Exploring the Potential in Practice for Opportunistic Networks amongst Smart Mobile Devices

Shu Liu, Aaron D. Striegel
Department of Computer Science and Engineering, University of Notre Dame
Notre Dame, IN, USA
sliu6@nd.edu, striegel@nd.edu

ABSTRACT

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. One of the key impediments towards the adoption of opportunistic communications has been the inability to demonstrate viability at scale, namely showing that sufficient opportunities exist and more importantly exist when needed to offer significant network performance gains. We demonstrate through a large-scale, longitudinal study of smartphone users that significant opportunities are indeed prevalent, are indeed stable, and end up being reasonably reciprocal both on short and long-term timescales. In this paper, we propose a framework dubbed PSR (Prevalence, Stability, Reciprocity) to capture key aspects that characterize the net potential for opportunistic networks which we feel merit significantly increased attention.

Categories and Subject Descriptors

C.4 [**PERFORMANCE OF SYSTEMS**]: Reliability, availability, and serviceability; C.5.3 [**COMPUTER SYSTEM IMPLEMENTATION**]: Microcomputers—*Portable devices*

Keywords

Relay; Opportunistic Networks; Proximity; Bluetooth; WiFi

1. INTRODUCTION

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 [1]. 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, the spectrum available for wireless data is finite and typically entails massive costs for acquisition and infrastructure deployment. A wide variety of solutions have emerged ranging from simpler solutions such as better WiFi offloading to much more complex solutions such as small heterogeneous cellular networks.

One rich category of work that is complementary to existing techniques is the concept of *opportunistic communication*. Opportunistic communication refers to the concept which allows nodes to leverage sporadic, intermittent contacts when two nodes come into direct radio communication range [2]. When cellular or WiFi links are not available or not strong enough, opportunistic relaying introduces an 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 [3–7]?

Although opportunistic networking has received more attention as of late with the rise of various point-to-point technologies (WiFiDirect, LTEDirect) 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 [8, 9].

A key aim of our work is to demystify the opportunities that exist for opportunistic relaying by bridging the gap with a large scale dataset across both a significant duration (15 months) and user pool (nearly 200 users). Critically, we demonstrate that not only is the opportunity for relaying more prevalent than expected, said 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 con-

MobiCom'13, September 30–October 4, Miami, FL, USA.

http://dx.doi.org/10.1145/2500423.2500430.

sidering energy levels of the involved devices. Specifically, the key contributions of our paper are as follows:

- *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 in our study could 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 exist on average for durations of 6 or more 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 PSR (Prevalence, Stability, Reciprocity) framework, a framework for systematically evaluating the potential and quality of the proximity of mobile network trace data. Particularly, PSR can serve as a broad foundation for others to explore larger coverage areas, alternative populations, and different environments.

The rest of this paper is organized as follows. Section 2 provides an overview of the related work and Section 3 introduces the NetSense dataset. Based on the NetSense dataset, the framework to evaluate the potential for relaying is proposed in Section 4. Finally, Section 5 concludes the paper commenting in particular on shortcomings of the work and future areas for research.

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 [3] and expanded up by numerous others [4–7]. A key foundation for the evaluation of opportunistic networks arises from the characterization of inter-contact times [8, 9]. The work in [8] was one of the first works to highlight the importance of inter-contact time for studying opportunistic networks and analyzed the features of aggregate inter-contact times. In the work of [9], 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.

Complementary to the characterization of inter-contact time is the ability to unobtrusively measure the proximity of mobile nodes to one another. 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. Various techniques have emerged ranging from both academia through the concept of Location-based Service (LBS) functions [10–12] 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). Drawing in part on the prevalence of WiFi deployments in indoor environments, WiFi triangulation can be employed in tandem with cellular signals to reasonably detect locations at a much lower cost than GPS. Without calculating absolute location, NearMe [13] 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, [14] uses GSM readings to explore the proximity of mobile targets. Conversely, the problem of inter-contact (proximity) detection can also be reduced to its most basic form simply asking whether two nodes are close rather than determining node proximity as a derivation from location. The MIT Reality Mining group was one of the pioneers in employing Bluetooth as a tool to infer the potential for social interactions among users capitalizing on the 10 m effective range of Bluetooth with agents on each mobile device [15, 16]. More recent works including the Nokia Data Challenge [17] also employed Bluetooth discoverable devices for proximity detection.

A key defining characteristic of the works adopting Bluetooth as a mechanism [15–18], for detecting proximity is the inclusion of a device-side agent for both detection and the enabling of devices as permanently discoverable with respect to Bluetooth. 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 [15], UCSD [19] and Dartmouth [20]), conference sites (Infocom [8]) and cities (Nokia [17]), most datasets are limited in size and scope in terms of capturing dyadic relationships, namely both sides of a potential proximity relationship for the purposes of truly

Table 1: 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.

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 workshop. 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 1 summarizes the most relevant studies and compares the works to our own reference study for this paper.

3. LARGE-SCALE PHONE DATASET

In August of 2011, two hundred participants were selected from the incoming freshmen class of Notre Dame 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.

The agent itself is a multi-threaded block of modular code installed via an APK with updates to the agent delivered via customized text messages to the user prompting installation from our website. 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 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. Critically, while the agent gathers various environmental data (application usage, wireless adapter tonnage, battery level, detected wireless and signal strengths, Bluetooth proximity) and communication patterns (SMS, e-mail, phone, and Facebook), the content of said communications are not recorded, only where, when, and with whom the communication occurred. All data recording is fully disclosed to the user upon entrance to the study with full approval of the university IRB for the procedure.

For the relevant portions of the paper, the critical aspects of the dataset are the respective wireless environmental readings, Bluetooth proximity, location (gleaned via Google Location services), device state (battery level, screen on / off), and data consumption. 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 dataset used in the paper 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 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 study reducing the data pool to 189 students.

Table 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 communica-

Table 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 >= -80 dBm)	131	106	<1	75
# of Detected Distinct WiFi APs	1218	1004	1777	1663
# of Detected Distinct WiFi APs in Campus	424	446	<1	433

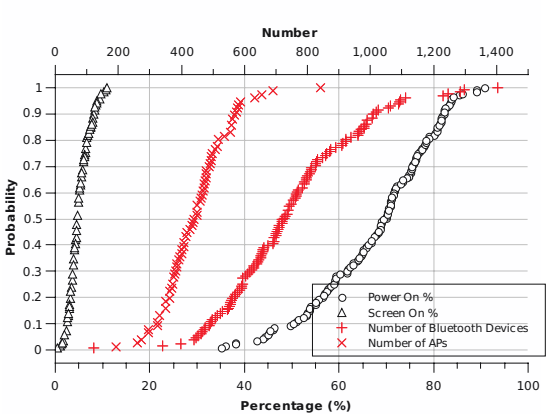

Figure 1: ECDF of Data in April 2012

tions (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 1 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. 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 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.

4. EVALUATING PRACTICAL RELAYING

The dataset outlined in the prior section 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 [21]) or striping across different wireless mediums (WiFi + cellular). Opportunistic communications would be provided through mobile-to-mobile communications (Bluetooth, WiFiDirect, LTEDirect). 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*?

To that end, we seek to answer the following questions in this section of the paper summarized in terms of metrics in Table 3:

- *Environment*: 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?

- *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 reciprocality disappear? Finally, reciprocality 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 being by defining several key attributes of the data. Each mobile node (MN_i) is

Table 3: Framework for Evaluating Prevalence, Stability, Reciprocity

Criteria		Term	Description
Prevalence	Sufficient Signal	SS	Time with a 'good signal' vs. time powered on
	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
	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
	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
	Node Strength	NS	Number of total appearances by n most common peers
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 $Assist$

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. We do not give priority to any of the above metrics since they exhibit different attributes in various scenarios while the missing of any of them might lead to the failure of relaying protocol in general environment.

We use Bluetooth as the proximity mechanism for multiple reasons. First, the effective range of Bluetooth typically is on the order of 10 m or less. In contrast with other point-to-point technologies such as WiFiDirect 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 10 m effective range of Bluetooth. Second, although Bluetooth is reasonably common amongst smart phones, the configuration of a device as being discoverable is fairly uncommon, largely for reasons of security. Despite the somewhat uncommon nature 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. For reference, we do explore the potential when proximity is relaxed to consider proximity as derived by sharing common detected WiFi APs similar to [19].

4.1 Prevalence

The first question to pose is to what extent that 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 PSR framework. We posit several questions that include: (1) the extent to which discoverable Bluetooth devices exist, (2) the extent to which diurnal (daily) patterns play a role in discoverable devices, (3) and the extent to which Bluetooth views are symmetric. 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).

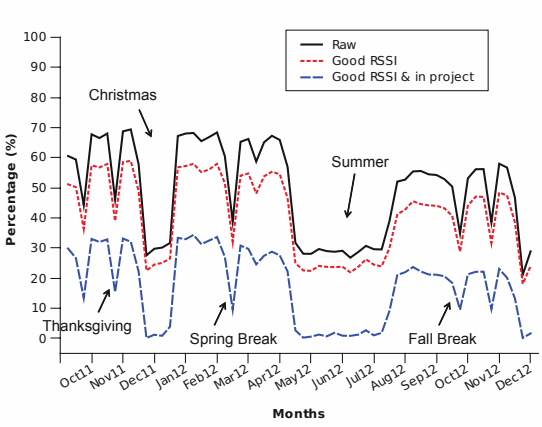

Figure 2: Bluetooth Proximity with Different Restrictions

Bluetooth Proximity

Figure 2 illustrates the average weekly Bluetooth proximity percentage with different restrictions across the study period. Various periods of interest representing various break periods are labeled in the graph. The percentage of time represents the percentage of time where an opportunity might exist for collaboration or relaying with a discoverable Bluetooth device. 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

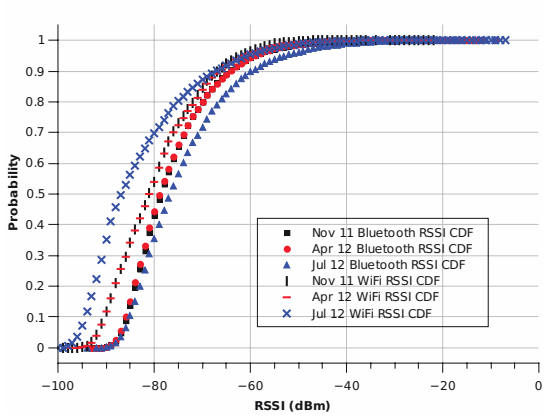

Figure 3: Bluetooth and WiFi RSSI Values ECDF

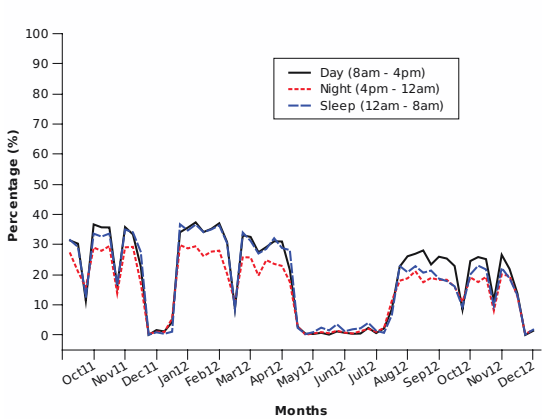

Figure 4: Diurnal Distribution of Bluetooth Proximity

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. The slight drop from the fall of 2011 to the fall of 2012 can largely be attributed to students moving dorms and joining their respective majors and having less shared coursework versus their freshmen year.

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 3 shows the ECDF of the Bluetooth and WiFi RSSI values in different months and more than 60% of the values are larger than -80 dBm. 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 [22], the RSSI value *-80 dBm* is used as a threshold to indicate good RSSI for direct mobile-to-mobile communication. As shown in Figure 2, even with the threshold of -80 dBm for the 'Good RSSI', there still exists more than 50% of the time slots, a fairly insubstantial drop from the raw detected Bluetooth devices.

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 4 shows the diurnal distribution with the twin restrictions of good RSSI and intra-study detected proximity only. The daytime proximity percentage from 8am to 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

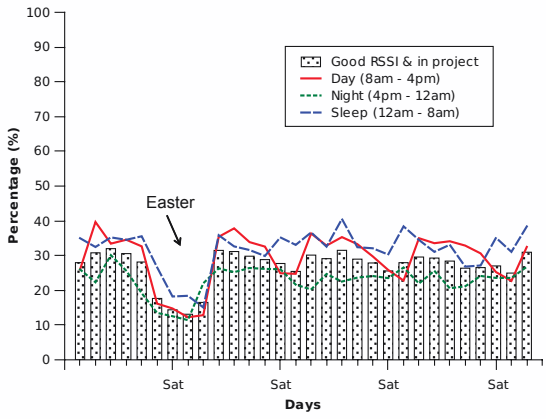

Figure 5: Diurnal Distribution of Bluetooth Proximity in April 2012

and Tuesday / Thursday courses. Similarly, weekends are much more likely to be quite diverse from normal weekday patterns. Figure 5 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. The percentage does not drop to zero since there are some international students in the project and not all students travel home for Easter. 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.

WiFi Proximity

While Bluetooth is used in the first part of the section to assess proximity, the next question to ask is to what extent could we infer proximity if we expanded our search to consider WiFi range via shared access points. For instance, in [19], WiFi signal strength is used to indicate proximity

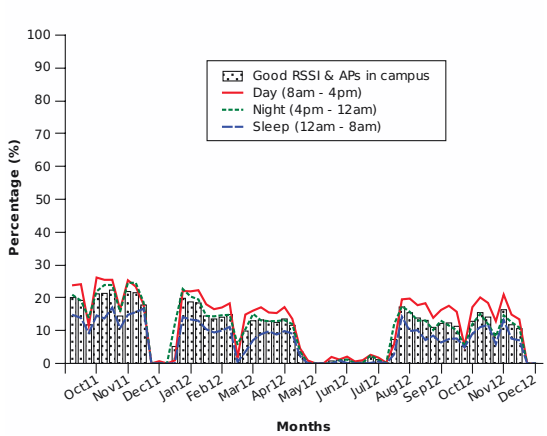

Figure 6: WiFi Proximity

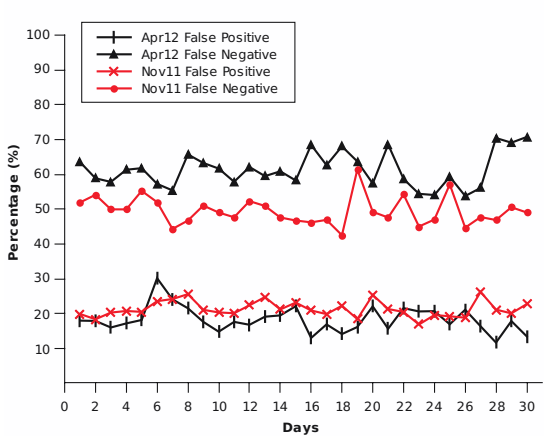

Figure 7: Comparison of Bluetooth Proximity and WiFi Proximity

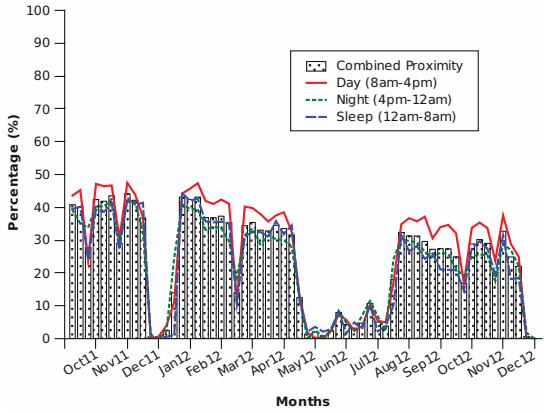

Figure 8: Combined Proximity (Fused Bluetooth / WiFi Detection)

when two devices are associated with the same AP. In our work, 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.

For appropriate range values within our study, we conducted experiments to evaluate the cutoff points for reasonable WiFi performance on the particular Android handset employed (Nexus S 4G). 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 one hundred instances for each particular 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 3, 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 for WiFi proximity.

For each device, 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 6 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.

Interestingly, the WiFi result conveys less proximity than the noted Bluetooth proximity. Notably when compared to reference laptops or tablets, our prior work in [23] notes a roughly 10 dBm signal penalty for the smartphone used in our study.[1] In the work of [6] the RSSI discrepancy of different WiFi adapters is discussed as well. 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 7, we compare the results of WiFi proximity and Bluetooth proximity in April 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 [23].

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

[1]Subsequent measurements indicated a 3-5dBm signal penalty in general for smartphone versus laptops.

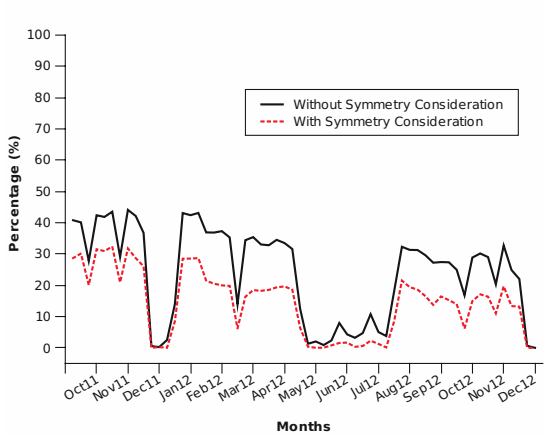

Figure 9: Combined Proximity with/without Symmetry

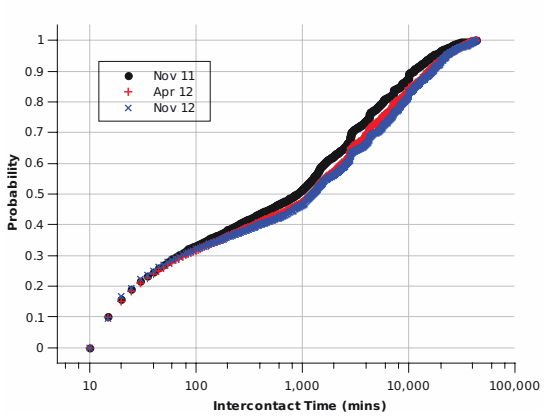

Figure 10: Inter-contact Time ECDF

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 8 illustrates such 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. In the following sections, the measurements combine the Bluetooth and WiFi together to indicate proximity.

Symmetry

For any two study participants that detect each other, there exists the opportunity to determine if symmetric detection 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)? Figure 9 compares the combined proximity with and without consideration of symmetry where symmetry is defined as both nodes seeing each other and having a 'Good RSSI.' Notably, roughly one third of the opportunities disappear in 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.

Inter-contact Time and Beyond

Inter-contact time is defined as the time elapsed between two successive contact periods for a given pair of devices [8] which is another important characteristic of prevalence. For each pair of devices, we compute the inter-contact time as the time taken before the pair meet again. Figure 10 exhibits the empirical distribution of the inter-contact times obtained among different months. On average, the inter-contact 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 inter-contact time, in large

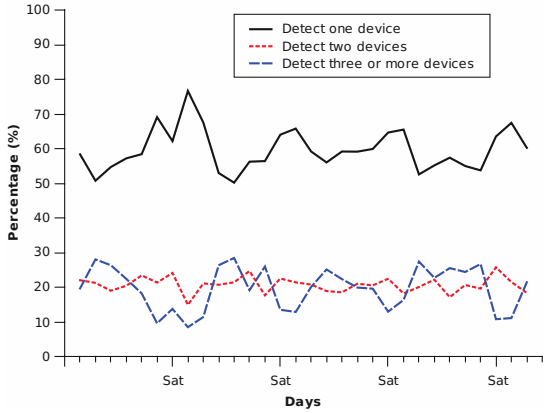

Figure 11: Distribution of Proximity Diversity (In-Study) in April 2012

part due to more shared coursework, reinforcing the findings from earlier graphs.

While inter-contact time 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. Relaying selection plays little role when there exists only a single device to choose from for relaying. Figure 11 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 12, 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.

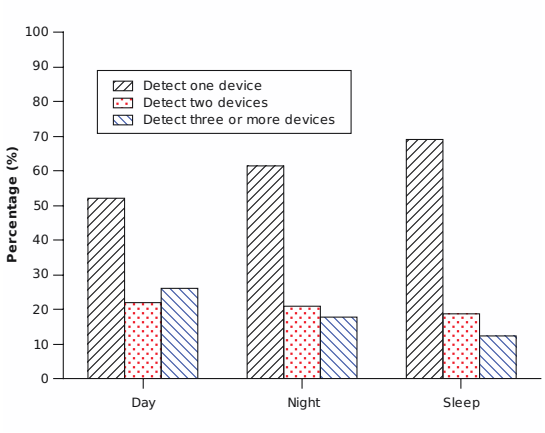

Figure 12: Diurnal Distribution of Proximity Diversity in April 2012

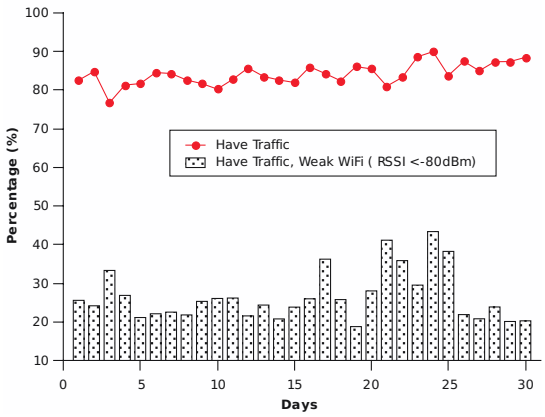

Figure 13: Weak WiFi Signal & Traffic

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.

The first context is when the device has traffic but does not have good WiFi connection. Figure 13 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 -80 dBm). Based on the results, there are more than 80% of time slots have traffic while 30% of time slots 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.

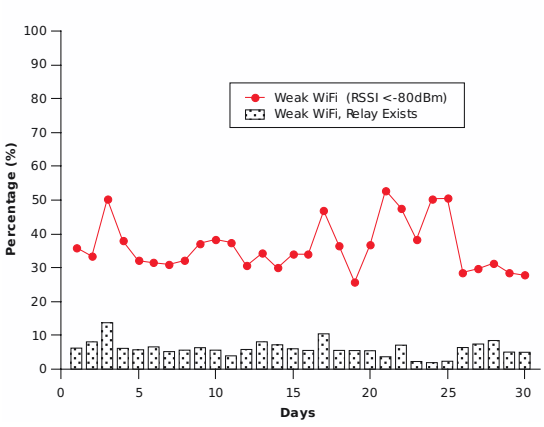

Figure 14: Relaying & Weak WiFi Signal

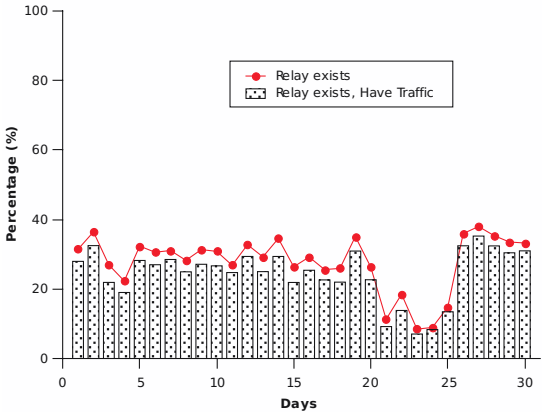

Figure 15: Traffic & Relaying

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 -80 dBm. In such case, relay is an option. In Figure 14, nearly 40% of the time the devices do not in good WiFi environment on average. There are around 10% of time slots do have relaying node(s) around while WiFi signal is not available or too weak.

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 results shown in Figure 15, the percentage is fairly high when relaying exists and there is traffic at the same time. Therefore, we posit that relaying can provide interesting opportunities for devices to send and receive traffic, not only in a passive way, but also in a proactive way.

4.2 Stability

Prevalence becomes of little use if the set of devices available for opportunistic communications are constantly in flux. While traditional opportunistic networking approaches as-

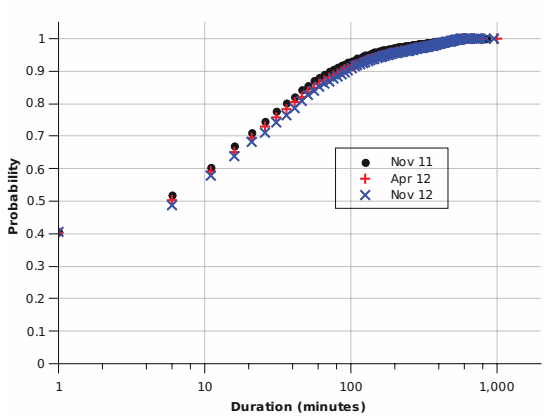

Figure 16: Proximity Duration ECDF

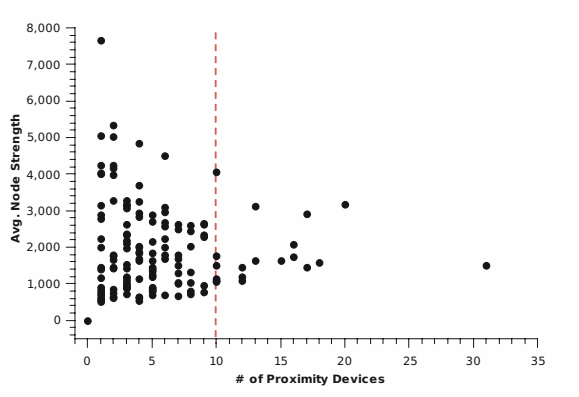

Figure 17: Number of Proximity Nodes vs. Node Strength

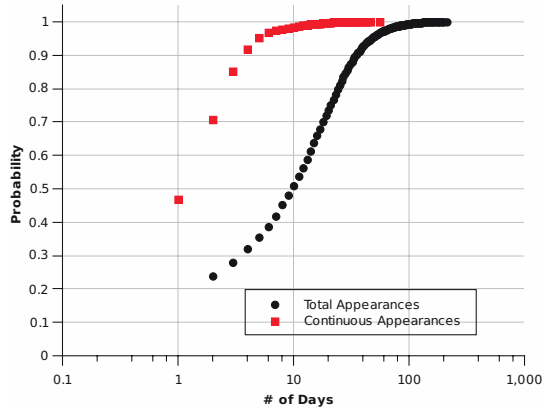

Figure 18: Total Appearances and Continuous Appearances in Day Counts ECDF

sume 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 *stabilty*, 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 local proximity. 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. 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 16 explores the most basic aspect of stability, namely the average duration of devices as present with regards to the locally detected nodes. Figure 16 exhibits the distribution of proximity durations of different months. Further filtering is applied where a node must have been seen at least once per day by that node for consideration for opportunistic communication. Nearly half of proximity lasts more than one time slot and the duration between 10 minutes and 45 minutes takes around 25%. 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.

In addition, we analyze the relationship between the number of relays and node strength in Figure 17. We define node strength 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 compute the node strength of proximity nodes for devices. Using the threshold of 500 for node strength (have been seen almost at least once per day), 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 18 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 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.

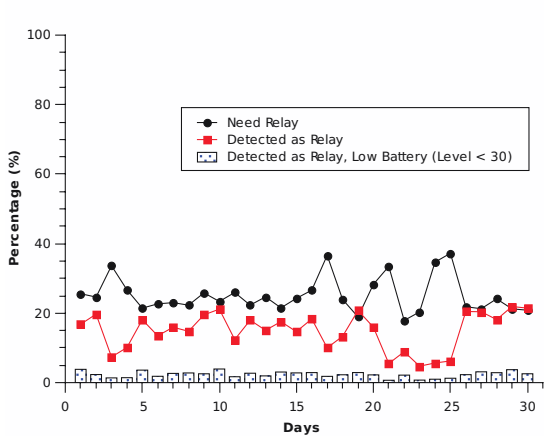

Figure 19: Distribution of Potential Interactions - Need vs. Serve (November 2012)

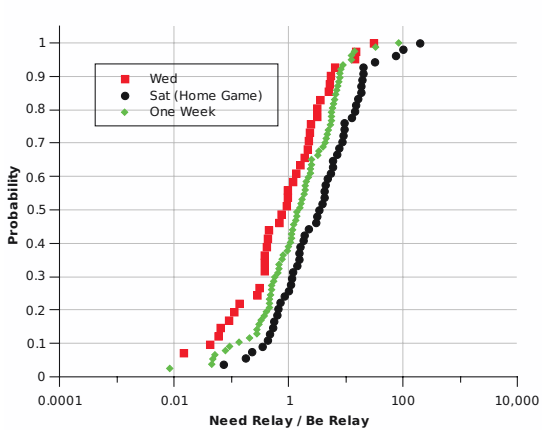

Figure 20: Ratio of Needing vs. Serving - One Week - November 2012

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 one day which means over 50% of the proximities did not happen continuously everyday.

4.3 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 19 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 20 captures this relationship on both short-term (single day) and longer-term (one week) timescales drawing from a single week from Figure 19. 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 have ratios much closer to one 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.

5. CONCLUSION

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 opportuni-

ties 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.

Acknowledgement

This work was funded in part by the National Science Foundation through grant IIS-0968529. We would also like to thank our collaborators, Dr. Christian Poellabauer, Dr. David Hachen, and Dr. Omar Lizardo. Special thanks to Sprint who provides us more than 200 phones and free data plan making the data gathering a reality.

6. REFERENCES

[1] Cisco, "Cisco VNI mobile data traffic forecast 2012-2017," Feburary 2013.

[2] L. Pelusi, A. Passarella, and M. Conti, "Opportunistic networking: data forwarding in disconnected mobile ad hoc networks," *Communications Magazine, IEEE*, vol. 44, no. 11, pp. 134–141, 2006.

[3] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *Information Theory, IEEE Transactions on*, vol. 50, no. 12, pp. 3062–3080, 2004.

[4] A. Sendonaris, E. Erkip, and B. Aazhang, "User cooperation diversity. part I. system description," *Communications, IEEE Transactions on*, vol. 51, no. 11, pp. 1927–1938, 2003.

[5] 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*, vol. 24, no. 3, pp. 659–672, 2006.

[6] M. Lu, P. Steenkiste, and T. Chen, "Design, implementation and evaluation of an efficient opportunistic retransmission protocol," in *Proc. of MobiCom*. ACM, 2009, pp. 73–84.

[7] 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)*, vol. 17, no. 4, pp. 1160–1171, 2009.

[8] 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*, vol. 6, no. 6, pp. 606–620, 2007.

[9] H. Cai and D. Eun, "Toward stochastic anatomy of inter-meeting time distribution under general mobility models," in *Proc. of MobiHoc*. ACM, 2008, pp. 273–282.

[10] G. Treu and A. Küpper, "Efficient proximity detection for location based services," in *Proc. of WPNC*, 2005.

[11] 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*, vol. 10, no. 3, pp. 1–12, 2006.

[12] L. Šikšnys, J. Thomsen, S. Šaltenis, and M. Yiu, "Private and flexible proximity detection in mobile social networks," in *Proc. of IEEE Mobile Data Management (MDM)*, 2010, pp. 75–84.

[13] J. Krumm and K. Hinckley, "The NearMe wireless proximity server," in *Proc. of UbiComp*. Springer, 2004, pp. 283–300.

[14] K. Li, T. Sohn, S. Huang, and W. Griswold, "Peopletones: a system for the detection and notification of buddy proximity on mobile phones," in *Proc. of MobiSys*. ACM, 2008, pp. 160–173.

[15] N. Eagle, A. Pentland, and D. Lazer, "Inferring social network structure using mobile phone data," *Proc. of the National Academy of Sciences (PNAS)*, vol. 106, no. 36, pp. 15 274–15 278, September 2009.

[16] N. Eagle and A. Pentland, "Social serendipity: Mobilizing social software," *Pervasive Computing*, vol. 4, no. 2, pp. 28–34, 2005.

[17] 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 *Proc. of Nokia Mobile Data Challenge Workshop*, 2012.

[18] S. Liu and A. Striegel, "Face-to-face proximity estimation using Bluetooth on smartphones," *Mobile Computing, IEEE Transactions on*, 2013.

[19] M. McNett and G. M. Voelker, "Access and mobility of wireless pda users," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 9, no. 2, pp. 40–55, 2005.

[20] T. Henderson, D. Kotz, and I. Abyzov, "The changing usage of a mature campus-wide wireless network," in *Proc. of MobiCom*. ACM, 2004, pp. 187–201.

[21] M.-H. Lu, P. Steenkiste, and T. Chen, "Time-aware opportunistic relay for video streaming over WLANs," in *Proc. of Multimedia and Expo*. IEEE, 2007, pp. 1782–1785.

[22] J. Polastre, R. Szewczyk, and D. Culler, "Telos: enabling ultra-low power wireless research," in *Proc. of IPSN*, 2005, pp. 364–369.

[23] S. Liu and A. Striegel, "Casting doubts on the viability of WiFi offloading," in *Proc. of CellNet*. ACM, 2012, pp. 25–30.