once the 10 seconds elapsed, the transportation mode started accelerating until it reached its average speed (see Table 4) and left the detection zone.
7. For experimental runs with a stop after the middle DCU, the transportation mode remained at the stopping point “B” located at the imaginary intersection (see Figure 9) for 10 seconds. Once the 10 seconds elapsed, the transportation mode started accelerating until it reached its average speed (see Table 4) and left the detection zone.
9. After passing all five DCUs, the transportation modes ended their run after the deceleration zone (i.e., 300 feet after the location of DCU1 or DCU5).
10. Once the transportation modes were at the end of the deceleration zone, all Bluetooth-enabled cell phones were turned OFF.

### 3.3.4 Fourth Data Collection Experiment

The objective of the fourth data collection experiment was to collect data from Bluetooth-enabled devices in an uncontrolled environment. Thus, a Bluetooth data collection system was setup in a four-way intersection controlled by traffic lights.

#### *3.3.4.1 General Description of the Experiment*

In this experiment, a Bluetooth-based data collection system composed of eight Bluetooth DCUs was utilized, as depicted in Figure 10. The four Bluetooth DCUs labeled “1”, “2”, “3” and “4” in Figure 10 were located in a straight line along the main road segment (i.e., Monroe Ave.) crossing the intersection. Four additional Bluetooth DCUs labeled “N”, “S”, “E”, and “W” were located to the north, south, east and west of the main Bluetooth DCU system to help in determining the direction of travel of the modes of transportation. Also, the latter four Bluetooth DCUs divided the study area into different sections (i.e., EW, SE, SW, NE, and NW) and were located far from each other so that there was not any overlap between them.

Table 11 shows the set of Bluetooth DCUs that were considered in each section when performing the data analyses. For example, Bluetooth DCUs “1”, “2”, “3” and “4” were used to calculate the statistics used for data analysis for section EW. The reason DCUs “E” and “W” were not included in this analysis was because these DCUs were located far from the intersection and the focus of the data collection experiment was to identify the modes of transportations *at* the intersection.


**Figure 10: Bluetooth Data Collection System for the Fourth Data Collection Experiment**

**Table 11: Bluetooth DCUs Used in Each Section for the Fourth Data Collection Experiment**

| Section | Bluetooth DCUs |
|---|---|
| **EW** | 1 , 2 , 3, and 4 |
| **SE** | 3, 4, and S |
| **SW** | 1, 2, and S |
| **NE** | 3, 4, and N |
| **NW** | 1, 2, and N |

Three video cameras were used to collect “ground truth” data. The video cameras were placed at the locations labeled “A,” “B,” and “C” in Figure 10. From these three locations, the vehicle, bicycle and pedestrian activities at the intersection could be captured completely.

All eight DCUs and the video cameras were time-synchronized before running the experiment. A laptop was used to connect to the coordinator DCU at location 2. In order to time-synchronize the video cameras with the coordinator DCU, two operators were responsible to set the current time manually on the coordinator DCU (using a laptop computer) and the video cameras exactly at the same time. As the experiment progressed, the operators responsible for the video cameras recorded the time displayed on a cell phone every five minutes. This time was used later as a reference during the data analysis phase.

Once the time was set on the coordinator DCU, this unit started sending a time reference to the router DCUs via a ZigBee wireless communications protocol. The DCUs acting as routers did not begin the inquiry process (i.e., the data collection process) until they had received the time information from the coordinator DCU. The time synchronization among all system components (i.e., coordinator DCU, router DCUs, and video cameras) was validated several times during the test.

The data collection process took place between 3:30 PM and 4:30 PM on Nov. 25$^{th}$, 2013. This time period was chosen because the intersection experienced reasonable traffic flow of pedestrians, bicyclists, and cars during this time.

Figure 11 depicts a snapshot taken from the video camera located at position "B" that shows all three modes of transportation at the same time traveling through the intersection.

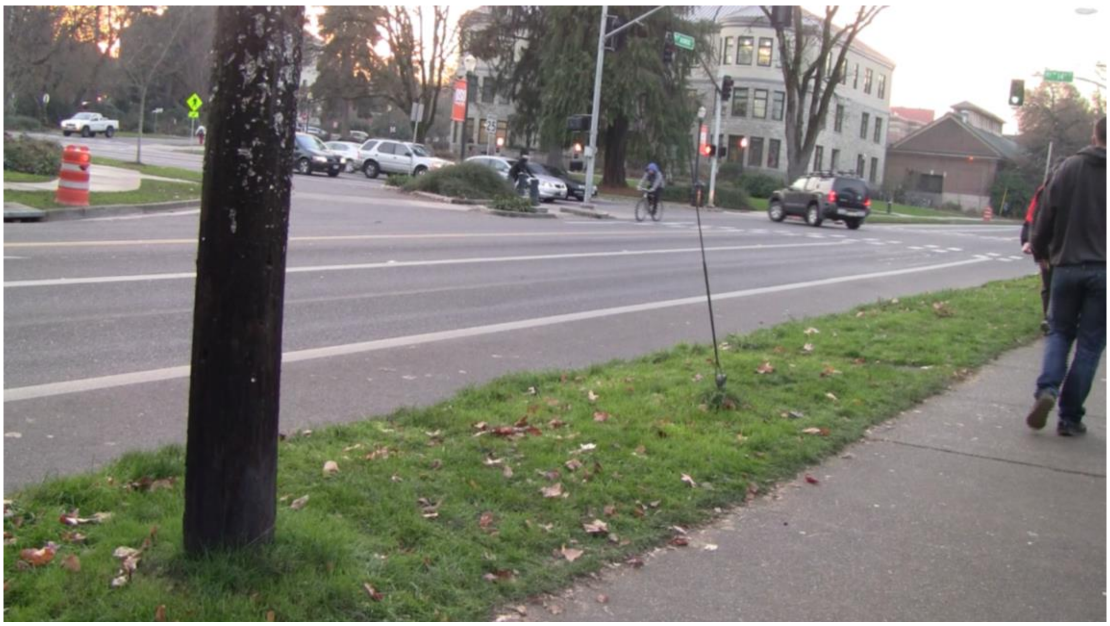

**Figure 11: Snapshot from the Fourth Data Collection Experiment**

### *3.3.4.2 Experiment Hardware and Personnel*

The following hardware was utilized to run the fourth data collection experiment:

- Eight Bluetooth DCUs
- Eight traffic drums, each equipped with a L-com 2.4 GHz 8dBi round patch antenna
- Three video cameras on tripods.

Two operators were responsible for time-synchronization of the Bluetooth DCUs and the video cameras at the beginning of the experiment and also for displaying the current time on a cell phone to video cameras every five minutes.

### 3.4 STATISTICS USED TO IDENTIFY MODES

The data collected by each Bluetooth DCU was a *group* of MAC address records. A *group* is defined as a collection of MAC address detections for the same MAC address that represents one trip (in a single direction) of the corresponding Bluetooth-enabled device along the length of road monitored by the Bluetooth DCUs (Saeedi et al., 2012). For a particular MAC address $k$:

- $n_i(k)$ = Number of MAC address records in a group detected by DCU $i$, $i$ = 1 to $N$.
- $t_{ij}(k)$ = Time stamp of the $j^{th}$ MAC address record in the group detected by DCU $i$, $j$=1 to $n_i(k)$.

The statistics that were used in the transportation mode identification analyses are the following:

- $d(k)$ = Total duration of travel for mode (in seconds) for all DCUs where,

$$d(k) = max_{i=1\,to\,N,\,j=1\,to\,n_i}\, t_{ij}(k) - min_{i=1\,to\,N,\,j=1\,to\,n_i}\, t_{ij}(k) \quad (1)$$

- $d_i(k)$ = Total duration of travel for mode (in seconds) for DCU $i$, where

$$d_i(k) = max_{j=1\,to\,n_i}\, t_{ij}(k) - min_{j=1\,to\,n_i}\, t_{ij}(k) \quad (2)$$

In the first three data collection experiments conducted in this research, the same Bluetooth-enabled devices were used with different modes of transportation. In these cases, the MAC addresses associated with different modes of transportation were considered as “different” addresses for analysis description purposes. In the fourth data collection experiment (i.e., application on a real road), the MAC addresses for different travel modes were all different.

As mentioned before, a mode of transportation was not always seen first by the first Bluetooth DCU and last seen by the last Bluetooth DCU in the Bluetooth data collection system. For example, if a car was traveling from the first Bluetooth DCU towards the last Bluetooth DCU, it could have been detected first by the third Bluetooth DCU and then detected last by the fourth Bluetooth DCU (in a five Bluetooth DCU data collection system).

## 3.5 DATA ANALYSIS METHODS

Cluster analysis and majority voting were the methods selected to identify different modes of transportation based on MAC address data.

### 3.5.1 Cluster Analysis

Cluster analysis "groups data objects based only on information found in the data that describes the objects and their relationships. The goal is that the objects within a group be similar (or related) to one another and different from (or unrelated to) the objects in other groups. The greater the similarity (or homogeneity) within a group and the greater the difference between groups, the better or more distinct the clustering" (Tan. PN, 2006). Table 12 presents the cluster analysis notation that will be used throughout this section.

**Table 12: Cluster Analysis Notation (Tan et al., 2006)**

| Symbol | Description |
|---|---|
| $x$ | An object |
| $C_i$ | The $i^{th}$ cluster |
| $c_i$ | The centroid of cluster $C_i$ |
| c | The centroid of all points |
| $m_i$ | The number of objects in the $i^{th}$ cluster |
| $m$ | The number of objects in the data set |
| $K$ | The number of clusters |

### *3.5.1.1 K-Means Clustering Method*

The first clustering method applied to the statistics generated from the MAC address data is known as k-Means. The first step in the k-Means algorithm is to define a value for $K$. An initial set of cluster centroids (i.e., $c_i \forall i = 1,...,K$ ) is typically selected at random from the data set $m$. Each object $x_i$ $\forall i = 1,...,m$ is then assigned to the "closest" centroid $c_i$ using a proximity measure (such as the Euclidean distance); each collection of objects assigned to a centroid is a cluster $C_i \forall i = 1,...,K$. The centroid of each cluster is then re-calculated based on the objects assigned to the cluster. The steps of assigning objects to clusters and updating the centroids of the clusters are repeated until no objects change cluster, or equivalently, until the centroids remain the same (Tan et al., 2006). The quality of the clustering is measured by calculating the sum of squared errors (SSE) using equation 3:

$$SSE = \sum_{i=1}^{K} \sum_{x \in C_i} (c_i - x)^2 \tag{3}$$

It can be shown that the centroid that minimizes the SSE is the mean. Therefore, equation 4 is used to calculate the centroids of the clusters in k-Means:

$$c_i = \frac{1}{m_i} \sum_{x \in C_i} x \tag{4}$$

Figure 12 depicts an example of how the k-Means clustering method would be applied to the total duration of travel data (i.e., *d(k)* data). The top graph labeled “k-Means Iteration 1” shows the values of the original *d(k)* data for three different modes of transportation. The data points displayed with a larger icon and a thicker black line in the top graph represent the initial centroids selected at random by the k-Means clustering method. In the bottom graph (labeled “k-Means Iteration N”), the original *d(k)* data is displayed sorted for clarity and the locations of the final centroids calculated by the k-Means clustering method are displayed with a black icon for each individual cluster.

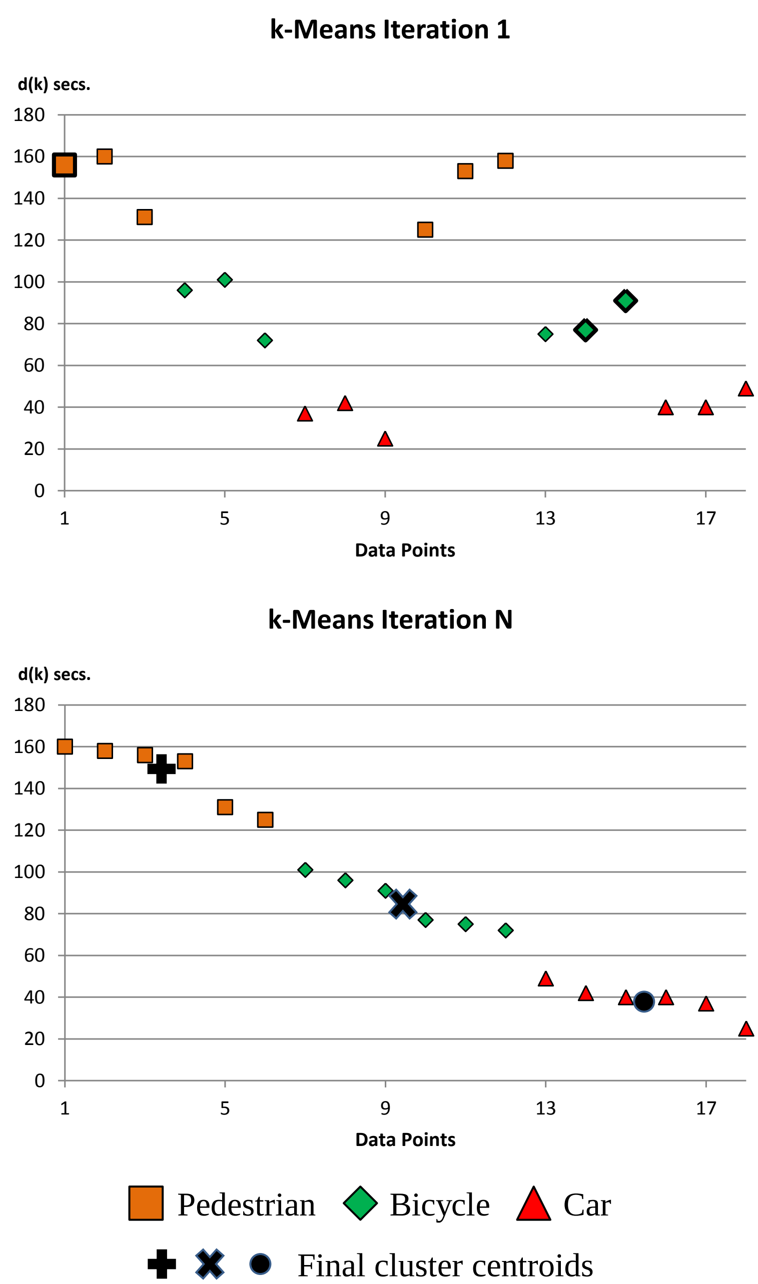


**Figure 12: Example of k-Means Clustering Method**

### *3.5.1.2 Fuzzy c-Means Clustering Method*

The second clustering method applied to the statistics generated from the MAC address data is known as Fuzzy c-Means (FCM). The first step in the FCM algorithm is to define a value for *K*. Then, each object $m_i$ $\forall i = 1,...,m$ is assigned a membership weight $w_{ij}$ $\forall i = 1,...,m; \ j = 1,...,K$; the set of weights $w_{ij}$ is referred to as the initial fuzzy pseudo-partition. The cluster centroids $c_i \forall i = 1,...,K$ are then computed and each object $x_i$ $\forall i = 1,...,m$ is then assigned to the "closest" centroid $c_i$ using a proximity measure (such as the Euclidean distance); each collection of objects assigned to a centroid is a cluster $C_i \forall i = 1,...,K$. After initialization, FCM repeatedly computes the centroids of each cluster $C_i \forall i = 1,...,K$ and the fuzzy pseudo-partition until the partition does not change. Two conditions are imposed in FCM (Tan et al., 2006):

1. All the weights for a given object, $x_i$, add up to 1.

$$\sum_{j=1}^{k} w_{ij} = 1$$

2. Each cluster, $C_j$, contains, with non-zero weight, at least one object, but does not contain, with a weight of one, all of the objects.

$$0 < \sum_{i=1}^{m} w_{ij} < m$$

As with k-Means, FCM can be interpreted as attempting to minimize the SSE, although FCM is based on a fuzzy version of SSE, as shown by equation 5:

$$SSE(C_1, C_2, \ldots, C_k) = \sum_{j=1}^{k}\sum_{i=1}^{m} w_{ij}^{p}\, dist(x_i, c_j)^2 \qquad (5)$$

Where $c_j$ is the centroid of the $j^{th}$ cluster and $p$, which is the exponent that determines the influence of the weights, has a value between 1 and ∞. Equation 6 shows the centroid calculation that minimizes the FCM SSE calculation:

$$c_j = \frac{\sum_{i=1}^{m} w_{ij}^{p} x_i}{\sum_{i=1}^{m} w_{ij}^{p}} \qquad (6)$$

The membership weights of a point in each cluster is "the reciprocal of the square of the distance between the point and the cluster centroid divided by the sum of all the membership weights of the point" (Tan et al., 2006). Equation 7 shows how membership weights get updated in each iteration when $p$ equals to 2:

$$w_{ij} = \frac{\dfrac{1}{dist(x_i, c_j)^2}}{\sum_{q=1}^{k} \left(\dfrac{1}{dist(x_i, c_q)}\right)^2} \qquad (7)$$

Figure 13 depicts an example of how the FCM clustering method would be applied to the total duration of travel data (i.e., *d(k)* data). The membership weights $w_{ij} \quad \forall i = 1,\ldots,m;\ j = 1,\ldots,K$ are shown in parenthesis above each data point. For instance, the membership weights for the data point labeled "1" are 0.78, 0.12, and 0.1, which represent the degree of membership of point "1" to the clusters motor vehicles, bicycles, and pedestrians, respectively. Since the highest membership weight belongs to the motor vehicles cluster, the *d(k)* of point "1" will be considered as an observation taken from a motor vehicle.

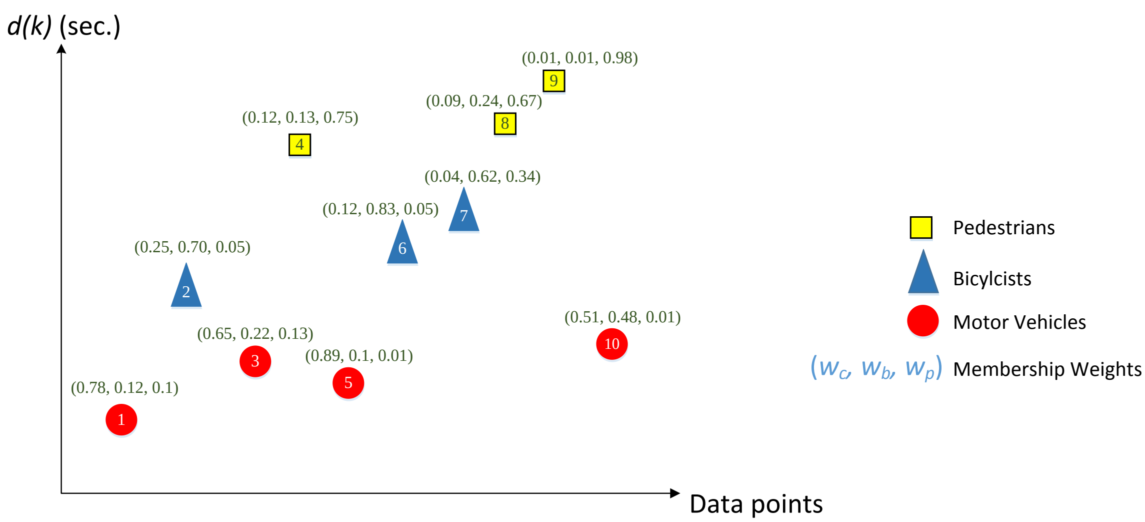


**Figure 13: Example of FCM Clustering Method**

#### *3.5.1.3 Partitioning Around Medoids (PAM)*

The third clustering method applied to the statistics generated from the MAC address data is known as Partitioning Around Medoids (PAM) or k-Medoids. The PAM algorithm is a clustering method related to k-Means. Both the k-Means and PAM algorithms are partitional algorithms which break the dataset up into groups. Contrarily to k-Means, PAM chooses actual data points as centers. Therefore, PAM could be more robust to noise and outliers as compared to k-Means. (Mirkes, 2011).

The first step in the PAM algorithm is to define a value for *K*. Next, *K* data points out of the *m* data points to be clustered are randomly selected as medoids. A medoid is the most centrally located point in the data set. For each data point, the cost of the configuration (i.e., the distance between the data point and the medoids) is calculated. Then, each point is associated to the closest medoid. The data point with the lowest configuration cost is selected as the new medoid. The steps of assigning data points to clusters and updating the medoids of the clusters are repeated until no data point changes cluster, or equivalently, until the medoids remain the same. The total configuration cost, which determines the quality of the clustering in each iteration, is calculated using equation 8:

$$cost(x,c) = \sum_{i=1}^{k}\sum_{j=1}^{m} | x_j - c_i | \tag{8}$$

Where $x$ is any data point, $c$ is the medoid, $m$ is the dimension of the data set, and $K$ is the number of clusters.

Figure 14 depicts an example of how the PAM clustering method would be applied to the total duration of travel data (i.e., *d(k)* data). The top graph labeled "PAM Iteration 1" shows the values of the original *d(k)* data for three different modes of transportation. The data points displayed with a larger icon and a thicker black line in the top graph represent the initial medoids selected at random by the PAM clustering method. In the bottom graph (labeled "PAM Iteration N"), the original *d(k)* data is displayed sorted for clarity and the final medoids selected by the PAM clustering method are again displayed with a larger icon and a thicker black line.

Table 13, Table 14, and Table 15 present a summary of the main characteristics of the three clustering methods discussed in this section.

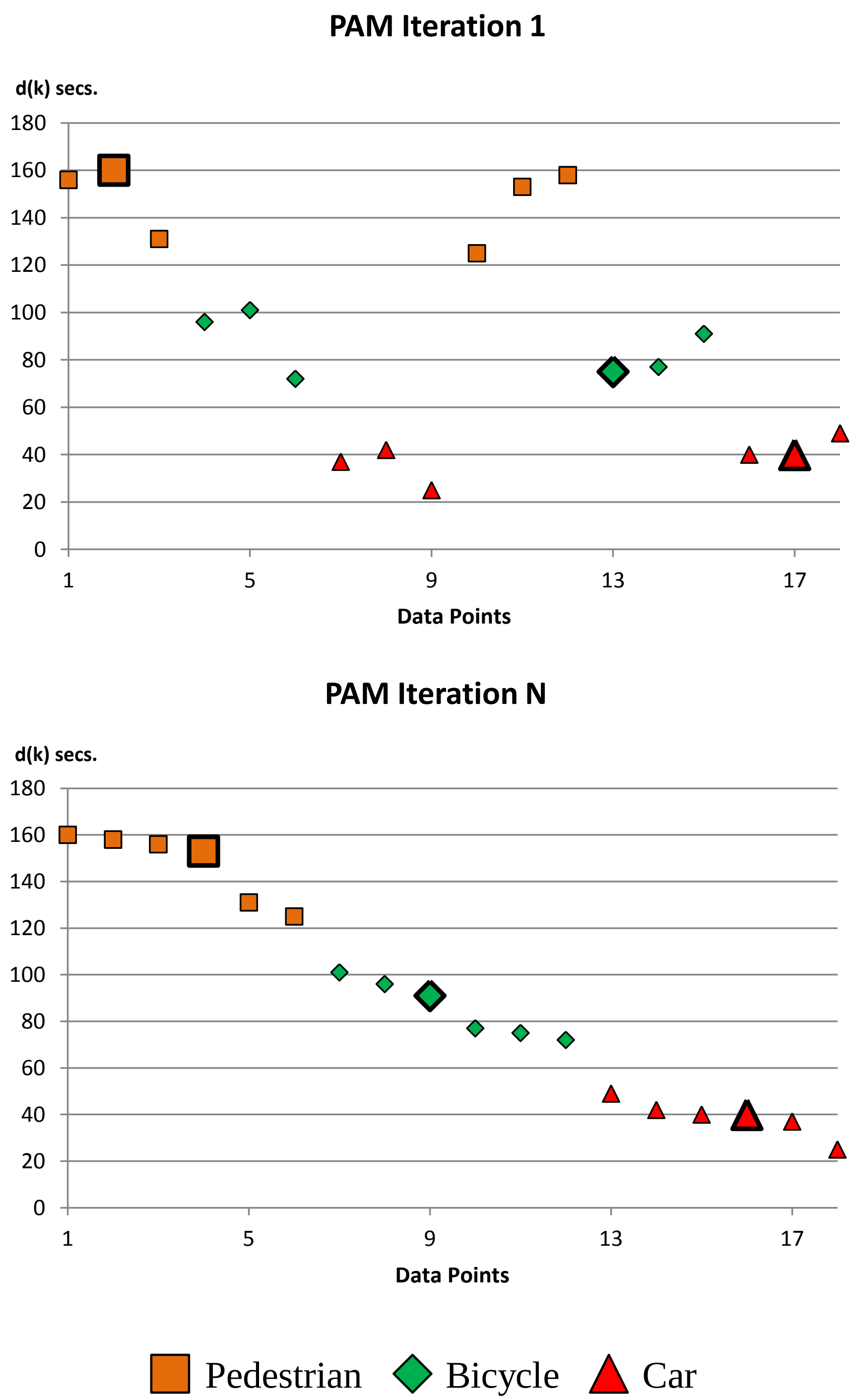


**Figure 14: Example of PAM Clustering Method**

**Table 13: Main Characteristics of the k-Means Clustering Method**

| Clustering | Key Parameters | Properties | Advantages | Disadvantages |
|---|---|---|---|---|
| **k-Means** | • Data points<br>• Number of clusters<br>• Distance metric<br>• Max number of iterations | • Must know the number of clusters before hand<br>• There is always at least one object in each cluster.<br>• The clusters are non-hierarchical<br>• Clusters do not overlap<br>• Results depend on initial choice for centers<br>• Every member of a cluster is closer to its cluster than any other cluster<br>• k-Means method is numerical, unsupervised, non-deterministic, and iterative.<br>• Should be run several times with different starting points | • Simplicity<br>• Efficiency<br>• Reasonably fast<br>• It is used as initial process in many other algorithms.<br>• With a large number of variables, k-Means may be computationally faster than hierarchical clustering (if *K* is small).<br>• k-Means may produce tighter clusters than hierarchical clustering | • *K* must be provided<br>• Difficulty in comparing quality of the clusters produced (e.g. different initial partitions or values of *K* affect outcome).<br>• Different initial partitions can result in different final clusters.<br>• It does not work on categorical data<br>• Not time-efficient |

**Table 14: Main Characteristics of the FCM Clustering Method**

| Clustering | Key Parameters | Properties | Advantages | Disadvantages |
|---|---|---|---|---|
| **FCM** | • Data points<br>• Number of clusters<br>• Fuzziness Parameter<br>• Distance metric<br>• Max number of iterations | • A object can be partially in a cluster<br>• A object can be in more than one cluster<br>• Total Cluster Memberships Sum to 1<br>• Hard clusters (all or nothing) are a special case of Fuzzy Clusters | • More generalized form of Hard Clusters. (e.g. k-Means)<br>• Handles data that are equidistant to multiple clusters<br>• Handles data that is not compact and well separated<br>• Easy to implement for programmers<br>• Simple to use (but not as simple as hard clustering) | • Can be slow to converge<br>• You should know the value of 'm' (Fuzziness Parameter)<br>• More suited to data that is more or less evenly distributed around the cluster centers<br>• Hard to cluster the very closed clusters together without the help of other mechanisms such as elimination of small clusters. |

**Table 15: Main Characteristics of the PAM Clustering Method**

| Clustering | Key Parameters | Properties | Advantages | Disadvantages |
|---|---|---|---|---|
| **PAM** | • Data points<br>• Number of clusters<br>• Distance metric | • Shares the same properties of k-Means<br>• Cluster each point based on the closest center<br>• Replace each center by the medoid of points in its cluster<br>• Centers are located among the data points themselves | • It is more robust than k-Means in the presence of noise and outliers. Because a medoid is less influenced by outliers or other extreme values than a mean<br>• It minimizes a sum of pairwise dissimilarities instead of a sum of squared Euclidean distances | • Computationally harder than k-Means<br>• It does not scale well to large datasets because of its computational complexity |

### 3.5.2 Majority Voting

In the first three experiment, a simple majority voting rule was applied to the individual *$d_i(k)$* data after applying either the k-Means, FCM or PAM clustering methods. The purpose of applying the majority voting rule was to further discriminate among the different modes of transportation.

For example, assume a run was conducted with a bicycle in the data collection system composed of three Bluetooth DCUs depicted in Figure 7. After the k-Means clustering method was applied to the *$d_i(k)$* data (i.e., the data collected from Bluetooth DCUs, $i$=1,2,3), this run could have been identified as a "Bicycle," "Bicycle," and "Pedestrian." By applying the majority voting rule to these data, this experimental run would be correctly identified as a "Bicycle" run.

# 4.0 RESULTS

This chapter presents the results of analyzing the media access control (MAC) address data collected from Bluetooth-enabled devices with the four data collection experiments performed in this research.

The rest of the chapter is organized as follows. Section 4.1 describes the data processing and analysis procedures followed in the first three data collection experiments, as well as the actual results obtained. Section 4.2 describes the data processing and analysis procedures for the fourth data collection experiment and its results.

## 4.1 DATA PROCESSING AND ANALYSIS PROCEDURES FOR THE FIRST THREE DATA COLLECTION EXPERIMENTS

The data collected by each Bluetooth DCU during the first three data collection experiments were saved in individual text files. Each of these text files contained thousands of records formatted as shown in Figure 4. Therefore, the first step was to copy all these data into separate Microsoft® Excel spreadsheets for further processing. The different elements of the data records collected by the Bluetooth DCUs were split into individual columns in the Excel spreadsheet, as shown in Table 16.

**Table 16: Example of Data Saved in an Excel File**

| Bluetooth-Enabled Device MAC Address | Date and Time | RSSI | Bluetooth DCU MAC Address |
|---|---|---|---|
| 7E:AC:1E:96 | 2013-11-25 15:25:00.276670 | -79 | 0:1:95:17:C9:A3 |
| 1C:13:32:AD | 2013-11-25 15:25:02.399782 | -86 | 0:1:95:17:C9:A3 |
| 8F:1D:CC:A3 | 2013-11-25 15:25:03.383903 | -83 | 0:1:95:17:C9:A3 |
| 81:8A:6F:36 | 2013-11-25 15:25:04.499770 | -83 | 0:1:95:17:C9:A3 |
| 7E:AC:1E:96 | 2013-11-25 15:25:05.881908 | -79 | 0:1:95:17:C9:A3 |
| 81:8A:6F:36 | 2013-11-25 15:25:06.873935 | -75 | 0:1:95:17:C9:A3 |

Next, data records were sorted based on the column “Date and Time” and then based on the column “Bluetooth-Enabled Device MAC Address.” Since several of the experimental runs conducted in the first three experiments utilized the same Bluetooth-enabled device (i.e., cell phone), groups of MAC addresses that corresponded to an individual run could not be identified using just a MAC address.

Instead, the group of MAC address records that corresponded to individual experimental runs performed by a mode of transportation were identified based on a time gap of at least 30 seconds between detections. The MAC address data shown in Table 17 is a representative example of how the 30-second gap was utilized.

The eight records shown in Table 17 belong to two separate experimental runs. However, since the MAC address shown in the column labeled “Bluetooth Enabled MAC Address” is the same, the corresponding groups of four MAC addressed per run cannot be distinguished. Therefore, the data shown in the column labeled “Date and Time” were used instead. Notice that records 4 and 5 are separated by a 142-second difference, thus indicating that the first four records should be assigned to the first run and the next four records (starting at time “12:55:26”) belong to a separate run.

**Table 17: Example of Group of Data Separated for Each Run**

| Record No. | Bluetooth-Enabled MAC Address | Date and Time | RSSI | Bluetooth DCU MAC Address |
|---|---|---|---|---|
| 1 | 7F:C9:F0:13 | 12:52:34 | -80 | 0:1:95:17:C9:7A |
| 2 | 7F:C9:F0:13 | 12:52:34 | -72 | 0:1:95:17:C9:7A |
| 3 | 7F:C9:F0:13 | 12:52:46 | -82 | 0:1:95:17:C9:7A |
| 4 | 7F:C9:F0:13 | 12:53:04 | -81 | 0:1:95:17:C9:7A |
| 5 | 7F:C9:F0:13 | 12:55:26 | -82 | 0:1:95:17:C9:7A |
| 6 | 7F:C9:F0:13 | 12:55:27 | -88 | 0:1:95:17:C9:7A |
| 7 | 7F:C9:F0:13 | 12:55:35 | -87 | 0:1:95:17:C9:7A |
| 8 | 7F:C9:F0:13 | 12:55:41 | -82 | 0:1:95:17:C9:7A |

Once the beginning and end of individual runs for each mode of transportation were defined, the statistics $d_i(k)$ and $d(k)$ were calculated using equations (1) and (2) (see section 3.4).

The implementations of the clustering methods k-Means, Fuzzy c-Means (FCM) and Partitioning Around Medoids (PAM) available in the R statistical software package (R Core Team, 2012) were applied to the $d_i(k)$ and $d(k)$ data to identify different transportation modes. The majority voting method was applied only to the clustering results obtained with the $d_i(k)$ data to further discriminate among the different modes of transportation.

A value of $K$=3 was used with all clustering methods. A value of $p$=2 was used with the FCM algorithm. The clustering methods were executed until the following two conditions were met:

- There was no change in the calculated centroids (or medoids), and
- The minimum possible value for the sum of squared errors (SSE) was obtained.

### 4.1.1 Results from the First Data Collection Experiment

The results of clustering the $d_i(k)$ and $d(k)$ data with the k-Means, FCM, and PAM algorithms for the first data collection experiment are shown in Table 18, Table 19 and Table 20, respectively. In all tables, column 1 identifies the MAC addresses captured during the experiment, whereas column 2 codes the type of experimental run conducted for a mode of transportation. The letters "P", "B" and "C" identify a mode of transportation (i.e., pedestrian, bicycle or car); the letter "t" identifies a through pass and the letter "s" identifies a run where the mode of transportation stopped at point "A." Columns 3, 4, and 5 show the total duration of travel for a mode (in seconds) as seen by each of the three Bluetooth DCUs that made up the Bluetooth data collection system in the first data collection experiment.

Columns 6, 7, and 8 in Table 18 show which cluster each run was assigned to after applying the k-Means algorithm to the $d_i(k)$ data. The results show that a run was clustered incorrectly (i.e., shaded cells) eight times by the k-Means algorithm. Column 9 shows that after applying majority voting to the clustering results of columns 6, 7, and 8, only two clustering errors remain. The total duration of travel for a mode (in seconds) for all Bluetooth DCUs (i.e., the $d(k)$ data) is shown in column 10. Column 11 shows that no clustering errors exist after applying the k-Means clustering method to these data.

Columns 6, 7, and 8 in Table 19 show which cluster each run was assigned to after applying the FCM algorithm to the $d_i(k)$ data. The results show that a run was clustered incorrectly (i.e., shaded cells) seven times by the FCM algorithm. Column 9 shows that after applying majority voting to the clustering results of columns 6, 7, and 8, only one clustering error remain. The total duration of travel for a mode (in seconds) for all Bluetooth DCUs (i.e., the *d(k)* data) is shown in column 10. Column 11 shows that no clustering errors exist after applying the FCM clustering method to these data.

Columns 6, 7, and 8 in Table 20 show which cluster each run was assigned to after applying the PAM algorithm to the $d_i(k)$ data. The results show that a run was clustered incorrectly (i.e., shaded cells) nine times by the PAM algorithm. Column 9 shows that after applying majority voting to the clustering results of columns 6, 7, and 8, two clustering errors remain. The total duration of travel for a mode (in seconds) for all Bluetooth DCUs (i.e., the *d(k)* data) is shown in column 10. Column 11 shows that no clustering errors exist after applying the PAM clustering method to these data.

**Table 18: k-Means and Majority Voting Results for the First Data Collection Experiment**

| | | | | | K-Means | | | | | K-Means |
|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | $d_1(k)$ | $d_2(k)$ | $d_3(k)$ | $d_1(k)$ | $d_2(k)$ | $d_3(k)$ | **Voting** | $d(k)$ | $d(k)$ |
| | ***Ps*** | 93 | 149 | 58 | P | P | ***B*** | P | 156 | P |
| | ***Pt*** | 96 | 157 | 91 | P | P | P | P | 165 | P |
| | ***Pt*** | 83 | 117 | 80 | P | P | P | P | 165 | P |
| | ***Ps*** | 116 | 134 | 83 | P | P | P | P | 168 | P |
| | ***Bt*** | 56 | 47 | 32 | B | B | B | B | 61 | B |
| | ***Bs*** | 74 | 112 | 73 | ***P*** | ***P*** | ***P*** | ***P*** | 112 | B |
| 1 | ***Bs*** | 71 | 90 | 77 | ***P*** | B | ***P*** | ***P*** | 93 | B |
| | ***Bt*** | 50 | 53 | 44 | B | B | B | B | 64 | B |
| | ***Ct*** | 4 | 21 | 6 | C | C | C | C | 25 | C |
| | ***Cs*** | 21 | 33 | 12 | C | C | C | C | 41 | C |
| | ***Ct*** | 3 | 14 | 7 | C | C | C | C | 25 | C |
| | ***Cs*** | 22 | 30 | 10 | C | C | C | C | 36 | C |
| | ***Ct*** | 4 | 15 | 12 | C | C | C | C | 22 | C |
| | ***Ps*** | 89 | 134 | 91 | P | P | P | P | 152 | P |
| | ***Pt*** | 74 | 130 | 69 | P | P | P | P | 149 | P |
| | ***Pt*** | 112 | 145 | 89 | P | P | P | P | 162 | P |
| | ***Ps*** | 91 | 152 | 113 | P | P | P | P | 173 | P |
| | ***Bt*** | 38 | 58 | 48 | B | B | B | B | 74 | B |
| | ***Bs*** | 38 | 63 | 75 | B | B | ***P*** | B | 97 | B |
| 2 | ***Bs*** | 60 | 66 | 68 | B | B | ***P*** | B | 72 | B |
| | ***Bt*** | 46 | 48 | 50 | B | B | B | B | 68 | B |
| | ***Ct*** | 11 | 13 | 9 | C | C | C | C | 21 | C |
| | ***Cs*** | 23 | 17 | 6 | C | C | C | C | 36 | C |
| | ***Cs*** | 29 | 9 | 14 | C | C | C | C | 31 | C |
| | ***Ct*** | 8 | 5 | 10 | C | C | C | C | 24 | C |
| | ***Ct*** | 10 | 17 | 10 | C | C | C | C | 22 | C |

**Mode of Transportation Codes:**
***P***: Pedestrian
***B***: Bicycle
***C***: Car

**Type of Run:**
***t**:* Through pass
***s**:* Mode stopped at "A"

**Table 19: FCM and Majority Voting Results for the First Data Collection Experiment**

| | | | | | FCM | | | | | FCM |
|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | $d_1(k)$ | $d_2(k)$ | $d_3(k)$ | $d_1(k)$ | $d_2(k)$ | $d_3(k)$ | **Voting** | $d(k)$ | $d(k)$ |
| 1 | ***Ps*** | 93 | 149 | 58 | P | P | ***B*** | P | 156 | P |
| | ***Pt*** | 96 | 157 | 91 | P | P | P | P | 165 | P |
| | ***Pt*** | 83 | 117 | 80 | P | P | P | P | 165 | P |
| | ***Ps*** | 116 | 134 | 83 | P | P | P | P | 168 | P |
| | ***Bt*** | 56 | 47 | 32 | B | B | B | B | 61 | B |
| | ***Bs*** | 74 | 112 | 73 | ***P*** | ***P*** | ***P*** | ***P*** | 112 | B |
| | ***Bs*** | 71 | 90 | 77 | B | B | ***P*** | B | 93 | B |
| | ***Bt*** | 50 | 53 | 44 | B | B | B | B | 64 | B |
| | ***Ct*** | 4 | 21 | 6 | C | C | C | C | 25 | C |
| | ***Cs*** | 21 | 33 | 12 | C | C | C | C | 41 | C |
| | ***Ct*** | 3 | 14 | 7 | C | C | C | C | 25 | C |
| | ***Cs*** | 22 | 30 | 10 | C | C | C | C | 36 | C |
| | ***Ct*** | 4 | 15 | 12 | C | C | C | C | 22 | C |
| 2 | ***Ps*** | 89 | 134 | 91 | P | P | P | P | 152 | P |
| | ***Pt*** | 74 | 130 | 69 | P | P | P | P | 149 | P |
| | ***Pt*** | 112 | 145 | 89 | P | P | P | P | 162 | P |
| | ***Ps*** | 91 | 152 | 113 | P | P | P | P | 173 | P |
| | ***Bt*** | 38 | 58 | 48 | B | B | B | B | 74 | B |
| | ***Bs*** | 38 | 63 | 75 | B | B | ***P*** | B | 97 | B |
| | ***Bs*** | 60 | 66 | 68 | B | B | ***P*** | B | 72 | B |
| | ***Bt*** | 46 | 48 | 50 | B | B | B | B | 68 | B |
| | ***Ct*** | 11 | 13 | 9 | C | C | C | C | 21 | C |
| | ***Cs*** | 23 | 17 | 6 | C | C | C | C | 36 | C |
| | ***Cs*** | 29 | 9 | 14 | C | C | C | C | 31 | C |
| | ***Ct*** | 8 | 5 | 10 | C | C | C | C | 24 | C |
| | ***Ct*** | 10 | 17 | 10 | C | C | C | C | 22 | C |

**Mode of Transportation Codes:**
***P***: Pedestrian
***B***: Bicycle
***C***: Car

**Type of Run:**
***t:*** Through pass
***s:*** Mode stopped at "A"

**Table 20: PAM and Majority Voting Results for the First Data Collection Experiment**

| | | | | | PAM | | | | | PAM |
|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | **Voting** | ***d(k)*** | ***d(k)*** |
| 1 | ***Ps*** | 93 | 149 | 58 | P | P | ***B*** | P | 156 | P |
| | ***Pt*** | 96 | 157 | 91 | P | P | P | P | 165 | P |
| | ***Pt*** | 83 | 117 | 80 | P | P | P | P | 165 | P |
| | ***Ps*** | 116 | 134 | 83 | P | P | P | P | 168 | P |
| | ***Bt*** | 56 | 47 | 32 | B | B | B | B | 61 | B |
| | ***Bs*** | 74 | 112 | 73 | ***P*** | ***P*** | ***P*** | ***P*** | 112 | B |
| | ***Bs*** | 71 | 90 | 77 | ***P*** | B | ***P*** | ***P*** | 93 | B |
| | ***Bt*** | 50 | 53 | 44 | B | B | B | B | 64 | B |
| | ***Ct*** | 4 | 21 | 6 | C | C | C | C | 25 | C |
| | ***Cs*** | 21 | 33 | 12 | C | C | C | C | 41 | C |
| | ***Ct*** | 3 | 14 | 7 | C | C | C | C | 25 | C |
| | ***Cs*** | 22 | 30 | 10 | C | C | C | C | 36 | C |
| | ***Ct*** | 4 | 15 | 12 | C | C | C | C | 22 | C |
| 2 | ***Ps*** | 89 | 134 | 91 | P | P | P | P | 152 | P |
| | ***Pt*** | 74 | 130 | 69 | P | P | P | P | 149 | P |
| | ***Pt*** | 112 | 145 | 89 | P | P | P | P | 162 | P |
| | ***Ps*** | 91 | 152 | 113 | P | P | P | P | 173 | P |
| | ***Bt*** | 38 | 58 | 48 | B | B | B | B | 74 | B |
| | ***Bs*** | 38 | 63 | 75 | B | B | ***P*** | B | 97 | B |
| | ***Bs*** | 60 | 66 | 68 | B | B | ***P*** | B | 72 | B |
| | ***Bt*** | 46 | 48 | 50 | B | B | B | B | 68 | B |
| | ***Ct*** | 11 | 13 | 9 | C | C | C | C | 21 | C |
| | ***Cs*** | 23 | 17 | 6 | C | C | C | C | 36 | C |
| | ***Cs*** | 29 | 9 | 14 | ***B*** | C | C | C | 31 | C |
| | ***Ct*** | 8 | 5 | 10 | C | C | C | C | 24 | C |
| | ***Ct*** | 10 | 17 | 10 | C | C | C | C | 22 | C |

**Mode of Transportation Codes:**
***P***: Pedestrian
***B***: Bicycle
***C***: Car

**Type of Run:**
***t:*** Through pass
***s:*** Mode stopped at "A"

### 4.1.2 Results from the Second Data Collection Experiment

Table 21, Table 22, and Table 23 show the results of clustering the $d_i(k)$ and $d(k)$ data with the k-Means, FCM, and PAM algorithms, respectively, for the second data collection experiment. In all tables, column 1 identifies the MAC addresses captured during the experiment, whereas column 2 codes the type of experimental run conducted for a mode of transportation. The letters "P", "B" and "C" identify a mode of transportation (i.e., pedestrian, bicycle or car). The letter "t" identifies a through pass; the letters "a" and "b" identify runs where the mode of transportation stopped at either point "A" or "B." Columns 3 through 7 show the total duration of travel for a mode (in seconds) as seen by each of the five Bluetooth DCUs that made up the Bluetooth data collection system in the second data collection experiment.

Columns 8 through 12 in Table 21 show which cluster each run was assigned to after applying the k-Means algorithm to the $d_i(k)$ data. The results show that a run was clustered incorrectly (i.e., shaded cells) 11 times by the k-Means algorithm. Column 13 shows that after applying majority voting to the clustering results of columns 8 through 12, only two clustering errors remain. The total duration of travel for a mode (in seconds) for all Bluetooth DCUs (i.e., the $d(k)$ data) is shown in column 14. Column 15 shows that no clustering errors exist after applying the k-Means clustering method to these data.

Columns 8 through 12 in Table 22 show which cluster each run was assigned to after applying the FCM algorithm to the $d_i(k)$ data. The results show that a run was clustered incorrectly (i.e., shaded cells) 12 times by the FCM algorithm. Column 13 shows that after applying majority voting to the clustering results of columns 8 through 12, only two clustering errors remain. The total duration of travel for a mode (in seconds) for all Bluetooth DCUs (i.e., the $d(k)$ data) is shown in column 14. Column 15 shows that no clustering errors exist after applying the FCM clustering method to these data.

Columns 8 through 12 in Table 23 show which cluster each run was assigned to after applying the PAM algorithm to the $d_i(k)$ data. The results show that a run was clustered incorrectly (i.e., shaded cells) 11 times by the PAM algorithm. Column 13 shows that after applying majority voting to the clustering results of columns 8 through 12, only two clustering errors remain. The total duration of travel for a mode (in seconds) for all Bluetooth DCUs (i.e., the $d(k)$ data) is shown in column 14. Column 15 shows that no clustering errors exist after applying the PAM clustering method to these data.

**Table 21: k-Means and Majority Voting Results for the Second Data Collection Experiment**

| | | | | | | | K-Means | | | | | | | K-Means |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | ***$d_1(k)$*** | ***$d_2(k)$*** | ***$d_3(k)$*** | ***$d_4(k)$*** | ***$d_5(k)$*** | ***$d_1(k)$*** | ***$d_2(k)$*** | ***$d_3(k)$*** | ***$d_4(k)$*** | ***$d_5(k)$*** | ***Voting*** | ***$d(k)$*** | ***$d(k)$*** |
| 1 | ***Pa*** | 113 | 123 | 136 | 125 | 132 | P | P | P | P | P | P | 156 | P |
| | ***Pb*** | 102 | 159 | 116 | 152 | 148 | P | P | P | P | P | P | 160 | P |
| | ***Pt*** | 69 | 99 | 109 | 113 | 126 | ***B*** | ***B*** | P | P | P | P | 131 | P |
| | ***Ba*** | 73 | 79 | 79 | 84 | 93 | B | B | B | B | B | B | 96 | B |
| | ***Bb*** | 76 | 80 | 86 | 84 | 99 | B | B | B | B | B | B | 101 | B |
| | ***Bt*** | 40 | 46 | 71 | 70 | 59 | ***C*** | ***C*** | B | B | ***C*** | ***C*** | 72 | B |
| | ***Ca*** | 22 | 17 | 28 | 22 | 34 | C | C | C | C | C | C | 37 | C |
| | ***Cb*** | 33 | 38 | 23 | 24 | 25 | C | C | C | C | C | C | 42 | C |
| | ***Ct*** | 10 | 16 | 21 | 9 | 16 | C | C | C | C | C | C | 25 | C |
| 2 | ***Pt*** | 84 | 93 | 84 | 102 | 106 | ***B*** | ***B*** | ***B*** | P | ***B*** | ***B*** | 125 | P |
| | ***Pa*** | 101 | 138 | 139 | 123 | 139 | P | P | P | P | P | P | 153 | P |
| | ***Pb*** | 151 | 139 | 122 | 154 | 110 | P | P | P | P | ***B*** | P | 158 | P |
| | ***Bt*** | 49 | 49 | 70 | 54 | 70 | B | ***C*** | B | B | B | B | 75 | B |
| | ***Ba*** | 51 | 63 | 60 | 50 | 77 | B | B | B | B | B | B | 77 | B |
| | ***Bb*** | 65 | 66 | 63 | 68 | 91 | B | B | B | B | B | B | 91 | B |
| | ***Ca*** | 29 | 32 | 30 | 24 | 40 | C | C | C | C | C | C | 40 | C |
| | ***Ct*** | 12 | 26 | 29 | 6 | 24 | C | C | C | C | C | C | 40 | C |
| | ***Cb*** | 28 | 16 | 39 | 22 | 49 | C | C | C | C | C | C | 49 | C |

**Mode of Transportation Codes:**
***P***: Pedestrian
***B***: Bicycle
***C***: Car

**Type of Run:**
***t****:* Through pass
***a****:* Mode stopped at "A"
***b****:* Mode stopped at "B"

**Table 22: FCM and Majority Voting Results for the Second Data Collection Experiment**

| | | | | | | | FCM | | | | | | | FCM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | $d_1(k)$ | $d_2(k)$ | $d_3(k)$ | $d_4(k)$ | $d_5(k)$ | $d_1(k)$ | $d_2(k)$ | $d_3(k)$ | $d_4(k)$ | $d_5(k)$ | ***Voting*** | $d(k)$ | $d(k)$ |
| 1 | ***Pa*** | 113 | 123 | 136 | 125 | 132 | P | P | P | P | P | P | 156 | P |
| | ***Pb*** | 102 | 159 | 116 | 152 | 148 | P | P | P | P | P | P | 160 | P |
| | ***Pt*** | 69 | 99 | 109 | 113 | 126 | ***B*** | ***B*** | P | P | P | P | 131 | P |
| | ***Ba*** | 73 | 79 | 79 | 84 | 93 | B | B | B | B | B | B | 96 | B |
| | ***Bb*** | 76 | 80 | 86 | 84 | 99 | B | B | B | B | B | B | 101 | B |
| | ***Bt*** | 40 | 46 | 71 | 70 | 59 | ***C*** | ***C*** | B | B | ***C*** | ***C*** | 72 | B |
| | ***Ca*** | 22 | 17 | 28 | 22 | 34 | C | C | C | C | C | C | 37 | C |
| | ***Cb*** | 33 | 38 | 23 | 24 | 25 | C | C | C | C | C | C | 42 | C |
| | ***Ct*** | 10 | 16 | 21 | 9 | 16 | C | C | C | C | C | C | 25 | C |
| 2 | ***Pt*** | 84 | 93 | 84 | 102 | 106 | ***B*** | ***B*** | ***B*** | ***B*** | ***B*** | ***B*** | 125 | P |
| | ***Pa*** | 101 | 138 | 139 | 123 | 139 | P | P | P | P | P | P | 153 | P |
| | ***Pb*** | 151 | 139 | 122 | 154 | 110 | P | P | P | P | ***B*** | P | 158 | P |
| | ***Bt*** | 49 | 49 | 70 | 54 | 70 | B | ***C*** | B | B | B | B | 75 | B |
| | ***Ba*** | 51 | 63 | 60 | 50 | 77 | B | B | B | B | B | B | 77 | B |
| | ***Bb*** | 65 | 66 | 63 | 68 | 91 | B | B | B | B | B | B | 91 | B |
| | ***Ca*** | 29 | 32 | 30 | 24 | 40 | C | C | C | C | C | C | 40 | C |
| | ***Ct*** | 12 | 26 | 29 | 6 | 24 | C | C | C | C | C | C | 40 | C |
| | ***Cb*** | 28 | 16 | 39 | 22 | 49 | C | C | C | C | C | C | 49 | C |

**Mode of Transportation Codes:**
***P***: Pedestrian
***B***: Bicycle
***C***: Car

**Type of Run:**
***t**:* Through pass
***a**:* Mode stopped at "A"
***b**:* Mode stopped at "B"

**Table 23: PAM and Majority Voting for the Second Data Collection Experiment**

| | | | | | | | PAM | | | | | | | PAM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | ***d₄(k)*** | ***d₅(k)*** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | ***d₄(k)*** | ***d₅(k)*** | ***Voting*** | ***d(k)*** | ***d(k)*** |
| | ***Pa*** | 113 | 123 | 136 | 125 | 132 | P | P | P | P | P | P | 156 | P |
| | ***Pb*** | 102 | 159 | 116 | 152 | 148 | P | P | P | P | P | P | 160 | P |
| | ***Pt*** | 69 | 99 | 109 | 113 | 126 | ***B*** | ***B*** | P | P | P | P | 131 | P |
| | ***Ba*** | 73 | 79 | 79 | 84 | 93 | B | B | B | B | B | B | 96 | B |
| 1 | ***Bb*** | 76 | 80 | 86 | 84 | 99 | B | B | B | B | B | B | 101 | B |
| | ***Bt*** | 40 | 46 | 71 | 70 | 59 | ***C*** | ***C*** | B | B | ***C*** | ***C*** | 72 | B |
| | ***Ca*** | 22 | 17 | 28 | 22 | 34 | C | C | C | C | C | C | 37 | C |
| | ***Cb*** | 33 | 38 | 23 | 24 | 25 | C | C | C | C | C | C | 42 | C |
| | ***Ct*** | 10 | 16 | 21 | 9 | 16 | C | C | C | C | C | C | 25 | C |
| | ***Pt*** | 84 | 93 | 84 | 102 | 106 | ***B*** | ***B*** | ***B*** | P | ***B*** | ***B*** | 125 | P |
| | ***Pa*** | 101 | 138 | 139 | 123 | 139 | P | P | P | P | P | P | 153 | P |
| | ***Pb*** | 151 | 139 | 122 | 154 | 110 | P | P | P | P | ***B*** | P | 158 | P |
| | ***Bt*** | 49 | 49 | 70 | 54 | 70 | B | ***C*** | B | B | B | B | 75 | B |
| 2 | ***Ba*** | 51 | 63 | 60 | 50 | 77 | B | B | B | B | B | B | 77 | B |
| | ***Bb*** | 65 | 66 | 63 | 68 | 91 | B | B | B | B | B | B | 91 | B |
| | ***Ca*** | 29 | 32 | 30 | 24 | 40 | C | C | C | C | C | C | 40 | C |
| | ***Ct*** | 12 | 26 | 29 | 6 | 24 | C | C | C | C | C | C | 40 | C |
| | ***Cb*** | 28 | 16 | 39 | 22 | 49 | C | C | C | C | C | C | 49 | C |

**Mode of Transportation Codes:**
***P***: Pedestrian
***B***: Bicycle
***C***: Car

**Type of Run:**
***t****:* Through pass
***a****:* Mode stopped at "A"
***b****:* Mode stopped at "B"

### 4.1.3 Results from the Third Data Collection Experiment

Table 24, Table 25, and Table 26 show the results of clustering the $d_i(k)$ and $d(k)$ data with the k-Means, FCM, and PAM algorithms, respectively, for the third data collection experiment. A total of 60 runs were conducted (i.e., 30 runs in each direction). In all tables, column 1 identifies the MAC addresses captured during the experiment, whereas column 2 codes the type of experimental run conducted for a mode of transportation. The letters "P", "B" and "C" identify a mode of transportation (i.e., pedestrian, bicycle or car). The letter "t" identifies a through pass; the letters "a" and "b" identify runs where the mode of transportation stopped at either point "A" or "B." Columns 3 through 7 show the total duration of travel for a mode (in seconds) as seen by each of the five Bluetooth DCUs that made up the Bluetooth data collection system in the third data collection experiment.

Columns 8 through 12 in Table 24 show which cluster each run was assigned to after applying the k-Means algorithm to the $d_i(k)$ data. The results show that a run was clustered incorrectly (i.e., shaded cells) 72 times by the k-Means algorithm. Column 13 shows that after applying majority voting to the clustering results of columns 8 through 12, 12 clustering errors remain. The total duration of travel for a mode (in seconds) for all Bluetooth DCUs (i.e., the $d(k)$ data) is shown in column 14. Column 15 shows that no clustering errors exist after applying the k-Means clustering method to these data.

Columns 8 through 12 in Table 25 show which cluster each run was assigned to after applying the FCM algorithm to the $d_i(k)$ data. The results show that a run was clustered incorrectly (i.e., shaded cells) 41 times by the FCM algorithm. Column 13 shows that after applying majority voting to the clustering results of columns 8 through 12, only four clustering errors remain. The total duration of travel for a mode (in seconds) for all Bluetooth DCUs (i.e., the $d(k)$ data) is shown in column 14. Column 15 shows that no clustering errors exist after applying the FCM clustering method to these data.

Columns 8 through 12 in Table 26 show which cluster each run was assigned to after applying the PAM algorithm to the $d_i(k)$ data. The results show that a run was clustered incorrectly (i.e., shaded cells) 22 times by the PAM algorithm. Column 13 shows that after applying majority voting to the clustering results of columns 8 through 12, only one clustering error remained. The total duration of travel for a mode (in seconds) for all Bluetooth DCUs (i.e., the $d(k)$ data) is shown in column 14. Column 15 shows that no clustering errors exist after applying the PAM clustering method to these data.

**Table 24: k-Means and Majority Voting Results for the Third Data Collection Experiment**

| | | | | | | | K-Means | | | | | | | K-Means |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | ***d₄(k)*** | ***d₅(k)*** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | ***d₄(k)*** | ***d₅(k)*** | ***Voting*** | ***d(k)*** | ***d(k)*** |
| 1 | ***Ct*** | 8 | 23 | 0 | 14 | 12 | C | C | C | C | C | C | 23 | C |
| | ***Ct*** | 2 | 0 | 0 | 9 | 19 | C | C | C | C | C | C | 19 | C |
| | ***Ct*** | 6 | 0 | 6 | 8 | 11 | C | C | C | C | C | C | 8 | C |
| | ***Ca*** | 19 | 22 | 29 | 21 | 28 | C | C | C | C | C | C | 30 | C |
| | ***Ca*** | 17 | 0 | 24 | 21 | 23 | C | C | C | C | C | C | 25 | C |
| | ***Pb*** | 174 | 176 | 170 | 135 | 170 | P | P | P | ***B*** | P | P | 177 | P |
| | ***Pa*** | 114 | 102 | 119 | 108 | 194 | P | ***B*** | P | ***B*** | P | P | 194 | P |
| | ***Ca*** | 6 | 18 | 21 | 19 | 21 | C | C | C | C | C | C | 24 | C |
| | ***Pt*** | 152 | 196 | 182 | 131 | 221 | P | P | P | ***B*** | P | P | 224 | P |
| | ***Pt*** | 118 | 154 | 163 | 154 | 166 | P | P | P | P | P | P | 174 | P |
| | ***Ct*** | 1 | 8 | 6 | 9 | 4 | C | C | C | C | C | C | 12 | C |
| | ***Cb*** | 12 | 17 | 14 | 22 | 22 | C | C | C | C | C | C | 26 | C |
| | ***Pt*** | 90 | 106 | 85 | 125 | 106 | ***B*** | ***B*** | ***B*** | ***B*** | ***B*** | ***B*** | 126 | P |
| | ***Ca*** | 17 | 21 | 26 | 24 | 23 | C | C | C | C | C | C | 32 | C |
| | ***Pa*** | 164 | 169 | 145 | 157 | 197 | P | P | P | P | P | P | 197 | P |
| | ***Pa*** | 157 | 165 | 170 | 157 | 170 | P | P | P | P | P | P | 183 | P |
| | ***Cb*** | 11 | 17 | 15 | 0 | 1 | C | C | C | C | C | C | 20 | C |
| | ***Ct*** | 0 | 18 | 4 | 0 | 5 | C | C | C | C | C | C | 20 | C |
| | ***Ca*** | 12 | 15 | 17 | 6 | 14 | C | C | C | C | C | C | 23 | C |
| | ***Ca*** | 23 | 23 | 26 | 29 | 18 | C | C | C | C | C | C | 29 | C |

| | | | | | | | K-Means | | | | | | | K-Means |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *k* | Pass Type | $d_1(k)$ | $d_2(k)$ | $d_3(k)$ | $d_4(k)$ | $d_5(k)$ | $d_1(k)$ | $d_2(k)$ | $d_3(k)$ | $d_4(k)$ | $d_5(k)$ | *Voting* | $d(k)$ | $d(k)$ |
| 1 | ***Ca*** | 20 | 11 | 12 | 20 | 23 | C | C | C | C | C | C | 27 | C |
| | ***Pt*** | 127 | 138 | 131 | 102 | 181 | P | P | P | ***B*** | P | P | 181 | P |
| | ***Pb*** | 132 | 139 | 142 | 221 | 172 | P | P | P | P | P | P | 222 | P |
| | ***Ct*** | 0 | 0 | 9 | 4 | 11 | C | C | C | C | C | C | 12 | C |
| | ***Pa*** | 130 | 123 | 158 | 112 | 133 | P | P | P | ***B*** | P | P | 158 | P |
| | ***Cb*** | 16 | 12 | 22 | 22 | 19 | C | C | C | C | C | C | 26 | C |
| | ***Ct*** | 0 | 5 | 2 | 9 | 7 | C | C | C | C | C | C | 11 | C |
| | ***Cb*** | 21 | 3 | 12 | 5 | 1 | C | C | C | C | C | C | 23 | C |
| | ***Ca*** | 22 | 17 | 10 | 18 | 9 | C | C | C | C | C | C | 27 | C |
| | ***Ct*** | 0 | 8 | 5 | 8 | 7 | C | C | C | C | C | C | 12 | C |
| 2 | ***Pt*** | 114 | 110 | 119 | 109 | 99 | P | ***B*** | P | ***B*** | ***B*** | ***B*** | 136 | P |
| | ***Pb*** | 143 | 0 | 72 | 111 | 124 | P | ***C*** | ***B*** | ***B*** | P | ***B*** | 147 | P |
| | ***Pb*** | 109 | 115 | 107 | 114 | 132 | P | ***B*** | P | ***B*** | P | P | 136 | P |
| | ***Pa*** | 139 | 142 | 140 | 162 | 182 | P | P | P | P | P | P | 182 | P |
| | ***Bt*** | 65 | 0 | 80 | 55 | 83 | B | ***C*** | B | ***C*** | B | B | 83 | B |
| | ***Bb*** | 85 | 93 | 68 | 90 | 90 | B | B | B | B | B | B | 93 | B |
| | ***Ba*** | 92 | 97 | 115 | 78 | 89 | B | B | ***P*** | B | B | B | 115 | B |
| | ***Ba*** | 66 | 50 | 59 | 42 | 64 | B | ***C*** | B | ***C*** | B | B | 69 | B |
| | ***Bb*** | 80 | 75 | 77 | 51 | 67 | B | B | B | ***C*** | B | B | 83 | B |
| | ***Bb*** | 78 | 68 | 87 | 88 | 87 | B | B | B | B | B | B | 93 | B |
| | ***Bt*** | 37 | 45 | 48 | 38 | 36 | ***C*** | ***C*** | B | ***C*** | ***C*** | ***C*** | 56 | B |
| | ***Bb*** | 44 | 53 | 40 | 59 | 59 | ***C*** | B | B | ***C*** | B | B | 64 | B |

| | | | | | | | K-Means | | | | | | | K-Means |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | ***$d_1(k)$*** | ***$d_2(k)$*** | ***$d_3(k)$*** | ***$d_4(k)$*** | ***$d_5(k)$*** | ***$d_1(k)$*** | ***$d_2(k)$*** | ***$d_3(k)$*** | ***$d_4(k)$*** | ***$d_5(k)$*** | ***Voting*** | ***$d(k)$*** | ***$d(k)$*** |
| 2 | ***Bt*** | 46 | 49 | 49 | 54 | 41 | ***C*** | ***C*** | B | ***C*** | ***C*** | ***C*** | 61 | B |
| | ***Bt*** | 40 | 41 | 45 | 42 | 42 | ***C*** | ***C*** | B | ***C*** | ***C*** | ***C*** | 52 | B |
| | ***Ba*** | 70 | 66 | 71 | 63 | 76 | B | B | B | ***C*** | B | B | 82 | B |
| | ***Bt*** | 24 | 29 | 46 | 53 | 55 | ***C*** | ***C*** | B | ***C*** | B | ***C*** | 62 | B |
| | ***Pb*** | 111 | 104 | 126 | 91 | 135 | P | ***B*** | P | ***B*** | P | P | 143 | P |
| | ***Pt*** | 122 | 111 | 120 | 103 | 128 | P | ***B*** | P | ***B*** | P | P | 146 | P |
| | ***Pt*** | 100 | 98 | 137 | 106 | 135 | ***B*** | ***B*** | P | ***B*** | P | ***B*** | 137 | P |
| | ***Pa*** | 129 | 132 | 142 | 89 | 121 | P | P | P | ***B*** | P | P | 156 | P |
| | ***Ba*** | 35 | 68 | 61 | 46 | 78 | ***C*** | B | B | ***C*** | B | B | 78 | B |
| | ***Bt*** | 41 | 44 | 78 | 27 | 53 | ***C*** | ***C*** | B | ***C*** | B | ***C*** | 78 | B |
| | ***Bb*** | 50 | 77 | 45 | 53 | 63 | B | B | B | ***C*** | B | B | 93 | B |
| | ***Bt*** | 55 | 29 | 50 | 50 | 56 | B | ***C*** | B | ***C*** | B | B | 58 | B |
| | ***Bt*** | 42 | 29 | 45 | 49 | 69 | ***C*** | ***C*** | B | ***C*** | B | ***C*** | 69 | B |
| | ***Bb*** | 64 | 60 | 58 | 55 | 72 | B | B | B | ***C*** | B | B | 72 | B |
| | ***Bb*** | 41 | 82 | 74 | 45 | 39 | ***C*** | B | B | ***C*** | ***C*** | ***C*** | 84 | B |
| | ***Pb*** | 85 | 142 | 119 | 107 | 129 | ***B*** | P | P | ***B*** | P | P | 142 | P |
| | ***Pt*** | 77 | 77 | 109 | 101 | 159 | ***B*** | ***B*** | P | ***B*** | P | ***B*** | 159 | P |
| | ***Bb*** | 55 | 64 | 65 | 42 | 63 | B | B | B | ***C*** | B | B | 72 | B |

**Mode of Transportation Codes:**

***P***: Pedestrian

***B***: Bicycle

***C***: Car

**Type of Run:**

***t****:* Through pass

***a****:* Mode stopped at "A"

***b****:* Mode stopped at "B"

**Table 25: FCM and Majority Voting for the Third Data Collection Experiment**

| | | | | | | | FCM | | | | | | | FCM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | ***d₄(k)*** | ***d₅(k)*** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | ***d₄(k)*** | ***d₅(k)*** | ***Voting*** | ***d(k)*** | ***d(k)*** |
| 1 | ***Ct*** | 8 | 23 | 0 | 14 | 12 | C | C | C | C | C | C | 23 | C |
| | ***Ct*** | 2 | 0 | 0 | 9 | 19 | C | C | C | C | C | C | 19 | C |
| | ***Ct*** | 6 | 0 | 6 | 8 | 11 | C | C | C | C | C | C | 8 | C |
| | ***Ca*** | 19 | 22 | 29 | 21 | 28 | C | C | C | C | C | C | 30 | C |
| | ***Ca*** | 17 | 0 | 24 | 21 | 23 | C | C | C | C | C | C | 25 | C |
| | ***Pb*** | 174 | 176 | 170 | 135 | 170 | P | P | P | P | P | P | 177 | P |
| | ***Pa*** | 114 | 102 | 119 | 108 | 194 | P | ***B*** | P | P | P | P | 194 | P |
| | ***Ca*** | 6 | 18 | 21 | 19 | 21 | C | C | C | C | C | C | 24 | C |
| | ***Pt*** | 152 | 196 | 182 | 131 | 221 | P | P | P | P | P | P | 224 | P |
| | ***Pt*** | 118 | 154 | 163 | 154 | 166 | P | P | P | P | P | P | 174 | P |
| | ***Ct*** | 1 | 8 | 6 | 9 | 4 | C | C | C | C | C | C | 12 | C |
| | ***Cb*** | 12 | 17 | 14 | 22 | 22 | C | C | C | C | C | C | 26 | C |
| | ***Pt*** | 90 | 106 | 85 | 125 | 106 | ***B*** | ***B*** | ***B*** | P | ***B*** | ***B*** | 126 | P |
| | ***Ca*** | 17 | 21 | 26 | 24 | 23 | C | C | C | C | C | C | 32 | C |
| | ***Pa*** | 164 | 169 | 145 | 157 | 197 | P | P | P | P | P | P | 197 | P |
| | ***Pa*** | 157 | 165 | 170 | 157 | 170 | P | P | P | P | P | P | 183 | P |
| | ***Cb*** | 11 | 17 | 15 | 0 | 1 | C | C | C | C | C | C | 20 | C |
| | ***Ct*** | 0 | 18 | 4 | 0 | 5 | C | C | C | C | C | C | 20 | C |
| | ***Ca*** | 12 | 15 | 17 | 6 | 14 | C | C | C | C | C | C | 23 | C |
| | ***Ca*** | 23 | 23 | 26 | 29 | 18 | C | C | C | C | C | C | 29 | C |
| | ***Ca*** | 20 | 11 | 12 | 20 | 23 | C | C | C | C | C | C | 27 | C |
| | ***Pt*** | 127 | 138 | 131 | 102 | 181 | P | P | P | P | P | P | 181 | P |
| | ***Pb*** | 132 | 139 | 142 | 221 | 172 | P | P | P | P | P | P | 222 | P |

| | | | | | | | FCM | | | | | | | FCM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | ***d₄(k)*** | ***d₅(k)*** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | ***d₄(k)*** | ***d₅(k)*** | ***Voting*** | ***d(k)*** | ***d(k)*** |
| 1 | ***Ct*** | 0 | 0 | 9 | 4 | 11 | C | C | C | C | C | C | 12 | C |
| | ***Pa*** | 130 | 123 | 158 | 112 | 133 | P | P | P | P | P | P | 158 | P |
| | ***Cb*** | 16 | 12 | 22 | 22 | 19 | C | C | C | C | C | C | 26 | C |
| | ***Ct*** | 0 | 5 | 2 | 9 | 7 | C | C | C | C | C | C | 11 | C |
| | ***Cb*** | 21 | 3 | 12 | 5 | 1 | C | C | C | C | C | C | 23 | C |
| | ***Ca*** | 22 | 17 | 10 | 18 | 9 | C | C | C | C | C | C | 27 | C |
| | ***Ct*** | 0 | 8 | 5 | 8 | 7 | C | C | C | C | C | C | 12 | C |
| 2 | ***Pt*** | 114 | 110 | 119 | 109 | 99 | P | ***B*** | P | P | ***B*** | P | 136 | P |
| | ***Pb*** | 143 | 0 | 72 | 111 | 124 | P | ***C*** | ***B*** | P | P | P | 147 | P |
| | ***Pb*** | 109 | 115 | 107 | 114 | 132 | P | ***B*** | P | P | P | P | 136 | P |
| | ***Pa*** | 139 | 142 | 140 | 162 | 182 | P | P | P | P | P | P | 182 | P |
| | ***Bt*** | 65 | 0 | 80 | 55 | 83 | B | ***C*** | B | B | B | B | 83 | B |
| | ***Bb*** | 85 | 93 | 68 | 90 | 90 | B | B | B | ***P*** | B | B | 93 | B |
| | ***Ba*** | 92 | 97 | 115 | 78 | 89 | B | B | ***P*** | B | B | B | 115 | B |
| | ***Ba*** | 66 | 50 | 59 | 42 | 64 | B | B | B | B | B | B | 69 | B |
| | ***Bb*** | 80 | 75 | 77 | 51 | 67 | B | B | B | B | B | B | 83 | B |
| | ***Bb*** | 78 | 68 | 87 | 88 | 87 | B | B | B | B | B | B | 93 | B |
| | ***Bt*** | 37 | 45 | 48 | 38 | 36 | ***C*** | ***C*** | B | B | ***C*** | ***C*** | 56 | B |
| | ***Bb*** | 44 | 53 | 40 | 59 | 59 | B | B | ***C*** | B | B | B | 64 | B |
| | ***Bt*** | 46 | 49 | 49 | 54 | 41 | B | B | B | B | ***C*** | B | 61 | B |
| | ***Bt*** | 40 | 41 | 45 | 42 | 42 | ***C*** | ***C*** | B | B | ***C*** | ***C*** | 52 | B |
| | ***Ba*** | 70 | 66 | 71 | 63 | 76 | B | B | B | B | B | B | 82 | B |
| | ***Bt*** | 24 | 29 | 46 | 53 | 55 | ***C*** | ***C*** | B | B | B | B | 62 | B |
| | ***Pb*** | 111 | 104 | 126 | 91 | 135 | P | ***B*** | P | P | P | P | 143 | P |
| | ***Pt*** | 122 | 111 | 120 | 103 | 128 | P | ***B*** | P | P | P | P | 146 | P |

| | | | | | | | FCM | | | | | | | FCM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | ***$d_1(k)$*** | ***$d_2(k)$*** | ***$d_3(k)$*** | ***$d_4(k)$*** | ***$d_5(k)$*** | ***$d_1(k)$*** | ***$d_2(k)$*** | ***$d_3(k)$*** | ***$d_4(k)$*** | ***$d_5(k)$*** | ***Voting*** | ***$d(k)$*** | ***$d(k)$*** |
| 2 | ***Pt*** | 100 | 98 | 137 | 106 | 135 | ***B*** | ***B*** | P | P | P | P | 137 | P |
| | ***Pa*** | 129 | 132 | 142 | 89 | 121 | P | P | P | ***B*** | ***B*** | P | 156 | P |
| | ***Ba*** | 35 | 68 | 61 | 46 | 78 | ***C*** | B | B | B | B | B | 78 | B |
| | ***Bt*** | 41 | 44 | 78 | 27 | 53 | ***C*** | ***C*** | B | ***C*** | B | ***C*** | 78 | B |
| | ***Bb*** | 50 | 77 | 45 | 53 | 63 | B | B | B | B | B | B | 93 | B |
| | ***Bt*** | 55 | 29 | 50 | 50 | 56 | B | ***C*** | B | B | B | B | 58 | B |
| | ***Bt*** | 42 | 29 | 45 | 49 | 69 | ***C*** | ***C*** | B | B | B | B | 69 | B |
| | ***Bb*** | 64 | 60 | 58 | 55 | 72 | B | B | B | B | B | B | 72 | B |
| | ***Bb*** | 41 | 82 | 74 | 45 | 39 | ***C*** | B | B | B | ***C*** | B | 84 | B |
| | ***Pb*** | 85 | 142 | 119 | 107 | 129 | ***B*** | P | P | P | P | P | 142 | P |
| | ***Pt*** | 77 | 77 | 109 | 101 | 159 | ***B*** | ***B*** | P | P | P | P | 159 | P |
| | ***Bb*** | 55 | 64 | 65 | 42 | 63 | B | B | B | B | B | B | 72 | B |

**Mode of Transportation Codes:**
***P***: Pedestrian
***B***: Bicycle
***C***: Car

**Type of Run:**
***t**:* Through pass
***a**:* Mode stopped at "A"
***b**:* Mode stopped at "B"

**Table 26: PAM and Majority Voting for the Third Data Collection Experiment**

| | | | | | | | PAM | | | | | | | PAM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | ***d₄(k)*** | ***d₅(k)*** | ***d₁(k)*** | ***d₂(k)*** | ***d₃(k)*** | ***d₄(k)*** | ***d₅(k)*** | ***Voting*** | ***d(k)*** | ***d(k)*** |
| 1 | ***Ct*** | 8 | 23 | 0 | 14 | 12 | C | C | C | C | C | C | 23 | C |
| | ***Ct*** | 2 | 0 | 0 | 9 | 19 | C | C | C | C | C | C | 19 | C |
| | ***Ct*** | 6 | 0 | 6 | 8 | 11 | C | C | C | C | C | C | 8 | C |
| | ***Ca*** | 19 | 22 | 29 | 21 | 28 | C | C | C | C | C | C | 30 | C |
| | ***Ca*** | 17 | 0 | 24 | 21 | 23 | C | C | C | C | C | C | 25 | C |
| | ***Pb*** | 174 | 176 | 170 | 135 | 170 | P | P | P | P | P | P | 177 | P |
| | ***Pa*** | 114 | 102 | 119 | 108 | 194 | P | ***B*** | P | P | P | P | 194 | P |
| | ***Ca*** | 6 | 18 | 21 | 19 | 21 | C | C | C | C | C | C | 24 | C |
| | ***Pt*** | 152 | 196 | 182 | 131 | 221 | P | P | P | P | P | P | 224 | P |
| | ***Pt*** | 118 | 154 | 163 | 154 | 166 | P | P | P | P | P | P | 174 | P |
| | ***Ct*** | 1 | 8 | 6 | 9 | 4 | C | C | C | C | C | C | 12 | C |
| | ***Cb*** | 12 | 17 | 14 | 22 | 22 | C | C | C | C | C | C | 26 | C |
| | ***Pt*** | 90 | 106 | 85 | 125 | 106 | ***B*** | P | ***B*** | P | ***B*** | ***B*** | 126 | P |
| | ***Ca*** | 17 | 21 | 26 | 24 | 23 | C | C | C | C | C | C | 32 | C |
| | ***Pa*** | 164 | 169 | 145 | 157 | 197 | P | P | P | P | P | P | 197 | P |
| | ***Pa*** | 157 | 165 | 170 | 157 | 170 | P | P | P | P | P | P | 183 | P |
| | ***Cb*** | 11 | 17 | 15 | 0 | 1 | C | C | C | C | C | C | 20 | C |
| | ***Ct*** | 0 | 18 | 4 | 0 | 5 | C | C | C | C | C | C | 20 | C |
| | ***Ca*** | 12 | 15 | 17 | 6 | 14 | C | C | C | C | C | C | 23 | C |
| | ***Ca*** | 23 | 23 | 26 | 29 | 18 | C | C | C | C | C | C | 29 | C |
| | ***Ca*** | 20 | 11 | 12 | 20 | 23 | C | C | C | C | C | C | 27 | C |
| | ***Pt*** | 127 | 138 | 131 | 102 | 181 | P | P | P | P | P | P | 181 | P |
| | ***Pb*** | 132 | 139 | 142 | 221 | 172 | P | P | P | P | P | P | 222 | P |

| | | | | | | | PAM | | | | | | | PAM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | ***$d_1(k)$*** | ***$d_2(k)$*** | ***$d_3(k)$*** | ***$d_4(k)$*** | ***$d_5(k)$*** | ***$d_1(k)$*** | ***$d_2(k)$*** | ***$d_3(k)$*** | ***$d_4(k)$*** | ***$d_5(k)$*** | ***Voting*** | ***$d(k)$*** | ***$d(k)$*** |
| 1 | ***Ct*** | 0 | 0 | 9 | 4 | 11 | C | C | C | C | C | C | 12 | C |
| | ***Pa*** | 130 | 123 | 158 | 112 | 133 | P | P | P | P | P | P | 158 | P |
| | ***Cb*** | 16 | 12 | 22 | 22 | 19 | C | C | C | C | C | C | 26 | C |
| | ***Ct*** | 0 | 5 | 2 | 9 | 7 | C | C | C | C | C | C | 11 | C |
| | ***Cb*** | 21 | 3 | 12 | 5 | 1 | C | C | C | C | C | C | 23 | C |
| | ***Ca*** | 22 | 17 | 10 | 18 | 9 | C | C | C | C | C | C | 27 | C |
| | ***Ct*** | 0 | 8 | 5 | 8 | 7 | C | C | C | C | C | C | 12 | C |
| 2 | ***Pt*** | 114 | 110 | 119 | 109 | 99 | P | P | P | P | P | P | 136 | P |
| | ***Pb*** | 143 | 0 | 72 | 111 | 124 | P | ***C*** | ***B*** | P | P | P | 147 | P |
| | ***Pb*** | 109 | 115 | 107 | 114 | 132 | P | P | P | P | P | P | 136 | P |
| | ***Pa*** | 139 | 142 | 140 | 162 | 182 | P | P | P | P | P | P | 182 | P |
| | ***Bt*** | 65 | 0 | 80 | 55 | 83 | B | ***C*** | B | B | B | B | 83 | B |
| | ***Bb*** | 85 | 93 | 68 | 90 | 90 | B | B | B | ***P*** | B | B | 93 | B |
| | ***Ba*** | 92 | 97 | 115 | 78 | 89 | B | B | ***P*** | B | B | B | 115 | B |
| | ***Ba*** | 66 | 50 | 59 | 42 | 64 | B | B | B | B | B | B | 69 | B |
| | ***Bb*** | 80 | 75 | 77 | 51 | 67 | B | B | B | B | B | B | 83 | B |
| | ***Bb*** | 78 | 68 | 87 | 88 | 87 | B | B | B | ***P*** | B | B | 93 | B |
| | ***Bt*** | 37 | 45 | 48 | 38 | 36 | ***C*** | B | B | B | ***C*** | B | 56 | B |
| | ***Bb*** | 44 | 53 | 40 | 59 | 59 | B | B | B | B | B | B | 64 | B |
| | ***Bt*** | 46 | 49 | 49 | 54 | 41 | B | B | B | B | ***C*** | B | 61 | B |
| | ***Bt*** | 40 | 41 | 45 | 42 | 42 | B | ***C*** | B | B | ***C*** | B | 52 | B |
| | ***Ba*** | 70 | 66 | 71 | 63 | 76 | B | B | B | B | B | B | 82 | B |
| | ***Bt*** | 24 | 29 | 46 | 53 | 55 | ***C*** | ***C*** | B | B | B | B | 62 | B |
| | ***Pb*** | 111 | 104 | 126 | 91 | 135 | P | P | P | P | P | P | 143 | P |
| | ***Pt*** | 122 | 111 | 120 | 103 | 128 | P | P | P | P | P | P | 146 | P |

| | | | | | | | PAM | | | | | | | PAM |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ***k*** | **Pass Type** | ***$d_1(k)$*** | ***$d_2(k)$*** | ***$d_3(k)$*** | ***$d_4(k)$*** | ***$d_5(k)$*** | ***$d_1(k)$*** | ***$d_2(k)$*** | ***$d_3(k)$*** | ***$d_4(k)$*** | ***$d_5(k)$*** | ***Voting*** | ***$d(k)$*** | ***$d(k)$*** |
| 2 | ***Pt*** | 100 | 98 | 137 | 106 | 135 | P | P | P | P | P | P | 137 | P |
| | ***Pa*** | 129 | 132 | 142 | 89 | 121 | P | P | P | P | P | P | 156 | P |
| | ***Ba*** | 35 | 68 | 61 | 46 | 78 | ***C*** | B | B | B | B | B | 78 | B |
| | ***Bt*** | 41 | 44 | 78 | 27 | 53 | B | B | B | ***C*** | B | B | 78 | B |
| | ***Bb*** | 50 | 77 | 45 | 53 | 63 | B | B | B | B | B | B | 93 | B |
| | ***Bt*** | 55 | 29 | 50 | 50 | 56 | B | ***C*** | B | B | B | B | 58 | B |
| | ***Bt*** | 42 | 29 | 45 | 49 | 69 | B | ***C*** | B | B | B | B | 69 | B |
| | ***Bb*** | 64 | 60 | 58 | 55 | 72 | B | B | B | B | B | B | 72 | B |
| | ***Bb*** | 41 | 82 | 74 | 45 | 39 | B | B | B | B | ***C*** | B | 84 | B |
| | ***Pb*** | 85 | 142 | 119 | 107 | 129 | B | P | P | P | P | P | 142 | P |
| | ***Pt*** | 77 | 77 | 109 | 101 | 159 | B | B | P | P | P | P | 159 | P |
| | ***Bb*** | 55 | 64 | 65 | 42 | 63 | B | B | B | B | B | B | 72 | B |

**Mode of Transportation Codes:**
***P***: Pedestrian
***B***: Bicycle
***C***: Car

**Type of Run:**
***t**:* Through pass
***a**:* Mode stopped at "A"
***b**:* Mode stopped at "B"

### 4.1.4 Percentage Error of Data Analysis Methods

Table 27 shows the percentage error resulting from applying the different analysis methods to the $d_i(k)$ and $d(k)$ data in first three data collection experiments. These values were calculated by dividing the total number of observations clustered incorrectly by the total available observations. For example, the value 10.26% was calculated as [8/(3*26)].

As the results show, the lowest percentage error is obtained when the clustering methods are applied to the $d(k)$ data only. The percentage error that results from applying the clustering methods to the $d_i(k)$ data combined with majority voting are better than those obtained when applying the clustering methods to the $d_i(k)$ data only, but neither are as good as the $d(k)$ results.

**Table 27: Percentage Error for Different Analyses Methods**

| | ***$d_i(k)$ only*** | | | ***$d_i(k)$ + Voting*** | | | ***$d(k)$ only*** | | |
|---|---|---|---|---|---|---|---|---|---|
| **Experiment** | **K-Means** | **FCM** | **PAM** | **K-Means** | **FCM** | **PAM** | **K-Means** | **FCM** | **PAM** |
| **1** | 10.26% | 8.97% | 11.54% | 7.69% | 3.85% | 7.69% | 0% | 0% | 0% |
| **2** | 12.22% | 13.33% | 12.22% | 11.11% | 11.11% | 11.11% | 0% | 0% | 0% |
| **3** | 24.00% | 13.67% | 7.33% | 20.00% | 6.67% | 1.67% | 0% | 0% | 0% |

## 4.2 Data Processing and Analysis Procedures for the Fourth Data Collection Experiment

The data collected by each Bluetooth DCU during the fourth data collection experiment were saved in individual text files. As explained in section 4.1, the data files were first copied and sorted in a Microsoft® Excel spreadsheet and different elements of the data records were split into individual columns, as shown in Table 16.

As explained in section 3.3.4, the test area used in the fourth data collection experiment was divided into several sections (i.e., EW, SE, SW, NE, and NW), and each of these sections was represented by a group of Bluetooth DCUs (see Table 11). In order to identify the experimental runs that corresponded to each section, the data files from the Bluetooth DCUs at the two ends of a section were combined and sorted in a single Excel spreadsheet. For example, to identify the experimental runs that corresponded to section EW, Bluetooth DCUs “1” and “E” were combined into an Excel spreadsheet. The reason why DCU “1” was used instead of DCU “W” was because DCU “W” did not collect data due to an undefined malfunction. The criterion applied to these data was that those MAC address records that were collected by Bluetooth DCUs “1” and “E” within a reasonable time difference (i.e., less than five minutes) belonged to section EW. The same criterion was applied to all sections.

In first three experiments, the results showed that applying the clustering methods alone or the clustering methods combined with majority voting to the $d_i(k)$ data proved to be less effective than applying the clustering methods to the $d(k)$ data. Therefore, once the MAC address records for each section were identified, only the statistic $d(k)$ was calculated using equation (1) (see section 3.4). Next, the implementation of the clustering methods k-Means, FCM, and PAM available in the R statistical software package (R Core Team, 2012) were applied to the $d(k)$ data to identify different transportation modes. A value of $K$=3 was used with all clustering methods. A value of $p$=2 was used with the FCM algorithm. The clustering methods were executed until the following two conditions were met:

- There was no change in the calculated centroids (or medoids)
- The minimum possible value for the sum of squared errors (SSE) was obtained.

The video files recorded by three video cameras were saved in a laptop computer. These video files were reviewed several times and the different modes of transportation travelling through each section were counted visually. The totals counts obtained from the videos were compared to the results obtained with the clustering methods.

The results obtained with the clustering methods for the data collected for the "East-West" section were validated against the video data for that same section. The validation process involved matching the recorded data runs by the Bluetooth DCUs with the corresponding modes of transportation captured in the video files during the same time frame. For example, if a run captured by the Bluetooth DCUs from west to east between times 14:33 and 14:35 was designated as a "Pedestrian" run by the clustering methods, the video files were reviewed during the same time interval to verify that in fact a pedestrian appeared on the video.

### 4.2.1 Result from the Fourth Data Collection Experiment

The results for the fourth data collection experiment are summarized in the Table 28, Table 29, Table 30, and Table 31.

The total observations for each mode of transportation (i.e., pedestrians, bicycles, and vehicles) travelling in different sections during the data collection process that were captured with video cameras are shown in Table 28. These data represent the "ground truth" in the fourth data collection experiment. The total number of runs from the data collected by the Bluetooth DCUs for each section separately is shown in Table 29.

The results of clustering the *d(k)* data with the clustering methods k-Means, FCM, and PAM for each section are shown in Table 30. The percentages of detections are shown in Table 31. These values were calculated by dividing the number of observations for each mode and section clustered in Table 30 by the ground truth values shown in Table 28. For example, the value 5.88% was calculated as [(19/323)*100].

As it was mentioned in Section 4.2, the results of the clustering methods for the section “East-West” were validated against the video data. The results indicate that all the runs were matched perfectly with the video files.

**Table 28: Ground Truth Data Collected from Video**

| | E-W | S-E | S-W | N-E | N-W |
|---|---|---|---|---|---|
| **Cars** | 323 | 220 | 262 | 17 | 29 |
| **Bicyclists** | 72 | 55 | 17 | 1 | 7 |
| **Pedestrians** | 76 | 54 | 32 | 7 | 25 |

**Table 29: Total Number of Runs in each Section Collected by Bluetooth DCUs**

| Sections | Number of Runs |
|---|---|
| **E-W** | 26 |
| **S-E** | 16 |
| **SW** | 27 |
| **N-E** | 4 |
| **N-W** | 9 |

**Table 30: Clustering Results for Fourth Data Collection Experiment**

| | | K-Means | FCM | PAM |
|---|---|---|---|---|
| **E-W** | **Cars** | 19 | 19 | 19 |
| | **Bicyclists** | 6 | 6 | 6 |
| | **Pedestrians** | 1 | 1 | 1 |
| **S-E** | **Cars** | 9 | 9 | 9 |
| | **Bicyclists** | 3 | 3 | 3 |
| | **Pedestrians** | 4 | 4 | 4 |
| **S-W** | **Cars** | 14 | 14 | 14 |
| | **Bicyclists** | 9 | 9 | 9 |
| | **Pedestrians** | 4 | 4 | 4 |
| **N-E** | **Cars** | 2 | 2 | 2 |
| | **Bicyclists** | 1 | 1 | 1 |
| | **Pedestrians** | 1 | 1 | 1 |
| **N-W** | **Cars** | 2 | 5 | 5 |
| | **Bicyclists** | 3 | 2 | 2 |
| | **Pedestrians** | 4 | 2 | 2 |

**Table 31: Percentage of Detections for Fourth Data Collection Experiment**

| | | K-Means | FCM | PAM |
|---|---|---|---|---|
| **E-W** | **Cars** | 5.88% | 5.88% | 5.88% |
| | **Bicyclists** | 8.33% | 8.33% | 8.33% |
| | **Pedestrians** | 1.32% | 1.32% | 1.32% |
| **S-E** | **Cars** | 4.09% | 4.09% | 4.09% |
| | **Bicyclists** | 5.45% | 5.45% | 5.45% |
| | **Pedestrians** | 7.41% | 7.41% | 7.41% |
| **S-W** | **Cars** | 5.34% | 5.34% | 5.34% |
| | **Bicyclists** | 52.94% | 52.94% | 52.94% |
| | **Pedestrians** | 12.50% | 12.50% | 12.50% |
| **N-E** | **Cars** | 11.76% | 11.76% | 11.76% |
| | **Bicyclists** | 100.00% | 100.00% | 100.00% |
| | **Pedestrians** | 14.29% | 14.29% | 14.29% |
| **N-W** | **Cars** | 6.90% | 17.24% | 17.24% |
| | **Bicyclists** | 42.86% | 28.57% | 28.57% |
| | **Pedestrians** | 16.00% | 8.00% | 8.00% |

# 5.0 CONCLUSIONS AND OPORTUNITES FOR FUTURE WORK

In this research, the feasibility of utilizing the information contained in data collected by Bluetooth-based data collection units (DCU) to automatically identify different modes of transportation travelling through an intersection was investigated. To accomplish this objective, four experiments were conducted where Bluetooth DCUs collected time-stamped media access control (MAC) address data from Bluetooth-enabled devices. The first three data collection experiments were controlled and were performed in a parking lot; the last data collection experiment was conducted in a real four-way intersection controlled by traffic lights. The time-stamped MAC address data were used to derive an overall duration of travel (i.e., $d(k)$) for individual runs performed by the different modes of transportation, as well as a duration of travel for each individual Bluetooth DCU (i.e., $d_i(k)$). Three different clustering methods and a majority voting procedure were applied to the $d(k)$ and the $d_i(k)$ data to distinguish among the modes of transportation. Video data collected in the fourth experiment served as "ground truth" to validate the performance of the Bluetooth DCUs in a real setting.

The rest of this chapter is organized as follows. Section 5.1 presents the conclusions reached in this study and section 5.2 discusses the opportunities for future work.

## 5.1 RESEARCH CONCLUSIONS

The results obtained in the four experiments performed in this research demonstrate that a Bluetooth-based data collection system is a viable approach for identifying different modes of transportation (i.e., vehicles, bicyclists, and pedestrians).

At the onset of the study, it was anticipated that pedestrians and bicycles would be more challenging to distinguish than vehicles, especially under the conditions of the third and fourth data collection experiments. Although this expectation was confirmed, the application of the clustering methods k-Means, Fuzzy c-Means (FCM) and Partitioning Around Medoids (PAM) to the statistic $d(k)$ proved to be extremely accurate and surpassed all expectations. As discussed in section 4.2, a majority voting procedure was applied to the statistic $d_i(k)$, but the results were not as successful as when the clustering methods were applied to the $d(k)$ data.

One of the expected challenges was that data collection would prove more difficult in a dense urban environment where speeds tend to be more uniform due to traffic control devices, which makes the classification of the modes of transportation more prone to errors. This was the main purpose of conducting the fourth data collection experiment. Although it cannot be stated that the intersection used in the fourth data collection experiment is truly representative of all the designs currently in operation, or that it sees many different kinds of traffic flows for all the modes of transportation, the results obtained are encouraging and are still beneficial for transportation agencies interested in understanding the traffic patterns of different modes of transportation at intersections.

As conceived, the Bluetooth-based data collection system used in this research is portable and significantly less expensive relative to alternative technologies used for similar purposes (e.g., inductive loop detectors, license plate recognition, and video). On the other hand, it is also important to note that the data collected with this system is only a sample of the actual traffic volume. Previous studies have shown that the percentage of detections for vehicles is about 6% (Haseman et al., 2010). This research was conducted in conjunction with a project sponsored by the Oregon Department of Transportation (ODOT) whose main objective was to design, integrate and test a specific type of Bluetooth-based DCUs. As part of the task conducted in this project, a data collection experiment was conducted to confirm the 6% sampling rate. The percentage of detections calculated in the fourth data collection experiment for the “East-West” and “South-West” sections were close to 6% (i.e., 5.88% and 5.34%). For the other sections, this percentage was smaller. These lower percentages could be attributed to differences in the experiment setup, antenna height, or the position of the traffic drums relative to the road. The percentage of detections can be used to derive the number of vehicles that actually passed that road segment during the time interval of the data collection experiment.

## 5.2 OPPORTUNITIES FOR FUTURE WORK

Following are potential research opportunities that can extend the work performed in this study to gain a better understanding of how a Bluetooth-based data collection system can be used to identify different modes of transportation:

- Received signal strength indicator (RSSI) data were analyzed at the beginning of this research to assess their suitability to serve as an alternative or complementary performance measure to the duration of travel (i.e., $d(k)$ and $d_i(k)$). However, no conclusive results were obtained due the variability observed in the RSSI data. Thus, only the duration of travel data were used as statistics in the data analyses. An opportunity still exists to explore RSSI data patterns more thoroughly and potentially utilize the results to increase the accuracy when identifying different modes of transportation.
- The two main data elements collected from Bluetooth-enabled devices during the different experiments were a time-stamped media access control (MAC) address and an RSSI value. Previous research has shown that the class of device (CoD) can also be collected during the Bluetooth inquiry process (Namaki Araghi et al., 2012). The CoD data element can provide additional insight into which mode of transportation is carrying the Bluetooth-enabled device. For example, it would be very unlikely that either a cyclist or a pedestrian would operate a geographic

positioning system (GPS) based device while moving; this would be more indicative of a motor vehicle.

- The L-com 2.4 GHz 8dBi round patch antenna used for Bluetooth communications was tested several times and compared with other types of antennae in data collection experiments involving *only* vehicles. However, this antenna type was never compared with other types of antennae in the experiments involving all three modes of transportation. Therefore, trying different types of antennae to assess how it could affect the results could be explored.
- The experiments conducted in this research could be expanded to try different installation heights for the antenna of the Bluetooth-based DCUs. In this research, the L-com 2.4 GHz 8dBi round patch antennae were always positioned at the height of approximately 3 feet from the ground. Therefore, the effect of different antenna heights could be further investigated.
- Three clustering methods (i.e., k-Means, FCM, and PAM) were used for data analysis. There are still opportunities for exploring alternative clustering methods with different features and apply them to the data collected from Bluetooth-enabled devices. Additionally, experiments should be conducted at intersections where the traffic volume of one of the three modes of transportation (e.g., pedestrians) is very low. In such

a case, the clustering methods should be modified to only cluster the data into two groups (i.e., bicyclists and motor vehicles).

# BIBLIOGRAPHY

Bisdikian, C. (2001). An Overview of the Bluetooth Wireless Technology. *IEEE Communications Magazine*. *39*(12), 86-94.

Blogg, M., Semler, C., Hingorani, M., & Troutbeck, R. (2010). Travel Time and Origin-Destination Data Collection using Bluetooth MAC Address Readers. *Australasian Transport Research Forum 2010 Proceedings*. Canberra, Australia: ARRB Group Limited.

Manning, C., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge: Cambridge University Press.

Dai, C., Zhen, Y., & Li, X. (2007). Pedestrian Detection and Tracking in Infrared Imagery Using Shape and Appearance. *Computer Vision and Image Understanding Archive*, *106*( 2-3), 288-299.

Duflot, M., Kwiatkowska, M., Norman, G., & Parker, D. (2006) A Formal Analysis of Bluetooth Device Discovery. *International Journal on Software Tools for Technology Transfer (STTT)*, *8*(6), 621-632.

Griffin, G., Nordback, K., Götschi, T., Stolz, E., & Kothuri, S. (2014). Monitoring Bicyclist and Pedestrian Travel and Behavior: Current Research and Practice. *Transportation Research Circular E-C18*. Retrieved March 25, 2014, from http://www.trb.org/main/blurbs/170452.aspx.

Mirkes, E. (2011). K-Means and K-medoids applet. *University of Leicester*. Retrieved February 10, 2014, from http://www.math.le.ac.uk/people/ag153/homepage/KmeansKmedoids/Kmeans_Kmedoids.html.

Fresnedo, O., Iglesia, D., & Escudero, C. (2007). Bluetooth inquiry procedure: Optimization and influence of the number of devices. *Proceedings of the Sixth IASTED International Conference on Communication Systems and Networks* (pp. 205-209). Palma de Mallorca, Spain: ACTA Press.

Fujisawa, S., Hasegawa, G., Taniguchi, Y., & Nakano, H. (2013). Pedestrian Counting in Video Sequences Using Optical Flow Clustering. *ACA'12 Proceedings of the 11th International Conference on Applications of Electrical and Computer Engineering* (pp. 51-56). Athens, Greece: WSEAS Press.

Google Nexus, 2010, Retrieved February 10, 2014, from http://www.google.com/nexus/.

Haghani, A., Hamedi, M., Farokhi Sadabadi, K., Young, S., & Tarnoff, P. (2010). Data Collection of Freeway Travel Time Ground Truth with Bluetooth Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2160, 60–68.

Haseman, R., Wasson, J., & Bullock, D. (2010). Real-Time Measurement of Travel Time Delay in Work Zones and Evaluation Metrics Using Bluetooth Probe Tracking. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2159, 40–53.

Jain, A. (2010). Data Clustering: 50 Years Beyond K-Means. *Proceeding of the 19th International Conference on Pattern Recognition* (pp. 651-666). Tampa, FL: IEEE.

Kim, D.S., Porter, J.D., Magana, M.E., Park, S., and Saeedi, A. (2012) Wireless Data Collection System for Travel Time Estimation and Traffic Performance Evaluation. *Final Report ODOT SPR 737 OTREC-RR-012-06*. http://www.oregon.gov/ODOT/TD/TP_RES/docs/Reports/2012/SPR737_Wireless.pdf.

LG (2008). Retrieved February 10, 2014, from http://www.lg.com/us/cell-phones/lg-CU720-Black-black-shine.

Malinovskiy, Y., Lee, U., Wu, Y., & Wang, Y. (2011). Investigation of Bluetooth Based Travel Time Estimation Error on a Short Corridor. *Proceedings of the 90th Annual Meeting of the Transportation Research Board*. Washington, DC: Omnipress.

Malinovskiy, Y., Saunier, N., & Wang, Y. (2012). Pedestrian Travel Analysis Using Static Bluetooth Sensors. *Proceedings of the 91st Annual Meeting of the Transportation Research Board*. Washington, DC: Omnipress.

Malinovskiy, Y., Wu, Y., Wang, Y., & Lee, U. (2010). Field Experiments on Bluetooth-based Travel Time Data Collection. *Proceeding of the 89th Annual Meeting of the Transportation Research Board*. Washington, DC: Omnipress.

Martchouk, M., Mannering, F., & Bullock, D. (2011). Analysis of Freeway Travel Time Variability Using Bluetooth Detection. *Journal of Transportation Engineering*, *137*(10), 697–704.

Martchouk, M., & Mannering, F. (2009). Analysis of Travel Time Reliability on Indiana Interstates. *USDOT Region V Regional University Transportation Center Final Report*. Retrieved July 9, 2013, from http://www.purdue.edu/discoverypark/nextrans/assets/pdfs/completedprojects/Final%20Report%20014.pdf.

Mei, Z., Wang, D., Chen, J. (2011). Investigation with Bluetooth Sensors of Bicycle Travel Time Estimation on a Short Corridor. *International Journal of Distributed Sensor Networks*, Vol. 2012. Retrieved July 9, 2013, from http://www.hindawi.com/journals/ijdsn/2012/303521/.

Namaki Araghi, B., Tørholm Christensen, L., Krishnan, R., & Lahrmann, H. (2012). Application of Bluetooth Technology for Mode-Specific Travel Time Estimation on Arterial Roads: Potentials and Challenges. *Proceedings of the 2012 Annual Transport Conference at Aalborg University*. Retrieved July 9, 2013. from http://www.trafikdage.dk/papers_2012/36_BaharNamakiAraghi.pdf.

Noyce, D., Gajendran, A., & Dharmaraju, R. (2006). Development of Bicycle and Pedestrian Detection and Classification Algorithm for Active-Infrared Overhead Vehicle Imaging Sensors. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1982, 202-209.

Peterson, B., Baldwin, R., & Kharoufeh, J. (2004). A Specification-Compatible Bluetooth Inquiry Simplification. *IEEE Proceedings of the 37th Annual Hawaii International Conference* (pp. 66-75). Big Island, HI: IEEE.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

Ringwald, M., & Romer, K. (2007). Practical Time Synchronization for Bluetooth Scatternets. *Fourth International Conference on Broadband Communications, Networks and Systems* (pp. 337-345). Raleigh, NC: IEEE.

Ronai, M., & Kail, E. (2003). A Simple Neighbour Discovery Procedure for Bluetooth Ad Hoc Networks. *IEEE Transactions on Parallel and Distributed Systems. 11*(5), 1028-1032.

Rourke, A., & Bell, M. (1994). An Image Processing System For Pedestrian Data Collection. *Proceedings of the 7th International Conference on Road Traffic Monitoring and Control* (pp. 123-126). London, UK: IEEE.

Saeedi, A., Park, S., Kim, D.S., and Porter, J.D. (2012) Improving the Accuracy and Precision of Travel Time Samples Collected at Signalized Arterial Roads Using Bluetooth Sensors. *Proceedings of the 91st Annual Meeting of the Transportation Research Board*. Washington, DC: Omnipress.

Schraml, S., Belbachir, A., & Brändle, N. (2010). A Real-time Pedestrian Classification Method for Event-based Dynamic Stereo Vision. *IEEE*

*Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 93-99). San Francisco, CA: IEEE.

Shladover, S., Kim, Z., Cao, M., Sharafsaleh, A., & Li, J. (2009). Bicyclist Intersection Crossing Times Quantitative Measurements for Selecting Signal Timing. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2128. 86-95.

Tan, P-N., Steinbach, M., & Kumar, V. (2006). *Introduction to Data Mining*. Boston: Pearson Addison-Wesley.

Wang, J., Chen, D., Chen, H., & Yang, J. (2011). On Pedestrian Detection and Tracking in Infrared Videos. *Pattern Recognition Letters*, *33*, 775-785.

Wheeler, N., Conrad, R., & Figliozzi, M.A. (2010). A Statistical Analysis of Bicycle Rider Performance: The impact of Gender on Riders' Performance at Signalized Intersections. *Proceedings of the 89th Annual Meeting of the Transportation Research Board*. Washington, DC: Omnipress.

Woodings, R., Joos, D., Clifton, T., & Knutson, C. (2002). Rapid Heterogeneous Connection Establishment: Accelerating Bluetooth Inquiry Using IrDA, *IEEE Wireless Communications and Networking Conference* (pp. 342-349). Orlando, FL: IEEE.

Guan, Y. (2006). *Large-scale Clustering: Algorithms and Applications*. Austin: ProQuest.

# APPENDIX

**FEATURES AND SPECIFICATIONS OF THE HARDWARE USED TO BUILD THE BLUETOOTH DATA COLLECTION UNITS**

1. Olimex iMX233-OLinuXino-MINI

**Figure 15: Olimex iMX233-OLinuXino-MINI (Olimex, 2013)**

**Table 32: iMX233-OLinuXino-MINI features and specifications (Olimex, 2013)**

| Features | Specification |
|---|---|
| Processor | iMX233 ARM926J at 454 MHz |
| RAM | 64 MB |
| Dimensions | 3.70" x 2.65" (94.0mm x 67.3mm) |
| DC Power Supply | 6V-16V |
| # of USB Hosts | 3 |
| GPIO Connector | 40 pins |
| Video Output | TV PAL/NTSC |
| # Buttons | 2 |
| Reset Button | Yes |
| SD Card Connector | Yes |
| UEXT Connector | Yes |
| Audio IN Connector | Yes |
| Audio IN Connector | Yes |

2. SENA Parani UD100 Bluetooth USB Adapter


**Figure 16: SENA Parani UD100 Bluetooth USB Adapter (SENA, 2011)**

**Table 33: Parani UD100 Bluetooth USB Adapter Specifications (SENA, 2011)**

| Features | Specification |
|---|---|
| Standards | • Bluetooth 2.0+EDR Class 1<br>• USB 2.0 |
| Max Transfer Rate | 3 Mbps (EDR) |
| Frequency Range | 2.402 ~ 2.480GHz |
| Transmit Output Power | +19dBm (+6dBm EDR) E.I.R.P |
| Receive Sensitivity | • Basic 1Mbps: -88 dBm<br>• EDR 2Mbps: -87dBm<br>• EDR 3Mbps: -82dBm |
| Antenna Connector | RP-SMA |
| Working Distance (In Open Field) | • Stub antenna – Stub antenna: 300 m<br>• Dipole (3 dBi) – Dipole (3 dBi) : 400 m<br>• Dipole (5 dBi) – Dipole (5 dBi): 600 m<br>• Patch antenna – Patch antenna: 1 km<br><br>* working distance can vary depending on install environment |
| Computer OS Support | • Windows XP/Vista/7 (32/64bit)<br>• Linux (3rd party driver required)<br>• MAC OS X (MAC OS X driver required) |
| Size | 72(L) x 22(W) x 10(H) mm |
| Operating Temperature | -20 ~ +70℃ |
| Storage Temperature | -40 ~ +85℃ |
| Humidity | 90% Non-condensing |
| Regulatory Approvals | FCC, CE, TELEC, KCC, IC, Bluetooth SIG |

3. MaxStream XBee Series 2 OEM RF Modules

**Figure 17: MaxStreamXBee Series 2 OEM RF Modules (MaxStream, 2006)**

**Table 34: XBee Series 2 OEM RF Modules Specifications (MaxStream, 2006)**

| Features | Specification |
|---|---|
| Indoor/Urban Range | up to 133 ft. (40 m) |
| Outdoor RF line-of-sight Range | up to 400 ft. (120 m) |
| Transmit Power Output | 2mW (+3dBm) |
| RF Data Rate | 250,000 bps |
| Serial Interface Data Rate | 1200 - 230400 bps (non-standard baud rates also supported) |
| Receiver Sensitivity | -95 dBm (1% packet error rate) |
| Supply Voltage | 2.8 – 3.4 V |
| Operating Current (Transmit) | 40mA (@ 3.3 V) |
| Operating Current (Receive) | 40mA (@ 3.3 V) |
| Power-down Current | < 1 uA @ 25◦C |
| Operating Frequency Band | ISM 2.4 GHz |
| Dimensions | 0.960” x 1.087” (2.438cm x 2.761cm) |
| Operating Temperature | -40 to 85º C (industrial) |
| Antenna Options | Integrated Whip, Chip, RPSMA, or U.FL Connector |
| Supported Network Topologies | Point-to-point, Point-to-multipoint, Peer-to-peer & Mesh |
| Number of Channels | 16 Direct Sequence Channels |
| Addressing Options | PAN ID and Addresses, Cluster IDs and Endpoints (optional) |

4. L-com 2.4 GHz 8 dBi Round Patch Antenna - 10in N-Female Connector

**Figure 18: L-com 2.4 GHz 8 dBi Round Patch Antenna**

**Table 35: 2.4 GHz 8 dBi Round Patch Antenna Electrical Specifications (L-com, 2005)**

| Features | Specification |
|---|---|
| Frequency | 2400-2500 MHz |
| Gain | 8 dBi |
| Horizontal Beam Width | 75 degrees |
| Vertical Beam Width | 65 degrees |
| Impedance | 50 Ohm |
| Max. Input Power | 25 Watts |
| Lightning Protection | DC Short |

**Table 36: 2.4 GHz 8 dBi Round Patch Antenna Mechanical Specifications (L-com, 2005)**

| Features | Specification |
|---|---|
| Weight | 0.3 lbs. (.13 Kg) |
| Dimensions | 4" (102 mm) Dia x 1" (25 mm) |
| Radome Material | UV-inhibited Polymer |
| Flame Rating | UL 94HB |
| Operating Temperature | -40° C to 85° C (-40° F to 185° F) |
| Mounting | Four ¼ in. (6.3 mm) Holes |
| Polarization | Horizontal or Vertical |
| Wind Survival | >150 MPH (241 KPH) |
| RoHS Compliant | Yes |

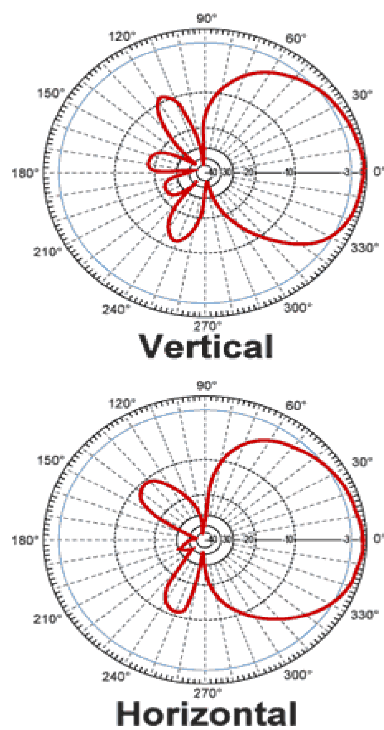


**Figure 19: L-com 2.4 GHz 8 dBi Round Patch Antenna Gain Patterns (L-com, 2005)**

5. L-com 2.4 GHz 5 dBi Rubber Duck Antenna - RP-SMA Plug Connector

**Figure 20: L-com 2.4 GHz 5 dBi Rubber Duck Antenna (L-com, 2012)**

**Table 37: 2.4 GHz 5 dBi Rubber Duck Antenna Electrical Specifications (L-com, 2012)**

| Features | Specification |
|---|---|
| Frequency | 2400-2500 MHz |
| Gain | 5.5 dBi |
| Impedance | 50 Ohm |
| VSWR | < 2.0 |

**Table 38: 2.4 GHz 5 dBi Rubber Duck Antenna Mechanical Specifications (L-com, 2012)**

| Features | Specification |
|---|---|
| Weight | 0.7 oz. (20 g) |
| Length | 8.2" (208 mm) |
| Diameter | 0.5" (13 mm) |
| Finish | Matte Black |
| Connector | Reverse Polarity SMA Plug |
| Operating Temperature | -40° C to 85° C (-40° F to 185° F) |
| Flame Rating | UL 94HB |
| Polarization | Vertical |
| RoHS Compliant | Yes |

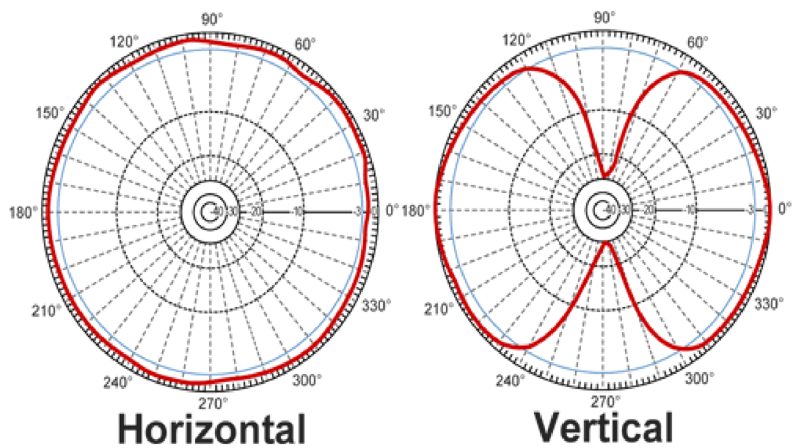


**Figure 21: L-com 2.4 GHz 5 dBi Rubber Duck Antenna Gain Patterns (L-com, 2012)**

6. Energizer XP18000 Universal AC Adapter

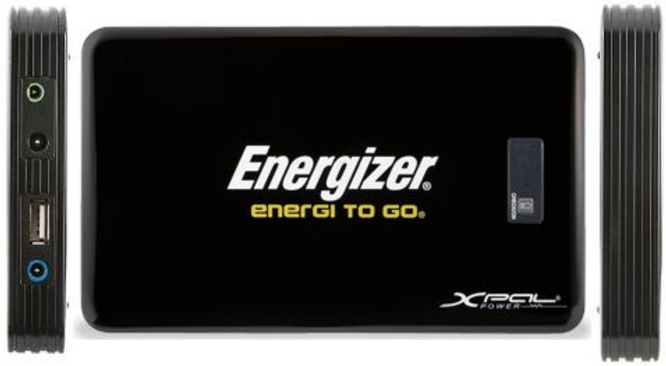


**Figure 22: Energizer XP18000 Universal AC adapter (Energizer, 2010)**

**Table 39: Energizer XP18000 Universal AC Adapter Specifications (Energizer, 2010)**

| Features | Specification |
|---|---|
| Battery Cell | Lithium Polymer |
| Power Capacity | 18000 mAh |
| Rated Input | DC 19V, Max 3500 mA |
| Rated Output | • DC 5V, 1000 mA<br>• DC 10.5V, 2000mA<br>• DC 19V, 3500mA |
| Recharge Time | ~4 Hrs |
| Weight | 0.52kgs / 1.14lb |
| Dimensions | 7.1(L) x 4.3(W) x 0.8(H) inches<br>(18(L) x11(W) x 2(H) cm) |

7. GlobalSat BU-353 USB GPS Receiver


**Figure 23: GlobalSat BU-353 USB GPS Receiver (GlobalSat, 2009)**

**Table 40: BU-353 USB GPS Receiver Specifications (GlobalSat, 2009)**

| Features | Specification |
|---|---|
| GPS Chipset | SiRF Star e/LP |
| Frequency | 1575.42 MHZ |
| C/A Code | 1.023 MHz chip rate |
| Channels | 20 all-in-view tracking |
| Sensitivity | -159 dBm |
| GPS Protocol | Default: NMEA 0183 (Secondary: SiRF Binary) |
| Operating | -40°~176°F (-40°~80°C) |
| Voltage | 4.5V-6.5V |
| Current | 50mA typical |
| Dimension | 2.08" diameter x 0.75" (53mm dia. X 19.2mm) |
| Weight | 2.2Ounces (67.37g) |

## APPENDIX REFERENCES

Energizer, 2010, Retrieved April 5, 2014, from:

http://www.energizerpowerpacks.com/it/products/xp18000/

GlobalSat, 2009, Retrieved April 5, 2014, from:

http://www.globalsat.com.tw/productspage.php?menu=2&gs_en_product_id=2&gs_en_product_cnt_id=28&img_id=409&product_cnt_folder=8

L-com, 2005, Retrieved April 5, 2014, from:

http://www.l-com.com/multimedia/datasheets/DS_RE08P-RSP.PDF

L-com, 2012, Retrieved April 5, 2014, from:

http://www.l-com.com/multimedia/datasheets/DS_HG2405RD-RSP.PDF

MaxStream, 2006, Retrieved April 5, 2014, from:

https://www.sparkfun.com/datasheets/Wireless/Zigbee/XBee-2.5-Datasheet.pdf

Olimex, 2013, Retrieved April 5, 2014, from:

https://www.olimex.com/Products/OLinuXino/iMX233/iMX233OLinuXino-MINI/resources/iMX233-OLINUXINO-MINI.pdf

SENA, 2011, Retrieved April 5, 2014, from:

http://www.sena.com/download/manual/manual_parani_ud100-v3.0.0.pdf