

Potential of Opportunistic Relaying
PERFORMANCE STUDY ACROSS WIRELESS INTERFACES ON SMARTPHONES

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy
in
Computer Science and Engineering

by
Shu Liu

Aaron Striegel, Director

Graduate Program in Computer Science and Engineering
Notre Dame, Indiana
January 2014

© Copyright by

Shu Liu

2014

All Rights Reserved

Potential of Opportunistic Relaying

PERFORMANCE STUDY ACROSS WIRELESS INTERFACES ON SMARTPHONES

Abstract

by

Shu Liu

To my husband Wei
and my loving parents
who inspires me to exceed my potential.

CONTENTS

| | |
|---|------|
| FIGURES | v |
| TABLES | vii |
| ACKNOWLEDGMENTS | viii |
| CHAPTER 1: INTRODUCTION | 1 |
| 1.1 Overview | 1 |
| 1.2 Contributions | 2 |
| CHAPTER 2: Large-scale Dataset | 5 |
| 2.1 Data Collection | 5 |
| 2.2 Dataset Overview | 7 |
| 2.3 Comparison | 10 |
| CHAPTER 3: Bluetooth Proximity | 13 |
| 3.1 Background | 13 |
| 3.2 Related Work | 14 |
| 3.3 Proximity Estimation Model | 16 |
| 3.3.1 Bluetooth RSSI vs. Distance | 17 |
| 3.3.2 Proximity Estimation Model | 22 |
| 3.3.3 Comparisons | 33 |
| 3.4 Case Study | 36 |
| 3.4.1 Proximity in large group | 37 |
| 3.4.2 Proximity in small group | 41 |
| 3.5 Summary | 46 |
| CHAPTER 4: WiFi Offloading | 47 |
| 4.1 Background | 47 |
| 4.2 Related work | 49 |
| 4.3 WiFi Offloading in Practice | 49 |
| 4.4 Comparison of User Behavior | 53 |
| 4.4.1 Phone Usage | 56 |
| 4.4.2 App Usage | 59 |
| 4.5 Summary | 61 |

| | |
|--|-----|
| CHAPTER 5: Opportunistic Relaying | 64 |
| 5.1 Background | 64 |
| 5.2 Related Work | 65 |
| 5.3 Evaluating Practical Relaying | 66 |
| 5.3.1 Availability | 69 |
| 5.3.1.1 Proximity | 69 |
| 5.3.1.2 Symmetry | 73 |
| 5.3.1.3 Diversity | 73 |
| 5.3.1.4 Residual Battery | 75 |
| 5.3.1.5 Evaluating Availability | 76 |
| 5.3.2 Prevalence | 77 |
| 5.3.2.1 Time in Proximity | 77 |
| 5.3.2.2 Intercontact Time and Beyond | 82 |
| 5.3.2.3 Measuring Effective Utility | 83 |
| 5.3.2.4 Evaluating Prevalence | 86 |
| 5.3.3 Stability | 87 |
| 5.3.4 Reciprocity | 92 |
| 5.4 Summary | 96 |
| CHAPTER 6: Energy Consumption | 98 |
| CHAPTER 7: Conclusion and Outlook | 99 |
| APPENDIX A: Appendix | 101 |
| BIBLIOGRAPHY | 103 |

FIGURES

| | | |
|------|--|----|
| 2.1 | PhoneMonitor Software Design | 7 |
| 2.2 | Monitoring System | 8 |
| 2.3 | Ecdf of Data in April 2012 | 11 |
| 3.1 | Bluetooth RSSI Values Distribution in One Week | 17 |
| 3.2 | Initial Indoor RSSI Values with Different Distances | 19 |
| 3.3 | Bluetooth RSSI vs. Distance - Theoretical, Indoor & Outdoor | 20 |
| 3.4 | Bluetooth RSSI vs. Distance Indoor Case | 21 |
| 3.5 | Symmetric RSSI Values | 21 |
| 3.6 | Bluetooth RSSI vs. Distance Outdoor Case | 22 |
| 3.7 | Walk Path for Real-world Scenario | 24 |
| 3.8 | Cdf of Light Sensor Data in Different Environments | 29 |
| 3.9 | Data in Real-world Scenario | 29 |
| 3.10 | Multiple Thresholds in Different Zones | 32 |
| 3.11 | Location Data Provided by WiFi and GPS | 36 |
| 3.12 | Light Sensor Data Distribution on Tuesday | 39 |
| 3.13 | Bluetooth RSSI Values Distribution on Tuesday | 40 |
| 3.14 | <i>S018</i> Data on Game Day | 41 |
| 3.15 | <i>S077</i> Data on Game Day | 42 |
| 3.16 | <i>S018</i> Data on Game Day with Proximity Estimation Model | 43 |
| 3.17 | Symmetry Analysis of Data on Game Day | 44 |
| 3.18 | <i>S018</i> Data on Weekday with Proximity Estimation Model | 45 |
| 3.19 | Fallbreak Data with Proximity Estimation Model | 45 |
| 4.1 | Average Downlink Traffic per Phone per Week | 50 |
| 4.2 | Switch between 3G and WiFi | 51 |
| 4.3 | WiFi Signal Strength (1700-1800h) | 52 |
| 4.4 | WiFi Signal Strength (2200-2300h) | 52 |
| 4.5 | Screen On Duration in One Day | 53 |

| | | |
|------|---|----|
| 4.6 | Distribution of Percent of WiFi Traffic | 54 |
| 4.7 | Numbers of Participant in Groups | 55 |
| 4.8 | Weekly Average Total Downlink Traffic per Phone | 56 |
| 4.9 | Avg. Daily Screen Usage of Groups | 57 |
| 4.10 | Avg. Screen session length of Groups | 58 |
| 4.11 | Avg. Screen on Duration in Different Time Slots | 58 |
| 4.12 | Weekly Application Downlink Traffic | 61 |
| 4.13 | Weekly Application Uplink Traffic | 62 |
| 5.1 | Bluetooth and WiFi RSSI Values ECDF | 70 |
| 5.2 | Bluetooth Proximity and WiFi Proximity ECDF in April 2012 | 72 |
| 5.3 | Comparison of Bluetooth Proximity and WiFi Proximity | 72 |
| 5.4 | Symmetry of Bluetooth Proximity | 73 |
| 5.5 | Distribution of Proximity Diversity (In-Study) in April 2012 | 74 |
| 5.6 | Diurnal Distribution of Proximity Diversity in April 2012 | 75 |
| 5.7 | Average Daily Low Residual Battery Time in April 2012 | 76 |
| 5.8 | Bluetooth Proximity with Different Restrictions | 79 |
| 5.9 | Diurnal Distribution of Bluetooth Proximity | 80 |
| 5.10 | Diurnal Distribution of Bluetooth Proximity in April 2012 | 80 |
| 5.11 | WiFi Proximity | 81 |
| 5.12 | Combined Proximity (Fused Bluetooth / WiFi Detection) | 82 |
| 5.13 | Intercontact Time ECDF | 83 |
| 5.14 | Weak WiFi Signal & Traffic | 85 |
| 5.15 | Relaying & Weak WiFi Signal | 85 |
| 5.16 | Traffic & Relaying | 86 |
| 5.17 | Proximity Duration ECDF | 88 |
| 5.18 | Number of Proximity Nodes vs. Node Strength | 89 |
| 5.19 | Total Appearances and Continuous Appearances in Day Counts ECDF | 90 |
| 5.20 | Weekly Cumulative Number and Active Number | 91 |
| 5.21 | Distribution of Potential Interactions - Need vs. Serve (November 2012) | 94 |
| 5.22 | Ratio of Needing vs. Serving (November 2012) | 95 |
| 5.23 | Diurnal Distribution of Ratio (November 2012) | 95 |

TABLES

| | | |
|------|---|-----|
| 2.1 | DATA TYPES COLLECTED BY PHONEMONITOR APPLICATION | 8 |
| 2.2 | MONTHLY DATA SUMMARY | 10 |
| 2.3 | DATASET COMPARISON | 12 |
| 3.1 | DIFFERENT PROXIMITY ESTIMATION TECHNIQUES COMPARISON | 16 |
| 3.2 | DISTANCE OF FACE-TO-FACE COMMUNICATION AROUND CAMPUS | 23 |
| 3.3 | ERROR RATE AGAINST REAL-WORLD DATA | 25 |
| 3.4 | IMPROVED ERROR RATE WITH MODIFIED THRESHOLD . . . | 26 |
| 3.5 | IMPROVED ERROR RATE WITH DATA SMOOTHING | 27 |
| 3.6 | ENVIRONMENT ESTIMATION WITH LIGHT SENSOR DATA . . | 30 |
| 3.7 | DEFINITION OF ZONES IN DIFFERENT ENVIRONMENTS . . . | 33 |
| 3.8 | BOUNDARY SUMMARY | 34 |
| 3.9 | ERROR RATE AGAINST REAL-WORLD DATA WITH PROXIMITY ESTIMATION MODEL | 35 |
| 3.10 | ACCURACY AND POWER CONSUMPTION COMPARISONS . . | 37 |
| 3.11 | PROXIMITY VARIATION IN A WEEK | 38 |
| 3.12 | PROXIMITY VARIATION ON TUESDAY DAYTIME | 40 |
| 4.1 | WIFI RSSI ON LAPTOP AND SMARTPHONE | 48 |
| 4.2 | NUMBER OF MALES AND FEMALES IN GROUPS | 55 |
| 4.3 | WEEKLY SMS/PHONE CALL COMPARISON | 59 |
| 4.4 | TOP APPLICATIONS CATEGORIES | 60 |
| 5.1 | FRAMEWORK FOR EVALUATING AVAILABILITY, PREVALENCE, STABILITY, RECIPROCITY | 67 |
| A.1 | DETAILED INFORMATION OF COLLECTED DATA | 102 |

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Aaron Striegel, who helped provide direction and guidance for this work. Without his trust and support, I would not have had the opportunity to work in such great research projects. I would also like to thank the members of my committee, Dr. Christian Poellabauer, Dr. David Hachen and Dr. Gregory Madey, who were more than generous with their expertise and precious time.

I also need to thank all my colleagues who create such a good atmosphere in the group; Andrew Blaich, Yingxin Jiang, Dirk Van Bruggen, Lei Meng, Xueheng Hu, Benjamin Bockdtege, and Rachael Purta. Also Margaret, Tylor and Michael, deserve a special thanks because their great help in the Netsense project. I need to further thank all my dear friends, Nikhil, Hongsheng, Chris, Pramita, Sal, Jun, Yihua and Mehrdad, for being there for me throughout the entire doctorate program.

CHAPTER 1

INTRODUCTION

1.1 Overview

Over the past few years, a vast array of wireless devices and services have emerged that are fundamentally transforming how we as a society gather and react to information. Furthermore, the new wireless ecosystem has increased wireless data consumption at phenomenal rates with the most popular cited estimates slating traffic to double every year for the next five years [10]. Dubbed the *wireless data tsunami*, the dominant question for wireless service providers (carriers) is how to meet what appears to be an insatiable need for wireless data. Unlike wired networks, spectrum available for wireless data is finite and typically entails massive costs for acquisition and infrastructure deployment. Although technologies such as LTE herald the arrival of fourth-generation (4G) wireless technology, the new speeds often only temporarily satiate the need for additional bandwidth. A wide variety of solutions have emerged ranging simpler solutions such as better WiFi offloading to much more complex solutions such as small heterogeneous cellular networks. For many cellular providers, WiFi offloading, i.e. the users receiving data from 802.11-based hotspots, offers a significant appeal by reducing the strain on the already overloaded cellular infrastructure. Recent studies such as the one in [27] points to offloading offering gains approaching 65% of the total traffic volume. There are other works such as [3, 11, 14, 46] that discuss the feasibility of WiFi offloading.

Besides WiFi offloading, a rich category of work that is complementary to existing techniques is the concept of *opportunistic communication*. Opportunistic communi-

cation refers to the concept which allows nodes to leverage sporadic, intermittent contacts when two nodes come into direct radio communication range [40]. When cellular or WiFi links are not available or not strong enough, opportunistic relaying introduces another alternative option for mobile device to get connected by working in tandem with one or more devices. From a conceptual standpoint for opportunistic networking, the design of the relaying protocol is critical, namely how does one select and manage appropriate relaying nodes as relays [2, 6, 24, 31, 42]. From a practical standpoint for opportunistic networking, it is essential to prove the potential for opportunistic relaying amongst actual mobile device users in real world.

The primary goal of my work is to gather high quality smartphone data and leverage the data for reinforcing analysis from a technical network system perspective with respect to network connectivity and performance enhancement. In particular, I answered three interrelated questions: *(i) Can existing wireless technologies on smartphone provide accurate face-to-face proximity estimation and to what extent are relationship formation and maintenance dynamics modified with the introduction of digital communication? (ii) Can the WiFi offloading be the ultimate solution for the predicted wireless data tsunami? (iii) Can the peer-to-peer communication between mobile devices in close proximity be a good candidate for offloading cellular systems?* The answers to these questions are fascinating to explore. The NetSense project is a study of first-year students tracked over a two-year period with respect to nearly all the smartphone information. The process of gathering data results in a rich pool of digital data which provides us the opportunities to analyze the technical dynamics of the network. These involve for instance, the wireless signal strengths can reflect the relative distance and connection status. Moreover, by collecting the information of traffic consumed by the phones, we are able to compare different types of traffic usage and understand their impacts on campus wireless networks.

1.2 Contributions

The key contributions of my dissertation is as follows:

- Smartphone monitoring system: By collecting data on Android smartphones, we get a fully anonymized dataset containing device information and digital communication data of participants. It is the foundation of the research and provides multiple dimensions of data for research in both network science and sociology.
- Bluetooth proximity: Under the premise that Bluetooth has better accuracy for short distance without the constrain of environment, we use Bluetooth signal strength as an indicator for relative distance between devices. In order to get accurate estimation of face-to-face distance, a proximity estimation model is proposed by leveraging the raw Bluetooth Received Signal Strength Indicator (RSSI) values and introducing multiple thresholds for different environments.
- WiFi offloading: For most cellular providers, WiFi offloading appeals to reduce the strain on the overloaded cellular infrastructure. However, based on our data, the WiFi traffic takes approximately 30% of the total data consumption which is much lower than expected. We analyze the possible reasons for this as well as the usage patterns of users in different categorization according to their WiFi traffic consumption.
- Proximity relay: With the support of peer-to-peer communication in smartphones, it is practical to leverage the device-to-device data transfers to bridge the coverage gap between cellular and WiFi infrastructure network. Based on our data, we will do the quantitative analysis of the potential relaying which is a key missing part in relaying related research and evaluate the benefits of relaying for cellular traffic offloading in real life.

The rest of this paper is organized as follows. Section 5.2 provides an overview of the related work and Section ?? introduces the dataset we have. Based on the dataset, the framework to evaluate the potential for relaying is proposed in Section 5.3. In the end, Section ?? concludes the paper.

CHAPTER 2

Large-scale Dataset

As espoused by the MIT Reality Mining group in [12, 37], the increasing availability and reliability of smartphones affords incredible opportunities for the unobtrusive gathering of data with regards to usage, location, and performance. Although quite expensive to gather versus user self-reporting or user volunteers, a fully instrumented smartphone represents a veritable wealth of data that can afford unique insight into the smartphone performance study.

In August of 2011, two hundred participants were selected from the incoming freshmen class of our university and received a free Android smartphone and plan in exchange for agreement to participate in the two-year data collection project. Each Android device was rooted and a custom ROM installed (Cyanogenmod) to enable the device being permanently Bluetooth discoverable. A user-level agent was installed on each of the smartphones that extracted a wide variety of environmental and usage data from the phone at periodic intervals. The monitoring agent was installed to start automatically when the smart phone power was turned on and ran passively in the background. Tuning was conducted on the agent to ensure that even with WiFi enabled and Bluetooth always discoverable for a battery life potential at distribution of roughly one and a half days.

2.1 Data Collection

In order to build a flexible and stable platform for delivering on-going subject queries in a secure and efficient way, we designed and implemented the monitoring

system in the Fall of 2010 and did a few serials of tests before the study began. The key part of the monitoring system is the customized agent installed on the smartphones named *PhoneMonitor* as illustrated in Figure 2.1. The Android platform was selected for its customization capabilities through normal API or rooted/customized interfaces with respect to hardware-level interactions. It collects a variety of types of data as illustrated in table 2.1 and each kind of data is invoked through the corresponding function calls. Separate threads are employed to compensate for the variety of speeds at which the respective functions retrieve relevant data. For example, the Bluetooth data includes the detailed values of timestamp, RSSI (Received signal strength indication), MAC address, and Bluetooth identifier (BTID) with a default sensing granularity of once per minutes. The wireless environmental data (primarily WiFi) has similar fields except the access point (AP) name is recorded and the granularity of sensing is three minutes. Traffic data includes breakdowns by application and wireless adapter (cell, WiFi) with respect to both downlink (Rx) and uplink (Tx) usage at intervals down to as low as once per minute. The light sensor data, which includes the timestamp and sensor values, can help us to determine whether the phone is sheltered (e.g. inside a backpack or in hand) and the surroundings (e.g. inside or outside buildings) during the daytime. It is recorded if the current value changes more than 30 compared with the last recorded value. For mail data, both sender and receiver email addresses are recorded without content snooping. More detailed information about collected data can be found in table A.1. Graphic User Interface (GUI) of the application allows participants to read the last check-in time as well as the status of the network connection and Bluetooth.

Figure 2.2 illustrates the framework of smartphone data collection, storage and query process. Data is locally spooled on the phone before being securely transmitted to one of two remote check-in servers. Data from the check-in server is then

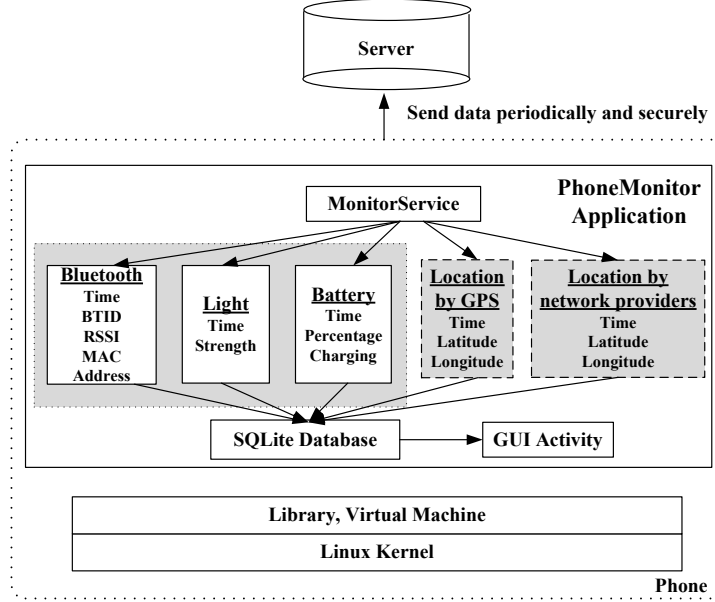


Figure 2.1. PhoneMonitor Software Design

spooled locally before being conveyed to a database server not directly connected to the Internet with all accesses strictly logged and validated to protect sensitive user data. From September 2011 to September 2013, we got more than 250G data in the database.

2.2 Dataset Overview

The dataset used in the dissertation covers 15 months data (Oct 2011 - Dec 2012) and there are around 41 million Bluetooth records, 50 million WiFi scans and 1 million SMS messages. Data from the first two months of the study is excluded to allow for the new freshmen to have settled into a new routine and social arrangements. A reasonable degree of co-location exists amongst students in the study with students selected clustered amongst six primary dormitories equally divided amongst male and female students. Once entered into the study, students were free to re-locate at

TABLE 2.1

DATA TYPES COLLECTED BY PHONEMONITOR APPLICATION

| | |
|------------------------------|--|
| Device Information | Bluetooth, WiFi, Cell, Location, Light Sensor, Screen On/Off, Battery Level, Connection Status, Network Traffic, Application Usage, Port Information, Running/Install Apps, Password, Contacts, Camera Usage |
| Digital Communication | Phone Call, SMS, Mail, Web Browser History, Music |

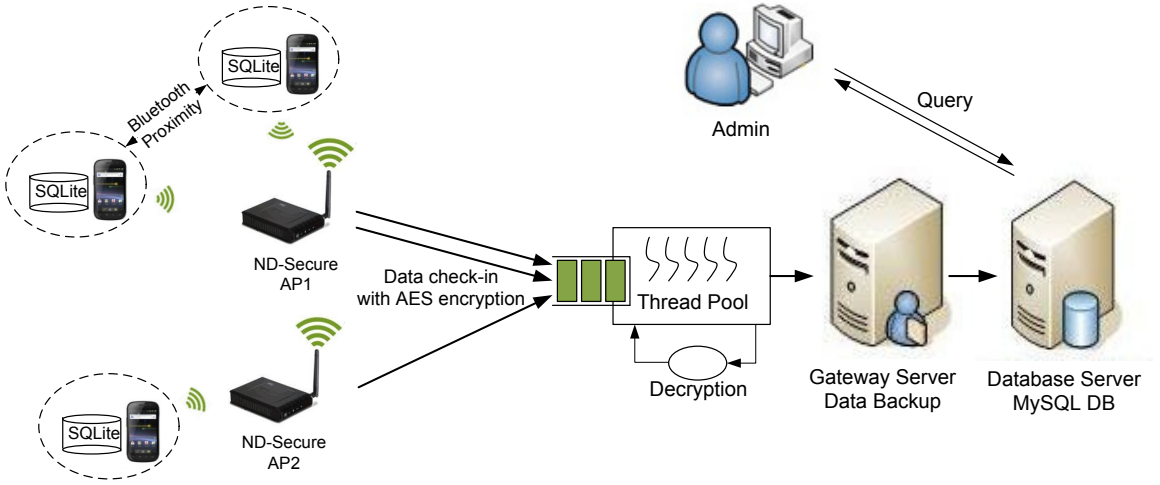


Figure 2.2. Monitoring System

either the beginning of the spring semester (2012) or the start of their sophomore year. Moreover, the respective distribution levels within individual dormitories made it reasonably close to random chance amongst incoming freshmen for that dormitory if two participants in the study were selected as roommates. At the onset of the monitored period for the paper, approximately 11 students had dropped from the student reducing the data pool to 189 students.

Table 2.2 summarizes the details of the data across different months. For each

of the respective metrics presented in the table, we reduce the sampling rate to five minutes slots scattered throughout the day where each day has the potential for 288 measurement points. The powered on percentage represents the number of slots when the phone monitor was active with a noticeable drop during later hours (12am - 8 am) but a notable uptick during the evening hours (4pm-12am). The screen on percentage is the duration when the phone screen is on and implies the average usage of the phone for either consuming data or conducting other communications (SMS, phone call, etc.). The monthly Bluetooth and WiFi records shows the average number of distinguished Bluetooth devices/WiFi APs detected per device across the month. University virtual APs are reduced to a single AP as each university router offers between two to four WiFi AP names.

Figure 2.3 exhibits the empirical distribution functions (ECDF) of different types of data in April 2012 to offer additional context for the data beyond the average values denoted in Table 2.2. Two percentage value ECDFs are plotted (power on, screen on) together with two detected environmental values (unique Bluetooth devices in the month, unique APs in the month). Scales for each of the two values are the lower axis for percentage and upper axis for raw numeric counts. Most notably, the average number of unique Bluetooth devices detected in April 2012 was 746 with some devices seeing as high as over 1000 devices and others seeing as low as slightly less than 200 unique devices. Table 2.2 also contains distinctions with respect to intra-study proximity and inter-study breakdowns of the various Bluetooth devices. As would be expected, intra-study detection dramatically trails off during summer break in tandem with detected university APs as the students leave campus.

TABLE 2.2

MONTHLY DATA SUMMARY

| Avg. Monthly Values per Phone | Nov 2011 | Apr 2012 | Jul 2012 | Nov 2012 |
|--|----------|----------|----------|----------|
| Powered On (%) | 73.7 | 70.6 | 51.3 | 61.2 |
| Powered On (%) (12am-8am) | 66.7 | 60.7 | 42.3 | 58.0 |
| Powered On (%) (8am-4pm) | 74.6 | 69.3 | 51.3 | 62.7 |
| Powered On (%) (4pm-12am) | 79.6 | 72.7 | 60.1 | 62.9 |
| Screen On (%) | 5.74 | 5.05 | 4.99 | 4.30 |
| Total Rx Traffic (MB) | 609.2 | 727.2 | 864.5 | 685.8 |
| Total Tx Traffic (MB) | 133.11 | 123.8 | 130.4 | 200.8 |
| # of Detected Distinct Bluetooth Devices | 961 | 746 | 310 | 681 |
| # of Detected Distinct Bluetooth Devices Within Project (RSSI \geq -80dBm) | 131 | 106 | <1 | 75 |
| # of Detected Distinct WiFi APs | 1218 | 1004 | 1777 | 1663 |
| # of Detected Distinct WiFi APs in Campus | 424 | 446 | <1 | 433 |

2.3 Comparison

The inclusion of a device-side agent on a live user device introduces notable complexity with respect to user privacy and IRB (Institutional Review Board) concerns. Moreover, such studies tend to be frequently expensive to conduct due to the cost of subsidizing access costs at a sufficient level to yield effective participatory compliance. Although there are several notable wireless datasets including university campuses (MIT Reality [37], UCSD [35] and Dartmouth [16]), conference sites (Infocom [9]) and cities (Nokia [25]), most datasets are limited in size and scope in terms of cap-

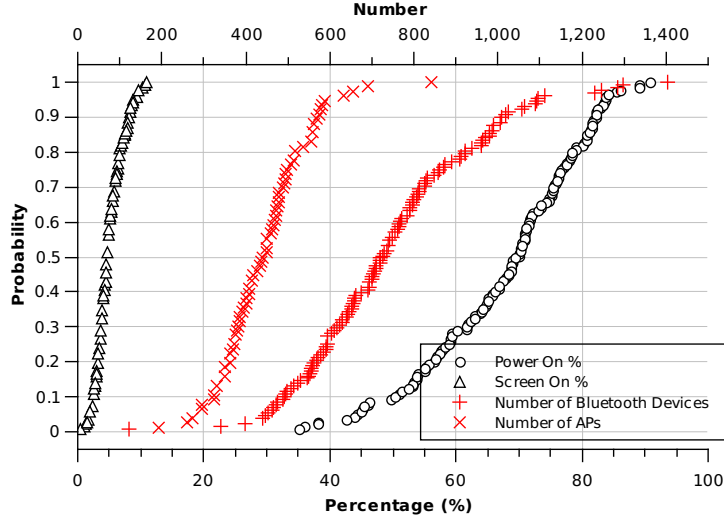


Figure 2.3. Ecdf of Data in April 2012

turing dyadic relationships, namely both sides of a potential proximity relationship for the purposes of truly evaluating opportunistic networking. For instance, both the MIT Reality and Infocom traces record when contact is detected by virtue of Bluetooth discovery, ample for characterizing the inter-contact times but not necessarily capturing energy levels nor traffic needs of the respective nodes. Alternatively, the UCSD and Dartmouth traces rely on WiFi for localization gathering either data via AP fingerprinting (UCSD) or SNMP logs directly from the AP (Dartmouth). The richest of the datasets is the data from the Nokia Data Challenge which includes both WiFi and Bluetooth data but the dataset is no longer public after the completion of the analysis. In contrast to prior works, our dataset includes nominal proximity (Bluetooth) / location data (WiFi, Google Location service) in addition to a complete view of the smartphone environment including WiFi signal strengths, energy levels, traffic demands, and the social context for the participants in the study (Facebook, SMS, etc.). Table 2.3 summarizes the most relevant studies and compares the works to our own reference study.

TABLE 2.3: DATASET COMPARISON

| Traces | Our Dataset | MIT Reality | UCSD | Dartmouth | Infocom | Nokia |
|-------------------------------|---------------------|-------------|----------|------------|-----------|-----------------|
| Device | Smartphone | Cell Phone | PDA | Laptop PDA | iMote | Smartphone |
| # of Devices | 189 | 97 | 275 | 6,648 | 41 | 185 |
| Network Type | Bluetooth/WiFi | Bluetooth | WiFi | WiFi | Bluetooth | Bluetooth/WiFi |
| Contact Type | Direct/AP-based | Direct | AP-based | AP-based | Direct | Direct/AP-based |
| Duration (days) | 458 | 246 | 77 | 114 | 4 | 210 |
| Granularity (seconds) | 60/300 | 300 | 120 | 300 | 120 | N/A |
| # of internal contacts | 3,616,184 | 54,667 | 195,364 | 4,058,284 | 22,459 | N/A |
| Internal pairwise contact/day | 0.221 | 0.022 | 0.034 | 0.008 | 3.4 | N/A |
| Other proximity related data | Cell, Traffic, etc. | Cell | N/A | N/A | N/A | Cell, etc. |

CHAPTER 3

Bluetooth Proximity

3.1 Background

Interactions are not limited to any particular area and can take place at a wide variety of locations, ranging from sitting and chatting in a Starbucks coffee shop to walking and chatting across a college campus. As will be explored later in the paper, for most face-to-face interactions, the approximate distance between individuals in casual conversation is within 0.5 to 2.5 meters. One of the solutions would seem to be location-based calculation which relies on location technologies such as WiFi triangulation [17], cell phone triangulation [38], GPS, or a combination of all three. However, none of these solutions are ideal or sufficient. Although WiFi triangulation can present a reasonable degree of accuracy, its accuracy in all but the most dense WiFi deployments is insufficient, ranging on the order of 3 to 30 meters [17]. Similarly, cell phone triangulation suffers from an even worse accuracy [38]. Moreover, while WiFi is reasonably pervasive, WiFi tends to generally be sparser in green spaces, i.e. outdoor spaces. Notably, GPS suffers from both an accuracy shortcoming (5-50m) as well as a lack of viability indoors [47].

However, it is important to note that face-to-face interaction does not demand an absolute position as offered by the previously mentioned schemes but rather requires a determination of *proximity*. With that important shift of the problem definition, Bluetooth emerges as a straightforward and plausible alternative, offering both accuracy (1-1.2m) [48] and ubiquity (most modern smartphones come with

Bluetooth) [33]. Although some prior work has attempted to use the detection of Bluetooth to indicate nearness [37], it is not enough for the face-to-face proximity estimation. The question addressed by this section is to what extent Bluetooth can be an accurate estimator of such proximity.

3.2 Related Work

Over the years, there has been a number of technologies proposed for proximity detection. Approaches such as those used by Meme Tags [7] provide good accuracy but require line of sight. Ultrasound approaches such as Activebadge [45] also provide good accuracy but they require infrastructure support. ZigBee technology is widely used in wireless sensor network to provide radio proximity estimation [4] in the environment where GPS is inoperative. Proximity can also be reported by sounds, and past work has shown audio to be effective for delivering peripheral cues [36]. However, it is untenable to expect the use of smartphones to reduce the unobtrusiveness of cues or increase comprehension. For the purposes of this paper, we are interested in techniques that are based on commonly available technologies in smartphones, i.e. GPS, Cell, WiFi and Bluetooth. Particularly, we are interested in techniques that can be applied at the smartphone itself without significant changes to the infrastructure.

Proximity detection is one of the advanced Location based Service (LBS) functions [22, 43, 44] to automatically detect when a pair of mobile targets approach each other closer than a predefined proximity distance (as in Location Alerts of Google Latitude). For realizing this function the targets are equipped with a cellular mobile device with an integrated GPS receiver, which passes position fixes obtained by GPS to a central location server. Most proposals for such services give low accuracy guarantees and incur high communication costs. Recently, 3-D optical wireless based location approach [5] is proposed which based on both GPS and triangulation technologies. It is another feasible way of utilizing GPS to get relative distance among

objects.

Some proximity estimation methods are based on Cell or WiFi signal. Using Place Lab [23], cell phones listen for the MAC address of fixed radio beacons, such as cell tower and wireless access points, and reference the beacons positions in a cached database. It provides adequate accuracy for detecting something like buddy proximity (e.g., median accuracy of 20-30 meters), but it requires a wardriving of the area to obtain pre-mapped cell towers and fixed wireless APs in the area. This can be quite costly, especially keeping the information up to date, as tower positions, etc. are updated on an annual basis. Without calculating absolute location, NearMe [21] explores the algorithm for detecting proximity using Wi-Fi signatures (WiFi APs and signal strengths), allowing it to work with no a priori setup. Similarly, [28] uses GSM readings to explore the proximity of mobile targets.

There are some proximity detection works using Bluetooth signal. From a specific work perspective, the works of Nathan et al. [12, 37] are highly relevant to the paper. In those studies, the authors use the ability to detect Bluetooth signals as indicators for people nearby within the Bluetooth range (around 10m). However, such indication does not meet the requirement of face-to-face proximity detection. In class, a student may discuss with others sitting beside him/her, but face-to-face talk is difficult with the students on the other side of the classroom even they are still in the Bluetooth range. Different from the above proximity detection method, our work is a fine grain Bluetooth-based proximity detection method which can provide adequate accuracy for face-to-face proximity estimation without environment limitations. Table 3.1 summarizes the differences of popular techniques, their prominent features and the performance [13, 29].

Based on the monitoring system described in Chapter 2, there are more than one million Bluetooth records collected per week. Figure 3.1 shows the distribution

TABLE 3.1
DIFFERENT PROXIMITY ESTIMATION TECHNIQUES
COMPARISON

| | Bluetooth | WiFi | GPS |
|-------------|-----------|--------------|----------------|
| HW costs | Medium | High | High |
| Coverage | High | High(Indoor) | High (Outdoor) |
| Power Usage | Medium | High | High |
| Accuracy | 1-4m | 2-30m | 5-50m |
| Security | High | High | Not applicable |

of the Bluetooth RSSI values collected from 196 phones in one week (more details will be discussed in Section 3.3 and Section 3.4). The data collected includes both indoor and outdoor environments. As it shows, the most prevalent value is around -76dBm which indicates much more than 5m indoor and nearly 5m outdoor as will be shown later. Therefore, an unfiltered detection method such as [12, 37] is not enough to estimate the face-to-face proximity and we use a more accurate method in Section 3.3 to solve this problem. Moreover, we introduce various smoothing effects and take advantage of empirical observations to function across a wide variety of typical environments.

3.3 Proximity Estimation Model

In this section, we explore the relationship between Bluetooth RSSI and distance in real world scenarios. The first method is using RSSI value threshold to determine whether two phones are in proximity or not. The second method introduces the light sensor data to determine whether the phone is indoors or outdoors, inside the

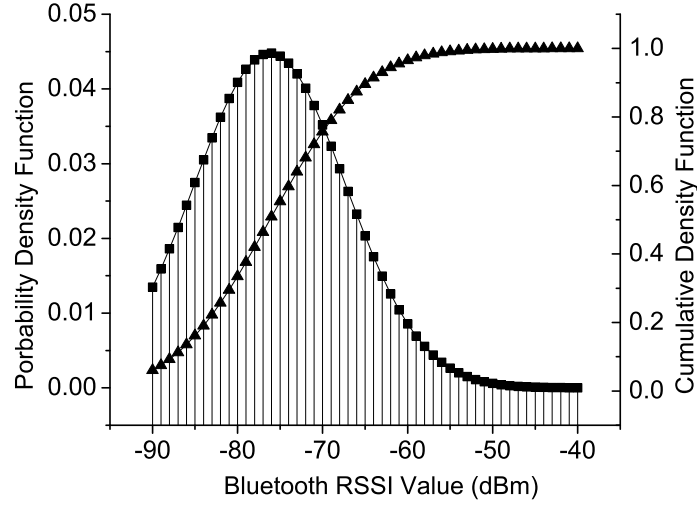


Figure 3.1. Bluetooth RSSI Values Distribution in One Week

backpack or in hand. By differentiating environments and smoothing data, a face-to-face proximity estimation model is outlined to improve the estimation accuracy in general scenarios. At the end of this section the proximity accuracy of Bluetooth, WiFi and GPS are analyzed and compared.

3.3.1 Bluetooth RSSI vs. Distance

Antti et al. presented the design and implementation of a Bluetooth Local Positioning Application (BLPA) [20] in which the Bluetooth received signal power level is converted to distance estimate according to a simple propagation model as follows:

$$\begin{aligned}
 RSSI &= P_{TX} + G_{TX} + G_{RX} + 20 \log \left(\frac{c}{4\pi f} \right) - 10n \log(d) \\
 &= P_{TX} + G - 40.2 - 10n \log(d)
 \end{aligned} \tag{3.1}$$

where P_{TX} is the transmit power; G_{TX} and G_{RX} are the antenna gains; G is the total antenna gain: $G = G_{TX} + G_{RX}$; c is the speed of light ($3.0 * 10^8 m/s$); f is the

central frequency (2.44 GHz); n is the attenuation factor (2 in free space); and d is the distance between transmitter and receiver (in m). d is therefore:

$$d = 10^{[(P_{TX} - 40.2 - RSSI + G)/10n]} \quad (3.2)$$

However, such a model can only be utilized as a theoretical reference. Due to reflection, obstacles, noise and antenna orientation, the relationship between RSSI and distance becomes more complicated. Our challenge was to assess how much impact these environmental factors have on Bluetooth RSSI values. Therefore, we carried out several experiments to understand how the Bluetooth indicators fade with distance under these environmental influences.

Indoor experiments were conducted in a noisy hallway (around seven other Bluetooth devices detected) in the campus engineering building. Outdoor experiments were conducted in the open area outside the building. In the measurement there were no obstacles between the two phones and the antennas of the phones were aligned towards each other. In such a way, we tried to build up a relatively simple and “ideal” environment where the possible impact factors are reflection and noise only. We repeated the measurements over the period of an hour with the distance being increased by 0.5 meters between each round. Figure 3.2 shows the initial fluctuations of indoor RSSI results with different distances. Although the data varies significantly even within the same distance, there is a noticeable gap exists between different distances. Such results further shed light on the viability of using Bluetooth RSSI to indicate the face-to-face proximity.

In Figure 3.3, we present indoor, outdoor, and theoretical results for Bluetooth across a variety of distances (0-5 meters). The theoretical values were predicted by the propagation model with $P_{TX} = 2.9 \text{ dBm}$ and $G = 4.82 \text{ dBi}$. These specific values are the average empirical results of our experiment phone. We calculated the

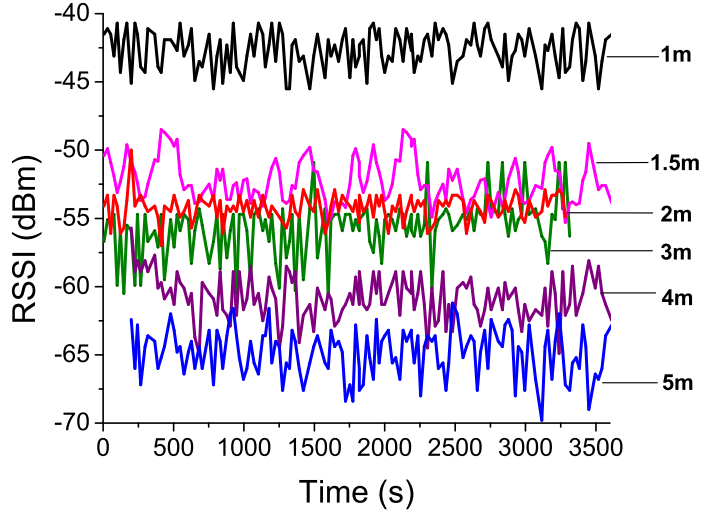


Figure 3.2. Initial Indoor RSSI Values with Different Distances

average RSSI from nearly 120 raw values for each distance. The indoor results were relatively close to the theoretical values. However, the results outside the building were much farther away from the theoretical reference and imply that these two kinds of environmental settings should be identified in the following measurements.

Furthermore, we performed similar experiments on two phones focusing on the indoor case but with different antenna orientation (e.g. in the same direction) and obstacles (e.g. put in a backpack or partitioned by cubicle) in order to discover the influence of these possible factors. Figure 3.4 illustrates the results with these impacts. The observations include the following: first, the change in orientation turns out to have little impact on the final results. As many smart phones cannot predict phone orientation, antenna design is typically optimized to account for this fact. Second, although we placed two phones on each side of a cubicle board, such an arrangement did not affect RSSI significantly. Third, the most important environmental issue came from the backpack. It may be because the signal of Bluetooth is disturbed or

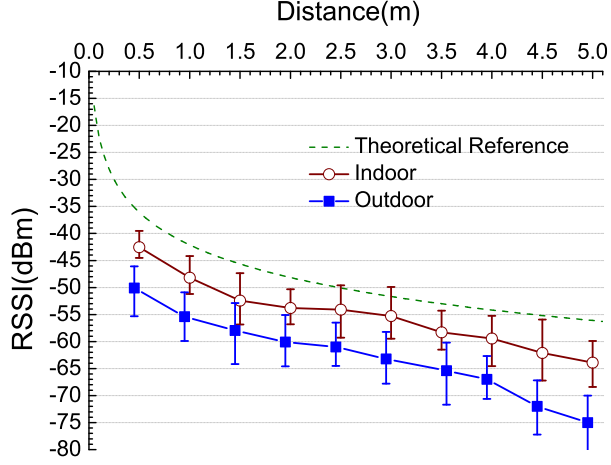


Figure 3.3. Bluetooth RSSI vs. Distance - Theoretical, Indoor & Outdoor

shielded in such a closed environment. As many individuals would be likely to carry their phone in a purse or backpack (particularly on a college campus), the backpack setting bears further investigation. We also recorded the data to check whether the RSSI values on phones are symmetric. Figure 3.5 shows the RSSI values on one phone are almost the same as the results on the other phone.

Using the same method, we measured the RSSI values outdoors with the consideration of the influence of a backpack. Figure 3.6 shows the results from those experiments. Similarly, the RSSI values become lower when the phones are in the backpack so it is a non-ignorable element in the following estimations, further reinforcing that detection of such an arrangement may be critical for proper distance estimation resolution.

Based on these indoor and outdoor results, there are two main environmental factors that may effect the RSSI values: inside/outside building and inside/outside a backpack. Besides those factors, it is also necessary to take multiple-phones scenario into consideration since phones with Bluetooth around may have interference on

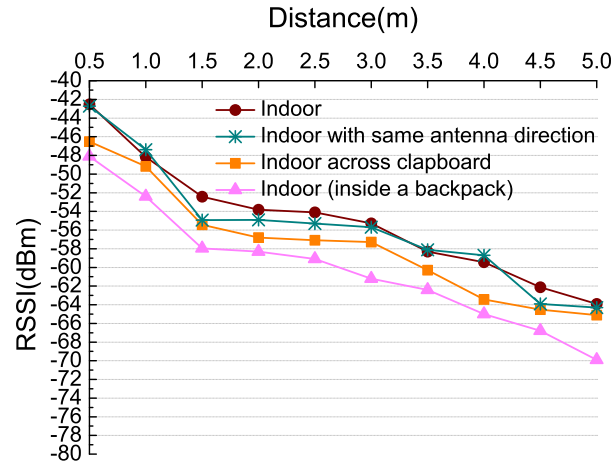


Figure 3.4. Bluetooth RSSI vs. Distance Indoor Case

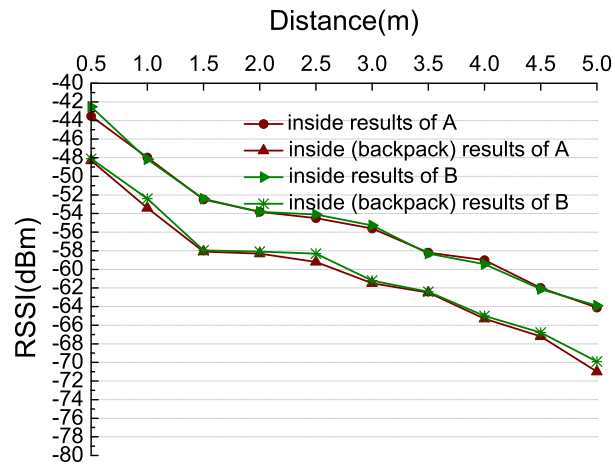


Figure 3.5. Symmetric RSSI Values

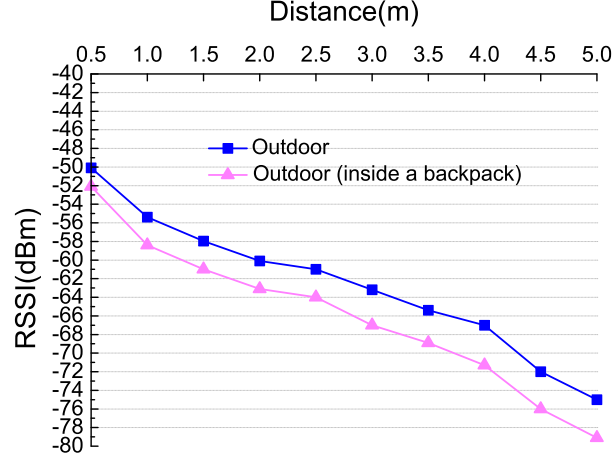


Figure 3.6. Bluetooth RSSI vs. Distance Outdoor Case

Bluetooth RSSI values.

3.3.2 Proximity Estimation Model

As mentioned in the beginning, we aim to provide an accurate proximity estimation for face-to-face communication. This raises a question: what is the face-to-face communication distance? In this subsection, we first define the face-to-face distance and then use the indoor results as a threshold to do the estimation in real world scenarios. Since the error rate of using a simple threshold is relatively high, we explore the possible reasons and propose a proximity estimation model with the introduction of light sensor values.

Distance of face-to-face communication: When we have dinner with our friends sitting at the same table, the conversation among us is called face-to-face communication; or when we talk with someone side by side, the distance between us is also called face-to-face communication. In other words, face-to-face communication happens when people are close enough to have conversations in a convenient manner.

People typically have such communication when they are sitting or walking together. Thus, we calculate the distance for this kind of communication by measuring distances across the campus (such as diagonal of desk in dinning hall, distance between desks in classrooms and etc.) and the average value is equal to 1.52m. The detailed samples are listed in Table 3.2.

TABLE 3.2
DISTANCE OF FACE-TO-FACE COMMUNICATION AROUND
CAMPUS

| Place | Distance(cm) |
|--|--------------|
| Diagonal of small square desk in La Fortune | 116 |
| Diagonal of large square desk in La Fortune | 160 |
| Diameter of round desk in La Fortune | 120 |
| Diagonal of desk in dinning hall | 250 |
| Distance between desks in Cushing office | 125 |
| Distance between desks in classroom | 155 |
| Diagonal of desk in discussion cubic | 220 |
| Distance between people walking side by side | 70 |

To conduct an evaluation of the accuracy of our Bluetooth method, we constructed a scenario that draws upon several likely occurrences in normal campus interactions. The scenario blends each of the earlier test cases and provides data to assess the accuracy in a real-world setting. The measurement was conducted as follows: two

people with two phones walked side by side from Cushing Hall to Grace Hall and then returned back (Figure 3.7).



Figure 3.7. Walk Path for Real-world Scenario

The whole process took 40 minutes and individuals were always within the distance for face-to-face communication. The Bluetooth update time interval was changed to 10 seconds temporarily in order to provide enough samples. During the first ten minutes (phone in hand) and last ten minutes (phone inside a backpack) individuals were inside Cushing Hall. When individuals were outside (the duration was 20 minutes), in the first ten minutes individuals held the phones in their hands and then put the phones in their backpack for the later ten minutes.

Single Threshold: After data collection, the corresponding RSSI value (-52dBm) of direct communication distance (152cm) based on the indoor measurements (Figure 3.4) was used as a threshold to estimate whether the individuals were in proximity. Accordingly, values less than -52dBm were considered as not in face-to-face proximity and labeled as a wrong estimation.

TABLE 3.3
ERROR RATE AGAINST REAL-WORLD DATA

| | |
|---|--------|
| Total samples | 246 |
| Total error rate | 72.8% |
| Indoor error rate (0 - 10 mins) | 14.3% |
| Outdoor error rate (10 - 20 mins) | 91.3% |
| Outdoor (inside a backpack) error rate (20 - 30 mins) | 100.0% |
| Indoor (inside a backpack) error rate (30 - 40 mins) | 85.0% |

Table 3.3 shows the results and error rate of this naïve method. It was found that both of the outdoor and backpack parts have extremely high error rates. After switching the threshold value to -58dBm which is the outdoor RSSI values with 152cm distance, the error rate was improved but still high as shown in Table 3.4.

In our opinion, the reasons for high error rates include:

- i) One fixed threshold is not enough as the indicator of correct or wrong estimation;
- ii) Only indoor or outdoor relationship was used to analyze the data without differ-

TABLE 3.4

IMPROVED ERROR RATE WITH MODIFIED THRESHOLD

| | |
|---|-------|
| Total samples | 246 |
| Total error rate | 48.4% |
| Indoor error rate (0 - 10 mins) | 4.9% |
| Outdoor error rate (10 - 20 mins) | 53.2% |
| Outdoor (inside a backpack) error rate (20 - 30 mins) | 85.5% |
| Indoor (inside a backpack) error rate (30 - 40 mins) | 49.2% |

entiation;

iii) The influence of backpack and other possible environmental interference were not taken into consideration;

iv) Each RSSI value was not smoothed to allow for environmental fluctuations.

Multiple Thresholds: According to the reasons for high error rate analyzed above, we introduce the proximity estimation model which is a multiple threshold-based method with the consideration of data smoothing and different environmental effects.

i) Data Smoothing

Since there is time delay during the data collection, we do smoothing on the data collection to avoid environmental fluctuation effects and there are several ways to achieve it. One way is using simple window function and each value $RSSI_i$ at time i is modified using the following function:

$$RSSI_i = a * RSSI_{i-1} + b * RSSI_i + c * RSSI_{i+1} \quad (3.3)$$

For the values of the parameters (a, b and c), several combinations such as (0.4, 0.6, 0), (0.3, 0.4, 0.3) and (0.2, 0.6, 0.2) are used in the following comparisons. Another

smoothing method is to utilize EWMA (exponentially weighted moving average) to analyze the dataset. Let E_i be the EWMA value at time i and s be the smoothing factor. The EWMA calculation is as follows:

$$E_i = s * RSSI_i + (1 - s)E_{i-1} \quad (3.4)$$

Based on real-world data as shown in Table 3.4, we combine the data smoothing method with signal threshold filter to analyze the effects of data smoothing and select the best smoothing function. In Table 3.5, we compare three different combinations for window function and two types of smoothing factors. While the combination (0.3, 0.4, 0.3) exhibits good improvement of error rate in different scenarios compared with results in Table 3.4, the EWMA methods with smoothing factor 0.5 is the best among the five options and we use it in the proximity estimation model.

TABLE 3.5
IMPROVED ERROR RATE WITH DATA SMOOTHING

| | Simple Window Function | | | EWMA | |
|------------------|------------------------|-----------------|-----------------|---------|---------|
| | (0.4, 0.6, 0) | (0.3, 0.4, 0.3) | (0.2, 0.6, 0.2) | s = 0.5 | s = 0.8 |
| Total | 49.6% | 28.3% | 37.1% | 26.8% | 38.2% |
| Indoor | 5.4% | 3.4% | 5.7% | 2.1% | 4.9% |
| Outdoor | 50.8% | 39.4% | 42.3% | 23.8% | 34.7% |
| Outdoor backpack | 82.9% | 59.3% | 76.8% | 61.2% | 79.1% |
| Indoor backpack | 50.7% | 30.2% | 45.7% | 27.5% | 48.3% |

ii) Light Sensor Data

As shown in Figure 3.4 and Figure 3.6, the Bluetooth RSSI values are much smaller than the indoor ones when the phone is in the backpack or outdoors. One of our observations is that it is possible to treat the light sensor data as an indicator of the environment. Figure 3.8 reveals the light sensor data distribution in different settings: during the daytime when the phone is inside the building the light sensor returns values between 225 to 1280; while this value comes up to larger than 1280 when phone is under daylight. When the phone is in the backpack, the light values are typically around 10. Therefore, when the light sensor value is in a range that indicates the phone is in a specific corresponding environment.

In Figure 3.9, we reviewed the distribution of the Bluetooth RSSI values collected in the walk experiment and the corresponding light sensor data got at the same time. As shown, there is RSSI data fluctuation even in the same setting due to the interference and noise. However, most Bluetooth RSSI values are larger than -55dBm when light sensor data is from 225 to 1280 (indicates indoor setting) while the RSSI values are smaller than -55dBm when light sensor data is in other zones (either in the backpack or outdoors).

Thus, light sensor data is introduced to differentiate the circumstances to improve the accuracy of distance estimation based on the rules in Table 3.6. Variations due to time of day (day vs. evening) may be accounted for using the smartphone time. Inevitably, accuracy will decrease during evening hours but we felt this is an adequate tradeoff for improved environmental detection. Unfortunately, the Nexus S 4G phone does not contain a pressure sensor which could be used to further improve results.

iii) Proximity Estimation Model

Based on the analysis of noise and interference, we use a multiple threshold model instead of a single threshold to do the proximity estimation. In Figure 3.3, the

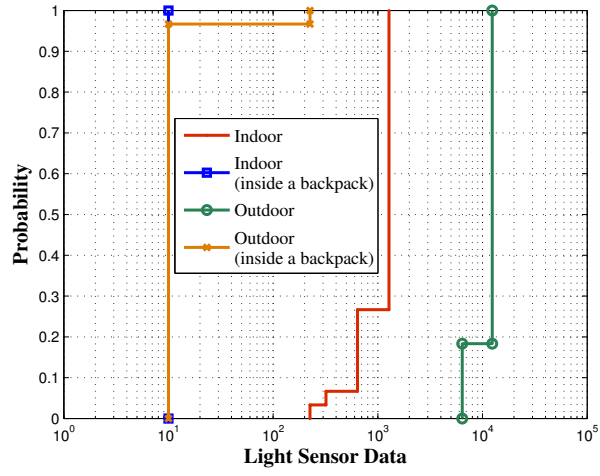


Figure 3.8. Cdf of Light Sensor Data in Different Environments

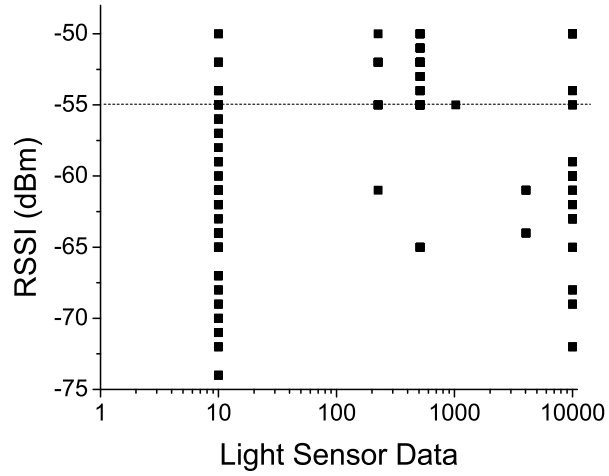


Figure 3.9. Data in Real-world Scenario

TABLE 3.6

ENVIRONMENT ESTIMATION WITH LIGHT SENSOR DATA

| Light Sensor Data | Environment Estimation |
|-------------------|-----------------------------|
| $(0, 100]$ | Inside backpack |
| $(100, 1280]$ | Indoor and out of backpack |
| >1280 | Outdoor and out of backpack |

corresponding indoor RSSI values of 2.5m (the maximum distance for face-to-face communication) is around -55dBm. It is obvious that when the received RSSI value is larger than -55dBm the two phones are in face-to-face proximity. Similarly, when the outdoor RSSI values is larger than -60dBm the two phones holder are close enough to have direct interaction. We call such data zone the “*Positive Zone*” denoting where two individuals are certainly within face-to-face interaction distance.

For the indoor data smaller than -55dBm, the model is constructed based on the following observation: in Figure 3.4 the smallest value for 5m distance in the test was -65dBm and it had a relatively low probability of occurrence (values larger than -65dBm is less than 20%) in Figure 3.1. Taking noise and interference into consideration, the value between -65dBm and -55dBm may also indicate a face-to-face proximity with high probability. Figure 3.1 includes both indoor and outdoor data and the most frequent data is -76dBm and the lowest values detected is -90dBm. When the indoor data is in (-76dBm, -65dBm), it is still possible to make a face-to-face communication but the probability is relatively low. The indoor data smaller than -76dBm implies that it is too far to have a direct communication and it is called the “*Negative Zone*”. Similarly, we set up several bounds for outdoor data: the smallest value for 5m is -75dBm. Therefore the zone(-75dBm, -60dBm) has a relatively high probability while the zone(-90dBm, -75dBm) is the low probability zone. When the outdoor data is

smaller than -90dbm, we strongly believe it is impossible to indicate a face-to-face proximity or even be detected.

Furthermore, we revisit the data regarding the inside backpack environment. Based on light sensor data, it is difficult to distinguish indoor or outdoor when the phone is inside the backpack. We noticed that when distance are 2.5m and 5m, the corresponding indoor Bluetooth RSSI values inside a backpack are -59dBm and -70dBm and the corresponding outdoor values are -64dBm and -79dBm. We defined the zone which is larger than -59dBm as “*Positive Zone*” for backpack data. Similarly, the zone which is smaller than -79dBm as “*Negative Zone*”. For the data between the two above bounds, we also define “*High Probability (HP) Zone*” (-70dBm, -59dBm) and “*Low Probability (LP) Zone*” (-79dBm, -70dBm). Therefore, for each type of environment, we have four zones as summarized in Table 3.7. Table 3.8 lists the corresponding multiple thresholds. Figure 4.6 illustrates the multiple thresholds of different zones in a more direct way.

For high and low probability zones, we name the minimum values in low probability zone as B_{min} , the maximum value in high probability zone as B_{max} and the range as B_{range} . Therefore $B_{range} = B_{max} - B_{min}$. To sum up, a proximity estimation model with multiple thresholds is proposed to improve the accuracy. For Bluetooth RSSI value x_i and the corresponding light sensor value y_i at time i , we calculate the probability of face-to-face proximity p_i as described in algorithm 1.

We define *Required Accuracy* (RA) as the lowest requirement for the probability to indicate two phones are in face-to-face proximity. With the improved estimation model, we analyze the data in the real-world scenario again and the error rate is improved greatly as shown in Table 3.9. Once p_i is higher than 45% ($RA = 45\%$), the two phones are considered to be in the face-to-face proximity. The bigger the RA value is, the more accurate face-to-face proximity we can obtain. We choose 45% as

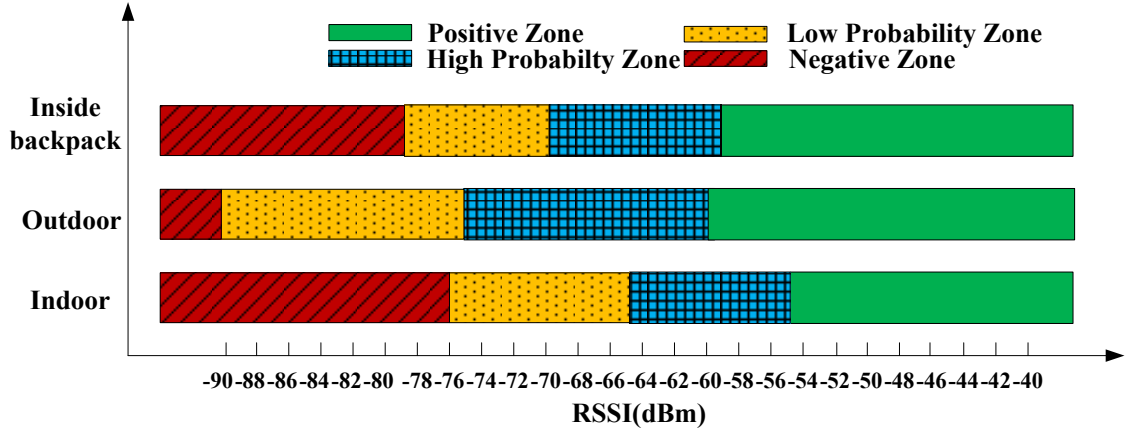


Figure 3.10. Multiple Thresholds in Different Zones

Algorithm 1 Estimate probability p_i of face-to-face proximity with Bluetooth RSSI value x_i and light sensor value y_i

```

 $x_i \leftarrow a * x_{i-1} + b * x_i + c * x_{i+1}$ 
determine the scenario depending on  $y_i$ 
if  $x_i$  is in positive zone then
     $p_i \leftarrow 1$ 
else if  $x_i$  is in probability zone  $[B_{min}, B_{max})$  then
     $p_i \leftarrow (x_i - B_{min}) / B_{range}$ 
else
     $p_i \leftarrow 0$ 
end if

```

TABLE 3.7

DEFINITION OF ZONES IN DIFFERENT ENVIRONMENTS

| Zones | Indoor | Outdoor | Inside Backpack |
|----------|-----------------------------|---------------------------|---------------------------------|
| Positive | $\geq B_{I_FTF}$ | $\geq B_{O_FTF}$ | $\geq B_{I_P_FTF}$ |
| HP | $[B_{I_5m}, B_{I_FTF})$ | $[B_{O_5m}, B_{O_FTF})$ | $[B_{I_P_5m}, B_{I_P_FTF})$ |
| LP | $[B_{frequent}, B_{I_5m})$ | $[B_{min}, B_{O_5m})$ | $[B_{O_P_5m}, B_{I_P_5m})$ |
| Negative | $< B_{frequent}$ | $< B_{min}$ | $< B_{O_P_5m}$ |

RA based on the following calculation: in each type of environment, the probability of the lowest possible value l in high probability zone equals to $(l - B_{min})$ divided by B_{range} . The smallest one in three types of environments equals to 45%.

3.3.3 Comparisons

WiFi triangulation/trilateration is a widely used method to do location indoors while GPS is perhaps the most popular way to do location outdoors. As summarized in Section 3.2, both of them have their own advantages and disadvantages. Here we use WiFi and GPS to do the face-to-face proximity estimation in order to compare the accuracy of them with the Bluetooth method we proposed. Together with the power consumption comparison in Section ??, the method of Bluetooth is proved to be an effective and efficient way in both aspects of accuracy and power usage.

We collected both network-provider location and GPS location data on the phone for the comparison. With the API provided by class *LocationManager* in Android SDK, we can get both kinds of location data by choosing different location providers with the frequency of three minutes. The GPS provider determines location using satellites while the network provider determines location based on availability of cell

TABLE 3.8

BOUNDARY SUMMARY

| Boundary | Detail | Values(dBm) |
|-----------------|--|-------------|
| B_{I_FTF} | Indoor fact-to-face (2.5m) | -55 |
| B_{O_FTF} | Outdoor fact-to-face (2.5m) | -60 |
| $B_{I_P_FTF}$ | Indoor inside backpack fact- to-face (2.5m) | -59 |
| $B_{O_P_FTF}$ | Outdoor inside backpack fact- to-face (2.5m) | -64 |
| B_{I_5m} | Indoor 5m | -65 |
| B_{O_5m} | Outdoor 5m | -75 |
| $B_{I_P_5m}$ | Indoor inside backpack 5m | -70 |
| $B_{O_P_5m}$ | Outdoor inside backpack 5m | -79 |
| $B_{frequent}$ | Most frequent value in the results | -76 |
| B_{min} | Minimum in the results | -90 |

tower and WiFi access points(APs). In the network provider method, the triangulation is used to get the location of the phone with the knowledge of cell towers' or APs' locations. When each phone's location is known, the relative distance as well as the accuracy is easy to calculate.

We conducted the experiment on a game day in the campus. Two students (A and B) went to Notre Dame Stadium to watch the game together. In Figure 3.11 the reported location data are marked. From 3pm to 7pm, the location data recorded by network provider are (41.699517, -86.232877) and (41.699203, -86.235269). During the same time period, the GPS provider collected the location as (41.699504,

TABLE 3.9

ERROR RATE AGAINST REAL-WORLD DATA WITH PROXIMITY
ESTIMATION MODEL

| | |
|---|------|
| Total samples | 246 |
| Total error rate | 4.3% |
| Inside error rate (0 - 10 mins) | 0.0% |
| Outside error rate (10 - 20 mins) | 4.5% |
| Outside (inside a backpack) error rate (20 - 30 mins) | 8.3% |
| Inside (inside a backpack) error rate (30 - 40 mins) | 6.2% |

-86.234624) and (41.699004, -86.235723). The corresponding distance between the reported data were 11.25 meters (443 inches) and 7.92 meters (312 inches) respectively. Based on the results reported by the phones, the accuracy of WiFi-based localization turns out to be around 10-15 meters and the one using GPS is around 10 meters.

Compared to the above WiFi triangulation and GPS methods, the Bluetooth-based method was more suitable for the face-to-face proximity estimation. As we mentioned before, there is no need to get the absolute location data to calculate the distance. Instead, we only need relative distance to do the estimation. From 3pm to 7pm on that day, we collected the Bluetooth RSSI on both phones and used the estimation model to do the analysis. When RA is 45%, we got the error rate around 6% with 554 sample data. Table 3.10 summarizes the comparison results of accuracy and power consumption percentage by invoking each method with similar frequency. Compared with WiFi and GPS, our method is accurate enough to indicate the face-to-face proximity. At the same time, the power consumption is at least 40% less than the other two technologies. These results are consistent with the data in Table 3.1



Figure 3.11. Location Data Provided by WiFi and GPS

and Bluetooth can definitely fulfill the requirements of proximity estimation in our system.

3.4 Case Study

While our experimental data shows the viability of Bluetooth as a proximity estimation tool, we examine the larger corpus of data from our smartphone study. We gathered high-fidelity data set by deploying the “PhoneMonitor” app on the Nexus S 4G android phones of 196 users. The participants were randomly chosen from the 2011 freshmen in the University of Notre Dame and were given the phones with unlimited voice, text and data plans. We encouraged the users to take advantage of all the features and services of the phone. The data set, including Bluetooth RSSI, WiFi RSSI, light sensor values as well as locations, was gathered between Sep and Oct 2011. With the data collected on these phones, we used the face-to-face proximity

TABLE 3.10

ACCURACY AND POWER CONSUMPTION COMPARISONS

| | Accuracy | Power consumption | Samples |
|------------|------------------|-------------------|---------|
| Our method | 1.5 - 2.5 meters | 15.7% | 554 |
| WiFi | 10-15 meters | 25.7% | 251 |
| GPS | 10 meters | 58.6% | 98 |

estimation model to get people who are in the direct communication distance with other participants. In the previous section, we showed that the proximity estimation model with multiple thresholds can increase the accuracy of proximity estimation effectively. In the following subsections, more cases and samples will be introduced to explore the Bluetooth-based method for proximity estimation in daily life.

3.4.1 Proximity in large group

With the data reported by 196 phones over two months, we analyzed the proximity among a large group. We first used Table 3.11 to show the proximity variation in one week (Oct 3rd - Oct 9th). There are three columns in the table: *Proximity Detected* column is the total number of devices which at least detected one of the other devices with face-to-face proximity probability larger than 45% ($RA = 45\%$); *Maybe Detected* column stands for the number of devices which detected other devices but most of them were in the low probability zone as shown in Table 3.7; *None* column is number of devices which do not report any data or the detected Bluetooth RSSI values were always in the negative zone. The number of samples may be varied from day to day.

Notably, the weekend included a home football game. Compared to the weekdays, more “Proximity Detected” cases are reported on Saturday since many students watched the game together and sit in the same student zone. On Sunday, we

observed significant "None" cases which may indicate students stayed in their room or went home instead of interacting.

TABLE 3.11

PROXIMITY VARIATION IN A WEEK

| | Num of Devices | Proximity Detected | Maybe Detected | None |
|-----|----------------|--------------------|----------------|------|
| Mon | 190 | 93 | 94 | 3 |
| Tue | 193 | 108 | 84 | 1 |
| Wed | 192 | 107 | 84 | 1 |
| Thu | 192 | 106 | 86 | 0 |
| Fri | 191 | 112 | 76 | 3 |
| Sat | 194 | 129 | 61 | 4 |
| Sun | 190 | 71 | 68 | 51 |

We look into the data on Tuesday in a more detailed way. Before we reveal the proximity variation on that day, we first plot the distribution of reported light sensor data and Bluetooth RSSI values in order to compare them with the former results we received in two-phones scenario. Figure 3.12 shows the trend of light sensor values in 24 hours. During the early morning and late night, most values were smaller than 1000 which means the participants are inside the buildings. From 8am to 8pm we noticed many values larger than 10000. At the same time, several values at the bottom (values vary from 10 to 100) indicate the phone is in the backpack. Since the light sensor values are not reliable to indicate indoor or outdoor during nighttime,

we focus on the proximity variation during the daytime. Using the same method as above, we summarize the proximity variation on that day in Table 3.12. The number of "Proximity Detected" cases during the class time (8am-5pm) is more than the one of after class time. Specially, the devices met much more other devices during the lunch time than the other time durations.

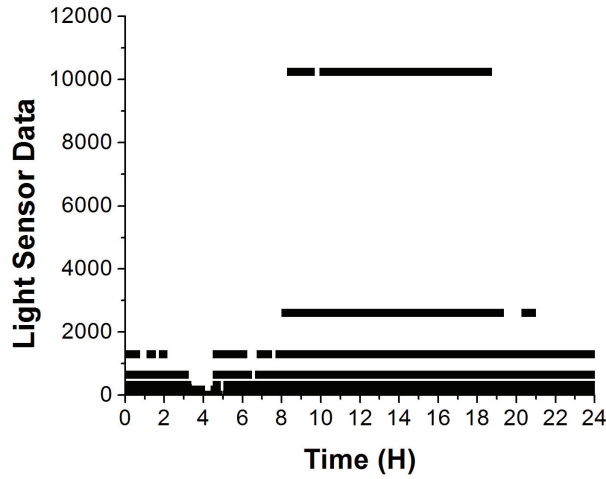


Figure 3.12. Light Sensor Data Distribution on Tuesday

Figure 3.13 reflects the distribution of Bluetooth RSSI values on that day. The most frequently value is around -75dBm which is similar as concluded in Figure 3.1. As revealed in Section 3.3, when RSSI value is smaller than -75dBm, it has a relatively low probability that the two phones are in face-to-face proximity. Critically, a method such as the one in [12, 37] would misdetect such interactions. Based on both Bluetooth RSSI values and light sensor values, our model can improve the accuracy

TABLE 3.12

PROXIMITY VARIATION ON TUESDAY DAYTIME

| | Proximity Detected | Maybe Detected | None |
|----------|--------------------|----------------|------|
| 8am-11am | 90 | 89 | 14 |
| 11am-2pm | 101 | 82 | 10 |
| 2pm-5pm | 89 | 97 | 7 |
| 5pm-8pm | 84 | 98 | 11 |

of face-to-face proximity indication.

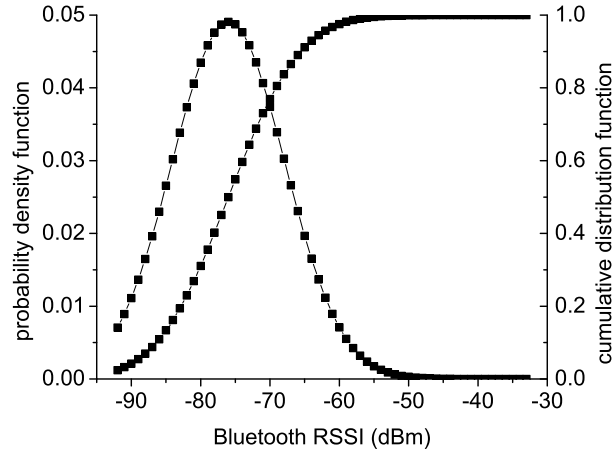


Figure 3.13. Bluetooth RSSI Values Distribution on Tuesday

3.4.2 Proximity in small group

Football Game Day: In 2011, the university had a football game with Air Force which lasted for four hours. There were 126 students among the 196 that watched the game in the stadium and we gathered 56710 Bluetooth records in the database. In Figure 3.14, we select one participant *S018* and show the detected phones around that student through Bluetooth during the 4 hour period. For every five minutes, if any other phone is detected we add its corresponding ID in that time slot. There are total 48 time slots and with several students always together with the student in the game. Similarly, we explore one of those nearby students *S077* to validate symmetric detection in Figure 3.15. These two figures further show that it is practical to use Bluetooth to detect people around.

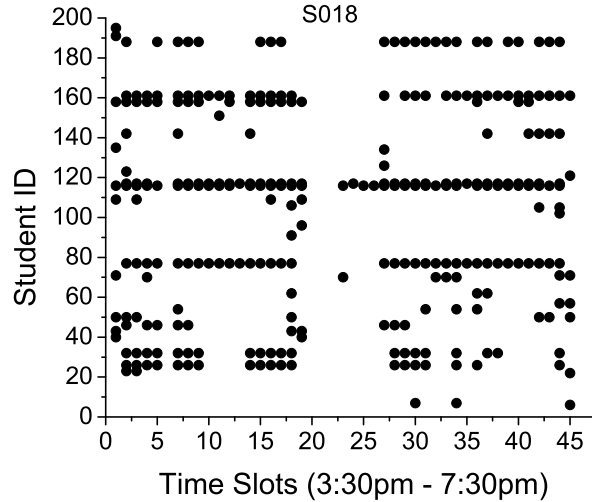


Figure 3.14. *S018* Data on Game Day

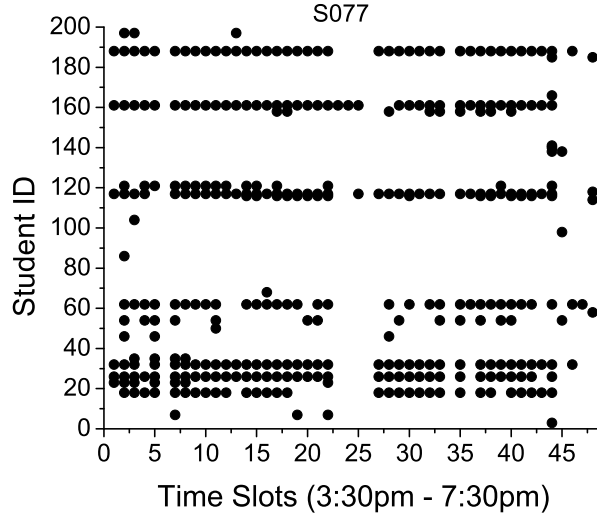


Figure 3.15. *S077* Data on Game Day

Figure 3.14 shows the detected students around without any restriction of Bluetooth RSSI values. In order to get list of the students in face-to-face proximity, we need to utilize the proximity estimation model to refine the results. Combined with light sensor data and method of data smoothing, the probability of proximity is calculated for the filtration. With RA of 45%, Figure 3.16 shows the filtered results which is more accurate to indicate the people who is in the face-to-face conversation range with the participant in the game. Compared with Figure 3.14, it is much more clear to find other devices kept close with the participant during the game.

We analyzed the symmetry between *S018* and *S077* in a more accurate way with proximity estimation model. In section 3.3 we discussed the symmetry of Bluetooth RSSI values between two phones and the values are almost the same when the noisy and interference is relatively low. Does such symmetry still exist when more than two phones are nearby? We look into the data reported on the game day again to check whether the symmetry between *S018* and *S077* still exists or not. Figure 3.17

includes the data from both *S018* and *S077* with RA equals to 45% and (018,077) means that *S018* detected *S077* was in the face-to-face range in the specific time slot. Due to the interference from other phones with Bluetooth, the values are not exactly symmetric in the four hours. There is nearly 40% of the time when such proximity detection is not symmetric.

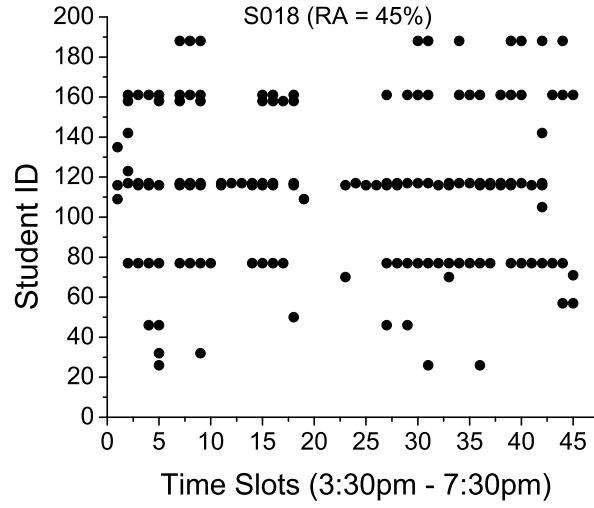


Figure 3.16. *S018* Data on Game Day with Proximity Estimation Model

Weekday: In this part, we analyzed the data recorded during weekdays, such as class time and lunch hours, and compared it with the data on game days. During the football game, most of the freshmen sit in the same student section and it is highly possible to detect more than 10 people around him/her at one time slot. However, when students are in class or having lunch, the data becomes reasonably sparse. Compared to more than fifty thousand records in the game, we recorded 6408

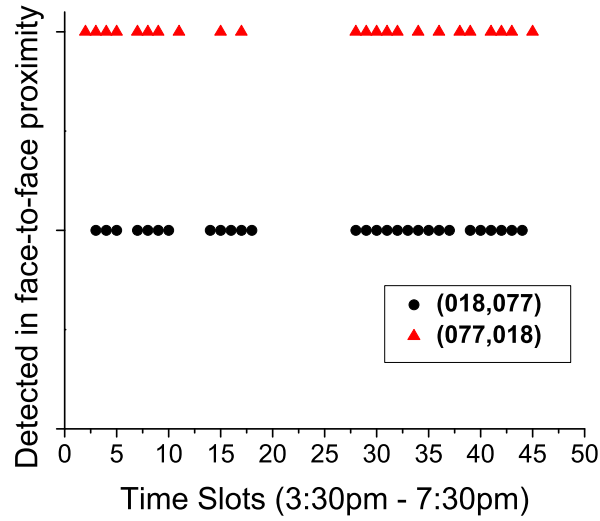


Figure 3.17. Symmetry Analysis of Data on Game Day

records on Oct 11th 2011 (Tuesday) from 9am to 1pm. Figure 3.18 illustrates the data reported by the same phone *S018* during this period. Obviously, the chance to meet other students (related to this project) in class or during the lunch time is relatively low. In *S018*'s case, he/she only met with two other students within a direct-communication distance in class and four during the lunch time.

Fall Break: During the fall break (from Oct 17th to Oct 23rd), most students went home and we got relatively less data. Take the data of Oct 19th for example, we got 4373 records in the whole day and only 26 devices detected other devices in the project in face-to-face proximity. We used the proximity estimation model and the same RA to analyze the results we got on Oct 19th between 10am and 10:05am. Figure 3.19 shows the proximity status among students in this specific time slot. There are in total 11 samples collected during the 5-minutes period and (035, 148) means device *S035* detected *S148* was in the face-to-face proximity. Since there is less interference, the symmetry is maintained as observed in Section 3.3. In

Figure 3.19, more than 70% values are symmetric. Compared with self-reporting method, our proximity estimation model is a more reliable and effective method to detect face-to-face proximity in daily life.

3.5 Summary

In this section, our presented work validates the usage of Bluetooth as a tool for face-to-face proximity detection and makes the following contributions:

- We demonstrate the viability of using Bluetooth for the purposes of face-to-face proximity estimation and propose a proximity estimation model with appropriate smoothing and consideration of a wide variety of typical environments.
- We study the relationship between the value of Bluetooth RSSI and distance based on empirical measurements and compare the results with the theoretical results using the radio propagation model.
- Based on the monitoring system, we are able to use the proximity estimation model across several real-world cases to provide high accurate determination of face-to-face interaction distance.

Based the empirical Bluetooth RSSI results, we further study the potential of opportunistic relaying when two devices are in close proximity in Chapter 5.

CHAPTER 4

WiFi Offloading

4.1 Background

With the advent of the smartphone, mobile data usage has exploded which in turn has created tremendous pressure on cellular data networks. A promising candidate to reduce the impact of cellular data growth is WiFi offloading. However, our dataset casts doubts on the viability of achieving such gains with WiFi offloading. In contrast to the prior work of [27], we have found that despite users operating in a dense university WiFi environment, the potential gains for the majority of users are quite muted. Rather than finding that WiFi usage dominates 3G usage, our study curiously finds that much of the user consumption of 3G dominates the consumption of WiFi.

At first glance, such a result would appear to counterintuitive. A dense WiFi deployment¹ would have ample WiFi access points placed throughout the buildings on the campus. The dormitory-oriented residential life of the campus meant that all study participants (all of whom were freshmen) would have ubiquitous WiFi coverage at night (dormitory) as well as during the day (classrooms/dorms). Coverage at the university is also frequently verified by employing roaming laptops by IT staff throughout campus. However, it is the usage of laptops to validate coverage as opposed to smartphones where the problem originates.

Consider the observation noted in Table 4.1 that compares the observed RSSI on a laptop versus the RSSI observed in the same time period and same location via a

¹The University of Notre Dame was ranked as one of the top 20 wireless campuses in 2010.

representative smartphone from the study. In the table, despite being in the same location, the smartphone observes a dramatically reduced signal strength (around 10dB) versus the laptop. The discrepancy between observed signal strength on the smartphone is not an isolated phenomenon to one particular smartphone but rather tends to be broadly indicative of many commercially available smartphones that are subject to marketing and development constraints. This is potentially significant for WiFi offloading as most dense WiFi environments (i.e. businesses) tend not to be designed at smartphones but rather tend to be designed at laptops.

TABLE 4.1

WIFI RSSI ON LAPTOP AND SMARTPHONE

| | RSSI on Laptop | | RSSI on Smartphone | |
|--------------|----------------|---------|--------------------|---------|
| Access Point | Avg. | Std Dev | Avg. | Std Dev |
| 1 | -75.27 | 2.76 | -83.32 | 2.72 |
| 2 | -75.22 | 2.79 | -85.04 | 3.39 |
| 3 | -75.40 | 2.87 | -84.94 | 3.04 |
| 4 | -71.99 | 2.06 | -79.52 | 3.51 |
| 5 | -72.05 | 2.06 | -80.18 | 3.80 |
| 6 | -73.88 | 5.13 | -80.02 | 3.34 |

The period in question analyzed in this chapter represents eight weeks of continuous data from the end of January 2012 to the end of March of 2012 in the spring semester. The time frame is selected to ensure that the students have fully adapted

to their phones and represent behaviors typical of their normal phone usage. The seventh week in the study represents spring break (March 10th - March 18th) which offers a likely period where students returned home for the week. In total, we selected 131 users out of the study (62 female, 69 male) that have continuous data throughout the entire eight weeks and have more than five minutes per day of usage on the phone. The graphs and tables in this chapter are based on the data collected from these 131 participants.

4.2 Related work

There have been several recent studies proposed to reducing the strain on the overloaded cellular infrastructure with WiFi offloading being one of the main candidate [3, 11, 14, 27, 46]. In [3], WiFi connectivity is used to reduce the pressure on 3G spectrum when possible for transferring data. In [11, 14, 27, 46], delayed WiFi offloading is introduced to migrate data traffic from cellular networks to WiFi access points. However, such methods are not practical for most access patterns (any interactive app including web browsing and most streaming) as the point of using the smartphone is for data access that moment. Conversely, the de facto solution is upgrading the network to the next generation networks. The configuration details of both WiMAX and LTE technologies are summarized in [19]. Similarly, [1] provided an overview of the 4G evolution and categorized how such technologies can accompany a more user focused world of wireless.

4.3 WiFi Offloading in Practice

Figure 4.1 shows the average 3G and WiFi downlink traffic per phone per week across the study period. On average, the WiFi traffic takes approximately 30% of the total data consumption. The average WiFi traffic statistics excludes the data of devices with no WiFi usage. The underlying 802.1X WiFi infrastructure (primarily

due to expired passwords) can prevent a user from using any WiFi traffic during that entire week. Once the password is updated (passwords are required to change every 90 days), the phone can correctly use WiFi once again. We note though that an inability to authenticate does not preclude monitoring signal strength (via beacons) and that users do not experience intermittent 802.1X issues (the phones either authenticate or do not contiguously).

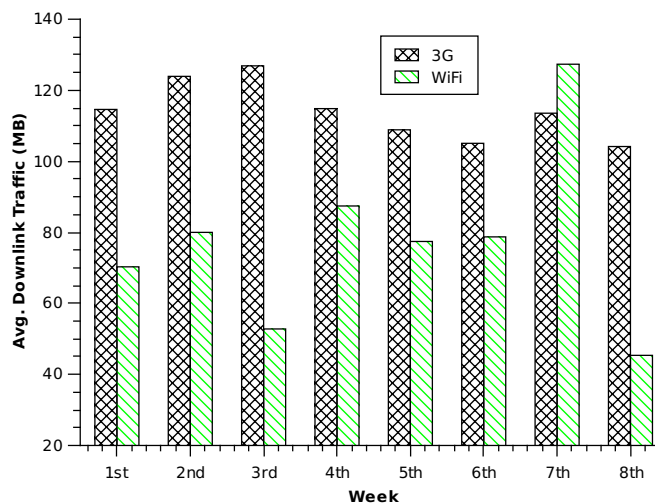


Figure 4.1. Average Downlink Traffic per Phone per Week

As noted in Section 4.1, the average WiFi usage is curiously lower than the average 3G usage aside from the week of spring break (likely on a single AP at home) where WiFi barely eclipses 3G usage. In Figure 4.2, we select a small subset of phones that exhibit a high degree of 3G usage versus WiFi and plot their respective average traffic patterns for a given day during the semester. From 5 PM to 6 PM (1700-1800h), the phones exhibit a degree of asymmetry, favoring 3G over WiFi. Figure 4.3 explores

this time period further noting the detected AP beacons but exceptionally low quality signal strengths over that time period. Hence, the phone naturally defaults to favor 3G over WiFi. Conversely, during the 10 PM to 11PM (2200-2300h) as noted in Figure 4.4 where the signal strength is better but not great, the ratio between 3G and WiFi changes to favor WiFi over 3G but not markedly so.

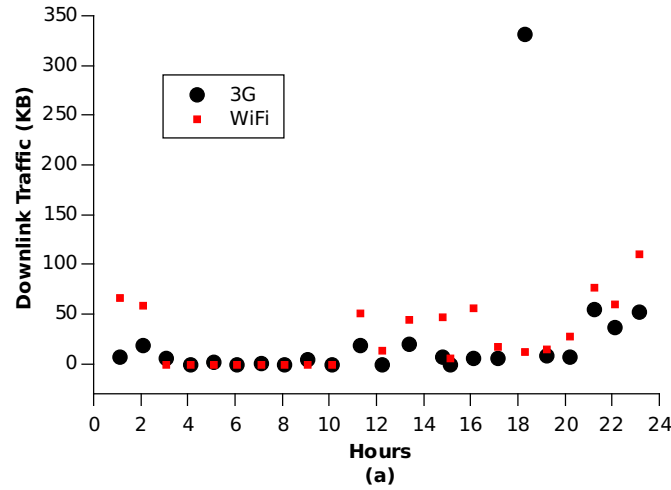


Figure 4.2. Switch between 3G and WiFi

To verify that the users in this subset were indeed active (using the screen), Figure 4.5 plots the average screen session lengths over that same selected day. The screen session time captures via event when the screen is turned on (timestamp recorded) and when the screen is turned off (timestamp recorded). While the users were more active in terms of the number of sessions during the later time period (10PM to midnight), the users actually used more data during the earlier time period. We also note that the spike did not correlate with a check in by the agent as such data has

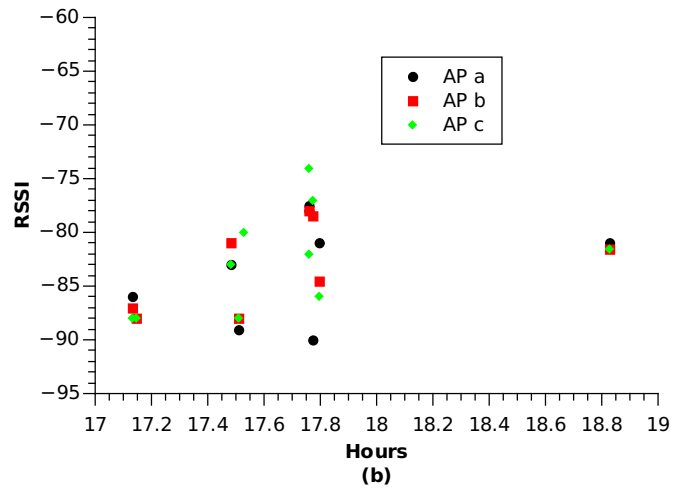


Figure 4.3. WiFi Signal Strength (1700-1800h)

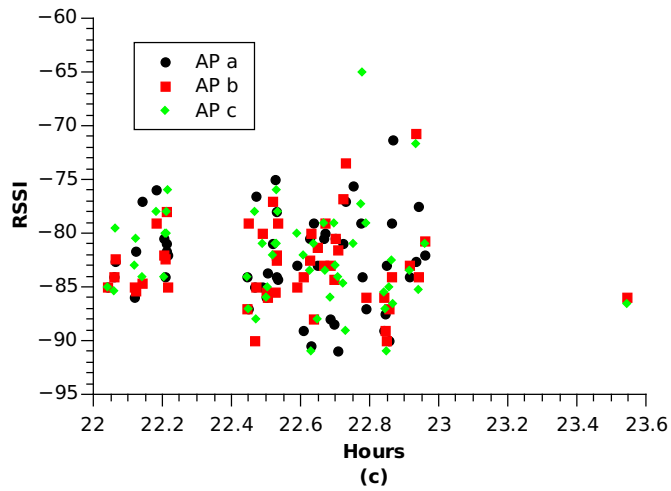


Figure 4.4. WiFi Signal Strength (2200-2300h)

been filtered from the results.

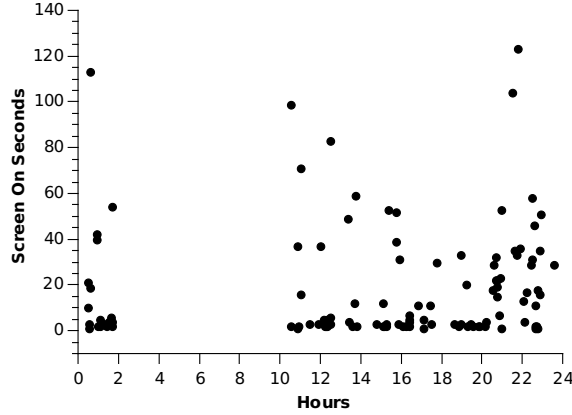


Figure 4.5. Screen On Duration in One Day

4.4 Comparison of User Behavior

We continue our analysis by exploring how user behavior changes if the user is able to get reasonable quantities of adequate WiFi smartphone coverage. Figure 4.6 shows the distribution of the percentage of WiFi downlink traffic as a ratio versus total traffic for one week (2nd week). The average WiFi consumption ratio is around 30%. From the figure, we see that there are approximately 30% of participants whose traffic is offloaded to WiFi for the week by more than 50%. Conversely, we see that there are nearly 20% of the participants that week who used no WiFi traffic for the reasons mentioned earlier (incorrect password, etc.).

For further analysis in the paper, we subdivide the 131 participants into three groups based on their WiFi downlink consumption ratio versus their total traffic

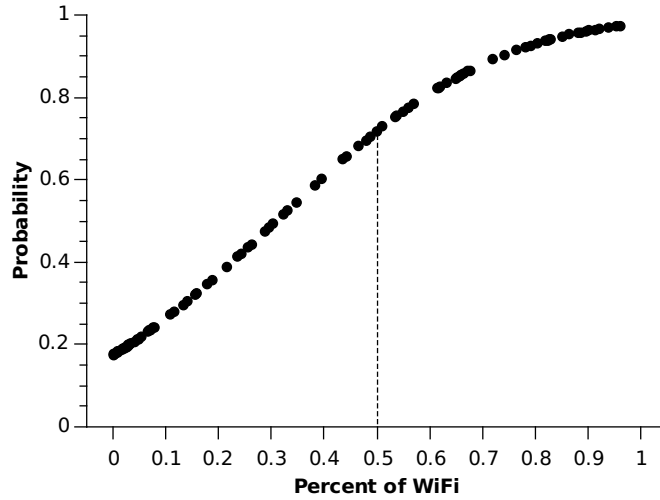


Figure 4.6. Distribution of Percent of WiFi Traffic

consumption: no WiFi, less than 50%, and more than 50%. We summarize how the participants break down into each group among the eight weeks in Figure 4.7. Similarly, the breakdown of males versus females is shown in Table 4.2. Although the numbers in the groups varied, the groups are reasonably well separated to analyze the participants in the various categories.

Figure 4.8 shows the average total downlink traffic per phone (3G+WiFi) of the different groups across the eight weeks. For the category of each particular participant, determination was computed on a weekly basis meaning that a user could move between categories. Interestingly enough, we begin to see a growing pattern of separation as the semester goes on and particularly so during the week of spring break. We posit that the week of spring break offered a much more consistent set of WiFi coverage versus what may occur on campus.

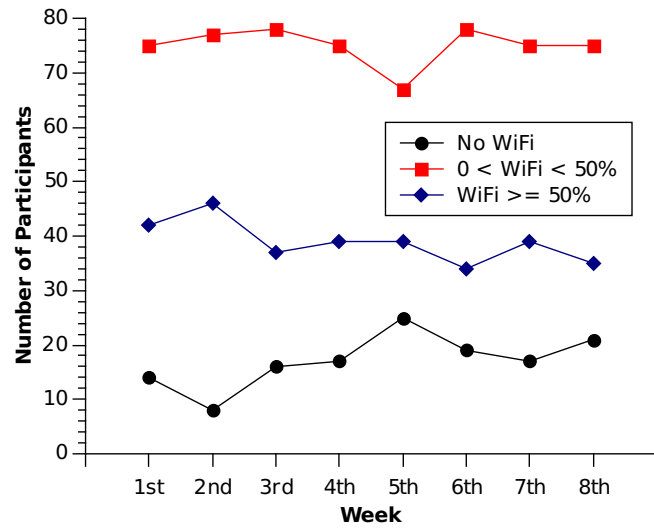


Figure 4.7. Numbers of Participant in Groups

TABLE 4.2

NUMBER OF MALES AND FEMALES IN GROUPS

| Week | No WiFi | | 0 < WiFi < 50% | | WiFi >= 50% | |
|------|---------|----|----------------|----|-------------|----|
| | M | F | M | F | M | F |
| 1st | 9 | 5 | 35 | 40 | 25 | 17 |
| 2nd | 6 | 2 | 35 | 42 | 29 | 17 |
| 3rd | 5 | 11 | 37 | 41 | 27 | 10 |
| 4th | 7 | 10 | 36 | 39 | 26 | 13 |
| 5th | 15 | 10 | 30 | 37 | 24 | 15 |
| 6th | 9 | 10 | 39 | 39 | 21 | 13 |
| 7th | 8 | 9 | 36 | 39 | 25 | 14 |
| 8th | 10 | 11 | 36 | 39 | 23 | 12 |

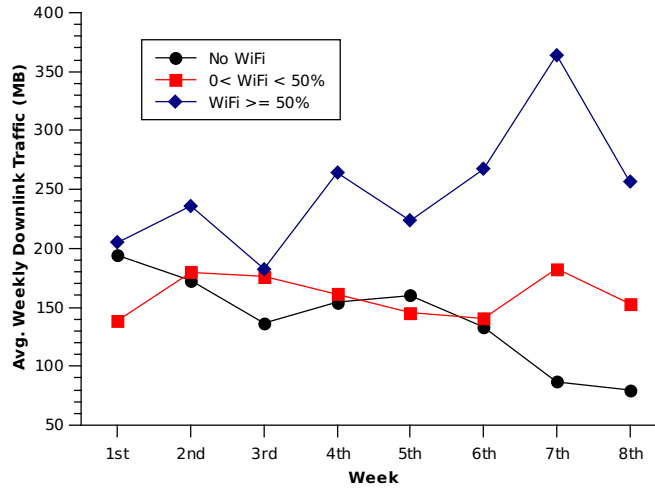


Figure 4.8. Weekly Average Total Downlink Traffic per Phone

4.4.1 Phone Usage

Related to usage, a secondary question arises with regards to the ratio of screen time to data consumption. The most straightforward way to indicate whether the person is using the phone is based on the phone screen status (on or off). When the phone screen is on, it is most likely because the person is actively using the phone (sending messages, checking e-mail, playing games, etc.). In order to explore the relationship between phone usage time and traffic usage, we collect the screen on session length on the phones by calculating the duration between screen on and screen off. On average, the screen session lengths vary from 10 seconds to more than 100 seconds. We categorize them into four ranges: (0, 30s), (30s, 60s), (60s, 90s) and more than 90s. Figure 5.17 lists the numbers of participants from three groups in different duration ranges. In the range of (60s, 90s) the number of people who use WiFi more than 50% is greater than the other two groups numbers. We calculate the weekly total screen on time and the corresponding daily average per phone as well. There are five ranges of daily average from less than half an hour per day to

more than 2 hours per day. Figure 4.9 gives an example (data from the 2nd week): compared with purely 3G users (no WiFi), the number of participants who use WiFi more than 50% in each range is much greater. The fact that the phone is more useful (faster access) may encourage such additional usage.

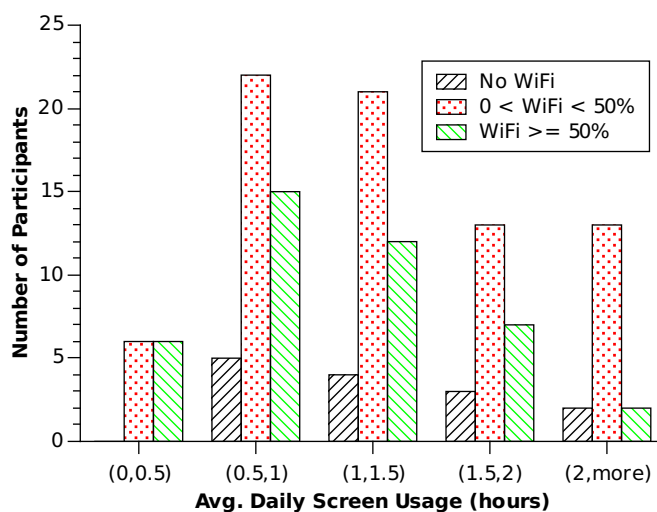


Figure 4.9. Avg. Daily Screen Usage of Groups

We further calculate the screen session duration across different time slots in a single day to analyze user behavior. As shown in Figure 4.11, we divide one day into four time slots: morning (7am-12pm), afternoon (12pm-5pm), night (5pm-10pm) and midnight (10pm-7am). As would be expected with a student population, usage in the morning is quite low relative to usage in the evening.

Finally, we explore the relationship of WiFi to 3G usage with respect to various other services on the phone including the number of text messages, total phone usage, average phone call length, etc. Table 4.3 shows the numbers across the entirety of the

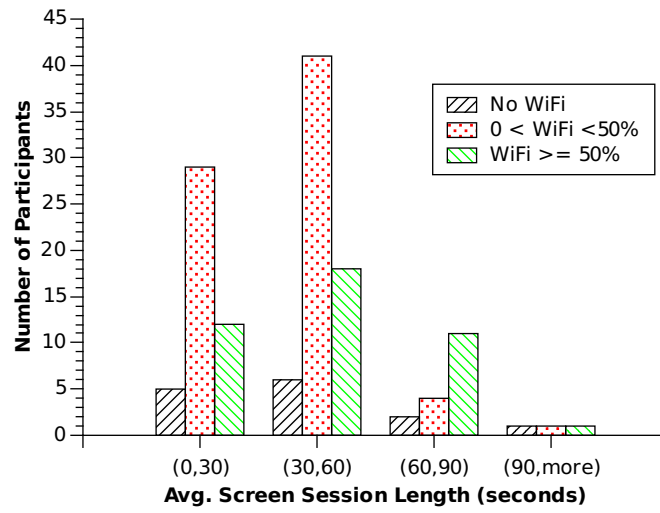


Figure 4.10. Avg. Screen session length of Groups

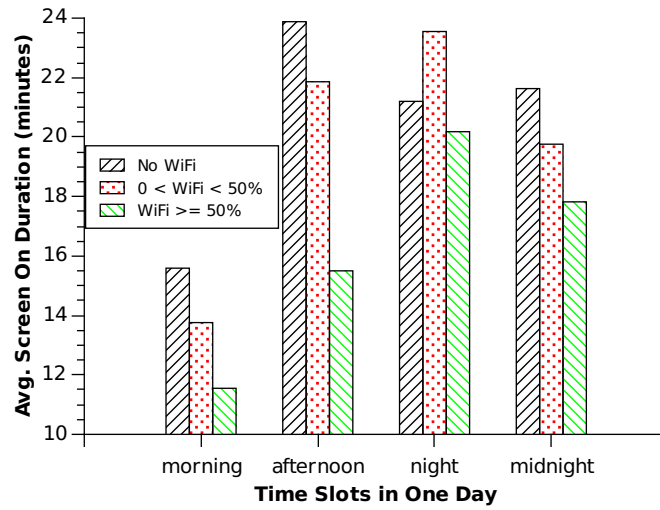


Figure 4.11. Avg. Screen on Duration in Different Time Slots

study with users categorized again on a weekly basis. While WiFi dominant users tended to use their screen for longer average periods of time, they tended to less frequently use text messages relative to the non-WiFi-dominant users (455 average text messages sent/received per week versus 522 average text messages sent/received per week). Text messages do not count against the data count for either 3G or WiFi usage.

TABLE 4.3

WEEKLY SMS/PHONE CALL COMPARISON

| Groups | No WiFi | 0 < WiFi < 50% | WiFi >= 50% |
|-----------------------------------|---------|----------------|-------------|
| Avg. Screen On Duration (seconds) | 39.33 | 43.93 | 48.76 |
| Avg. SMS All | 436 | 522 | 455 |
| Avg. SMS Sent | 209 | 249 | 225 |
| Avg. Number of Phone Calls | 26 | 31 | 24 |
| Avg. Phone Call Duration (hours) | 1.03 | 1.60 | 1.27 |
| Avg. Email All | 89 | 95 | 90 |
| Avg. Number of Browser Sessions | 102 | 81 | 67 |

4.4.2 App Usage

The network bandwidth and throughput has a profound influence on application usage preference. When the network provides better service, users tend to use more intensive services. We look into the data of installed application and application

traffic which are logged by the service as part of the agent (hourly application data usage).

TABLE 4.4

TOP APPLICATIONS CATEGORIES

| | No WiFi | $0 < \text{WiFi} < 50\%$ | $\text{WiFi} \geq 50\%$ |
|--------------|--------------|--------------------------|-------------------------|
| Top Downlink | Browser | Browser | Netflix |
| | Facebook | Facebook | Browser |
| | Zynga Words | Pandora | Pandora |
| | Amazon | Zynga Words | Facebook |
| | Twitter | Twitter | Dictionary |
| Top uplink | Google Maps | Google Maps | Pandora |
| | PhoneMonitor | Pandora | Google Maps |
| | Browser | PhoneMonitor | Glu Games |
| | Facebook | Browser | PhoneMonitor |
| | Gmail | Facebook | Browser |

In Table 4.4 we summarize the top 5 applications appeared in the each group based on their total downlink or uplink traffic. For the phones using WiFi more than 50%, streaming apps such as Netflix and Pandora always are the top applications. For the participants using 3G only, their main activities on the phone are web browsing and email checking. Figure 4.12 demonstrates the weekly downlink traffic of the top 10 apps in different categories which include video & audio (Netflix, Youtube, Pandora

and Google Music), social (Facebook and Twitter), tools (Browser, Gmail, Google Maps, Dictionary) and games (Zynga and Glu applications). Similarly, we present the results of uplink traffic in Figure 4.13. In both graphs, the traffic of video and audio applications increases dramatically when the participants use more WiFi. An interesting question emerges if the streaming consumption is user specific or rather becomes enabled by WiFi speeds implying that WiFi with continuous LTE would show the same usage patterns.

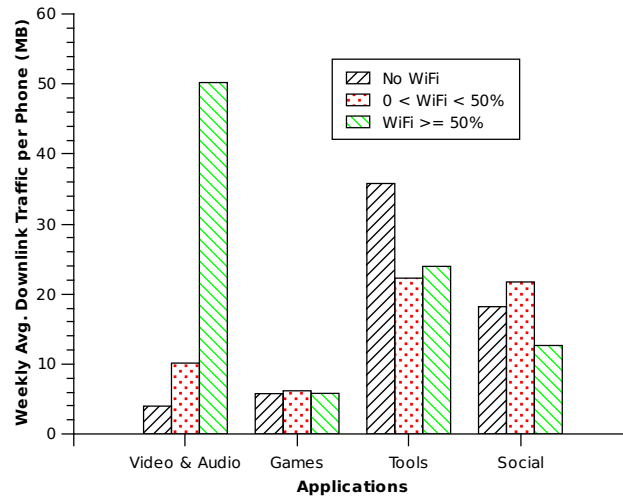


Figure 4.12. Weekly Application Downlink Traffic

4.5 Summary

To that end, we make the following contributions in this chapter:

- The net result of our findings over an eight week period of our smartphone

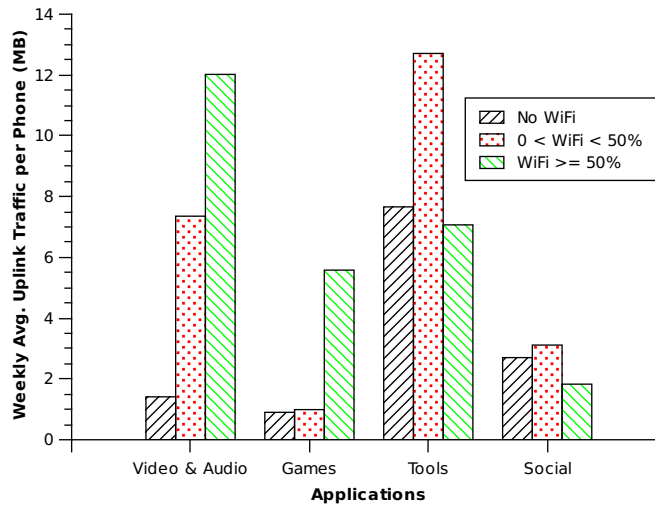


Figure 4.13. Weekly Application Uplink Traffic

usage data shows that the gaps in WiFi coverage for smartphones temper the perceived potential gains by WiFi offloading. More attention should be paid to WiFi coverage to examine said coverage through the lens of a typical smartphone rather than a typical laptop or tablet device.

- We explore the relationship of WiFi data consumption and phone usage time deduced by screen session length. Moreover, our work includes accommodations for considering SMS/phone call/browser impacts with regards to phone usage.

Although individual handsets (ex. iPhone) may possess better WiFi characteristics, the heterogeneity of available handsets implies that it is exceptionally likely there will be significant variations in the ability of phones to take advantage of WiFi offloading. While certainly any offloading is desperately appreciated due to capacity shortages, WiFi offloading may be further clouded by bottlenecks in the next hop following the WiFi link as well. Therefore, we believe that additional scrutiny is needed with respect to the end benefits that can arise from WiFi offloading. In next chapter, we discuss the potential of opportunistic relaying which is complementary to WiFi

offloading and demonstrate that the concept of opportunistic networks is powerful in face of data tsunami.

CHAPTER 5

Opportunistic Relaying

5.1 Background

Wireless network providers are under tremendous pressure to deliver unprecedented amounts of data to a variety of mobile devices. A powerful concept that has only gained limited traction in practice has been the concept of opportunistic networks whereby nodes opportunistically communicate with each other when in range to augment or overcome existing wireless systems. Although opportunistic networking has received more attention as of late with the rise of various point-to-point technologies (WiFiDirect, LTEDirect) directly embedded in user devices, traction in terms of significant adoption remains elusive as the mere existence of point-to-point wireless technology is insufficient. Rather, infrastructure support (albeit at a software or protocol level) is still required that is impeded by the lack of shared, longitudinal data taking a serious look at the potential for opportunistic relaying amongst actual mobile device users to justify said investment in infrastructure. The actual gathering of said data though is in and of itself quite difficult, requiring one to overcome numerous issues with respect to privacy, cost, scale, and quite simply the amenability of the device to acquire said data. Thus, in the absence of real data and the difficulty in acquiring such data, opportunistic networking research has largely espoused synthetic or theoretical explorations of user mobility. In this Chapter, we demystify the opportunities that exist for opportunistic relaying by bridging the gap with our large scale dataset.

5.2 Related Work

The notion of what constitutes an *opportunistic communication* encompasses a wide variety of research and standardization efforts within the networking community. From the more traditional perspective, opportunistic relaying represents wireless nodes taking advantage of emergent opportunities to relay data towards its eventual destination originally espoused by [24] and expanded up by numerous others [2, 6, 31]. More recently, considerable efforts have emerged where IEEE 802.11-based networks (WiFi) are viewed through an opportunistic viewpoint, espousing the delay of packets to favor the potentially faster WiFi over the congested cellular link. More aggressive variants include ‘Opp-Off’ as proposed in [15] which intentionally delays the delivery of information over cellular networks and offload it through the ‘free’ opportunistic communications. Similarly, in [11], another delay-tolerant network (DTN)-like architecture called ‘MADNet’ is proposed by integrating WiFi networks and mobile-to-mobile Pocket Switched Networks (PSN) with cellular networks. With the ability to leverage the availability of inexpensive 802.11, both Opp-Off and MADNet are able to conduct real-world experiments on live systems, albeit on a limited scale and time period.

A key foundation for the evaluation of opportunistic networks arises from the characterization of inter-contact times [8, 9, 18, 26, 39]. The work in [9] was one of the first works to highlight the importance of inter-contact for studying opportunistic networks and analyze the features of aggregate inter-contact times. In the work of [8, 26], the mobility models proposed for opportunistic networks aim at reproducing the aggregate power-law distributions. The mobility models in turn provide powerful abstractions for the evaluation of various theoretical properties of opportunistic networks that is critical for understanding general limits of the overarching relay protocols. In contrast from these studies on statistical patterns of human mobility, we focus on the analysis to take in not only the proximity into consideration but also

other elements such as traffic needs and battery influences which have to the best of our knowledge, not explored in a unified manner in the literature.

5.3 Evaluating Practical Relaying

The dataset outlined in Chapter 2 provides a fascinating opportunity to evaluate the actual prevalence with respect to opportunistic relaying in the ‘wild’ evaluating not only the prevalence itself but when intra-study proximity is involved, significant explorations with respect to the mutual benefits of relaying or collaborative efforts. For the purposes of evaluation, we are concerned with two types of opportunistic networking, namely relaying and collaboration. In the first case of relaying, a mobile node (MN_i) might relay the traffic from mobile node (MN_j) when MN_j has a weak / non-existent wireless signal or MN_i has a good signal to a preferred wireless medium (ex. WiFi) while MN_j does not. In the second case of collaboration, MN_i and MN_j work together to overcome a lossy channel either by the use of striping across both nodes (ex. auxiliary relaying of TCP ACKs [32]) or striping across different wireless mediums (WiFi + cellular). Opportunistic communications would be provided through mobile-to-mobile communications (Bluetooth, WiFiDirect, LTE-Direct). Both approaches to opportunistic relaying would require modifications to the wireless infrastructure as well as security concerns to establish any mobile-to-mobile communication which are beyond the scope of this paper. Rather, we focus on the more fundamental question of *does enough opportunity exist* and if so, *is it of reasonable quality to make it worthwhile?*

TABLE 5.1: FRAMEWORK FOR EVALUATING AVAILABILITY, PREVALENCE, STABILITY, RECIPROCITY

| Criteria | | Term | Description |
|--------------|------------------------|----------|--|
| Availability | Sufficient Signal | SS | ‘good signal’ to indicate proximity |
| | Symmetry | Sym | Time when proximity is symmetric, has SS vs. time powered on |
| | Diversity | Div_n | Time with at least n nodes with SS detected vs. time nodes detected |
| | Residual Battery | RB | Time with proximity of other nodes with SS and enough battery to do relaying vs. time with proximity |
| Prevalence | Time in Proximity | TIP | Time with proximity vs. time powered on |
| | Intercontact Time | IT | Time between two successive contact/proximity periods |
| | Effective Utility | EU | Time with proximity of other nodes with SS , Sym , traffic to send vs. time with traffic to send |
| Stability | Duration | D | Number of consecutive contact durations vs. instances of contact at MN_i |
| | Node Strength | NS | Number of total appearances by n most common peers |
| | Total Appearances | TA | Number of appearances of a device vs. total device appearances at MN_i |
| | Continuous Appearances | CA | Number of consecutive appearances of a device vs. appearances of multiple days at MN_i |
| Reciprocity | Need Service | $Serv$ | Time when no auxiliary WiFi is present and traffic demand exists vs. time powered on |
| | Offer Assistance | $Assist$ | Time ween detected with SS vs. time powered on |
| | Ratio Need vs. Assist | R_{NA} | Time of $Serv$ vs. Time of $Assist$ |

To that end, we seek to answer the following questions in this section of the paper summarized in terms of metrics in Table 5.1:

- *Availability*: How is proximity defined and what criterion should exist to enable opportunistic communication with respect to inter-node communications? What can be considered a ‘good’ opportunistic link in terms of the underlying physical link quality and at what granularity is such data recorded? How does the residual battery influence the potential for relaying?
- *Prevalence*: What is the frequency that a mobile device detect other devices in proximity? To what extent is the detection symmetric for intra-study participants? How does the inter-contact times of nodes compare to prior work? Do opportunities exist when needed (traffic demand) making them useful versus simply available?
- *Stability*: To what extent are discovered opportunities stable enough that an overarching security mechanism could complete or relevant data exchanges could take place? Are there consistent patterns with regards to nodes appearing more common than others allowing expedited trust?
- *Reciprocity*: Even if opportunities are shown to be prevalent and stable, to what extent are the relationships likely to be reciprocal, namely both MN_i and MN_j will benefit on average equally in terms of both receiving assistance and giving assistance? When energy is factored in as a consideration for nodes not being willing participants, does any sort of prevalence or reciprocity disappear? Finally, reciprocity if it exists, does it exist exclusively in only longer-time scales or does it exist on reasonably short-time scales as well?

For the wireless network, we first begin by defining several key attributes of the data. Each mobile node (MN_i) is considered to have a discrete set of samples at

periodic intervals capturing the proximity of nearby nodes (Bluetooth), the signal strength from each detected Bluetooth node as received at MN_i , (ex. $MN_j \rightarrow MN_i$), each detected AP and also the signal strength for each AP as detected by MN_i by virtue of the beacon signal strength. Each time slot is assumed to be five minutes long (for the purposes of assessing if the phone was on or off) though durations of contact are calculated using per-minute measurements. The term *Good RSSI* refers to thresholds for RSSI values that indicate the potential for excellent performance although in practice such performance may vary. Performance with respect to each node is normalized unless explicitly noted to the actual time that the phone was on (powered on, agent running) for that particular day or period.

5.3.1 Availability

The first question to pose is how to define proximity. We begin with the most basic question with regards to the signal strength of discoverable Bluetooth devices and detected APs and further refine our queries to explore various aspects of availability through the Availability criterion of the APSR framework.

5.3.1.1 Proximity

As demonstrated in Chapter ??, we use Bluetooth as the proximity mechanism for multiple reasons. First, the effective range of Bluetooth typically is on the order of 10m or less. In contrast with other point-to-point technologies such as Wi-FiDirect and LTEDirect which have the potential for significantly longer ranges, Bluetooth in effect represents a minimum or floor to the potential for opportunistic communications as the various *Direct technologies would easily be able to cover the 10m effective range of Bluetooth. Second, although Bluetooth is reasonably common amongst smart phones, the locking of a device as being discoverable is fairly uncommon, largely for reasons of security. Despite the somewhat uncommon nature

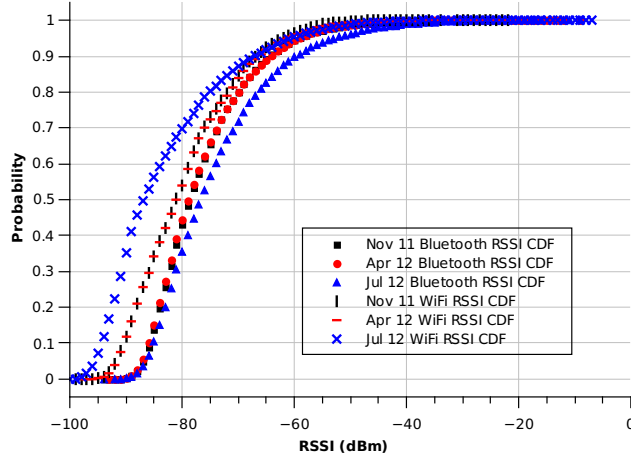


Figure 5.1. Bluetooth and WiFi RSSI Values ECDF

of discoverable Bluetooth (most devices require specific action to make themselves discoverable), a robust finding even with the limits of Bluetooth discoverability represents again a floor or minimum potential available. Critically, the act of gathering Bluetooth discoverable devices is reasonably power efficient allowing one to gather available devices at a reasonable pace (with associated signal strength for Bluetooth) without pairing and while still having a reasonable device battery life for the study.

Notably, the mere existence of the device being discoverable does not necessarily imply that the point-to-point link will be suitable for opportunistic communications. Figure 5.1 shows the ECDF of the Bluetooth RSSI values in different months and more than 60% of the values are larger than -80dBm. In particular, the data of July 2012 consider all the records including the Bluetooth devices outside the project and APs not under university control. Based on the work in [41], the RSSI value $-80dBm$ is used as a threshold to indicate good Bluetooth RSSI for direct mobile-to-mobile communication.

While Bluetooth is used in the first part of the section to assess proximity, to what extent could we infer proximity if we expanded our search to consider common

detected WiFi APs. For instance, in [35], WiFi signal strength is used to indicate proximity when two devices are associated to the same AP. We further restrict the characterization to infer proximity if two mobile nodes detect two or more same access points. Barring the two APs being situated nearly on top of one another (note that university APs are filtered down to disambiguate virtual APs), one can reasonably infer that two nodes sharing the same two or more APs could be in WiFiDirect and / or LTEDirect range. Figure 5.2 shows the ECDF of Bluetooth proximity percentage and WiFi proximity percentage. The percentage of time represents the percentage of time where an opportunity might exist for collaboration either within Bluetooth proximity or WiFi proximity. Interestingly, the WiFi result conveys less proximity than the noted Bluetooth proximity. The discrepancy can be traced to one primary culprit, namely the poor signal reception of smartphone WiFi adapters as originally noted by [30]. Notably when compared to reference laptops or tablets, the work in [30] notes a roughly 10 dBm signal penalty for the smartphone. The net result is in addition to the detection of WiFi APs being hampered, reasonable placement of APs and auto-tuning of AP strength would result in significantly reduced probabilities of multiple mobile devices being able to detect one or more APs. In Figure 5.3, we compare the results of WiFi proximity and Bluetooth proximity in April of 2012 and November of 2011. The false positive means the device detects other(s) in WiFi proximity which are not detected by Bluetooth proximity. On the other hand, the false negative means that the device does not detect some device in WiFi proximity but does detect the device in Bluetooth proximity. Notably, the false positive in the context of WiFi may not be a false positive but the false negative most certainly represents an incorrect result due to the aforementioned signal holes noted in [30].

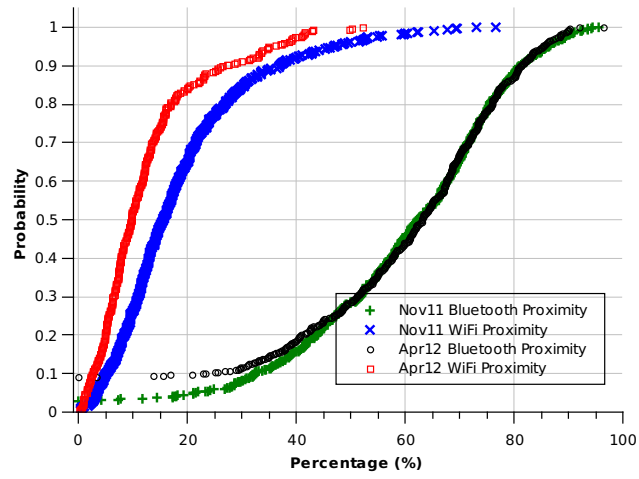


Figure 5.2. Bluetooth Proximity and WiFi Proximity ECDF in April 2012

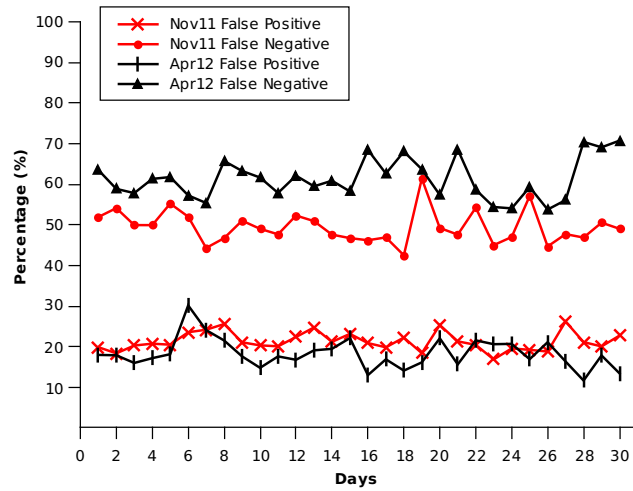


Figure 5.3. Comparison of Bluetooth Proximity and WiFi Proximity

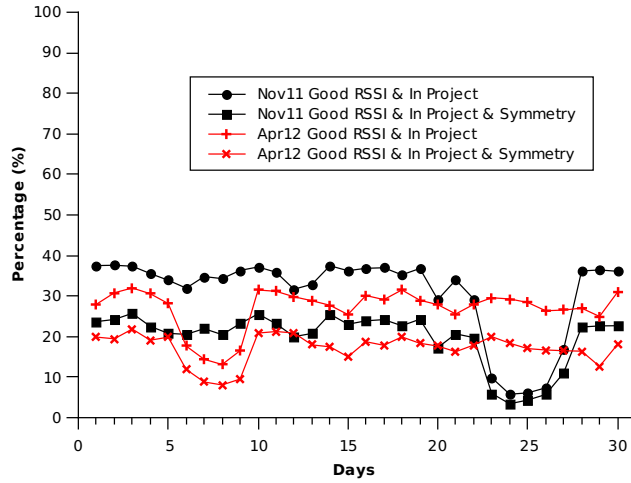


Figure 5.4. Symmetry of Bluetooth Proximity

5.3.1.2 Symmetry

Symmetry is defined as both nodes seeing each other in proximity. For WiFi proximity, such symmetry is 100% based on its definition. However, for any two study participants in Bluetooth proximity, there exists the opportunity to determine if detection symmetric exists, namely do both nodes detect each other and even if both nodes detect each other, do both nodes have a ‘Good RSSI’ to one another ($MN_i \rightarrow MN_j$ and vice versa)? In Figure 5.4, the symmetry of Bluetooth proximity within study with ‘Good RSSI’ is illustrated by using the data in November 2011 and April 2012. Notably, roughly one third of the opportunities disappear due to the consideration of symmetry. Many reasons may cause the asymmetry such as environment inference and it is quite expectable in practice.

5.3.1.3 Diversity

While symmetry represents just one part of the puzzle, an related factor for the opportunity to be useful is with regards to the diversity of opportunities available.

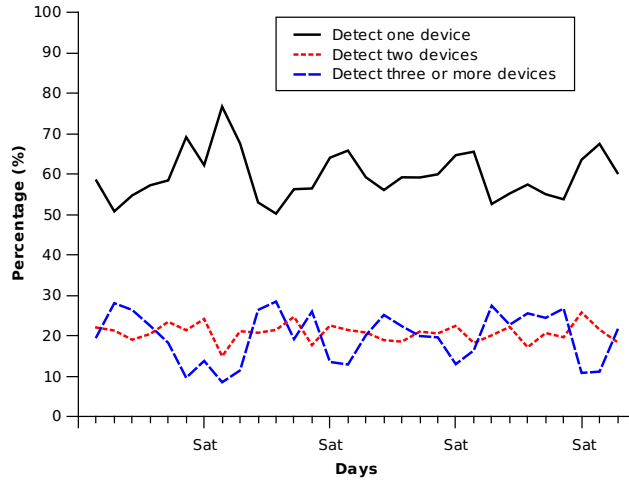


Figure 5.5. Distribution of Proximity Diversity (In-Study) in April 2012

Relaying selection plays little role when there exists only a single device to choose from for relaying. Figure 5.5 shows the breakdown of detected devices in proximity through both Bluetooth and WiFi in April 2012 representing only cases where proximity occurred. In the figure, nearly 60% of time has only one peer is detected among all the possible opportunistic time slots. Meanwhile, there is nearly 20% of the time when two opportunities are detected and finally 20% when three or more opportunities are detected. In Figure 5.6, more details about the number of detected devices in April 2012 is illustrated with a diurnal distribution across the same month breaking down only cases where proximity was detected, i.e. 50% of daytime detections involved only a single device in the study. During the daytime, the chance to detect two or more devices in proximity is higher than other durations. Again, day time offers an increased potential for diversity in large part due to the increased mobility around the campus.

As mentioned before, it is fairly uncommon making the device Bluetooth discoverable. In order to reveal the diversity of potential relaying nodes, we investigate the

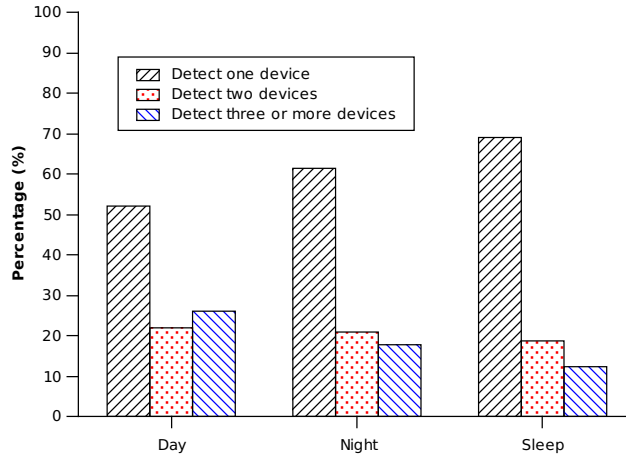


Figure 5.6. Diurnal Distribution of Proximity Diversity in April 2012

types of available Bluetooth devices in proximity based on the Bluetooth data across the study. There are around 49% of distinguished Bluetooth discoverable devices are smartphones. Interestingly, 10% of the devices are Apple products: 80% of them are iPhone/iPod/iPad, and rest of them are Macbook laptops.

5.3.1.4 Residual Battery

The availability of opportunistic communication depends on not only the existence of devices in proximity, but also the battery status of detected devices, i.e. making sure the relaying devices have enough residual battery for the following communication. As discussed in [34, 49], residual battery is an important consideration factor for relaying selection. Even the device is detected in proximity, it cannot be utilized for relaying if the battery is too low to do transmission. At the same time, it is the basic premise for device having enough energy for communication when it detects other devices in proximity. Due to the battery data collection limitation, we focused on the devices within the project and analyzed the percentage of Bluetooth

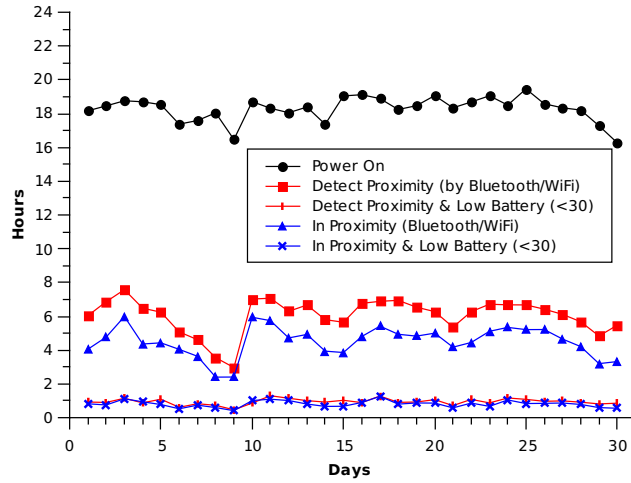


Figure 5.7. Average Daily Low Residual Battery Time in April 2012

proximity when the battery level is low (i.e. level value smaller than 30) and the result is shown in Figure 5.7. In April 2012, daily power on time is around 18 hours on average and the low battery time when the devices are in proximity scenarios is nearly 1 hour. Meanwhile, the average time of detecting proximity and being detected as proximity is around 7 and 5 hours. Therefore, the percentage of low battery time is relatively small and

5.3.1.5 Evaluating Availability

To summarize, we conclude availability of opportunistic communication with a brief discussion of metrics used to evaluate availability.

- **Sufficient Signal:** While the use of all available proximity potentially may be of interest, only useable mobile-to-mobile links are of interest. An explicit filter likely unique to each particular handset or device should be employed.
- **Symmetry:** Filtering the data further, agents monitoring the signal strength on both sides of the dyadic pair can assess if symmetry exists both with respect

to detection (MN_i detects MN_j , vice versa) and sufficient signal strength.

- Diversity: When there are more than one relaying nodes nearby, it is easier for the device to choose one of the most appropriate nodes with the consideration of benefits on both sides.
- Residual Battery: As an important criterion for relaying selection in various relaying protocol, battery of mobile devices is essential for availability. When the residual battery of mobile devices is low, it is difficult for the devices to do opportunistic communication.

5.3.2 Prevalence

Availability analysis exhibits the possibility to detect other devices in proximity. A more important question is to what extent such opportunistic communication opportunities exist in practice. We begin with the most basic question with regards to the prevalence of discoverable Bluetooth devices and further refine our queries to explore various aspects of prevalence through the Prevalence criterion of the APSR framework. We posit several questions that include: (1) the extent to which discoverable Bluetooth devices exist, (2) the extent to which WiFi proximity exist, (3) and the extent to which diurnal (daily) patterns play a role in discoverable devices. We further continue the explorations examining the prevalence of proximity to determine if the prevalence of such opportunities are still useful, namely (1) are such opportunities only available when there is no traffic and (2) how frequent do opportunities exist to augment access to ‘better’ wireless channels (ex. WiFi).

5.3.2.1 Time in Proximity

Based on the 15 months data, Figure 5.8 illustrates the average weekly Bluetooth proximity percentage with different restrictions across the study period. Various pe-

periods of interest representing various break periods are labeled in the graph. Further restrictions are placed on the data to filter for ‘good RSSI’ and candidates for opportunistic communications to exist only within the project (study). Notably, when completely unrestricted opportunities exist on the average of nearly 60% of the time when on campus, falling only slightly when good signal strength is added as a restriction. Even intra-study opportunities are quite prolific despite the fact that the 200 study participants represent less than one tenth of the freshmen class and less than 2.5% of the overall university student population. With the threshold of -80dBm for the ‘Good RSSI’, there still exists more than 50% of the time slots, a fairly insubstantial drop from the raw detected Bluetooth devices. The slight drop from the fall of 2011 to the fall of 2012 can largely be attributed to students joining their respective majors and having less shared coursework versus their freshmen year. Moreover, we compare Bluetooth proximity with and without consideration of symmetry. Similarly as the results in Figure 5.4, the one third drop rate due to asymmetry is quite stable across the study. Although the lack of symmetry does temper the earlier findings of prevalence, we note that Bluetooth is most appropriately viewed as a floor for potential interactions. The potential for asymmetry though between nodes does represent a consideration that merits further attention.

Given that the students were selected from a core group of six dormitories, a natural skepticism should emerge from the data with regards to the diurnal effects of proximity, namely proximity does little good if it only occurs at night when the phone is not otherwise being used (ex. sleeping in the dorm room). Hence, we analyze the diurnal distribution of Bluetooth proximity by dividing one day into three parts: day time (8am to 4pm), night time (4pm to 12am) and sleep (12am to 8am). In this way, we are able to investigate the impacts of time durations on such proximity. Figure 5.9 shows the diurnal distribution with the twin restrictions of good RSSI and intra-study detected proximity only. The daytime proximity percentage from 8am to

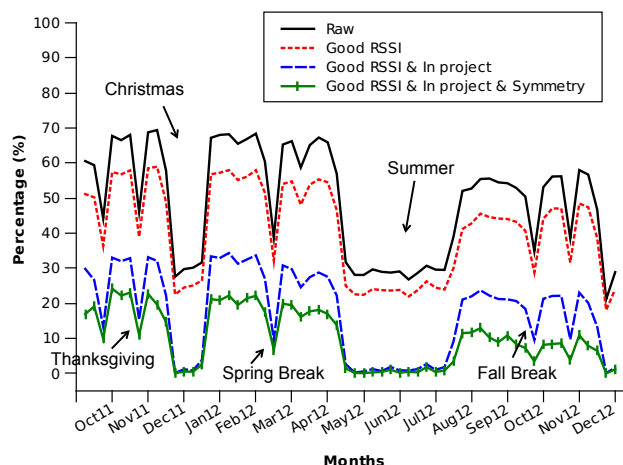


Figure 5.8. Bluetooth Proximity with Different Restrictions

4pm is larger than the value during nighttime and almost equal to the value of sleep time. The separation from in the fall of 2012 largely follows the classroom / major separation as noted in the raw and restricted graphs.

Beyond coarse diurnal metrics, significant day of week effects can also exist. Course patterns can vary dramatically between the Monday / Wednesday / Friday courses and Tuesday / Thursday courses. Similarly, weekends are much more likely to be quite diverse from normal weekday patterns. Figure 5.10 shows the daily Bluetooth proximity and diurnal distribution in April 2012 for a more detailed comparisons among the weekdays with each respective Saturday noted on the x axis. The peak values always appear at the beginning of the week and decrease slowly during the week. During the Easter holiday (April 6th - April 9th), the opportunities for proximity shrink dramatically as the students travel home for Easter break. Meanwhile, more Bluetooth proximity appeared in the later hours (broadly defined as sleep time) than daytime hours during weekend which indicates the participants spent more time in closer proximity to others during those time periods largely as a byproduct of social relationships.

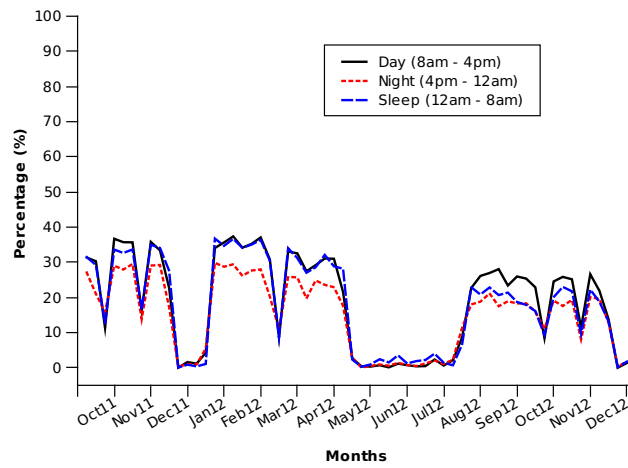


Figure 5.9. Diurnal Distribution of Bluetooth Proximity

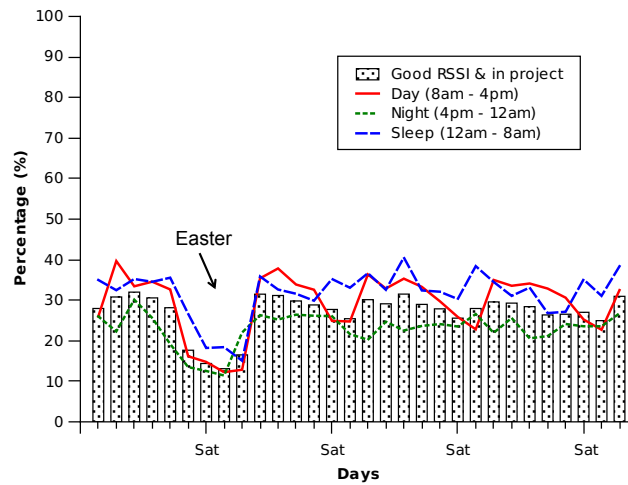


Figure 5.10. Diurnal Distribution of Bluetooth Proximity in April 2012

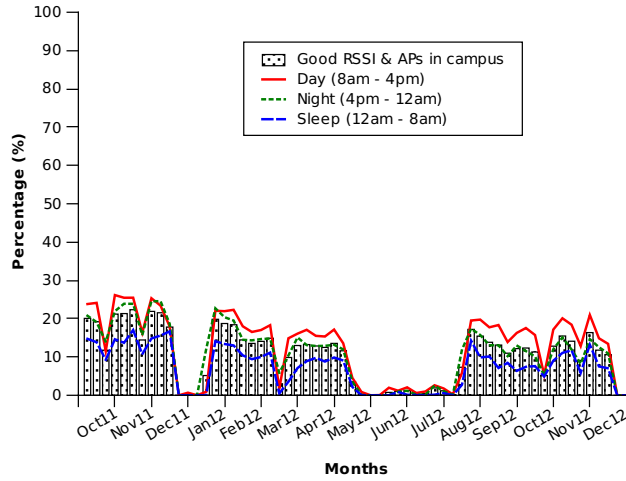


Figure 5.11. WiFi Proximity

For WiFi proximity, we analyze the number of time slots when the device shares at least two of the same access points with another device (in-study devices only) and we calculate the percentage of such time slots. Figure 5.11 demonstrates the WiFi proximity and its diurnal distribution across the study period. During the holidays and summer, the percentage is almost zero as only university APs are considered for WiFi performance. Compared to other periods, WiFi proximity during daytime from 8am to 4pm represents a slight uptick though not statistically significant variation over other periods.

Taking a bit more of an optimistic stance, we combine both WiFi proximity detection and Bluetooth proximity detection to create a combined proximity. Note that earlier proximity measurements used either exclusively Bluetooth or WiFi. Broadly defined, in one specific time slot, if the device can either detect other devices within the project with good Bluetooth RSSI or share at least two same APs in the campus with good WiFi RSSI, this time slot is counted as one of the slots when the device is in proximity with another device (intra-study only). Figure 5.12 illustrates such

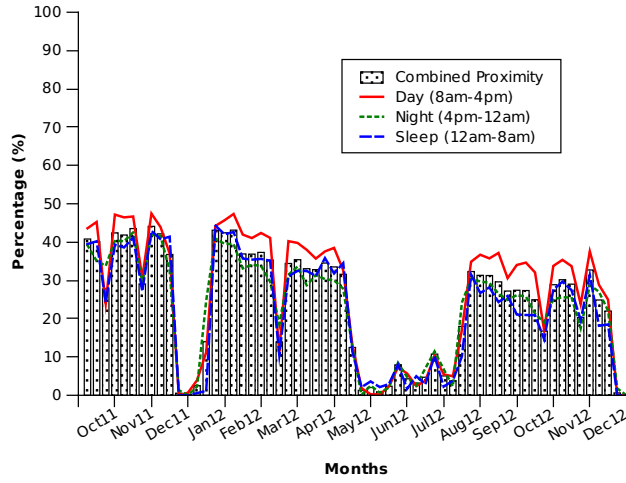


Figure 5.12. Combined Proximity (Fused Bluetooth / WiFi Detection)

proximity across months. Compared to Bluetooth proximity only or WiFi proximity only, the percentage is increased to more than 40% trailing to roughly 30% by the end of the study, a nearly 10% increase over Bluetooth only considerations. The diurnal distribution keeps the same trend and the proximity percentage during the daytime being is the highest among the respective time periods.

5.3.2.2 Intercontact Time and Beyond

Intercontact time is defined as the time elapsed between two successive contact periods for a given pair of devices [9] and it is another important characteristic of prevalence. For each pair of devices, we compute the intercontact time as the time takes before the pair meet again. Figure 5.13 exhibits the empirical distribution of the intercontact times obtained among different months. On average, the intercontact time is around 1000 minutes which is more than 16 hours. The distribution varies only slightly with the earlier time periods (Nov 2011) exhibiting a shorter intercontact time, in large part due to more shared coursework, reinforcing the findings

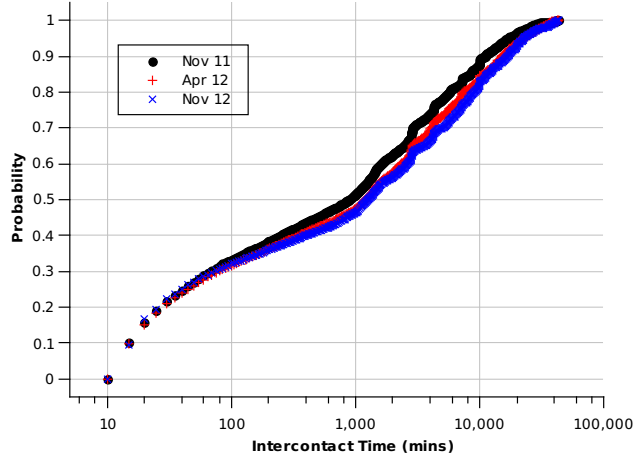


Figure 5.13. Intercontact Time ECDF

from earlier graphs.

5.3.2.3 Measuring Effective Utility

The final aspect for prevalence is the notion of effective utility, namely that the opportunity must occur when it is actually needed. From a general perspective, that need can happen when the following two conditions are satisfied: a) the node MN_i has some traffic to receive or send but no direct / poor connection is available b) there is some other node MN_j in proximity which can work together with the node. We analyze how often the devices really need relaying in the following three types of context.

In order to find the appropriate range values for reasonable WiFi performance, we conducted experiments to evaluate the cutoff points of WiFi RSSI on the particular Android handset employed (Nexus S). Channels were selected to be orthogonal to campus WiFi deployments (via a specially marked research channel space for our building) with distance, orientation, and other factors varied to get a wide variety of channel characteristics. Downloads were conducted over 100 times for each particu-

lar configuration and the results explored for throughput. Based on the experiment results of AP RSSI values and the corresponding throughputs, we note that -80 dBm can be the threshold to indicate good link quality with throughputs dramatically ramping up shortly after -80 dBm. At the same time, according to the RSSI distribution in Figure 5.1, there are more than 40% of records showing that APs in campus having RSSI values larger than -80 dBm. Therefore, we use -80 dBm as the good RSSI threshold to indicate good WiFi connection.

The first context is when the device has traffic but does not have good WiFi connection. Figure 5.14 includes two sets of results in Nov 2012: one is the percentage of having traffic and the other is the percentage of having traffic but the device does not have good WiFi connection (all the detected access points have signal strength less than -80dBm). Based on the results, there are more than 80% of time slots have traffic while 30% among them do not have good WiFi connection to do the transmission. Without WiFi, the traffic goes through the mobile link which further imposes pressure on mobile networks. However, relaying can work as an alternative to do the traffic offloading.

The second context is from the aspect of WiFi RSSI. With the prevalence of WiFi, it is interesting to investigate how often devices are not be fully covered by WiFi, i.e., detect no WiFi access points with RSSI larger than -80dBm. In such case, relay is an option. In Figure 5.15, nearly 40% of the time the devices do not in good WiFi environment on average. Among these time slots, there are around 20% of them do have relaying node(s) around.

In the third context, we consider the further benefits which relaying may produce. Besides WiFi and mobile network, relaying provides the third option to transmit traffic. When there is traffic need, the device may use relaying even it has established WiFi connection. Such cases happen when the relaying nodes have the content the device requests or the relaying communication has better link quality. Based on the

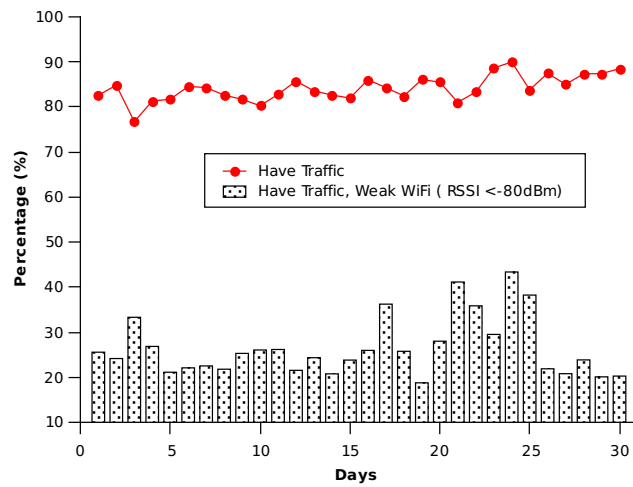


Figure 5.14. Weak WiFi Signal & Traffic

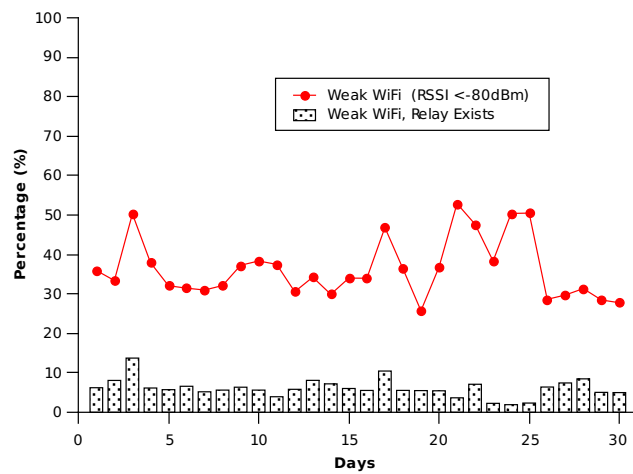


Figure 5.15. Relaying & Weak WiFi Signal

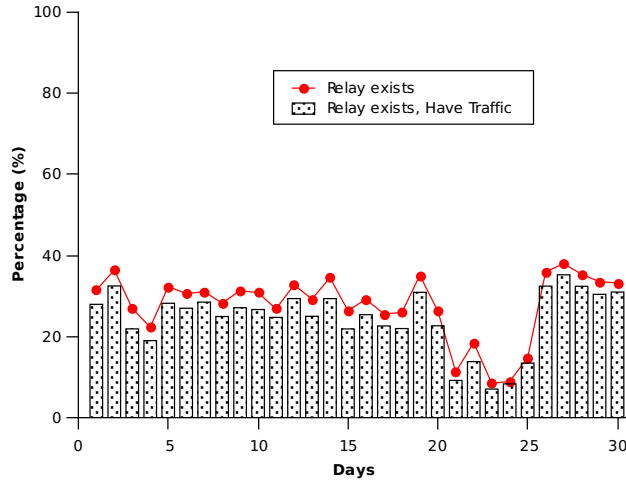


Figure 5.16. Traffic & Relaying

results shown in Figure 5.16, the percentage is pretty high when relaying exists and there is traffic at the same time. Therefore, relaying can provide great opportunities for devices to send and receive traffic, not only in a passive way, but also in a proactive way.

5.3.2.4 Evaluating Prevalence

To summarize, we conclude our explorations on prevalence with a brief discussion of metrics and conditions used to evaluate prevalence, i.e. the potential for opportunistic communications.

- Time in Proximity: It is the most important metric for prevalence and the premise for opportunistic communication. Without enough time in proximity, any relaying protocols cannot be deployed in practice. Through the analysis of data collected from nearly 200 smartphones in the campus environment, we demonstrate such prevalence is significant.
- Intercontact time: As mentioned in the last section, intercontact time is the

time duration between two continuous meets for a given pair of devices. It is another characterization to indicate the frequency with which data can be transferred between networked devices. While relevant for approaches such as DTNs, intercontact time has less relevance for single hop relaying or collaborative opportunistic communications.

- **Effective Utility:** The prevalence of opportunities offer little utility to the handset if said prevalence occurs when the device has or no data to send, i.e. those opportunities need to occur where and when you need them. Effective utility measures the cases where traffic demand exists and appropriate conditions exist to deliver improved wireless performance.

5.3.3 Stability

Prevalence becomes of little use if the set of devices available for opportunistic communications are constantly in flux. While traditional opportunistic networking approaches assume a degree of shared trust, there exists a non-zero cost for the establishment of secure point-to-point channels or at a minimum, the assurance that the device being considered for point-to-point communications is indeed legitimate. To that end, we evaluate the data through the lens of *stability*, namely to what extent are the devices available detected for communication likely to be available for a reasonable duration of time. Hence, the primary measure of stability lies in the distribution of contact durations with considerations for sufficient signal strength and symmetric connectivity detection. Put simply, candidates for opportunistic communications must exist for a reasonable duration of time on the order of minutes, not simply seconds to amortize the cost of point-to-point channel negotiation over the lifetime of the opportunity.

In a complementary sense, trust between various devices can be realized not only through duration but also through the repeated appearances in the local proximity.

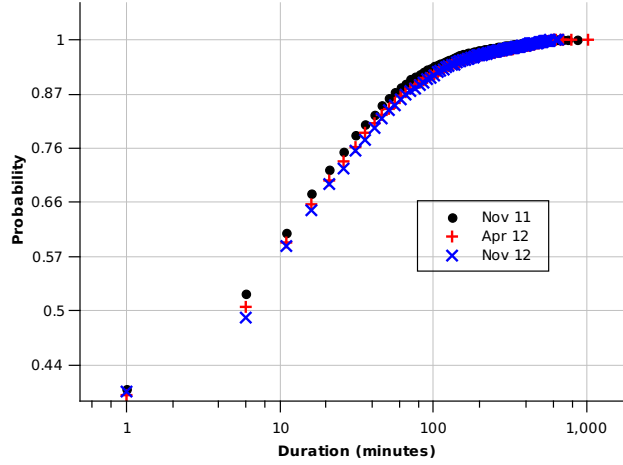


Figure 5.17. Proximity Duration ECDF

Although an individual node may have long inter-contact times individually, the fact that a node appears consistently at multiple times per day or multiple times per week for a reasonably stable period of time implies a likely social construct or external relationship between the devices. Hence, a secondary metric for stability is the degree to which the most frequent nodes churn, i.e. are the extent to which nodes appear for opportunities uniformly distributed and infrequent or are there nodes that appear dramatically more often than other nodes, potentially allowing for extended trust to be constructed by virtue of repeat appearances (or at a minimum, the trust exchange accelerated due to prior exchanges).

Hence, Figure 5.17 explores the most basic aspect of stability, namely the average duration of devices as present with regards to the locally detected nodes. Further filtering is applied where a node must have been seen at least once per day by that node for consideration for opportunistic communication. Critically, the average stability of a connection falls at roughly 6 minutes, a considerable time period for conducting the security negotiations at the front of the communication.

Moreover, node strength is another metric used to evaluate the stability of prox-

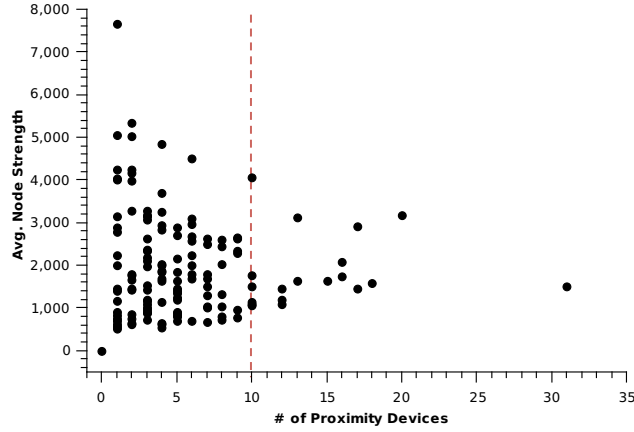


Figure 5.18. Number of Proximity Nodes vs. Node Strength

imity nodes and it is defined as follows: In one time slot, if the device detects a proximity device then the total appearance times of this peer is increased by one. Node strength is the total number of appearances of a proximity device in the period. Based on the data across 15 months, we get the node strength of proximity nodes for devices and analyze the relationship between number of relays and node strength in Figure 5.18. Using the threshold of 500 for node strength, we remove those infrequent proximity nodes and calculate the average node strength for each device. It is interesting that most of the nodes have less than 10 frequent proximity nodes (loosely correlating with their social circles) and the corresponding average node strength is relatively higher than those with more than 10 frequent relays.

While node strength captures the raw magnitude of total time spent together and indirectly infers longitudinal behavior, Figure 5.19 takes the concept further by examining the ‘streakiness’ over multiple days by which a potential peer is likely to appear across the entire duration of the study. The total appearances represents the distribution of how frequently a node is likely to appear whereby if the prospective peer is seen at least once in the day, it counts as being successfully seen for the

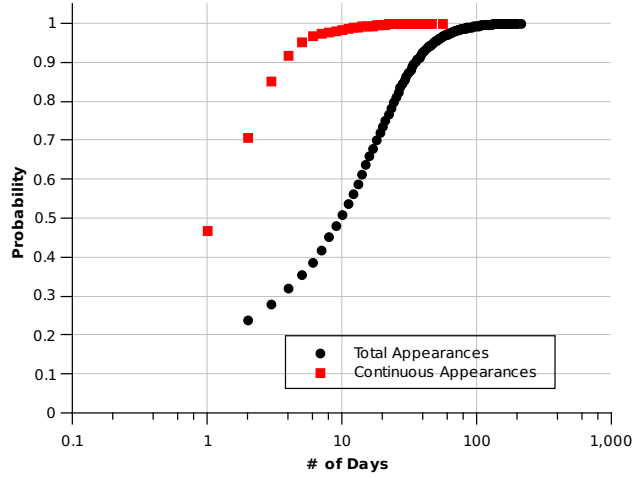


Figure 5.19. Total Appearances and Continuous Appearances in Day Counts ECDF

purposes of consistency. Note that for the purposes of the graph, all nodes are considered rather than only the nodes with a reasonable degree of consistency to capture the full longitudinal nature of the study. No filtering is done either for sufficiency of signal strength. Interestingly, several nodes eclipse nearly two-thirds of the time in the study in terms of consistency of meeting. It is those nodes that appear consistently that represent the stable foundation from which the prevalence can be effectively leveraged. The other plot in the graph captures the distribution of continuous days, namely if a node appears more than one day in a row, how long is the streak likely to continue with subsequent appearances over the next few days? The average number of continuous days is 1 day which means over 50% of the proximities did not happen continuously by day.

From the perspective of number of proximity devices, we calculate the cumulative number and compare it with the indeed active number to analyze the stability of prospective relays. The cumulative number is defined as the total number of distinguished relays in history while the weekly active number is the total number of relays

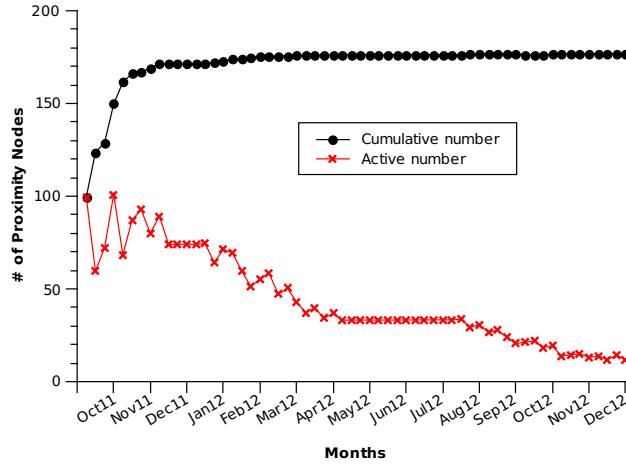


Figure 5.20. Weekly Cumulative Number and Active Number

which appear in the current week and were active in the past one month. Figure 5.20 illustrates the weekly cumulative number and active number across 15 months. The cumulative number increases to a limit (nearly the total number of devices within the project) and such value is quite stable in the following months. Meanwhile, the active number decreases across the study and big decreases happen when semesters change.

Therefore, there are four important points for the evaluation of stability:

- **Duration:** The duration of relaying connection is essential for successful traffic transmission between the device and its relaying nodes. If two devices come across each other and separate soon, such proximity is not suitable for the establishment of usable connection. We roughly define a mean average of duration as needing to be an order of magnitude or better versus trust establishment for the mobile-to-mobile link.
- **Node Strength:** The usage of node strength can capture the degree to which a trusted subset of mobile peers emerge that have the potential for expedited

trust for opportunistic efforts (i.e. negotiate heavy once, light weight for the next N times).

- **Total Appearances:** While a purely opportunistic network where a de facto sense of trust exists can tolerate frequent dynamics in terms of random nodes serving as relay or collaboration partners, consistency of the peering partner for a node can both elicit trust by virtue of familiarity (repetition) which also has roots in normative social structures.
- **Continuous Appearances:** Although trust can be more easily inferred by virtue of repetition, consistent trust over continuous days likely implies a stronger degree of trust (not a guarantee, simply more likely). The continuous consistency ($\geq 1/\text{day}$, $\geq 1/\text{week}$) can serve as a complementary metric to the earlier aforementioned inter-contact time espoused in traditional opportunistic network evaluations.

5.3.4 Reciprocity

Finally, a key metric for evaluating the potential opportunistic communication is the degree to which reciprocity exists, ideally converging to an equal ratio of service versus need. Although altruism amongst nodes typically leads to a healthier network performance as a whole, the potential for increased energy consumption is of little consolation if the device is always helping but never benefitting. To that end, we posit that reciprocity amongst nodes must be taken into account when considering the overall utility of an opportunistic solution. Whereas much of the prior work is limited to evaluating only some aspects of prevalence and stability, interactions among pairs (dyads) in the study afford us the ability to evaluate the extent to which reciprocity exists both in short and long-term time scales with actual traffic patterns and actual energy constraints.

To that end, Figure 5.21 shows the daily average percentages with respect to needing assistance (relaying to another node) versus offering assistance (relaying for another node). The data is drawn strictly from intra-study interactions across the month of November 2012. A mobile node is defined as needing assistance if it has traffic but detects no WiFi access points with a sufficient signal strength ($> -80dBm$). A mobile node is defined as serving (offering assistance) if it has a good WiFi connection and has been detected by other nodes within the study. Hence, a value of 25% with respect to being detected as a relay denotes that another node (MN_j) in the study detected the node (MN_i), MN_j had a good signal strength to the node MN_i , and MN_i had a good signal strength for WiFi for 25% of the time that MN_i was on for that day. Critically, a node needing relaying but detecting no nodes is similar to a node being willing to serve but yet no nodes needing its service. A further sub-division is done to breakdown the nodes filtered if battery level were taken into consideration, i.e. what percentage could no longer offer assistance if the battery level fell below 30% which barely has an impact, typically only a reduction of 1-3% for the number of times a node could offer assistance but would not due to energy constraints.

For a reasonable portion of the time, most nodes on average have an increased need for assistance versus the ability to serve. The primary exceptions emerge on weekends which also happen to correspond with football games, campus events, and travel. Figure 5.22 captures this relationship on both short-term (single day) and longer-term (one week/one month) timescales drawing from a single week from Figure 5.21. The figure plots the distribution of the node having the potential to serve versus needing assistance. A ratio of one implies that the node had an equal number of time slots where the node could serve as a relay (detected, good WiFi) versus the node needing a relay. Whereas the traffic from the weekend shows much stronger need (due in part to the lack of WiFi at the football stadium), the nodes on the long-term scale

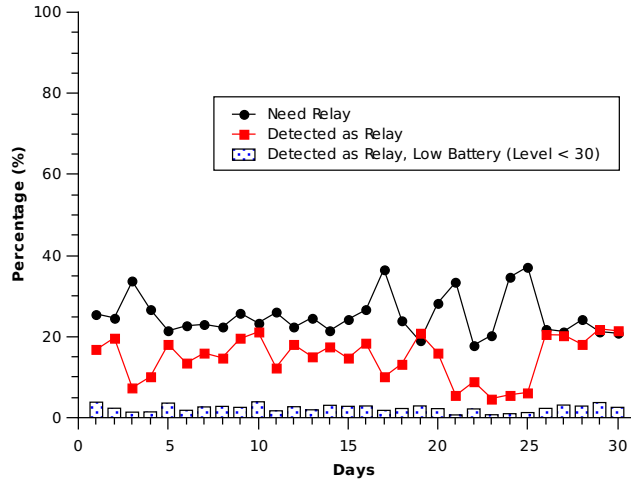


Figure 5.21. Distribution of Potential Interactions - Need vs. Serve (November 2012)

have ratios much closer to 1 representing a reasonable reciprocity. Moreover, the typical weekday (as opposed to the game day scenario with limited WiFi) gravitates more closely to a balanced distribution as well. Similarly as diurnal analysis in Section 5.3.2, the diurnal distribution of such ratio is shown in Figure 5.23. Notably, the ratio is most balanced in daytime while the needing is more than offering due to the lack of WiFi during the night and sleep time.

To close, we list the essential criteria related to reciprocity and their formal definitions as follows:

- Needs Service: A critical first step to establishing the likely reciprocity is to characterize the share of time that a node is in the need service state, namely does the node either have poor / none WiFi or alternatively poor / no cellular coverage. The need service case is where the node will perceive value to the underlying opportunistic communications.
- Offers Assistance: In contrast to needing service, the ability to offer service

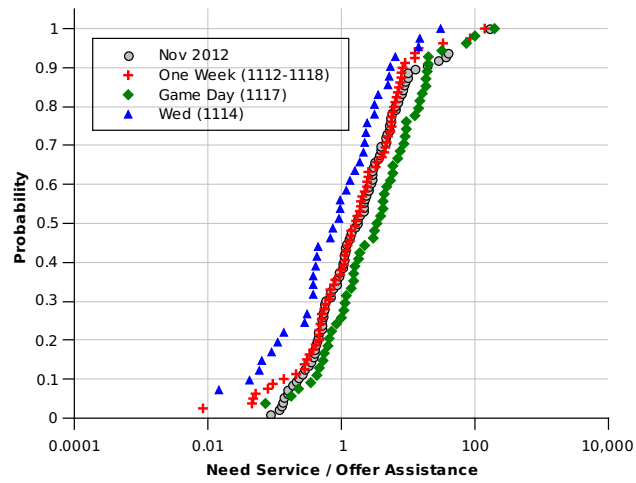


Figure 5.22. Ratio of Needing vs. Serving (November 2012)

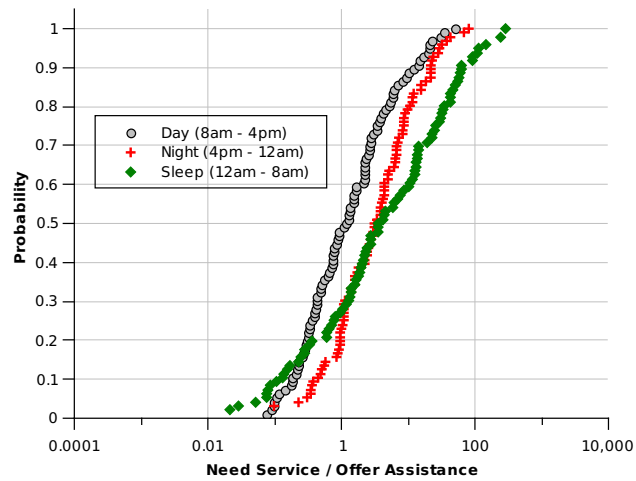


Figure 5.23. Diurnal Distribution of Ratio (November 2012)

represents the altruistic aspects of needing service, namely when can the node assist other nodes? The share of when a node can offer relaying / collaboration must be considered along with energy filters excluding devices with low energy levels.

- Service vs. Assistance: The final and most critical component is the actual ratio of reciprocity itself, ex. how often is the device likely to serve versus the device need service? While atypical situations can exacerbate sharing (everyone needs help), the normative situation should be one where reciprocity is preserved both on short-term timescales (daily) and long-term timescales (weekly, monthly).

5.4 Summary

Critically, we investigate the availability of opportunity for relaying and demonstrate that not only is such opportunity more prevalent than expected, the opportunities are stable enough to be worthwhile to establish even with auxiliary security constraints and the opportunities are reciprocal across both short-term and long-term time scales even when considering energy levels of the involved devices. Specifically, the key contributions are as follows:

- *Investigate the availability of opportunity for relay:* Proximity among mobile devices is the key to determine whether relaying can happen or not. Based on Bluetooth and WiFi detections, we are able to use appropriate signal strength to reflect the relative distance of devices. Moreover, power is essential for mobile devices therefore residual battery is another important fact to be considered for availability.
- *Demonstrate the opportunity (prevalence) for relaying is indeed significant:* Through the analysis of a 15 month dataset of detailed smartphone data gleaned from

nearly 200 smart phone users, we demonstrate that a typical user can find devices amenable to relaying averaging nearly 60% of the time. To the best of our knowledge, we believe our study is the first to conduct a longitudinal study of relaying prevalence with demonstrating ample opportunities with respect to both raw relaying prevalence and useful relaying prevalence .

- *Demonstrate that said opportunities for relaying are not only prevalent but stable:* We show with our study data that relaying opportunities existing on average for durations of 6+ minutes. Moreover, we show that the opportunities for relaying tend to be highly focused on a select subset of devices enabling enhanced levels of trust versus random, intermittent opportunities as espoused in the literature.
- *Demonstrate that said opportunities would largely be reciprocal:* In addition to opportunities being prevalent and stable (useful), we show that interactions for relaying would be overwhelmingly reciprocal in both short-term and long-term time windows, even when accounting for the energy levels of both parties involved in the relaying. We demonstrate this through the inclusion of actual traffic demands on fine temporal granularities, a unique feature of our dataset.
- *Propose a framework to evaluate the relaying potential for network traces:* While we believe our work is one of the first of its kind to fuse fine-grained proximity and traffic data over such a long period, we hope that others will embark on similar efforts to broaden the effective data pool available to the community. To that end, we view all of our data through the lens of what we dub the APSR (Availability, Prevalence, Stability, Reciprocity) framework, a framework for systematically evaluating the potential and quality of the proximity of mobile network trace data.

CHAPTER 6

Energy Consumption

CHAPTER 7

Conclusion and Outlook

In summary, our work is one of the first longitudinal and fine-grained studies that explores the practical potential for opportunistic communications with respect to smart devices. Whereas the natural intuition would be cast skepticism towards the potential for opportunistic communications (relaying or collaborative), we demonstrate through our explorations that opportunities for enhancements are not only prevalent but also sufficient, symmetric, and utilizable in our studied environment. Moreover, we find that such opportunities are stable in terms of having sufficient duration, low churn in peers, and relative consistency of peers. Finally, and perhaps most intriguingly, we show that the opportunities exhibit a reasonable degree of reciprocity for the short-term across weekdays and over the long-term when viewed across the entire week.

While we believe our work shows promising capabilities for practical mobile-to-mobile optimizations, the work is merely a first step that deserves further attention from the research community. Notably, several key aspects need to be explored including: (1) to what extent could one instrument a community outside of a campus drawing on either the Nokia Data Challenge or new efforts to explore a variety of audiences; (2) to what impact might the scale of coverage play a role, for instance would doubling or tripling the population yield considerably better results; And would other techniques besides Bluetooth and WiFi introduce more proximity opportunities (3) are there opportunities to improve content selection and distribution be that through Content Centric Networks (CCN) or other techniques (redundancy elimination); and

(4) to what extent is there a ‘critical mass’ where opportunistic networks would seem useful either by analyzing subsets of the data or through theoretical explorations.

APPENDIX A

Appendix

TABLE A.1

DETAILED INFORMATION OF COLLECTED DATA

| Type | Collected Information | Frequency |
|-----------------|--------------------------------------|--------------------------------|
| Location | Provider/Latitude/Longitude/Accuracy | 10 min/100 meters |
| Bluetooth | ID/MAC/RSSI | 1 min |
| WiFi | AP name/MAC/RSSI | 3 min |
| Cell | | when RSSI changes |
| Battery | Charge method/Level | when level changes at least 1 |
| Light | Level | when level changes at least 30 |
| Screen | On/Off | when screen on/off |
| Camera | Picture/Video name | when created |
| Network Traffic | Type/Cumulative bytes | 1 min |
| App Traffic | Type/Name/Cumulative bytes | 1 min |
| App Status | Name/Permission/Total number | 1 day |
| Port Traffic | Protocal/State/Pid/Src/Dst/Name | 15 min |
| Password | Type/Encrypted pwd/Salt | 1 hour |
| Contacts | Phone number/Email/Total number | 1 day |
| Connection | Status | when status changes |
| Storage | Total vol/Free vol | 12 hours |
| Version | Version number | when phone starts |
| Phone Call | Number/Duration | 1 hour |
| SMS | Number/Length | 1 hour |
| MMS | Number/Type | 1 hour |
| Mail | Sender/Receiver/ | 6 min |
| Music | Album name/Display name/Size | 3 min |
| Browser History | Title/Url/Visited Times | 1 hour |

BIBLIOGRAPHY

1. M. Arshad, A. Farooq, and A. Shah. Evolution and development towards 4th generation (4G) mobile communication systems. *Journal of American Science*, 6(12):63–68, 2010.
2. P. Bahl, R. Chandra, P. Lee, V. Misra, J. Padhye, D. Rubenstein, and Y. Yu. Opportunistic use of client repeaters to improve performance of wlans. *IEEE/ACM Transactions on Networking (TON)*, 17(4):1160–1171, 2009.
3. A. Balasubramanian, R. Mahajan, and A. Venkataramani. Augmenting mobile 3G using WiFi. In *MobiSys’10*, pages 209–222. ACM, 2010.
4. C. Baouche, A. Freitas, and M. Misson. Radio proximity detection in a wsn to localize mobile entities within a confined area. *Journal of Communications*, 4(4):232–240, 2009.
5. M. Bilgi, M. Yuksel, and N. Pala. 3-d optical wireless localization. In *GLOBE-COM Workshops (GC Wkshps)*, 2010 IEEE, pages 1062–1066. IEEE, 2010.
6. A. Bletsas, A. Khisti, D. Reed, and A. Lippman. A simple cooperative diversity method based on network path selection. *Selected Areas in Communications, IEEE Journal on*, 24(3):659–672, 2006.
7. R. Borovoy, F. Martin, S. Vemuri, M. Resnick, B. Silverman, and C. Hancock. Meme tags and community mirrors: moving from conferences to collaboration. In *CSCW’98*, pages 159–168. ACM, 1998.
8. H. Cai and D. Eun. Toward stochastic anatomy of inter-meeting time distribution under general mobility models. In *MobiHoc’08*, pages 273–282. ACM, 2008.
9. A. Chaintreau, P. Hui, J. Crowcroft, C. Diot, R. Gass, and J. Scott. Impact of human mobility on opportunistic forwarding algorithms. *Mobile Computing, IEEE Transactions on*, 6(6):606–620, 2007.
10. Cisco. Cisco VNI mobile data traffic forecast 2012-2017. February 2013.
11. S. Dimatteo, P. Hui, B. Han, and V. Li. Cellular traffic offloading through WiFi networks. In *MASS’11*. IEEE, 2011.
12. N. Eagle and A. Pentland. Social serendipity: Mobilizing social software. *Pervasive Computing*, 4(2):28–34, 2005.

13. J. Figueiras and S. Frattasi. Mobile positioning and tracking. Technical report, Wiley Online Library, 2010.
14. B. Han, P. Hui, and A. Srinivasan. Mobile data offloading in metropolitan area networks. *ACM SIGMOBILE Mobile Computing and Communications Review*, 14(4):28–30, 2011.
15. B. Han, P. Hui, V. A. Kumar, M. V. Marathe, J. Shao, and A. Srinivasan. Mobile data offloading through opportunistic communications and social participation. *Mobile Computing, IEEE Transactions on*, 11(5):821–834, 2012.
16. T. Henderson, D. Kotz, and I. Abyzov. The changing usage of a mature campus-wide wireless network. In *MobiCom’04*, pages 187–201. ACM, 2004.
17. F. Izquierdo, M. Ciurana, F. Barcelo, J. Paradells, and E. Zola. Performance evaluation of a TOA-based trilateration method to locate terminals in WLAN. In *Wireless Pervasive Computing, 2006 1st International Symposium on*, pages 1–6, jan. 2006. doi: 10.1109/ISWPC.2006.1613598.
18. T. Karagiannis, J. Le Boudec, and M. Vojnovic. Power law and exponential decay of intercontact times between mobile devices. *Mobile Computing, IEEE Transactions on*, 9(10):1377–1390, 2010.
19. A. Khan, M. Qadeer, J. Ansari, and S. Waheed. 4G as a next generation wireless network. In *ICFCC’09*, pages 334–338. IEEE, 2009.
20. A. Kotanen, M. Hannikainen, H. Leppakoski, and T. Hamalainen. Experiments on local positioning with Bluetooth. In *ITCC 2003: Information Technology: Coding and Computing*, pages 297–303, april 2003. doi: 10.1109/ITCC.2003.1197544.
21. J. Krumm and K. Hinckley. The nearme wireless proximity server. *UbiComp’04*, pages 283–300, 2004.
22. A. Küpper and G. Treu. Efficient proximity and separation detection among mobile targets for supporting location-based community services. *ACM SIGMOBILE Mobile Computing and Communications Review*, 10(3):1–12, 2006.
23. A. LaMarca, Y. Chawathe, S. Consolvo, J. Hightower, I. Smith, J. Scott, T. Sohn, J. Howard, J. Hughes, F. Potter, et al. Place lab: Device positioning using radio beacons in the wild. *Pervasive Computing*, pages 301–306, 2005.
24. J. Laneman, D. Tse, and G. Wornell. Cooperative diversity in wireless networks: Efficient protocols and outage behavior. *Information Theory, IEEE Transactions on*, 50(12):3062–3080, 2004.
25. J. K. Laurila, D. Gatica-Perez, I. Aad, J. Blom, O. Bornet, T.-M.-T. Do, O. Dousse, J. Eberle, and M. Miettinen. The mobile data challenge: Big data

- for mobile computing research. In *Mobile Data Challenge by Nokia Workshop, in conjunction with Int. Conf. on Pervasive Computing, Newcastle, UK*, 2012.
26. K. Lee, S. Hong, S. Kim, I. Rhee, and S. Chong. Slaw: A new mobility model for human walks. In *INFOCOM'09*, pages 855–863. IEEE, 2009.
 27. K. Lee, I. Rhee, J. Lee, S. Chong, and Y. Yi. Mobile data offloading: how much can WiFi deliver? In *Co-Next'10*, pages 26–38. ACM, 2010.
 28. K. Li, T. Sohn, S. Huang, and W. Griswold. Peopletones: a system for the detection and notification of buddy proximity on mobile phones. In *MobiSys'08*, pages 160–173. ACM, 2008.
 29. H. Liu, H. Darabi, P. Banerjee, and J. Liu. Survey of wireless indoor positioning techniques and systems. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(6):1067–1080, November 2007.
 30. S. Liu and A. Striegel. Casting doubts on the viability of wifi offloading. In *Proc. of CellNet*, pages 25–30. ACM, 2012.
 31. M. Lu, P. Steenkiste, and T. Chen. Design, implementation and evaluation of an efficient opportunistic retransmission protocol. In *MobiCom'09*, pages 73–84. ACM, 2009.
 32. M.-H. Lu, P. Steenkiste, and T. Chen. Time-aware opportunistic relay for video streaming over wlans. In *Proc. of IEEE Multimedia and Expo*, pages 1782–1785. IEEE, 2007.
 33. A. O. M. Raento and N. Eagle. Smartphones: An emerging tool for social scientists. *Sociological Methods Research*, 37(3):426–454, 2009.
 34. R. Madan, N. Mehta, A. Molisch, and J. Zhang. Energy-efficient cooperative relaying over fading channels with simple relay selection. *Wireless Communications, IEEE Transactions on*, 7(8):3013–3025, 2008.
 35. M. McNett and G. M. Voelker. Access and mobility of wireless pda users. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(2):40–55, 2005.
 36. E. Mynatt, M. Back, R. Want, and R. Frederick. Audio aura: Light-weight audio augmented reality. In *Proceedings of the 10th annual ACM symposium on User interface software and technology*, pages 211–212. ACM, 1997.
 37. A. P. Nathan Eagle and D. Lazer. Inferring social network structure using mobile phone data. *Proc. of the National Academy of Sciences (PNAS)*, 106(36):15274–15278, September 2009.
 38. V. Otsason, A. Varshavsky, A. LaMarca, and E. De Lara. Accurate GSM indoor localization. *UbiComp 2005: Ubiquitous Computing*, pages 903–921, 2005.

39. A. Passarella and M. Conti. Characterising aggregate inter-contact times in heterogeneous opportunistic networks. *NETWORKING'11*, pages 301–313, 2011.
40. L. Pelusi, A. Passarella, and M. Conti. Opportunistic networking: data forwarding in disconnected mobile ad hoc networks. *Communications Magazine, IEEE*, 44(11):134–141, 2006.
41. J. Polastre, R. Szewczyk, and D. Culler. Telos: enabling ultra-low power wireless research. In *IPSN'05*, pages 364–369. IEEE, 2005.
42. A. Sendonaris, E. Erkip, and B. Aazhang. User cooperation diversity. part i. system description. *Communications, IEEE Transactions on*, 51(11):1927–1938, 2003.
43. L. Šikšnys, J. Thomsen, S. Šaltenis, and M. Yiu. Private and flexible proximity detection in mobile social networks. In *Mobile Data Management (MDM), 2010 Eleventh International Conference on*, pages 75–84. IEEE, 2010.
44. G. Treu and A. Küpper. Efficient proximity detection for location based services. In *WPNC'05*, 2005.
45. R. Want, A. Hopper, V. Falcão, and J. Gibbons. The active badge location system. *ACM Transactions on Information Systems (TOIS)*, 10(1):91–102, 1992.
46. S. Wiethölter, M. Emmelmann, R. Andersson, and A. Wolisz. Performance evaluation of selection schemes for offloading traffic to IEEE 802.11 hotspots. In *ICC'12*. IEEE, 2012.
47. V. Zeimpekis, G. M. Giaglis, and G. Lekakos. A taxonomy of indoor and outdoor positioning techniques for mobile location services. *SIGecom Exch.*, 3:19–27, December 2002.
48. S. Zhou and J. Pollard. Position measurement using Bluetooth. *Consumer Electronics, IEEE Transactions on*, 52(2):555–558, May 2006. ISSN 0098-3063. doi: 10.1109/TCE.2006.1649679.
49. Y. Zou, J. Zhu, and R. Zhang. Exploiting network cooperation in green wireless communication. *Communications, IEEE Transactions on*, 61(3):999–1010, 2013.