Analyzing the Impact of Proximity, Location, and Personality on Smartphone Usage

Lei Meng, Shu Liu, Aaron D. Striegel
Department of Computer Science and Engineering
University of Notre Dame
Notre Dame, IN 46556, USA
Email: {*lmeng,sliu6,striegel*}*@nd.edu*

***Abstract*—Over the past few years, mobile devices, particularly smartphones have seen dramatic increases in data consumption. The significant increases in data usage have placed tremendous strain on the wireless infrastructure necessitating research across a variety of optimization, efficiency, and capacity improvements. Complementary to those research efforts is the acquisition of a better understanding what aspects drive user smartphone usage. In this paper, we leverage the unique characteristics of the NetSense study to demonstrate how proximity, location, and individual differences (e.g., personality) can play an important role in understanding user behavior.**

I. Introduction

With the increasing popularity of smartphones and the wide range of applications available on mobile devices, wireless traffic has experienced tremendous growth in the past few years [1]. Users are able to browse the web, access e-mails, participate in on-line gaming, and enjoy high definition multimedia all from their mobile devices. Moreover, the innate mobility of said devices affords users the freedom to experience pervasive connectivity across a variety of contexts and environments. As a result, mobile devices are carried nearly ubiquitously from the moment one wakes up to when one goes to sleep all the while consuming ever-increasing quantities of data. The net result is a seemingly unrelenting demand for data that places tremendous strain on the existing wireless infrastructure [2].

While the process of observing cellular data traffic has gained in popularity significantly in the research community [3]–[5], the extraction of useful design parameters requires an understanding that can only be gained by observing the overall context of usage. Notably, it is the device usage as impacted by non-smartphone aspects including internal factors (goals, personality) and external factors (friendship, environment, nearby users) that are difficult to monitor but yet tremendously informative [5], [6].

To that end, we aim to leverage the unique characteristics of our NetSense dataset, a longitudinal study of a large cohort of smartphone users at the University of Notre Dame, to offer insight on said aspects. Our dataset and the associated survey / metadata allow us to explore the interplay of user proximity, friendship, location, and personality type on device and application usage. In this paper, we explore the active usage of a cohort of seventy-eight users over an entire semester from the Spring of 2012.

The key contributions of our paper are as follows:

- We analyze the impact of user density (i.e., the number of nearby mobile devices) on data consumption and application usage. Further, we assess a variety of friendship filtering methods ranging from simple proximity [7] to more sophisticated friendship criteria (prior SMS interactions, prior Facebook interactions, etc.). We find that the context of user proximity plays a significant role.
- We study the impact of personality on traffic consumption and application distributions. We find that personality type tends to influence not only the top application use but also the degree to which various applications are used. For instance, we find that extroverts consume more data than introverts but yet extroverts use popular social applications such as Facebook less than introverts.
- We analyze the impact of location and how location interplays with application usage / consumption. While it is not surprising that application usage tends to vary with location, the notable finding is that the impact due to the proximity of other users can vary significantly based on location.

The remainder of this paper is organized as follows. We first introduce the work related to our research in Section II. Next, Section III describes the NetSense data source and presents several basic summary characteristics of the dataset. Section IV represents the key contributions of our paper and Section V concludes our study.

II. Related Work

The study of phone and user characteristics has been seen considerable growth over the past few years with numerous studies and smartphone testbeds [8]–[12]. Similar to our own work, many studies have explored a variety of contextual implications for usage including location, proximity, and user interest. One of the most notable studies was the MIT Reality Mining project which was one of the first studies to explore the interplay of applications and social interactions [7]. Later studies by those such as Do, Blom, and Gatica-Perez subdivided usage further into hourly assessments and further refined proximity through improved use of Bluetooth density [3]. Conversely, others such as Xu et al. aggregated significantly larger datasets to explore application usage at a national level

978-1-4799-3088-3/14/$31.00 ©2014 IEEE

[12]. Shafiq et al. continued the similar trend by looking at geographical dynamics of application usage [13]. In contrast, our own dataset draws from the strength of not only fine-grained longitudinal data (order of minutes, nearly complete contextual logging) but also from significant improvements to Bluetooth proximity [14] and detailed survey / metadata monitoring from social networks (Facebook, contact lists) and survey data (alter disambiguation, etc.).

Besides external factors, individual difference such as personality is also non-trivial for data analysis. Although less explored than smartphone context / usage, there are several works that have begun to explore linkages between smartphone usage and personality [5], [15]. Chittaranjan et al. continued the earlier work of Do et al. to analyze the relationships between smartphone usage such as applications usage (frequency), calls, SMS, and Big-Five personality traits. They were able to infer the personality type of a user based on aggregated creatures obtained from the smartphone usage data significantly above chance and up to 75.9% accuracy. In the work by Butt and Phillips [15], the researchers found that disagreeable extroverts reported spending more time calling, changing ring tones, and changing phone wallpaper. Unconscientious, emotionally unstable, disagreeable extroverts also reported spending more time sending text message. Critically, all data reported by the study was gathered via self-reported surveys in contrast to our own which was directly logged on the device. The key contribution of our work relative to the works by Chittaranjan [5] and Butt [15] is to bring location context and user proximity into consideration.

III. Data

Our data comes from the NetSense study that we launched in August of 2011 and consisted of 200 incoming freshmen at the University of Notre Dame. Each participant in the NetSense study was provided with a smartphone (Nexus S) with unlimited data, unlimited texting, and unlimited mobile-to-mobile minutes in exchange for complete monitoring access of the phone. As part of the study, we deployed a user-level agent to collect a wide variety of information including network traffic, application traffic, location, phone call, text message, email, browser history, screen usage time, and various other aspects. We note that actual data contents were not logged but rather that only the handset environmental view and who/where/when of phone communications were recorded. From an operational perspective, the agent locally spools data onto the phone flash and then periodically relays the data across a secure connection to the server. The full details with regards to the collection mechanisms can be found in [14]. The data relevant to this paper includes:

- *Proximity*. A key differentiation of the NetSense study is the active recording of proximity gleaned through Bluetooth discovery. Each study participant must leave their device as permanently Bluetooth discoverable, allowing for low-power discerning of relative proximity. Proximity is recorded once per minute.
- *Phone usage*. Phone usage is defined as the time when the phone screen is active, i.e., the user is highly likely to be actively using the phone. Phone usage is recorded by a trigger capturing the exact time when the screen activates until when the screen shuts off. With this data, we are able to identify if the user is actually using the phone instead of leaving the apps running in the background.
- *Network usage*. Students in this study may consume mobile data either through campus-wide WiFi (802.11n) or 3G EVDO (Sprint). We log four types of traffic once per minute: total downlink traffic, total uplink traffic, 3G downlink traffic, and 3G uplink traffic.
- *App usage*. In order to understand network usage, we log the usage of each application installed on the phone including the uplink and downlink traffic once per minute.
- *Location data*. Location is recorded via the Google Android Location Service triggered either by movement out of a perceived 100m area or if 10 minutes has elapsed since the last recording. Due to power constraints, GPS is not used but rather WiFi fingerprints and cellular triangulation ascertain an approximate position[1].

Users are filtered to include only users who were actively using their phone and present on campus for the majority of the week. If a user u in week x has seen at least thirty distinct Bluetooth devices on average per day (typically the number is much higher) and has turned on the screen for more than thirty minutes on average per day, the user u is included for the week. Break weeks such as spring break and Easter break are not included. The data filtering resulted in seventy-eight distinct good users from January 2012 through April 2012.

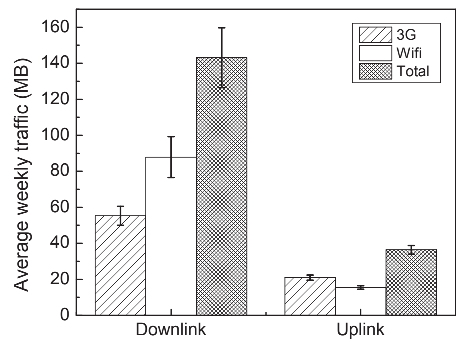


Fig. 1: Average weekly traffic per user

To start, we give an overall view on the traffic usage from the observed week of traffic. Figure 1 shows the average weekly traffic consumption and 95% CI over the observed period. The traffic is broken down by the *network adapter*: 3G and WiFi. On average, the total downlink traffic per week is around 140 MB, with roughly 65% coming from WiFi and 35% coming from 3G. The uplink total traffic (including our agent traffic) per week is 38 MB, 0.27 of the downlink traffic. Due to the majority of the traffic consumption coming from the downlink, we mainly discuss downlink traffic for the remainder in this paper.

As mentioned before, traffic consumption may occur anytime and at any places due to the pervasive nature of wireless

[1] Position (latitude, longitude) as well as estimated error are recorded.

access for the smartphone. Different location categories have different usages. For example, the dormitory (dorm) is a place to sleep / interact socially, the classroom is place to study / have lecture, and the dining halls are places to eat. The different functionality of the locations impacts the smartphone usage and thus the traffic consumption. Therefore, it is important to understand the impact of location. We identified and aggregated the two top location categories - *dorm* and *classroom*, where our participants consume traffic. Figure 2 shows the percentage of traffic used in the dorm and classroom. The two locations constitute over 70% of the total traffic. 56.4% of the traffic comes from dorm while 14.3% comes from the classroom. We will analyze and discuss location impact in details in Section IV-C.

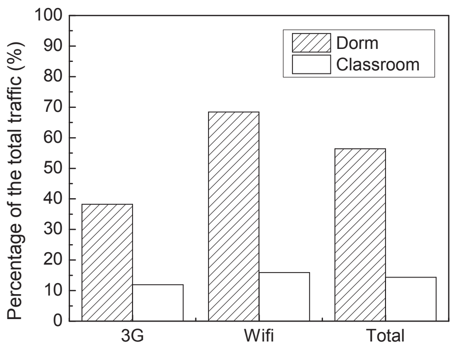


Fig. 2: Percentage of traffic in dorm and classroom (downlink)

Additionally, the study asked all the participants to fill out a survey each semester since Aug 2011 until Aug 2013. The survey includes general demographic information, prior education, personality, emotional state as well as cultural and political viewpoints.

IV. Results

In this section, we present and discuss the impact on traffic consumption from three different perspectives: proximity density, personality and location categories.

A. *Proximity Density*

The context that leads to proximity of individuals can vary significantly across a college campus. For those who are perhaps a bit more old-fashioned, proximity tends to be for direct face-to-face interaction (e.g., talking) or common activities. To that end, we aim to analyze if the proximity density within certain distance might impact user smartphone behavior in terms of traffic consumption. Critically, the impact may vary between when a close friend nearby versus when a stranger sits over at the next table. Thus, we analyze the influence from two perspectives: 1) distance 2) social relationship.

Distance is referred based on the signal strength attached to the detected device. Here we choose two signal strength thresholds *close proximity* (-65 dBm, $\leq$2.5m) and *near proximity* (-80 dBm, $\leq$5.5m) [14] to restrict the detection area. We present the proportion of different proximity densities occurred across the day and compare the results between *close proximity* (-65 dBm) and *near proximity* (-80 dBm). Figure 3 and Figure 4 show the normalized frequency (i.e., number of time slots) of each proximity density occurred during different time of a day for two types of proximity respectively. None (i.e., zero), single (i.e., one) and multiple (i.e., more than one) represent the proximity density which is the number of distinct devices detected in one time slot (5 mins) within a given threshold. The normalized frequency at hour h is calculated by dividing the frequency F_h at hour h with the accumulated frequency $\sum F_i$ ($0 \leq i \leq 23$) over all day. For close proximity (-65 dBm), the frequency of each proximity density case does not vary significantly across the day. The proximity density of none occurred five times as much as the single case and thirty times as much as the multiple case on average. For near proximity (-80 dBm), the frequency of single and multiple cases is higher than near proximity (-65 dBm) and the none case is smaller. This is due to the fact that the detection area is enlarged. We can also observe the diurnal pattern that the frequency starts to increase after 8 A.M. for multiple cases and drops after 12 A.M. (e.g., sleep time). For the remainder of the paper, we do not include the time period between 12 A.M. and 8 A.M. in the following analysis to avoid introducing bias where significant portions of the participants may be sleeping.

Next, we analyze the traffic consumption at different proximity densities. For a given proximity density and threshold, the average traffic consumption is calculated by averaging the total traffic consumed when the screen is active over all the corresponding time slots. The filtering of active screen traffic helps to characterize active use despite nearby users rather than simple background traffic (e.g., Facebook newsfeed). Figure 5 shows the results of proximity on active usage. We can see that there is not a statistically significant difference between the thresholds. For near proximity (-80 dBm), when proximity density increases from cases of no nearby users to one user, the average traffic decreases significantly ($p < 0.01$). When the proximity density change from the case of one user to multiple users, there is not a statistically significant difference in traffic consumption. For close proximity (-65 dBm), no statistically significant change is observed. Since there is not a significant impact of either of the two thresholds, we choose to use near proximity (-80 dBm) in the following analysis due to its larger sample size for the cases of single user and multiple users.

We further refine proximity to 'meaningful proximity' from a social relationship perspective. We define two types of social relationship: SMS contacts and Facebook friends. Two people are SMS contacts if they have sent or receive at least one text message per month. Facebook friends mean that the two people are on each other's Facebook friend list. Since we only record SMS/Facebook information for our participants, the proximity density for friends could only be measured for in-study devices, i.e., the number of in study devices detected within a certain area. In order to ensure that measuring only based on in-study devices would not introduce bias, we compare the results with measuring based on all devices as shown in Figure 6. We can see that there is no significant difference between the two types of proximity density measure. Figure 6 also shows the results computed by measuring proximity density based on SMS contacts and Facebook friends within the study. When greater than two SMS contacts or Facebook friends are around, traffic consumption significantly decreases. Table I shows the frequency of different proximity densities in each case. SMS contacts and Facebook friends have a much smaller sample size than the other cases when single and multiple users nearby but still provides significant results.

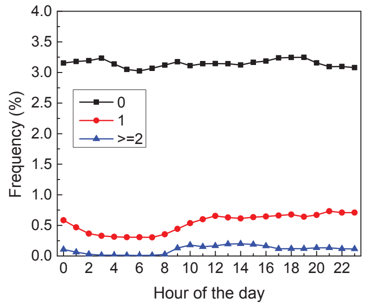


Fig. 3: Frequency of different proximity densities over time (close proximity,-65 dBm)

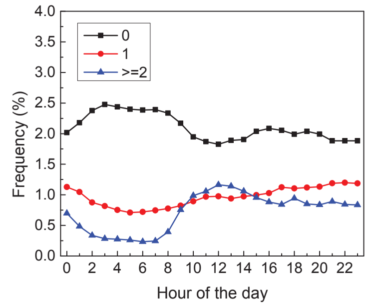


Fig. 4: Frequency of different proximity densities over time (near proximity,-80 dBm)

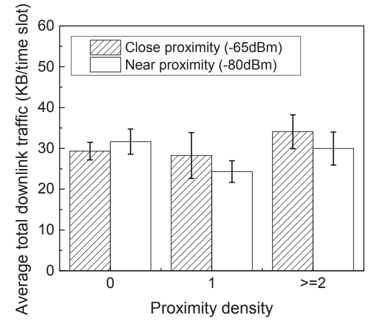


Fig. 5: Average traffic consumption by different proximity densities when screen is active

B. Big-Five Personality Traits

Next, we analyze if individual difference (i.e., personality) would influence user's behavior in terms of traffic usage. We measure the Big-Five personality traits based on the questionnaire although it is important to note that personality tends to be remarkably invariant over time. Big-Five personality traits are five broad categories that are used to describe human personality (see details in Table II). Each trait represents a range between two extremes of zero to five with most individuals tending away from the extremes. We categorize participants based on scores of the Big-Five Personality traits. Table III shows the number of participants in each category.

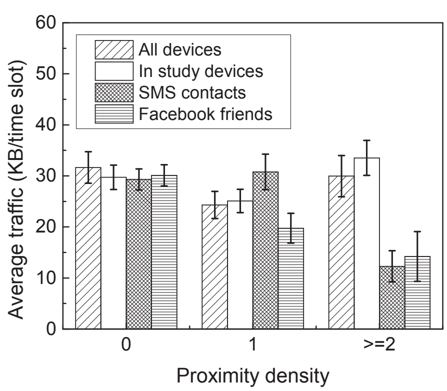


Fig. 6: Average traffic by social relationship when screen is active (-80 dBm)

# Devices	All devices	In-study devices	SMS friends	Facebook friends
0	350462	551118	652659	642544
1	181157	95274	35811	42042
$\geq$2	159136	44363	2285	6169
Total	690755	690755	690755	690755

TABLE I: Total frequency of different proximity densities measured on four types of social relations

Figure 7 shows the average screen on traffic for each of the five personality with different scores. We can see that extroversion, neuroticism, and openness have a significant impact on the total traffic consumption. An extroversion score of greater than or equal to four tends to have significant more screen on traffic than an extroversion score of three and less than three. For neuroticism, a score of less than or equal to two tends to consume more traffic than score of greater than or equal to three. For openness, a score of greater than or equal to four tends to have higher traffic consumption than score of less than or equal to three. Personality traits such as agreeableness and conscientiousness do not appear to have a statistically distinguishable impact.

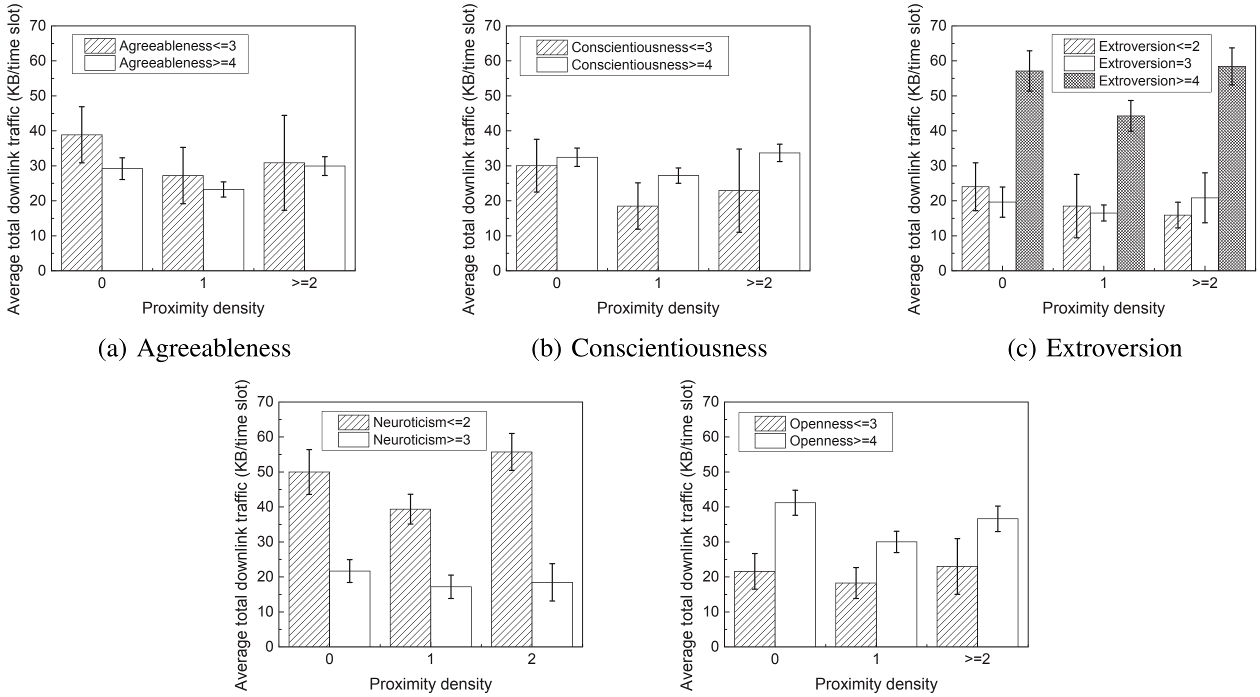
Trait	Description
Agreeableness	Affable, tolerant, sensitive, trusting, kind, and warm
Extraversion	Outgoing, talkative, sociable
Conscientiousness	Organized, systematic, punctual, achievement oriented
Neuroticism	Anxious, irritable, temperamental, and moody
Openness	Curious, original, intellectual, creative

TABLE II: Big-Five Personality description [16]

Personality traits	Score				
	1	2	3	4	5
Agreeableness	0	0	22	56	0
Extroversion	1	16	39	18	4
Conscientiousness	0	0	28	50	0
Neuroticism	0	25	51	2	0
Openness	0	1	37	40	0

TABLE III: Number of participants in different scores for the Big-Five personality traits

In order to analyze the reason why some personality traits have a significant impact on the traffic consumption, we list the top five applications used by participants with different scores for a given personality. Due to limited space, we present the results for extroversion as shown in Table IV. The third column is the traffic consumed out of the total traffic aggregated over all the applications. The fourth column is the traffic consumption in MB per week per user. Each application has at least five users for the application. Unsurprisingly, the core browser ranks first in both score categories. However, more extroverted participants (i.e., score greater than or equal to four) tend use much the browser significantly more than less extroverted participants (i.e., score less than four). Browser traffic constitutes nearly 70% percent of the total traffic for more extroverted participants while 45% for less extroverted participants. More extroverted participants also tend to use more of the Google Play Store and Google Maps (Friend service) while less extroverted participants tends to generate more Facebook traffic. The top five application list of more extroverted people constitute 96.57% of the total while the ones from less extroverted constitute 77.59% of the total. The



(a) Agreeableness (b) Conscientiousness (c) Extroversion

(d) Neuroticism (e) Openness

Fig. 7: Average screen on traffic for Big five personality traits

number of total distinct applications that generate traffic from less extroverted people is nearly 1.6 times of the number from more extroverted people.

Ranking	Applications	Traffic (%)	Traffic per week per user (MB)
1	Android Browser	45.62	337.68
2	Facebook	14.30	107.52
3	Android Weather	9.23	67.2
4	Google Maps (Friend service)	5.16	36.96
5	Google Play Store	3.28	23.52
Total		77.59	

(a) Extroversion score ≤ 3

Ranking	Applications	Traffic (%)	Traffic per week per user (MB)
1	Android Browser	69.04	994.56
2	Google Play Store	14.69	208.32
3	Google Maps (Friend service)	9.25	131.04
4	Android Weather	2.07	0.18
5	Google Maps (Google Location service)	1.52	0.13
Total		96.57	

(b) Extroversion score ≥ 4

TABLE IV: Top five apps ranked by the traffic when screen is active

C. Location categories

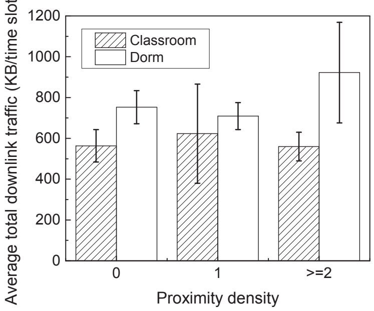
Finally, we break down the traffic by application and study whether users have preferences with regards to applications at different locations. We select several representative/popular applications in different genres to study. The chosen applications include Facebook, the browser, ESPN, Pandora, and Gmail. As we mentioned before, we choose the top two ranked locations of the classroom and dormitory as an example to analyze the results (see Figure 8). We can see that Facebook, ESPN, and Gmail tend to consume more traffic in the dormitory than in the classroom while the browser and Pandora tend to be used more in the classroom. To analyze the impact of proximity density, we compute pairwise significance among none, single and multiple cases. Table V shows only the significant pairs. Notably, proximity density does not have a statistically significant impact on Facebook or ESPN but does have an impact on the browser, Pandora, and Gmail. For the browser, the traffic drops in the dorm from the case of zero user to one user nearby in the classroom and drops in the classroom from the case of zero user to one and more users nearby. For Pandora, the traffic drops from zero to more than zero nearby in both dorm and classroom. For Gmail, the traffic drops from the cases of zero user to one user nearby in the dorm.

Apps	Dorm	Classroom
Browser	None and single ($p < 0.001$)	None and single ($p < 0.05$) None and multiple ($p < 0.001$)
Pandora	None and single ($p < 0.05$) None and multiple ($p < 0.05$)	None and single ($p < 0.001$) None and multiple ($p < 0.001$)
Gmail	None and single ($p < 0.05$)	

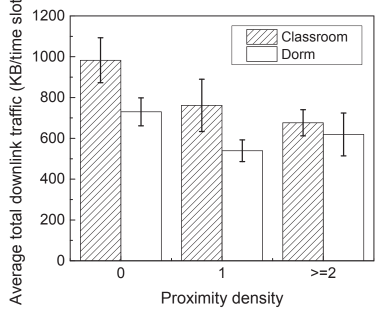
TABLE V: Significantly different pairs of different proximity densities

V. Conclusion

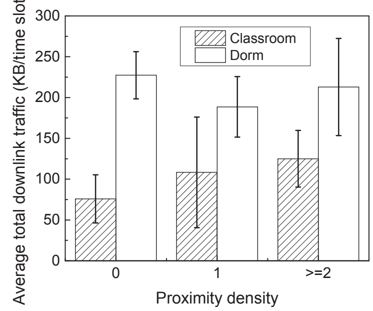
In this paper, we analyzed the impact of proximity density, location, and personality on smartphone traffic consumption. Our results shows that 1) friendship (i.e., SMS contacts and Facebook friendship) in proximity has a significant impact on



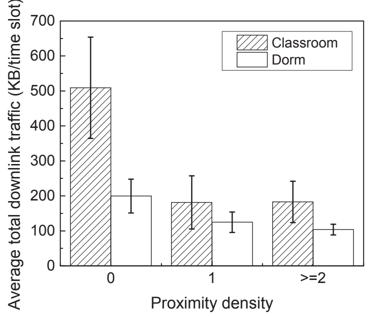
(a) Facebook



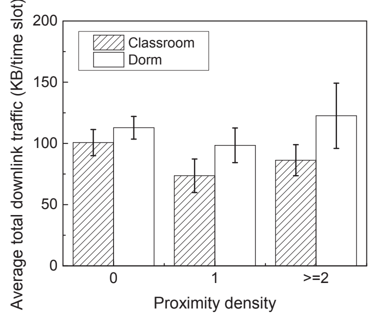
(b) Browser



(c) ESPN



(d) Pandora



(e) Gmail

Fig. 8: Average screen on traffic for selected applications (-80 dBm, downlink)

traffic consumption; 2) personality tends to impact application preference / consumption; and 3) applications can have significantly different contextual usages based on the location. We believe our study raises the importance of the entirety of proximity, personality, and location context as future data relevant for the purposes of assessing user data consumption.

ACKNOWLEDGMENT

This work was funded in part by the National Science Foundation through grant IIS-0968529.

REFERENCES

[1] Y. Jin, N. Duffield, A. Gerber, P. Haffner, W.-L. Hsu, G. Jacobson, S. Sen, S. Venkataraman, and Z.-L. Zhang, "Characterizing data usage patterns in a large cellular network," in *Proceedings of the 2012 ACM SIGCOMM Workshop on Cellular Networks: Operations, Challenges, and Future Design*, ser. CellNet '12. New York, NY, USA: ACM, 2012, pp. 7–12.

[2] Cisco, "Cisco visual networking index: Global mobile data traffic forecast update," *White Paper*, Februry 2012.

[3] T. M. T. Do, J. Blom, and D. Gatica-Perez, "Smartphone usage in the wild: a large-scale analysis of applications and context," in *ICMI '11*. New York, NY, USA: ACM, 2011, pp. 353–360.

[4] B. Karlsson, K. Okada, and T. Noleto, "A mobile-based system for context-aware music recommendations," in *Artificial Intelligence Applications and Innovations*, ser. IFIP Advances in Information and Communication Technology, L. Iliadis, I. Maglogiannis, H. Papadopoulos, K. Karatzas, and S. Sioutas, Eds. Springer Berlin Heidelberg, 2012, vol. 382, pp. 520–529.

[5] G. Chittaranjan, J. Blom, and D. Gatica-Perez, "Who's who with big-five: Analyzing and classifying personality traits with smartphones," in *Proceedings of the 2011 15th Annual International Symposium on Wearable Computers*, ser. ISWC '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 29–36.

[6] J. Blom, J. Chipchase, and J. Lehikoinen, "Contextual and cultural challenges for user mobility research," *Commun. ACM*, vol. 48, no. 7, pp. 37–41, Jul. 2005.

[7] N. Eagle and A. (Sandy) Pentland, "Reality mining: sensing complex social systems," *Personal Ubiquitous Comput.*, vol. 10, no. 4, pp. 255–268, Mar. 2006.

[8] A. Misra and R. K. Balan, "Livelabs: Initial reflections on building a large-scale mobile behavioral experimentation testbed," *SIGMOBILE Mob. Comput. Commun. Rev.*, vol. 17, no. 4, pp. 47–59, Dec. 2013.

[9] A. Nandugudi, A. Maiti, T. Ki, F. Bulut, M. Demirbas, T. Kosar, C. Qiao, S. Y. Ko, and G. Challen, "Phonelab: A large programmable smartphone testbed," in *Proceedings of First International Workshop on Sensing and Big Data Mining*, ser. SENSEMINE'13. New York, NY, USA: ACM, 2013, pp. 4:1–4:6.

[10] Y. Zhang and A. Arvidsson, "Understanding the characteristics of cellular data traffic," in *CellNet '12*. New York, NY, USA: ACM, 2012, pp. 13–18.

[11] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin, "Diversity in smartphone usage," in *MobiSys '10*. New York, NY, USA: ACM, 2010, pp. 179–194.

[12] Q. Xu, J. Erman, A. Gerber, Z. Mao, J. Pang, and S. Venkataraman, "Identifying diverse usage behaviors of smartphone apps," in *IMC '11*. New York, NY, USA: ACM, 2011, pp. 329–344.

[13] M. Shafiq, L. Ji, A. Liu, J. Pang, and J. Wang, "Characterizing geospatial dynamics of application usage in a 3g cellular data network," in *INFOCOM*, 2012, pp. 1341–1349.

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

[15] S. Butt and J. G. Phillips, "Personality and self reported mobile phone use," *Computers in Human Behavior*, vol. 24, no. 2, pp. 346–360, Mar. 2008.

[16] R. R. McCrae and O. P. John, "An introduction to the five-factor model and its applications," *Journal of Personality*, vol. 60, pp. 175–215, 1992.