Diversity in Smartphone Usage

Hossein Falaki
CENS, UCLA

Ratul Mahajan
Microsoft Research

Srikanth Kandula
Microsoft Research

Dimitrios Lymberopoulos
Microsoft Research

Ramesh Govindan
CENS, USC

Deborah Estrin
CENS, UCLA

Abstract – Using detailed traces from 255 users, we conduct a comprehensive study of smartphone use. We characterize intentional user activities – interactions with the device and the applications used – and the impact of those activities on network and energy usage. We find immense diversity among users. Along all aspects that we study, users differ by one or more orders of magnitude. For instance, the average number of interactions per day varies from 10 to 200, and the average amount of data received per day varies from 1 to 1000 MB. This level of diversity suggests that mechanisms to improve user experience or energy consumption will be more effective if they learn and adapt to user behavior. We find that qualitative similarities exist among users that facilitate the task of learning user behavior. For instance, the relative application popularity for can be modeled using an exponential distribution, with different distribution parameters for different users. We demonstrate the value of adapting to user behavior in the context of a mechanism to predict future energy drain. The 90th percentile error with adaptation is less than half compared to predictions based on average behavior across users.

Categories and Subject Descriptors
C.4 [Performance of systems] Measurement techniques
General Terms
Measurement, human factors
Keywords
Smartphone usage, user behavior

1. INTRODUCTION

Smartphones are being adopted at a phenomenal pace but little is known (publicly) today about how people use these devices. In 2009, smartphone penetration in the US was 25% and 14% of worldwide mobile phone shipments were smartphones [23, 16]. By 2011, smartphone sales are projected to surpass desktop PCs [25]. But beyond a few studies that report on users' charging behaviors [2, 17] and relative power consumption of various components (e.g., CPU, screen) [24], many basic facts on smartphone usage are unknown: *i*) how often does a user interact with the phone and how long does an interaction last? *ii*) how many applications does a user run and how is her attention spread across them? *iii*) how much network traffic is generated?

Answering such questions is not just a matter of academic interest; it is key to understanding which mechanisms can effectively improve user experience or reduce energy consumption. For instance, if user interactions are frequent and the sleep-wake overhead is significant, putting the phone to sleep aggressively may be counterproductive [8]. If the user interacts regularly with only a few applications, application response time can be improved by retaining those applications in memory [7]. Similarly, if most transfers are small, bundling multiple transfers [1, 22] may reduce per-byte energy cost. Smartphone usage will undoubtedly evolve with time, but understanding current usage is important for informing the next generation of devices.

We analyze detailed usage traces from 255 users of two different smartphone platforms, with 7-28 weeks of data per user. Our traces consist of two datasets. For the first dataset we deploy a custom logging utility on the phones of 33 Android users. Our utility captures a detailed view of user interactions, network traffic, and energy drain. The second dataset is from 222 Windows Mobile users across different demographics and geographic locations. This data was collected by a third party.

We characterize smartphone usage along four key dimensions: *i*) user interactions; *ii*) application use; *iii*) network traffic; and *iv*) energy drain. The first two represent intentional user activities, and the last two represent the impact of user activities on network and energy resources. Instead of only exploring average case behaviors, we are interested in exploring the range seen across users and time. We believe that we are the first to measure and report on many aspects of smartphone usage of a large population of users.

A recurring theme in our findings is the diversity across users. Along all dimensions that we study, users differ by one or more orders of magnitude. For example, the mean number of interactions per day for a user varies from 10 to 200; the mean interaction length varies from 10 to 250 seconds; the number of applications used varies from 10 to 90; and the mean amount of traffic per day varies from 1 to 1000 MB, of which 10 to 90% is exchanged during interactive use. We also find that users are along a continuum between the extremes, rather than being clustered into a small number of groups.

The diversity among users that we find stems from the fact that users use their smartphones for different purposes and with different frequencies. For instance, users that use

MobiSys'10, June 15–18, 2010, San Francisco, California, USA.

	#users	Duration	Platform	Demographic info	Information logged
Dataset1	33	7-21 weeks/user	Android	16 high school students, 17 knowledge workers	Screen state, applications used network traffic, battery state
Dataset2	222	8-28 weeks/user	Windows Mobile	61 SC, 65 LPU, 59 BPU, 37 OP Country: 116 USA, 106 UK	Screen state, applications used

Table 1: An overview of the datasets in our study.

games and maps applications more often tend to have longer interactions. Our study also shows that demographic information can be an unreliable predictor of user behavior, and usage diversity exists even when the underlying device is identical, as is the case for one of our datasets.

Among the many implications of our findings, an overriding one is that mechanisms to improve user experience or energy consumption should not follow a one-size-fits-all mindset. They should instead adapt by learning relevant user behaviors; otherwise, they will likely be only marginally useful or benefit only a small proportion of users.

We show that despite quantitative differences qualitative similarities exist among users, which facilitates the task of learning user behavior. For several key aspects of smartphone usage, the same model can describe all users; different users have different model parameters. For instance, the time between user interactions can be captured using the Weibull distribution. For every user, the shape parameter of this model is less than one, which implies that the longer it has been since the user's last interaction, the less likely it is for the next interaction to start. We also find that the relative popularity of applications for each user follows an exponential distribution, though the parameters of the distribution vary widely across users.

We demonstrate the value of adapting to user behavior in the context of a mechanism to predict future energy drain. Predicting energy drain is an inherently challenging task. Bursty user interactions at short time scales combined with diurnal patterns at longer time scales lead to an energy consumption process that has a very high variance and is seemingly unpredictable. We show, however, that reasonably accurate predictions can be made by learning the user's energy use signature in terms of a "trend table" framework. For predicting the energy use one hour in the future, our predictor's 90th percentile error is under 25%. Without adaptation and basing the predictions on average behavior, the 90th percentile error is 60%.

2. DATA COLLECTION

Our work is based on two sets of data. The first is a high-fidelity data set that we gathered by deploying a custom logger on the phones of 33 Android users. The second data set consists of 222 Windows Mobile users across different demographics. Together, these data sets provide a broad and detailed view of smartphone usage. We leave for the future the task of studying other smartphone platforms such as iPhone and BlackBerry. The characteristics of our datasets are summarized in Table 1.

2.1 Dataset1

Our first set of traces is from 33 Android users. These users consisted of 17 knowledge workers and 16 high school students. Knowledge workers were computer science researchers and high school students were interns in a single organization and were recruited by a third person on our behalf. As stated in our study consent form, the users' identities were not revealed to us. The participants were given HTC Dream smartphones with unlimited voice, text and data plans. We encouraged the users to take advantage of all the features and services of the phones.

The data was collected using a custom logging tool that we developed and deployed on the smartphones. The logger runs in the background and records a highly detailed view of smartphone use, including the state of the smartphone screen, start and end of incoming and outgoing voice calls, the time the user spends interacting with each application, the network traffic sent and received per application, and the battery level. The Android OS provides mechanisms to access this information. The logger keeps data records in a local SQLite database on the phone and uploads them only when the phone is plugged to the charger, to minimize the impact on the phone battery. Our logging utility is available to other researchers by request.

The data was gathered between May and October 2009. There is 7-21 weeks of data per user, with the average being 9 weeks.

2.2 Dataset2

Our second data set was collected by an organization that was interested in investigating smartphone usability and application popularity. This organization provided 222 users with Windows Mobile smartphones from different hardware vendors. It also paid for their voice and data plans. For representativeness, the users were drawn from different demographics as shown in Table 1. The demographic categories were defined based on what users stated as the primary motivation for using a smartphone. Social Communicators (SC) wanted to "stay connected via voice and text." Life Power Users (LPU) wanted "a multi-function device to help them manage their life." Business Power Users (BPU) wanted "an advanced PC-companion to enhance their business productivity." Organizer Practicals (OP) wanted "a simple device to manage their life." The subjects were asked about their intended use of the phone before the study and were categorized based on their answers. To our knowledge, the results of this study are not public.

Traces were collected using a logger that recorded start and end time of each application. This information was logged using the ActiveApplication API call of the OS, which reports on the executable that currently has the foreground window (with a callback for changes) Other details that our custom logger in §2.1 records (e.g., network traffic and battery level) were not logged in this study. Thus, this dataset has lower fidelity than the first one, but it provides a view of smartphone usage across a broader range of users.

The traces were collected between May 2008 and January 2009. There is 8-28 weeks of data per user, with the average being 16 weeks.

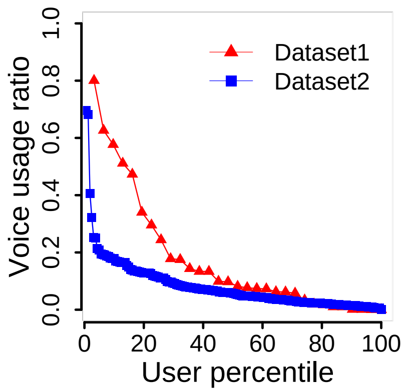


Figure 1: Ratio of voice usage to total usage. The x-axis is user percentile and users are sorted in decreasing order of ratio.

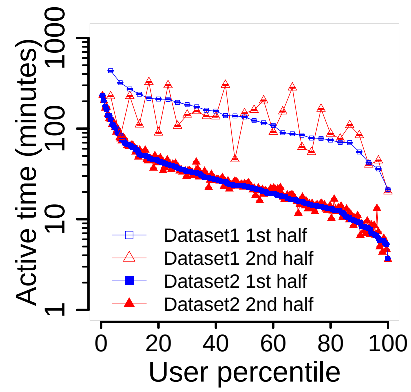


Figure 2: Total interaction per day during the first and second halves of study for each user. Users within each dataset are sorted based on the interaction time in the first half. The y-axis is log scale.

2.3 Representativeness of conclusions

An important concern for user studies such as ours is whether the resulting conclusions represent the entire population. There are two potential sources of bias in our data: i) the users are not representative; and ii) the measured usage is not representative. We believe that our conclusions are general. The first concern is alleviated by the fact that aside from some quantitative differences, we find remarkable consistency among users in the two datasets. This consistency suggests generality given that the two datasets are gathered independently, on different platforms, and Dataset2 was professionally designed to be representative.

The second concern stems from the possibility that users may not be using the monitored smartphones as their primary devices or that the usage during the monitoring period may not be normal. All users in Dataset2 used the provided smartphones as their primary devices. We do not know this aspect with certainty for Dataset1, but we understand from anecdotal evidence that some users used these devices as their only phones and others took advantage of the unlimited minutes and text plans. We study voice usage as indicative of the extent to which users relied on the monitored devices. Higher voice usage suggests use as primary phones. Figure 1 shows the ratio of time users spent in phone calls to total time spent interacting with the phone (§3). We see that the overall voice usage in Dataset1 was higher than that in Dataset2 in which all users used the phone as their primary device.

Given that the monitored devices tended to be primary and the long duration of the monitoring interval, we conjecture that our traces predominantly capture normal usage. Earlier work has pointed to the possibility of an initial adoption process during which usage tends to be different than long-term usage [19]. To show that our traces are not dominated by the initial excitement of users or other special events that cause usage to be appreciably different from the normal usage, Figure 2 shows the average interaction time per day (§3) in the first and second halves of the datasets for each user. We see roughly similar usage in the two halves. Detailed investigation shows that the visible differences in the averages of the two halves, especially in Dataset1, are not statistically significant. Other measures of usage (e.g., network activity) look similar. We do not claim that instances of abnormal usage are absent in the datasets, but the monitored period was long enough for our results to not be impacted by such instances.

3. USER INTERACTIONS

We begin our analysis by studying how users interact with their smartphones, independent of the application used. We characterize application use in the next section, and the impact of user actions on network traffic and energy drain in the following sections.

We define an interaction interval, also referred to as a session in this paper, differently for each dataset. In Dataset1, we deem a user to be interacting with the phone whenever the screen is on or a voice call is active. In Dataset2, an interaction is defined as the interval that an application is reported to be on the foreground. This includes voice calls because on Windows Mobile a special program ("cprog.exe") is reported in the foreground during voice calls.

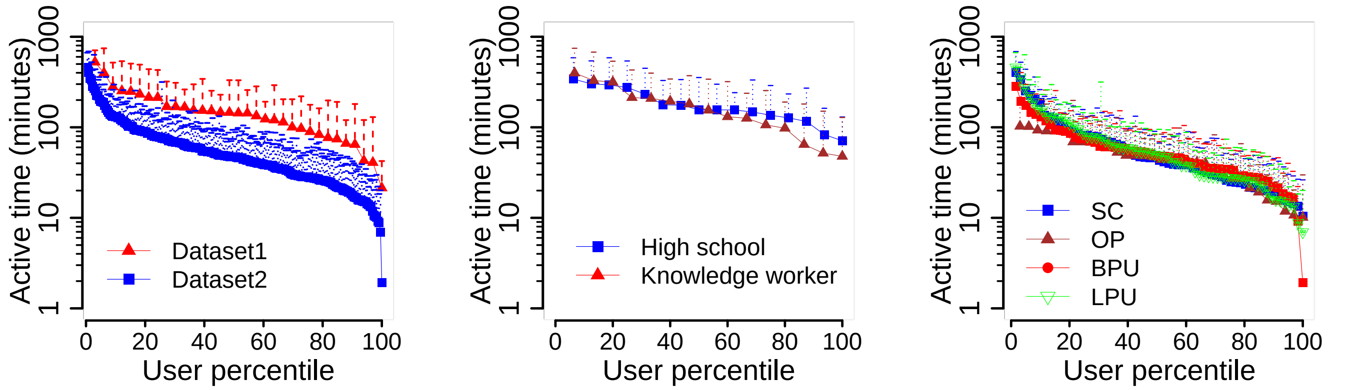
3.1 Interaction Time

Figure 3(a) shows a basic measure of user interaction—the number of minutes in a day that a user interacts with the smartphone. The plot shows the mean and the standard deviation of this number for each user. For visual clarity, in such graphs, we plot only the upper end of the standard deviation; plotting both ends occludes the other curves. The interested reader can estimate the lower end since standard deviation is symmetric around the mean.

Dataset1 users tend to have more interaction minutes because, as we show later, they tend to have longer interaction sessions while having a similar number of sessions. Within each dataset, however, there is an order of magnitude difference among users. In Dataset1, the lower end is only 30 minutes in a day. But the high end is 500 minutes, which is roughly eight hours or a third of the day. We are surprised by this extremely high level of usage.

Figure 3(a) also shows that users cover the entire range between the two extremes and are not clustered into different regions. The lack of clusters implies that effective personalization will likely need to learn an individual user's behavior rather than mapping a user to one or a few predefined categories.

We examine two factors that can potentially explain the extent to which a user interacts with the phone but find that



(a) All users (b) Dataset1 (c) Dataset2

Figure 3: The mean and the upper end of the standard deviation of interaction minutes per day. (a) All users in each dataset. (b)&(c) Different demographics in the two datasets. The y-axes are log scale.

neither is effective. The first is that heavier users use different types of applications (e.g., games) than lighter users. But, we find that the relative popularity of application types is similar across classes of users with different interaction times (§4.2). The second is user demographic. But, as Figures 3(b) and 3(c) show, the interaction times are similar across the different demographics in the two datasets. Within each demographic, user interaction times span the entire range. In §4.2, we show that user demographic does not predict application popularity either.

To understand the reasons behind diversity of user interaction times, we study next how user interaction is spread across individual sessions. This analysis will show that there is immense diversity among users in both the number of interaction sessions per day and the average session length.

3.2 Interaction Sessions

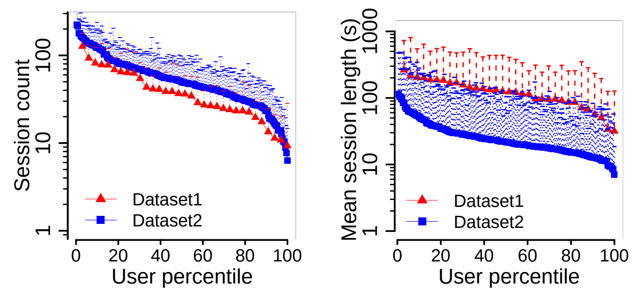
Interaction sessions provide a detailed view of how a user interacts with the phone. Their characteristics are important also because energy use depends not only on how long the phone is used in aggregate but also on the usage distribution. Many, short interactions likely drain more energy than few, long interactions due to the overheads of awakening the phone and radio. Even with negligible overheads, battery lifetime depends on how exactly energy is consumed [20]. Bursty drain with high current levels during bursts can lead to a lower lifetime than a more consistent drain rate.

Figure 4(a) shows the number of sessions per day for different users. We again see a wide variation. Individual users interact with their smartphone anywhere between 10 to 200 times a day on average.

Figure 4(b) shows the mean and standard deviation of interaction session lengths. Dataset1 users tend to have much longer sessions than Dataset2 users. Given that they have roughly similar number of interactions per day, as seen in Figure 4(a), their longer sessions explain their higher interaction time per day, as seen in Figure 3(a).

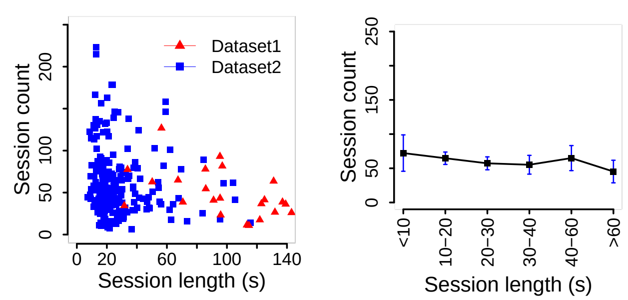
Within each dataset, the mean session length varies across users by an order of magnitude. Across both datasets, the range is 10-250 seconds.

Explaining the diversity in session lengths: Several hypothesis might explain the differences in different users'



(a) Number of sessions (b) Session length

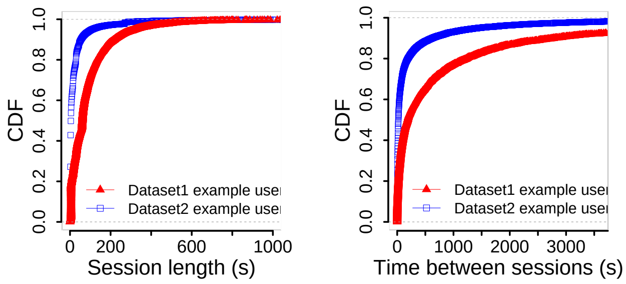
Figure 4: The mean and the upper end of the standard deviation for the number of sessions per day and the session length. The y-axes are log scale.



(a) (b)

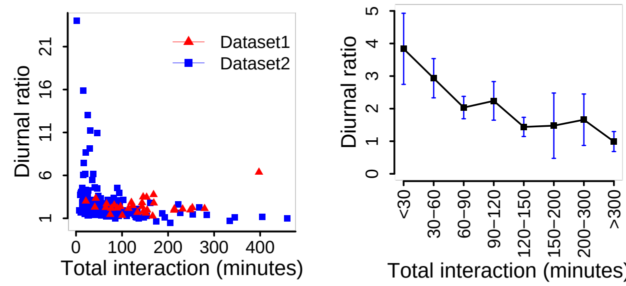
Figure 5: (a) Scatterplot of session count per day and mean session length of various users. (b) The mean and 95% CI of session count per day for users in Dataset2 with different mean session lengths.

session lengths. One hypothesis is that users with longer sessions concentrate their smartphone usage in fewer sessions. Figure 5 shows, however, that there is little correlation between users' number of sessions and session length. Figure 5(a) shows a scatterplot of session count versus mean length for different users. There is one data point for each user. Figure 5(b) shows the dependence of session count on session length by aggregating data across Dataset2 users. It plots the observed mean and 95% confidence interval (CI) for



(a) Session lengths (b) Time between sessions

Figure 6: CDFs of session length and time between sessions for two example users. The x-axis ranges in the graphs are different.



(a) (b)

Figure 8: (a) Scatterplot of diurnal ratio of interaction time per hour and interaction minutes per day. (b) The mean and 95% CI of diurnal ratio vs. total interaction time per day.

session counts per day for users with different mean session lengths. The differences in the session counts are not statistically significant. In other words, it is not the case that users who have longer sessions have fewer or more sessions.

Our other hypotheses are related to application use. The second hypothesis is that users run varying numbers of applications during an interaction, and users that tend to use more applications per session have longer sessions. The third hypothesis is that users run different applications and some applications, such as maps, have longer sessions than others. The fourth one is that even for the same application, users have different session lengths.

Our analysis of application use in §4 reveals that the second hypothesis is not explanatory, as users overwhelmingly use only one application per session. It also reveals that the third and fourth hypotheses are likely contributors to diversity in session lengths. Note that the inability of application types to explain interaction time per day, which we mention in the previous section, is different from their ability to explain session lengths.

Distribution of a single user's sessions: We find that for any given user, most of the sessions are short but some are very long. Figure 6(a) shows the CDF of session lengths for two example users. The median session length is less than a minute but some are longer than an hour (not shown in the graph). A similar skewed distribution can be seen for all users in our datasets, albeit with different median and mean session length values. This highly skewed distribution also explains why the standard deviations in Figure 4(b) are high relative to the mean. In §8.1, we show how session lengths depend on the screen timeout values.

Figure 6(b) shows that the time between sessions, when the phone is not used, also has a skewed distribution. Most are short (relative to the mean) but some are very long. We show later that these off periods have the property that the longer a user has been in one of them, the greater the chance that the user will continue in this state.

3.3 Diurnal Patterns

We now study diurnal patterns in interaction. The presence of such patterns has several consequences. For instance, the length of time a given level of remaining battery capacity lasts will depend on the time of day.

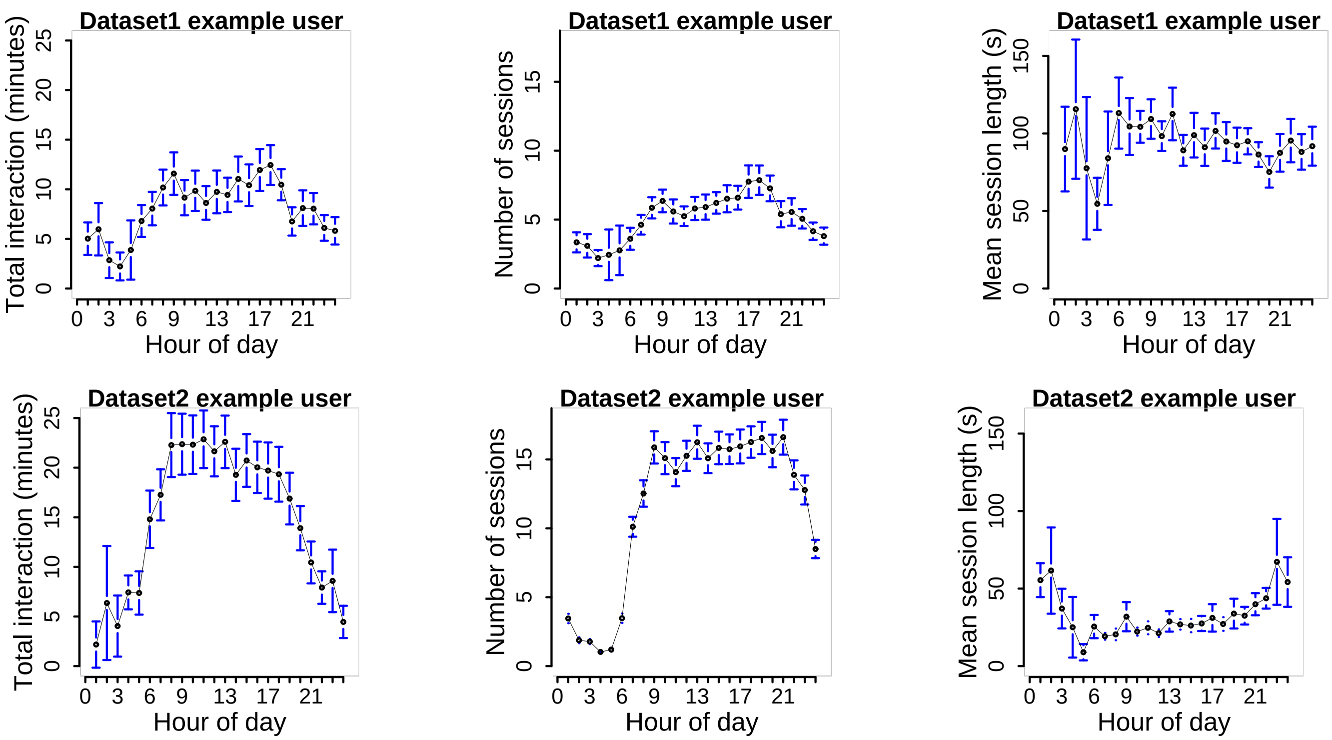
Figure 7 shows for two example users that, as expected, strong diurnal patterns do exist. As a function of the hour of the day, Figure 7(a) plots the mean number of interaction minutes per hour. It also plots the 95% confidence interval (CI) around the mean, which can be used to judge if the differences in the means are statistically significant. We see a clear pattern in which daytime use is much higher than nighttime use, though the exact pattern for different users is different.

Figure 7(a) also shows that usage at hours in the night is low but not completely zero. We believe that this non-zero usage stems from a combination of irregular sleeping hours and users using their devices (e.g., to check time) when they get up in the middle of the night.

To capture the significance of the diurnal pattern for a user, we define the *diurnal ratio* as the ratio of the mean usage during the peak hour to the mean usage across all hours. A diurnal ratio of one implies no diurnal pattern, and higher values reflect stronger patterns. Figure 9(a) plots the diurnal ratio in interaction time for all users. It shows that while diurnal ratios vary across users, roughly 70% of the users in each dataset have a peak hour usage that is more than twice their mean usage.

Explaining the diversity in diurnal patterns: To help explain the variability among users' diurnal ratios, in Figure 8 we study its dependence on interaction time. Figure 8(a) shows a scatterplot of the diurnal ratio and the mean interaction time per day. We see that the diurnal ratio tends to be inversely correlated with interaction time. Figure 8(b) shows this negative correlation more clearly, by aggregating data across users. It plots the mean and 95% CI of the diurnal ratio of total interaction time per day for users with different total interaction times. The diurnal ratio decreases as interaction time increases. This inverse relationship suggests that heavy users tend to use their phone more consistently during the day whereas light users tend to have concentrated use during certain hours of the day.

Understanding the source of diurnal patterns: The variation in interaction time of a user across the day can result from variation in the number of interaction sessions or the length of individual sessions. We find that both factors contribute. Users tend to have different number of sessions as well as different session lengths at different hours of the day. Figures 7(b) and 7(c) illustrate this point for two example users. They plot the mean number of sessions and the mean session length for each hour of the day.



(a) Interaction time (b) Number of sessions (c) Session length

Figure 7: The mean and 95% CI of interaction time, number of sessions, and session length during each hour of the day for an example user from each dataset.

Figures 9(b) and 9(c) show the strength of the diurnal pattern for the number of sessions and session length for all the users. Observe that compared to interaction time and session length, the diurnal ratio of the number of sessions tends to be lower.

4. APPLICATION USAGE

We now study the applications that users run when they interact with their smartphones. Unlike previous attempts to understand mobile application usage [5, 19, 26] that use diaries and interviews, we rely on the mobile phone's OS to report application usage. We define an application as any executable that the OS reports. On Windows Mobile, we get timestamped records of start and end times of application executions in the foreground. On Android, we log usage counters that are updated by the OS. Every time the OS calls the *onStart*, *onRestart* or *onResume* method of an Android application it starts a timer. The timer stops when the *onPause*, *onStop*, or *onDestroy* method is called. We record periodically the cumulative value of the timer for each installed application. This information on the extent of application use is not as accurate as the equivalent information on Windows Mobile, but it helps us understand relative time spent by the user in each application.

4.1 Number of applications

Figure 10 shows the number of applications used by users in each dataset over the length of their trace. We see that this number varies significantly, from 10 to 90, across users. The median is roughly 50. We are surprised by this high number given that the iPhone, which is reported to have thousands of applications, is not part of our study. Our results show that avid use of smartphone applications is a trait shared by Android and Windows Mobile users as well.

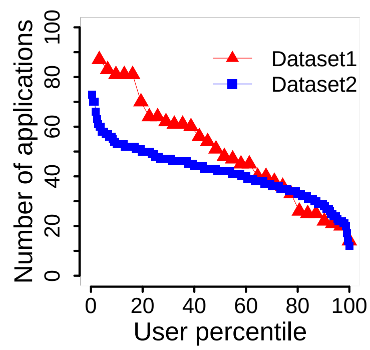
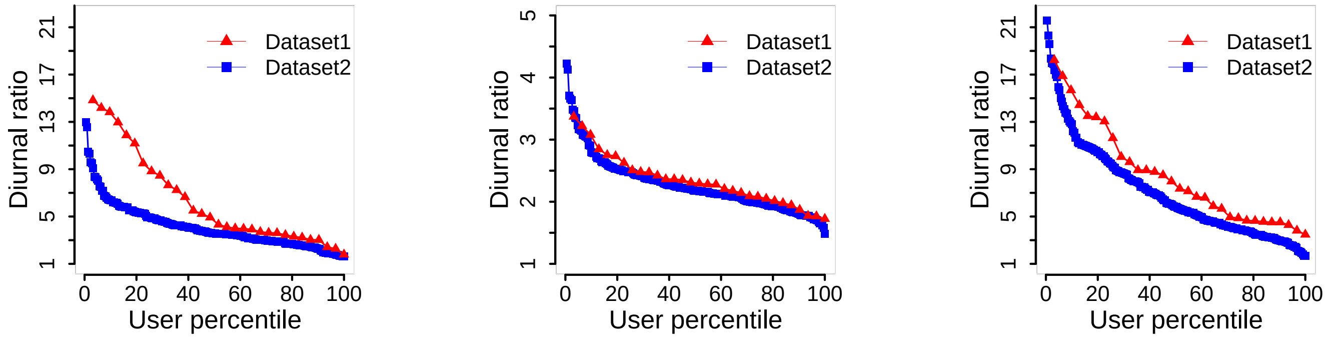


Figure 10: Number of applications installed and used by users of each dataset.

4.2 Application Popularity

The large number of applications installed by the users does not mean that they use them equally. We find that users devote the bulk of their attention to a subset of applications of their choice. Figure 11 illustrates this popularity bias for example users in each dataset. It plots the relative popularity of each application, that is, the ratio of the time spent interacting with the application and the total time spent interacting with the smartphone. The bars show the popularity PDF for the top 20 applications, and the inset shows the semi-log plot for all applications. Because they



(a) Interaction time (b) Number of sessions (c) Session length

Figure 9: Diurnal ratio of the interaction time, the number of sessions, and the session length for different users. The y-axis ranges for the number of sessions is different.

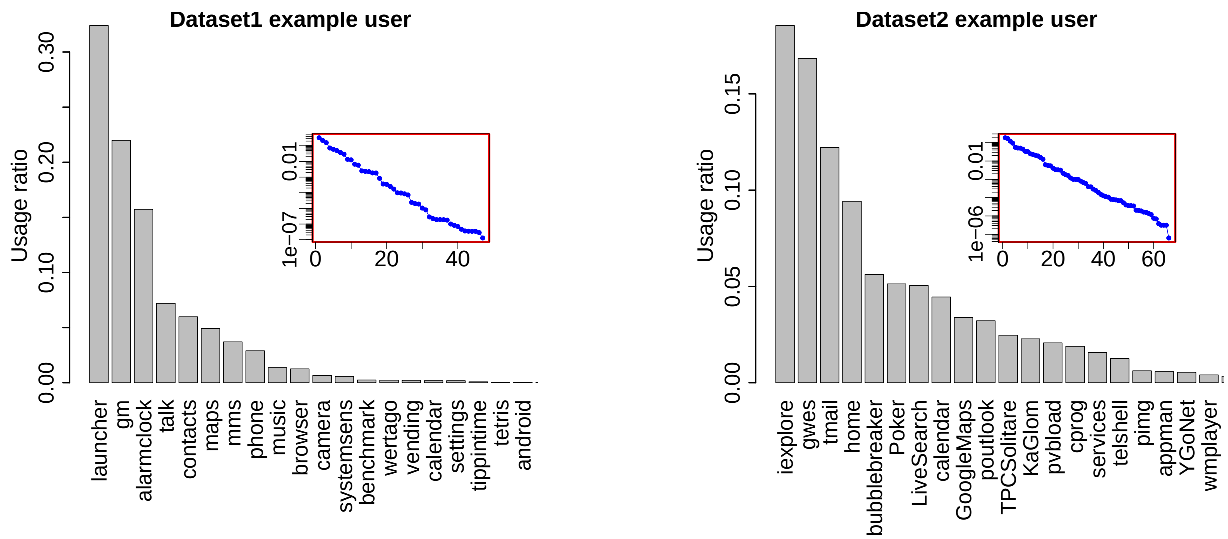


Figure 11: Relative time spent running each application for example users in each dataset. Inset is the semi-log plot of application popularity.

are binary names, even some popular applications may appear unfamiliar. For instance, in Dataset1, "launcher" is the default home screen application on Android; in Dataset2, "gwes" (Graphical, Windows, and Events Subsystem) is the graphical shell on Window Mobile.

The graphs show that relative application popularity drops quickly for both users. In §8.3, we show that for all users application popularity can be modeled by an exponential distribution.

Diurnal patterns: Interestingly, application popularity is not stable throughout the day, but has a diurnal pattern like the other aspects of smartphone use. That is, the relative popularity of an application is different for different times of the day. Figure 12 illustrates this for an example user in Dataset2. We see, for instance, that tmail.exe, which is a messaging application on Windows Mobile, is more popular during the day than night. Time dependent application popularity was recently reported by Trestian *et al.*, based on an analysis of the network traffic logs from a 3G provider [27]. Our analysis based on direct observation of user behavior confirms this effect.

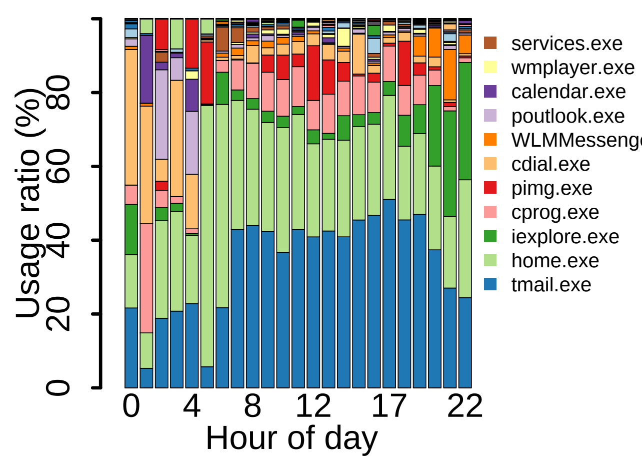


Figure 12: Relative time spent with each application during each hour of the day for a sample user and her top applications.

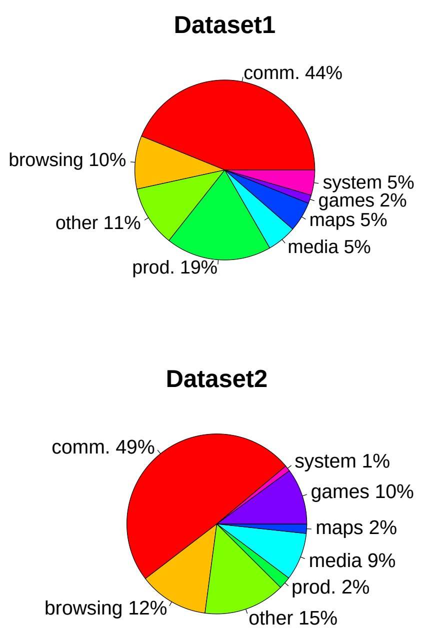


Figure 13: Relative popularity of each application category across all users in each dataset.

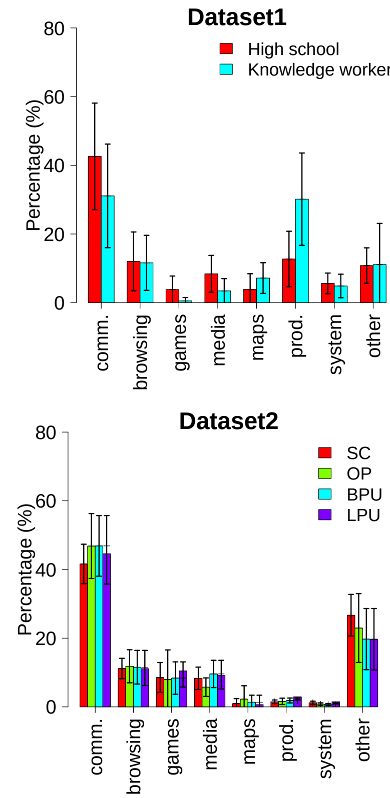


Figure 14: The mean and 95% CI of relative popularity of application categories among users of different demographics.

Aggregate view application popularity: To provide an aggregate view of what users use their smartphones for, we categorize applications into eight distinct categories: *i*) *communication* contains applications for exchanging messages (e.g., email, SMS, IM) and voice calls; *ii*) *browsing* contains Web browser, search, and social networking applications; *iii*) *media* contains applications for consuming or creating media content (e.g., pictures, music, videos); *iv*) *productivity* contains applications for calendars, alarms, and for viewing and creating text documents (e.g., Office, PDF reader); *v*) *system* contains applications for changing user preferences and viewing system state (e.g., file explorer); *vi*) games; *vii*) maps; and *viii*) *other* contains applications that we could not include in any of the categories above, e.g., because we did not know their function.

Figure 13 shows the mean relative popularity of each application category across all users in each dataset. While the results are not identical across the two datasets, they are similar to a first order. Communication dominates in both. Browsing is another major contributor in both datasets. Maps, media, and games have a comparatively lower but nevertheless substantial share of user attention.

Relationship to user demographic: To understand the extent to which user demographic determines application popularity, Figure 14 shows the mean and 95% CI of relative popularity of each application category for different user demographics. As for interaction time (§3.1), we see that the impact of user demographics in our datasets is minimal. In Dataset2, the relative popularity of various application types is similar for each of the four demographics. In Dataset1, there are noticeable differences in the mean for communication, games and productivity applications. High school students use communication and games applications more, while knowledge workers use productivity applications more. However, considering the overlap of confidence intervals, these differences in application popularity are not statistically significant.

From this result and the earlier one on the lack of dependence between user demographic and interaction time (§3.1), we conclude that user demographic, at least as defined in our datasets, cannot reliably predict how a user will use the phone. While demographic information appears to help in some cases (for e.g., the variation in usage of productivity software in Dataset1), such cases are not the norm, and it is hard to guess when demographic information would be useful. Pending development of other ways to classify users such that these classes more predictably explain the variation incorporating factors specific to a user appear necessary. This insensitivity to user demographic has positive as well as negative implications. A negative is that personalization is more complex; we cannot predict a users' behavior by knowing their demographic. A positive implication is that the range of user behaviors along many dimensions of interest can be found in several common demographics. This simplifies the task of uncovering the range because recruiting subjects across multiple demographics tends to be difficult.

Relationship to interaction time: We also study if users that interact more with their phones tend to use different applications. For each dataset, we sort users based on their average interaction time per day and partition them into different classes. For Dataset1, we use two classes, one each for the top and the bottom half of the users. For Dataset2, which has more users, we use three classes for the

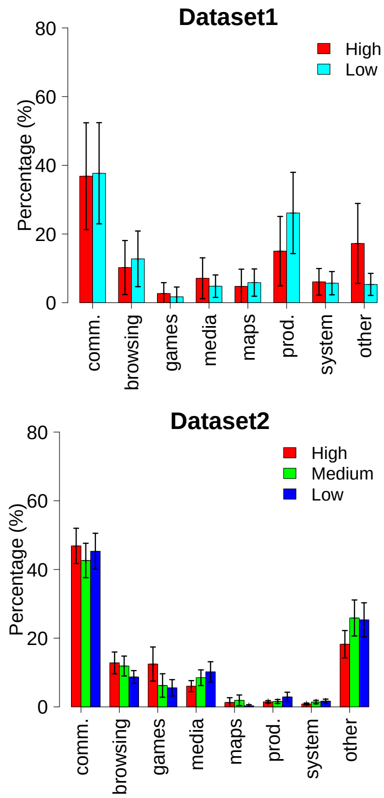


Figure 15: The mean and 95% CI of relative popularity of application categories among different classes of users based on interaction time per day.

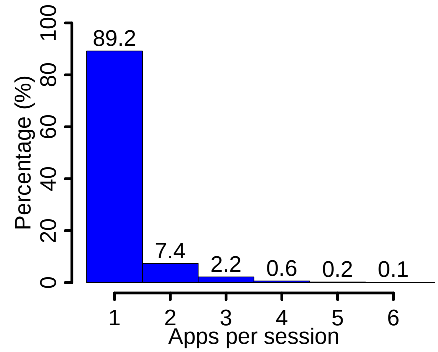


Figure 16: Histogram of the number of applications called during each interaction session for all the users in Dataset2.

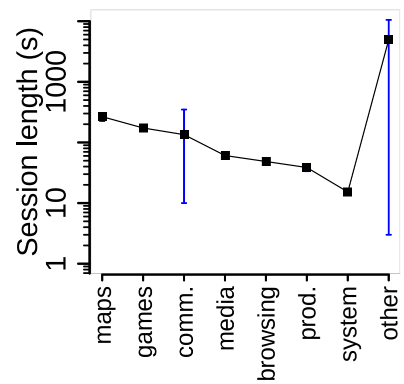


Figure 17: The mean and 95% CI of session lengths of different application categories. The y-axis is log scale.

top, middle, and bottom third of the users. Figure 15 shows the mean and 95% CI for relative time spent with each application category by each user class. We see that users in different classes have similar application usage. Thus, we cannot explain why some users use the phone more simply based on the applications that they use.

4.3 Application Sessions

We now study the characteristics of application sessions. We conduct this analysis only for Dataset2, based on timestamps for when an application is started and ended; Dataset1 does not contain this information precisely. Because applications can run in the background, start and end refer to the period when the application is in the foreground.

Applications run per interaction: We begin by studying the number of applications that users run in an interaction session. Figure 16 shows that an overwhelming majority, close to 90%, of interactions include only one application. This graph aggregates data across all users. We did not find statistically significant differences between users. A large fraction of sessions of all users have only one application.

That interaction sessions very often have only one application session suggests that users tend to interact with their smartphone for one task (e.g., reading email, checking calendar, etc.) at a time, and most of these tasks require the use of only one application.

Application session lengths: Because interaction sessions are dominated by those with only one application, the overall properties of application sessions, such as their lengths, are similar to those of interaction sessions (§3).

However, studying the session lengths of applications separately reveals interesting insights. Different application types have different session lengths, as shown in Figure 17, for the categories defined earlier. Interactions with maps and games tend to be the longest and those related to productivity and system tend to be the shortest.

Further, given an application, different users run them for different times. Figure 18 shows this effect for a messaging application, tmail.exe, and a browsing application, iexplore.exe. For each application, the mean session lengths of users differ by more than two orders of magnitude.

These observations help explain why users have different session lengths (§3.2). They prefer different applications and those application sessions tend to have different lengths. Further analysis confirms this phenomenon. For instance, if we divide users into two classes based on their mean session lengths, the popularity of games is twice as high in the class with high session lengths.

5. TRAFFIC

In this section, we investigate traffic generated by smartphones. Unlike interaction events and application use, network traffic is not an intentional user action but a side-effect of those actions. Most users are likely oblivious to how much

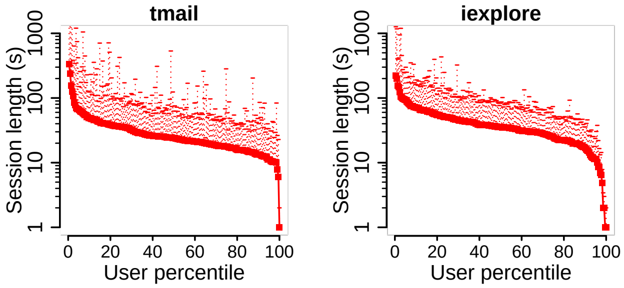


Figure 18: The mean and the upper end of the standard deviation of session lengths of two applications.

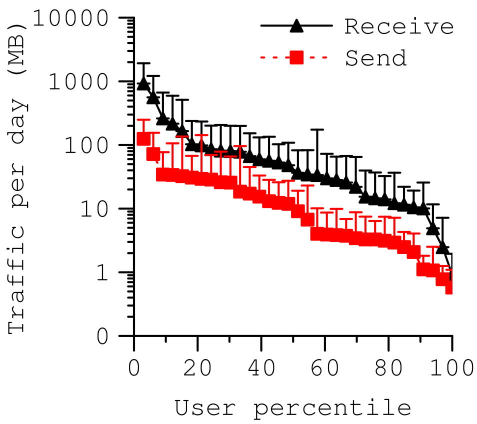


Figure 19: The mean and the upper end of the standard deviation of the traffic sent and received per day by users in Dataset1.

traffic they generate. We show that the diversity and diurnal patterns of this side-effect match those of user actions themselves.

The analysis in this section includes only Dataset1; we do not have traffic information for Dataset2. In Dataset1, we record all of the data sent (or received) by the phone except for that exchanged over the USB link, i.e., the traffic herein includes data over the 3G radio and the 802.11 wireless link.

5.1 Traffic per day

Figure 19 shows that the amount of traffic sent and received per day differs across users by almost three orders of magnitude. The traffic received ranges from 1 to 1000 MB, and the traffic sent ranges from 0.3 to 100 MB. The median values are 30 MB sent and 5 MB received.

Our results indicate that traffic generated in a day by smartphone users is comparable to traffic generated by PCs a few years ago. A study of a campus WiFi network revealed that on average users were generating 27 MB of traffic in 2001 and 71 MB in 2003 [12]. A study of Japanese residential broadband users in 2006 revealed that on average users generate 1000 MB of traffic per day [11]. This high level of traffic has major implications for the provisioning of wireless carrier networks as smartphone adoption increases.

Relationship to application types: To investigate if certain types of applications are favored more by users that generate more traffice, we divide the users into two equal classes based on their sum of sent and received traffic per day. Figure 20 shows mean and 95% CI of relative popularity of each application category for each user class. Expectedly,

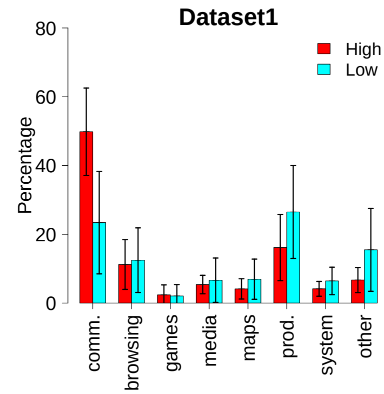


Figure 20: The mean and 95% CI of relative popularity of each application category among high and low traffic consumers.

it shows that communication applications are more popular among users that consume more traffic.

5.2 "Interactive" traffic

We next estimate what fraction of the total traffic is "interactive," that is, has timeliness constraints. Approaches to reduce power consumption of data transfers often advocate rescheduling network transfers, for instance, by delaying some transfers so that multiple of them may be bundled [1, 22]. Such policies are difficult to implement for interactive traffic without hurting user response time, and thus are likely to be less valuable if the bulk of the traffic is interactive.

We classify traffic as interactive if it was generated when the screen is on. This classification method might classify some background traffic as interactive. We expect the error to be low because the screen is on for a small fraction of the time for most users. Because certain user interactions with the phone begin right after traffic exchange (e.g., a new email is received), we also consider traffic received in a small time window (1 minute) before the screen is turned on as having timeliness constraints. The results are robust to the exact choice of time window. Indeed some of the traffic that we classify as interactive might be delay tolerant, e.g., a new email to be sent might be deferred for a little while. However, the volume of traffic received by the phone, which users would rather see immediately, dominates by one order of magnitude the volume that is sent and delay-tolerant messaging applications such as email contribute only a small chunk of all traffic.

Figure 21 shows the fraction of interactive traffic for each user. We see that for about 90% of the users, over 50% of the traffic is interactive but for the rest almost none of their traffic is interactive. Stated differently, for different smartphone users, almost all to almost none of the traffic is generated by applications in the background. The extremes represent disparate ways in which people use smartphones and which applications on the smartphone generate most traffic. Our results imply that the energy savings that can be had by rescheduling network activity will vary across users.

5.3 Diurnal patterns

Figure 22(a) shows the diurnal pattern for an example user, with a sharp decline at night. Figure 22(b) shows the strength of the diurnal pattern for individual users by

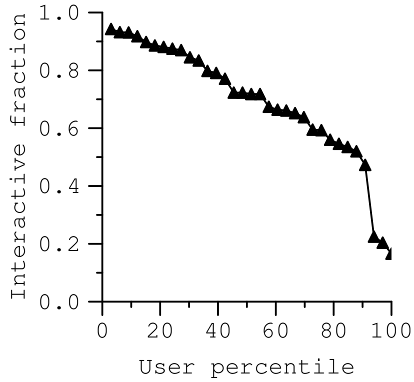


Figure 21: The fraction of interactive traffic.

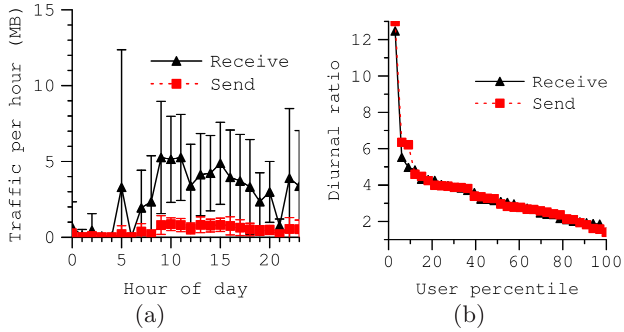


Figure 22: (a) The mean and 95% CI for traffic generated per hour by an example user. (b) The diurnal ratio of traffic per hour for all users.

plotting the diurnal ratio of traffic. As defined in §3.3, the diurnal ratio reflects how much higher the mean during the peak hour is to the overall mean.

We see that the diurnal ratio varies across users but most have a strong diurnal behavior. 80% of them generate over twice their average amount of traffic in their peak hour. This behavior is likely a direct result of the high proportion of interactive traffic for most users and that user interactions themselves have a diurnal pattern.

6. ENERGY CONSUMPTION

The final aspect of smartphone usage that we investigate is energy consumption. Energy drain depends on two factors: i) user interactions and applications; and ii) platform hardware and software. If the second factor dominates, the energy drain of various users with the identical smartphone will be similar. Otherwise, the energy drain will be as diverse as user behaviors.

We estimate the amount of energy drain based on the remaining battery indicator which varies between 0 and 100%. If the battery indicator has gone down by X% in a time period for a battery with capacity Y mAh, we compute the energy drain in that period to be $X \cdot Y$ mAh [1]. Given that batteries are complex electro-chemical devices [14, 20], this computation is approximate. It assumes that the battery level indicator is linear with respect to energy drain.

[1]mAh is technically a unit of charge, yet is commonly used to indicate energy drain because battery voltage during normal operations is typically constant. For phones in our dataset, this is 4V, so multiplying a mAh reading by 4 would yield an accurate energy reading in milli-watt-hours.

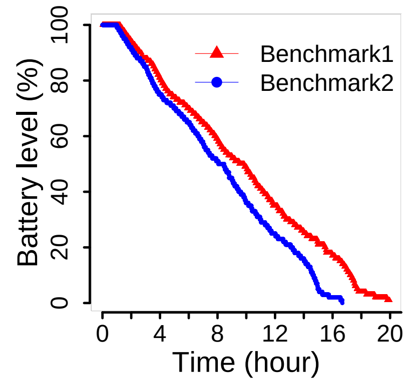


Figure 23: Timelapse of the remaining battery level indicator in controlled experiments with two different workloads at room temperature.

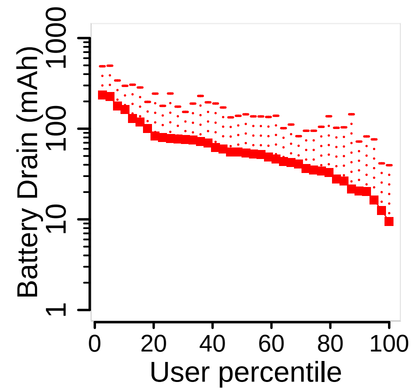
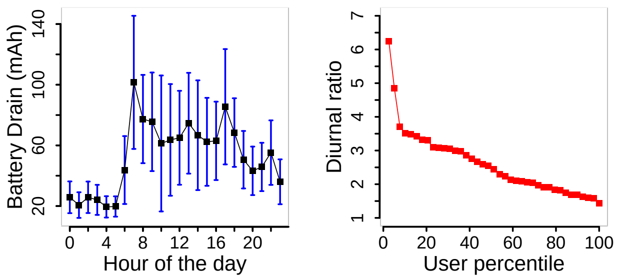


Figure 24: The mean and the upper end of the standard deviation of one hour energy drain for Dataset1 users during discharge periods.

Controlled experiments suggest that the linearity assumption holds to a first order. We run a benchmark load that drains the battery at a fixed rate in room temperature. Under this benchmark, if the battery level indicator decreases linearly with time, it must be linear with respect to energy drain. Figure 23 shows that the level indicator decreases roughly linearly for two different benchmarks. Benchmark1 turns the screen on and off periodically. Benchmark2 computes and idles periodically. We conclude thus that the level indicator can be used to estimate energy drain.

Figure 24 shows the mean and standard deviation of energy that users drain in an hour. This graph is computed using only periods in which the battery is not charging because energy drain in those periods are of primary interest. We see a two orders of magnitude difference among users. While heaviest users drain close to 250 mAh the lightest of users drain only 10 mAh. If the battery capacity is 1200 mAh, this leads to a lifetime variation from about 4 to 120 hours.

Figure 25(a) shows for an example user that the drain is not the same throughout the day but has diurnal variations in which more energy is consumed during the day than during the night. For this user, the level of energy consumed changes by roughly a factor of five. Figure 25(b) plots the diurnal ratio of energy use for all users. It shows that diurnal variations occur, with different strengths, for all users.



(a) (b)

Figure 25: (a) The mean and 95% CI of energy drain of an example Dataset1 user. (b) Diurnal ratio of all users in Dataset1.

Our results show that user activities contribute heavily towards energy drain; users in Dataset1, who have identical smartphones, drain energy at different rates, and energy drain has diurnal patterns. In the future, we will develop methods to accurately quantify the energy consumption of the platform from that due to user-induced workload.

7. IMPLICATIONS OF USER DIVERSITY

We uncover a surprising level of diversity among smartphone users. For almost every aspect of usage that we study, we find one or more orders of magnitude difference between users. Our findings strongly motivate the need for customizing smartphones to their users. We believe that this need is greater than that for customizing ordinary cellphones or laptops. Ordinary cellphones do not have as rich an application environment. Laptops are not as portable and are more resource rich. For example, many users plug-in their laptops while using them.

Customization can help at all levels. Consider something as low-level as the battery. Suppose we want batteries to be both lightweight and last for at least a day with a high probability. Meeting the latter goal for all users of a given platform will require catering to the heaviest users. But that will lead to unnecessarily heavy batteries for many users. (Higher capacity batteries are heavier.) Offering multiple types of batteries with different lifetime-weight tradeoffs provides a way out of this bind.

At levels where intelligent mechanisms to improve user experience or reduce energy consumption reside, user diversity motivates adapting to the smartphone user. Driving these mechanisms based on average case behaviors may not be effective for a large fraction of the users.

The ease and utility of customization depends on two properties of user behavior. First, despite quantitative differences, there must be qualitative similarities among users. For instance, we should be able to describe the behavior of all users with the same model. Different users may have different parameters of this model, which will then lead to quantitative differences among them. The presence of qualitative similarities imply that users are not arbitrary points in space, and it significantly simplifies the task of learning user behavior. Second, user behavior in the past must also be predictive of the future. Otherwise, customization based on past behaviors will be of little value in the future.

In the next two sections, we present evidence that these properties hold for several key aspects of smartphone usage.

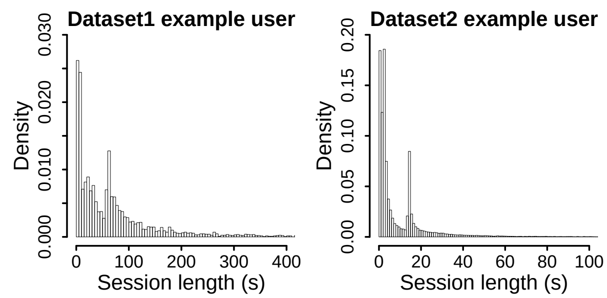


Figure 26: The PDF of session length for sample users of each dataset.

In §8, we show that user sessions and relative application popularity can be described using simple models. In §9, we show that energy drain can be described as well as predicted using a "trend trable" framework.

8. SMARTPHONE USAGE MODELS

In this section, we develop simple models that describe three aspects of smartphone usage – session lengths, interarrival time between sessions, and application popularity. The models are common across users but have different parameters for different users. While they do not completely describe user behavior, they capture first order factors and represent a first step towards more complete modeling of smartphone usage. More importantly, along with the results of the next section, they show that qualitative similarities do exist among users.

8.1 Session Lengths

We first consider the statistical properties of session length distributions of users. We find that session length values tend to stationary. With the KPSS test for level stationarity [13], 90% of the users have a p-value greater than 0.1. The presence of stationarity is appealing because it suggests that past behavior is capable of predicting the future.

We also find that session lengths are independent, that is, the current value does not have a strong correlation with the values seen in the recent past. With the Ljung-Box test for independence [15], 96% of the users have a p-value that is greater than 0.1.

Stationarity and independence, considered together, suggest that session length values can be modeled as i.i.d samples from a distribution. Choosing an appropriate distribution, however, is complicated by the nature of the session lengths. Most sessions are very short and the frequency drops exponentially as the length increases. However, inconsistent with exponential behavior, there are some very long sessions in the tail for each user.

We find that a mixture of exponential and Pareto distributions can model both ends of the spectrum. The former captures short sessions and the latter captures long sessions. That is, session lengths can be described by the following mixture model:

$$r \cdot Exp(\lambda) + (1 - r) \cdot Pareto(x_m, \alpha)$$

In this equation, r is the relative mix of the two distributions, λ is the rate of the exponential, and x_m and α are the location and shape parameters of the Pareto distribution.

The location for a Pareto distribution represents the minimum possible value of the random variable. The location

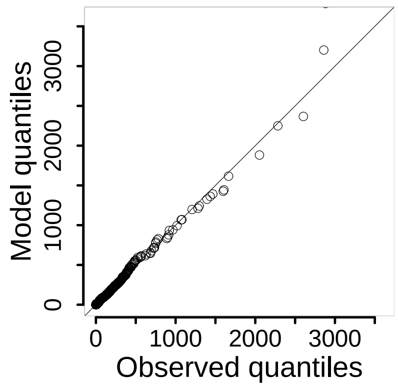
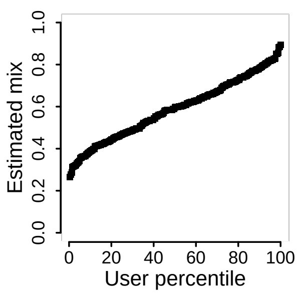
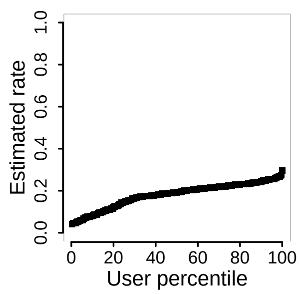


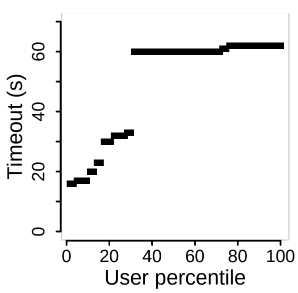
Figure 27: QQ plot of session lengths model for a sample user



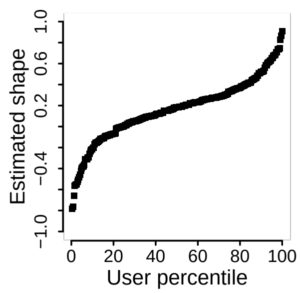
(a) Relative mix (r)



(b) Exponential rate (λ)



(c) Pareto location (x_m)



(d) Pareto shape (α)

Figure 28: Distribution of inferred model parameters that describe session length values of users in both datasets.

value that offers the best fit is the screen timeout value of the user, because the session length PDF has a spike at this value. The spike corresponds to short sessions that are ended by the timeout (when the user forgets to switch the screen off); we confirm this using controlled experiments with different timeout values. Figure 26 shows this spike, at 60 and 15 seconds, for example users from each dataset. The timeout provides a natural division between the two component distributions. We automatically infer its approximate value using a simple spike detection algorithm.

We use the EM algorithm to infer the maximum likelihood estimation (MLE) of the remaining three parameters [6]. Figure 27 shows the quality of this fit for an example user using the QQ plot [3]. Almost all quantiles are along the $y = x$ line, indicating a good fit.

Figure 28 shows the four inferred parameters for various users. While users can be modeled using the same mixture model, the parameters of this model vary widely across users. Because of the way we construct our model, the distribution of the parameter r and x_m also provide insight into how frequently users' screen is switched off by the timeout and the relative popularity of different timeout values. 60 seconds is the most popular value, likely because it is the most common default timeout and many users never change the default.

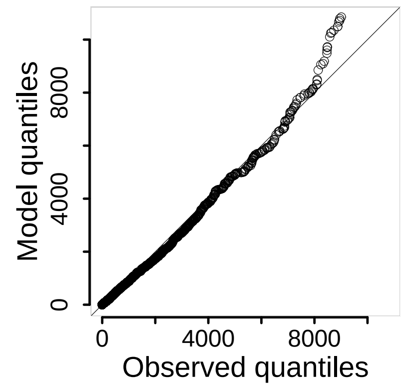
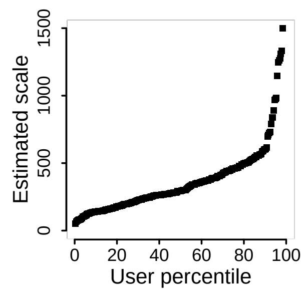
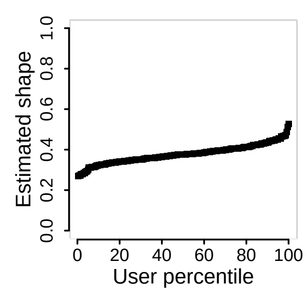


Figure 29: QQ plot of session offtime model for a sample user



(a) Weibull scale



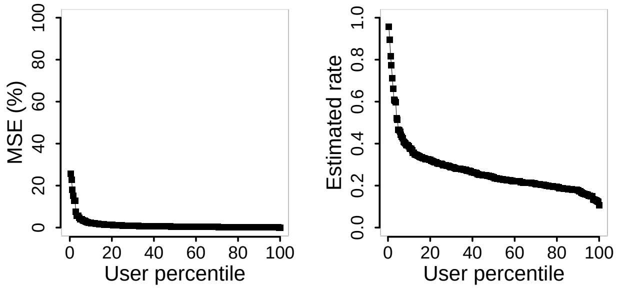
(b) Weibull shape

Figure 30: Distribution of inferred model parameters that describe the distribution of time between sessions for users in both datasets.

8.2 Time between Sessions

We find that the Weibull distribution can explain the screen off times. This distribution has two parameters referred to as its scale and shape. We find the MLE for these parameters for each user. From the QQ-plot in Figure 29, we notice that the model predicts a greater probability of seeing some very large offtimes than are observed in the datasets. However, the probability of seeing these large offtimes is small; there are 2.7% data points that have a y-value greater than 8000 in that graph. Hence, we believe that Weibull provides a good fit for the length of intervals between interactions.

Figure 30 shows the distribution of the estimated shape and scale of the fitted Weibull distributions. Interestingly, the shape is consistently less than one. Weibull shape values less than one indicate that the longer the screen has been off, the less likely it is for it to be turned on by the user. This behavior has interesting implications for power saving policies. For instance, periodic activities such as checking for email when the screen has been off for a while may be deferred or rescheduled if needed without hurting user experience.



(a) MSE (b) Rate parameter

Figure 31: (a) The mean square error (MSE) when application popularity distribution is modeled using an exponential. (b) The inferred rate parameter of the exponential distribution for different users.

8.3 Application Popularity

We find that for each user the relative popularity of applications can be well described by a simple exponential distribution. This qualitative invariant is useful, for instance, to predict how many applications account for a given fraction of user attention. For the example users in Figure 11, this facet can be seen in the inset plots; the semi-log of the popularity distribution is very close to a straight line.

Figure 31(a) shows that this exponential drop in application popularity is true for almost all users; the mean square error between modeled exponential and actual popularity distribution is less than 5% for 95% of the users.

Figure 31(b) shows the inferred rate parameter of the application popularity distribution for various users. We see that the rate varies by an order of magnitude, from 0.1 to almost 1. The value of the rate essentially captures the pace of the drop in application popularity. Lower values describe users that use more applications on a regular basis. One implication of the wide range is that it may be feasible to retain all popular applications in memory for some users and not for others.

9. PREDICTING ENERGY DRAIN

In this section, we demonstrate the value of adapting to user behaviors in the context of a mechanism to predict future energy drain on the smartphone. Such a mechanism can help with scheduling background tasks [9], estimating battery lifetime, and improving user experience if it is estimated that the user will not fully drain the battery until the next charging cycle [2, 21].

Despite its many uses, to our knowledge, none of the smartphones today provide a prediction of future energy drain. Providing an accurate prediction of energy drain is difficult.[2] User diversity of energy use makes any static prediction method highly inaccurate. Even for the same user there is a high variance in energy usage. Figure 32 shows this variance by plotting the ratio of the standard deviation to the mean energy drain over several time periods for users in Dataset 1. The standard deviation of 10-minute windows is higher than three times the mean for a fifth of the users. The

[2]Energy drain prediction on smartphones is more challenging than what is done for laptops. Laptops provide a prediction of battery lifetime only for active use. Smartphone predictions on the other hand must cover multiple active, idle periods to be useful.

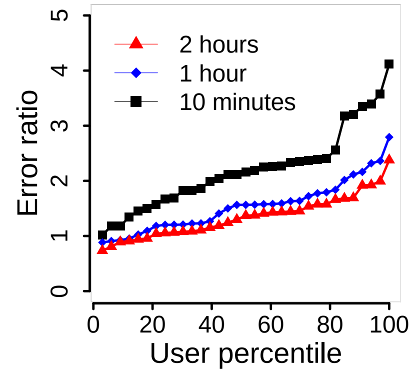


Figure 32: Battery drain predictors have to cope with the high variance in drain

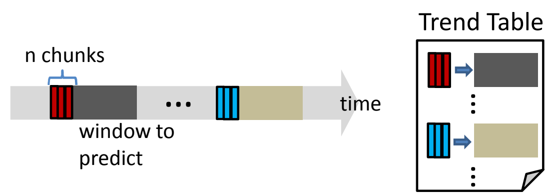


Figure 33: A user's trend table captures how battery drain in n consecutive chunks relates to drain in the subsequent window.

bursty nature of user interactions is one cause of this high variance. The variance is high even at longer time scales. The standard deviation for two hour windows is larger than 150% of the mean for half the users. The diurnal patterns that we showed earlier are one cause of high variance over such large time scales.

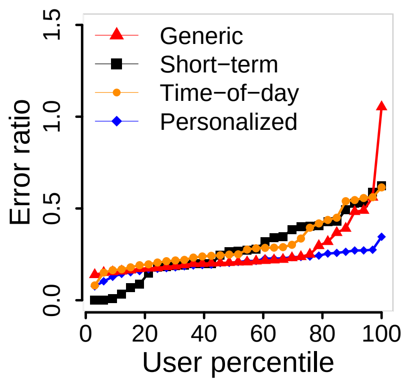
Although the variance is high, there are patterns in a user's behaviors. For instance, we showed earlier that a phone that has been idle for a while is likely to remain idle in the near future, and it will thus continue to draw energy at a similar rate.

We hypothesize that a predictor tuned to user behavior can be accurate. We present a simple personalized energy drain predictor that incorporates various user-specific factors such as length of idle periods, and different types of busy periods. Instead of explicitly identifying these factors, our predictor captures them in a "trend table" framework.

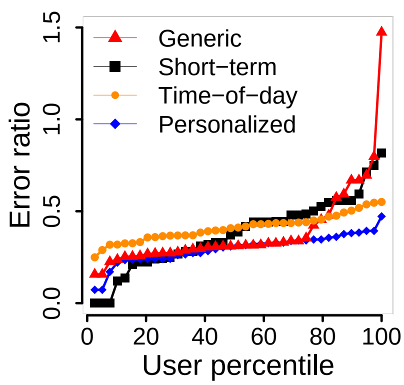
Our framework is shown in Figure 33. Each entry in the trend table is indexed by an n-tuple that represents energy usage readings from adjacent chunks of size δ each. This index points to energy usage statistics across all time windows of size w in the history that follow the n chunk values represented by that index. Thus, the trend table captures how the energy usage in n consecutive chunks is related to the energy use in the subsequent time windows of size w.

In this paper, we use $n = 3, \delta = 10$ minutes and maintain different trend tables for each window size w for which the prediction is required. To keep the trend table small, we quantize the energy readings of chunks. We maintain the mean and standard deviation for all subsequent windows that map to that index.

To make a prediction, we use the energy readings from the n immediately preceding chunks. Let these readings be $(x_1 \ldots x_n)$. Then, we search for indices in the table that differ from $(x_1 \ldots x_n)$ by no more than a threshold s and



(a) 1-hour window



(b) 2-hour window

Figure 34: The ratio of error to true battery drain for different energy drain predictors.

predict based on them, such that similar indices have higher relative weight. More precisely, the prediction is

$$\sum_{e\ s.t. \forall i=1\ldots n \left|h_i^e - x_i\right| < s} \frac{k^e \times \hat{u}^e}{\sum k^e},$$

where e iterates over all tuples in the trend table, h_i^e is the value of the chunk i of the tuple e, $\hat{u}^e$ represents the statistics stored for that index, and the relative weight $k^e = s - \max_{i=1\ldots n} |h_i^e - x_i|$. For results shown here, we use $s = 0.5\%$.

Figures 34 shows how well our *Personalized* predictor works for predicting usage for two different future time windows. There is one point per user in each graph, which corresponds to the median error seen across many predictions for that user. We see that the 90th percentile error is 25% for 1-hour window and 40% for 2-hour window.

To place this level of accuracy in context, the figure also shows the accurary of three alternative predictors. The *Generic* predictor uses a trend table but instead of building a user-specific table, it builds one that combines all users. We see that its 90th percentile error is roughly twice that of the Personalized predictor. Its predictions are worse for half of the users.

The *Short-term* predictor predicts energy drain simply as the amount drained in the immediately preceding window of the same size. For both time windows, its 90th percentile error is no better than that of the Generic predictor.

The *Time-of-day* predictor predicts energy drain based on values observed for the user at the same time of the day in the past, similar to that proposed by Banerjee *et al.* [2]. Thus, this predictor has some degree of personalization, but it learns user behavior in less detail than the Personalized predictor. We see that its performance for 1-hour window is similar to the *Short-term* predictor. For 2-hour window, its error is higher than the Personalized predictor across the board, though its worst case error is lower than the Short-term and Generic predictors.

Overall, these results demonstrate the value of appropriately learning user behaviors to design intelligent mechanisms on smartphones.

10. RELATED WORK

In a range of domains, there is a rich history of work that characterizes user workloads. However, because smartphone adoption has gathered pace relatively recently, our work represents one of the few to study how people use smartphones.

Along with other recent works, it helps complete the picture. Banerjee *et al.* and Rahmati *et al.* report on battery charging behaviors [2, 18]. Like us, they find considerable variation among users. Banerjee *et al.* also propose a predictor that estimates the excess energy of the battery at the time of charging, using a histogram of past battery usage [2]. Shye *et al.* study the power consumption characteristics of 20 users [24]. They infer properties such as which components in the phone consume most power and explore optimizations based on these characteristics. Rahmati and Zhong study 14 users of a particular demographic to study which applications are popular in that demographic, where the phone is used, and how it is shared among users [19].

In contrast to these works, we focus on understanding different aspects of smartphone use (e.g., interactions and traffic) and on exposing the diversity of user behaviors, instead of only the average behavior. Our study also entails an order of magnitude more users than previous efforts.

Like us, other researchers have developed logging utilities. MyExperience is one such early utility [10], and the works above involve custom utilities as well. Our focus in this paper is not on the development of logging tools but on analyzing their data to gain insight into user behavior.

There is a body of work in modeling the aggregate behavior of mobile users. Using traces from a large cellular operator some network related aspects of mobile usage have been modeled. Halepovic *et al.* and Williamson *et al.* report that call arrivals are bursty and present diurnal patterns [28]. Willkomm *et al.* and Brown *et al.* report that mobile users call duration can be approximately modeled by a lognormal distribution [4, 29]. We use traces collected on the mobile device itself and focus on modeling the interactions of individual users instead of the aggregate behavior.

11. CONCLUSIONS

By studying 255 users of two different smartphone platforms, we comprehensively characterized user activities and their impact on network and battery. We quantify many hitherto unknown aspects of smartphone usage. User diversity is an overarching theme in our findings. For instance, different users interact with their phones 10-200 times a day on average; the mean interaction length of different users is 10-250 seconds; and users receive 1-1000 MB of data per day, where 10-90% is received as part of active use.

This extent of user diversity implies that mechanisms that work for the average case may be ineffective for a large fraction of the users. Instead, learning and adapting to user behaviors is likely to be more effective, as demonstrated by our personalized energy drain predictor. We show that qualitative similarities exist among users to facilitate the development of such mechanisms. For instance, the longer the user has not interacted with the phone, the less likely it is for her to start interacting with it; and application popularity for a user follows a simple exponential distribution.

Our study points to ways in which smartphone platforms should be enhanced. Effective adaptation will require future platforms to support light-weight tools that monitor and learn user behaviors *in situ*. It will also require them to expose appropriate knobs to control the behavior of lower-level components (e.g., CPU or radio).

12. ACKNOWLEDGEMENTS

We are grateful to the users in Dataset1 that ran our logger and to the providers of Dataset2. Elizabeth Dawson provided administrative support for data collection. Jean Bolot, Kurt Partridge, Stefan Saroiu, Lin Zhong, and the MobiSys reviewers provided many comments that helped improve the paper. We thank them all. This work is supported in part by the NSF grants CCR-0120778 and CNS-0627084.

13. REFERENCES

[1] Balasubramanian, N., Balasubramanian, A., and Venkataramani, A. Energy consumption in mobile phones: A measurement study and implications for network applications. In *IMC* (2009).

[2] Banerjee, N., Rahmati, A., Corner, M. D., Rollins, S., and Zhong, L. Users and batteries: Interactions and adaptive energy management in mobile systems. In *Ubicomp* (2007).

[3] Becker, R., Chambers, J., and Wilks, A. *The new S language*. Chapman & Hall/CRC, 1988.

[4] Brown, L., Gans, N., Mandelbaum, A., Sakov, A., Shen, H., Zeltyn, S., and Zhao, L. Statistical analysis of a telephone call center. *Journal of the American Statistical Association 100*, 469 (2005).

[5] Church, K., and Smyth, B. Understanding mobile information needs. In *MobileHCI* (2008).

[6] Dempster, A., Laird, N., and Rubin, D. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological) 39*, 1 (1977).

[7] Esfahbod, B. Preload: An adaptive prefetching daemon. Master's thesis, University of Toronto, 2006.

[8] Falaki, H., Govindan, R., and Estrin, D. Smart screen management on mobile phones. Tech. Rep. 74, Center for Embedded Networked Sesning, 2009.

[9] Flinn, J., and Satyanarayanan, M. Managing battery lifetime with energy-aware adaptation. *ACM TOCS 22*, 2 (2004).

[10] Froehlich, J., Chen, M., Consolvo, S., Harrison, B., and Landay, J. MyExperience: A system for in situ tracing and capturing of user feedback on mobile phones. In *MobiSys* (2007).

[11] Fukuda, K., Cho, K., and Esaki, H. The impact of residential broadband traffic on Japanese ISP backbones. *SIGCOMM CCR 35*, 1 (2005).

[12] Kotz, D., and Essien, K. Analysis of a campus-wide wireless network. *Wireless Networks 11*, 1–2 (2005).

[13] Kwiatkowski, D., Phillips, P., Schmidt, P., and Shin, Y. Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics 54*, 1-3 (1992).

[14] Lahiri, K., Dey, S., Panigrahi, D., and Raghunathan, A. Battery-driven system design: A new frontier in low power design. In *Asia South Pacific design automation/VLSI Design* (2002).

[15] Ljung, G., and Box, G. On a measure of lack of fit in time series models. *Biometrika 65*, 2 (1978).

[16] Smartphone futures 2009-2014. http://www.portioresearch.com/Smartphone09-14.html.

[17] Rahmati, A., Qian, A., and Zhong, L. Understanding human-battery interaction on mobile phones. In *MobileHCI* (2007).

[18] Rahmati, A., and Zhong, L. Human–battery interaction on mobile phones. *Pervasive and Mobile Computing 5*, 5 (2009).

[19] Rahmati, A., and Zhong, L. A longitudinal study of non-voice mobile phone usage by teens from an underserved urban community. Tech. Rep. 0515-09, Rice University, 2009.

[20] Rao, R., Vrudhula, S., and Rakhmatov, D. Battery modeling for energy aware system design. *IEEE Computer 36*, 12 (2003).

[21] Ravi, N., Scott, J., Han, L., and Iftode, L. Context-aware battery management for mobile phones. In *PerCom* (2008).

[22] Sharma, A., Navda, V., Ramjee, R., Padmanabhan, V., and Belding, E. Cool-Tether: Energy efficient on-the-fly WiFi hot-spots using mobile phones. In *CoNEXT* (2009).

[23] Sharma, C. US wireless data market update - Q3 2009. http://www.chetansharma.com/usmarketupdateq309.htm.

[24] Shye, A., Sholbrock, B., and G, M. Into the wild: Studying real user activity patterns to guide power optimization for mobile architectures. In *Micro* (2009).

[25] Snol, L. More smartphones than desktop PCs by 2011. http://www.pcworld.com/article/171380/more_smartphones_than_desktop%_pcs_by_2011.html.

[26] Sohn, T., Li, K., Griswold, W., and Hollan, J. A diary study of mobile information needs. In *SIGCHI* (2008).

[27] Trestian, I., Ranjan, S., Kuzmanovic, A., and Nucci, A. Measuring serendipity: Connecting people, locations and interests in a mobile 3G network. In *IMC* (2009).

[28] Williamson, C., Halepovic, E., Sun, H., and Wu, Y. Characterization of CDMA2000 Cellular Data Network Traffic. In *Local Computer Networks* (2005).

[29] Willkomm, D., Machiraju, S., Bolot, J., and Wolisz, A. Primary users in cellular networks: A large-scale measurement study. In *DySPAN* (2008).