The Challenges in Large-Scale Smartphone User Studies*

Earl Oliver
David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
eaoliver@uwaterloo.ca

ABSTRACT

We present preliminary results of a large-scale smartphone user study that examines how users interact with and consume energy on their personal mobile devices. Our dataset consists of over one millennium of user interaction traces from over 17300 BlackBerry users. Despite the scale and detail of the dataset, there are many research questions that it cannot answer; further user studies are therefore needed. We detail our insight into the major challenges in conducting a large-scale user study on BlackBerry devices.

1. INTRODUCTION

The symbiotic relationship that exists between humans and their person mobile devices is an interesting phenomenon. As the user of mobile devices continue to grow, the study of this relationship can provide insight the daily lives of mobile users and benefit many research areas. Modelling users' interaction time can, for example, impact software interaction design and non-technical domains such as addictology [5], health [1], polysomnography [13], psychology [12], and business [6]. Identifying energy consumption patterns among smartphone users can also affect the design of smartphone energy systems or the implementation of energy intensive mobile application [9].

We present preliminary results of a large-scale smartphone user study that examines how users interact with and consume energy on their personal mobile devices. As of May 1, 2010, our study has collected over one millennium of traces from over 17300 BlackBerry users around the world. Among many interesting results, we were surprised to find that users interact with their device on average 1.68 hours per day, that approximately 80% of device interactions are less than 90 seconds in length, and that 50% of interactions occur within 115 seconds of each other. We also observed wide variation among between user's daily interaction.

Despite the scale and breadth of our dataset, there are many questions that it cannot answer. For example, our dataset does not provide context to user interaction, nor does does it reveal how memory, CPU, or storage resources are consumed by the user. Further measurement studies are required to answer these questions. We therefore detail the major challenges in conducting a large-scale smartphone user study. We believe that our insight into these challenges [9] will benefit other researchers conducting similar large-scale experiments. Although our experiences are with BlackBerry smartphones, most of the following challenges are isomorphic to the challenges faced on other smartphone platforms [7].

We begin with an overview of our smartphone user study and a presentation of preliminary results of an analysis of smartphone user interaction. Section 3 then describes the major challenges in conducting similar large-scale smartphone user studies. We conclude in Section 4.

2. SMARTPHONE USER STUDY

The goal of our user study is to examine how users interact with their personal mobile devices. To achieve this goal we have developed a BlackBerry application that executes both as an independent application or as a library that can be embedded into third-party applications. Figure 1 illustrates a screen capture from the logger. On BlackBerry devices, user interaction can be inferred by exploiting two criteria. First, the device's LCD backlight being enabled is a necessary condition of user interaction. Second, the BlackBerry OS also maintains an idle time, which is reset every time the user presses a button on the device. An *interaction ses-*

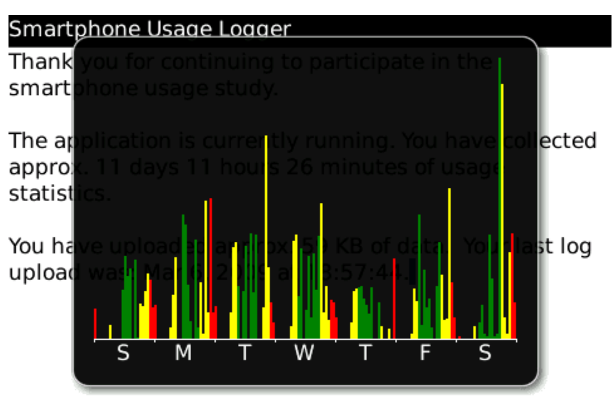


Figure 1: Screen capture of the BlackBerry logging application.

*Invited talk

ACM HotPlanet June 15, 2010, San Francisco, California, USA.

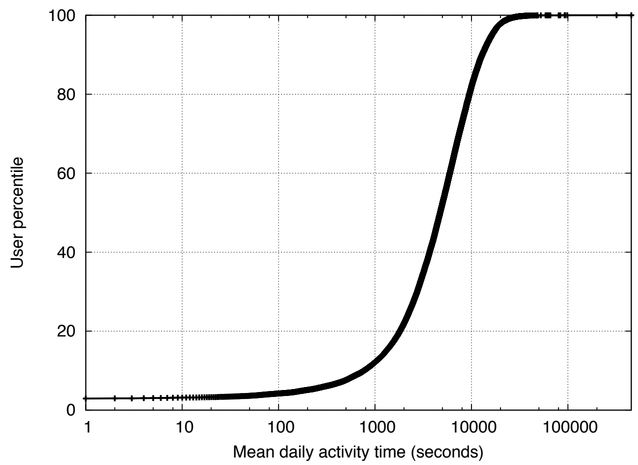


Figure 2: Distribution of the mean daily activity time.

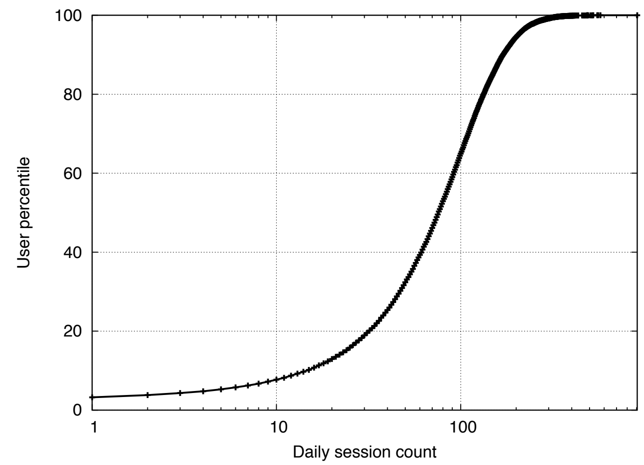


Figure 3: Distribution of the mean number of daily sessions.

sion is therefore the duration that the LCD backlight was enabled less the time that the user was not interacting with the device (and resetting the idle time). For a detailed discussion of the measurement driver implementation, we refer the interested reader to [9].

As of May 1, 2010 our user study has produced a dataset that consists of over one millennium of cumulative user interaction and energy consumption behaviour from a total of 17300 BlackBerry users. The cumulative total is the sum duration of all users' log contributions. The average user contributed approximatley 17 days of user interaction and energy consumption data. These users span 23 timezones and use all BlackBerry device types released since early 2006 [11]. Our preliminary user interaction analysis considers three basic criteria: total daily interaction with the device, the characteristics of daily interaction sessions, and diurnal usage patterns. The remainder of this section is organized to complement recent work by Falaki et. al that examines smartphone usage of both Android and Windows Mobile smartphone users [4].

2.0.1 Interaction Time

The daily *interaction time* of a user is the sum of the duration of daily interaction sessions. The cumulative amount of time that a user interacts with the device can serve as a coarse grained classifier. Users that use their device very little may appear to be uninteresting; however, these users' devices could be prime candidates for running resource intensive mobile applications. For example, memory and CPU bound operations, such as opportunistic communication [8], would be transparent to these users. Conversely, heavy users are more likely to be aware of their device's state and less tolerant to resource demands by third party applications.

The mean daily interaction time for all users in our study is illustrated in Figure 2. Surprisingly, BlackBerry users spend an average of 1.68 hours on their devices each day. This value is dominated by a small proporotion of users that interact with their device for several hours each day. We observed a median total interaction time of 1.31 hours.

2.0.2 Interaction Sessions

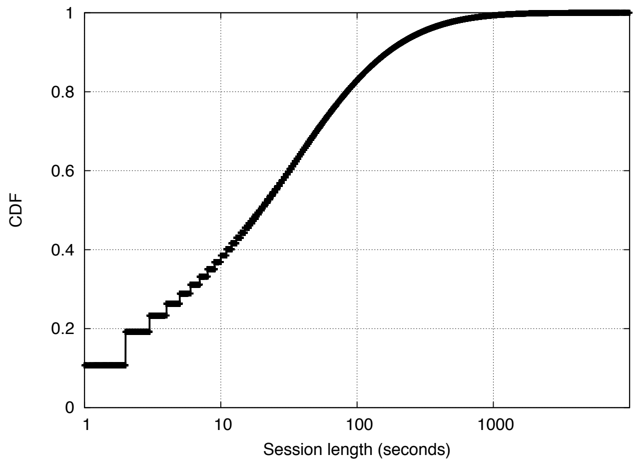
Interaction sessions provide a fine-grained view of how users interact with their device. We observed that BlackBerry users interact with their device an average of 86.9 times per day. As before, this quantity is dominated by the upper 10th percentile that interact with their device more than 200 times per day. The median number of daily interactions was 76. These results are illustrated in Figure 3.

The length of user interaction sessions varies widely over the sample population. We illustrated the CDF of session lengths in Figure 4(a). The median session length is 20.0 seconds. The mean session length is 68.4 seconds. The significant gap between the median and mean is due relatively few sessions that last as long as 2.53 hours. The time between sessions is also highly variable. Figure 4(b) illustrates the CDF of time between sessions for the sample population. We observe a mean duration of 682.5 seconds and a median duration of 114.8 seconds. This surprising rapid rate could be indicative of technology addiction and/or obsessive compulsive behaviour [5] in a subset of the population. We plan to explore this interesting statistic in future work.

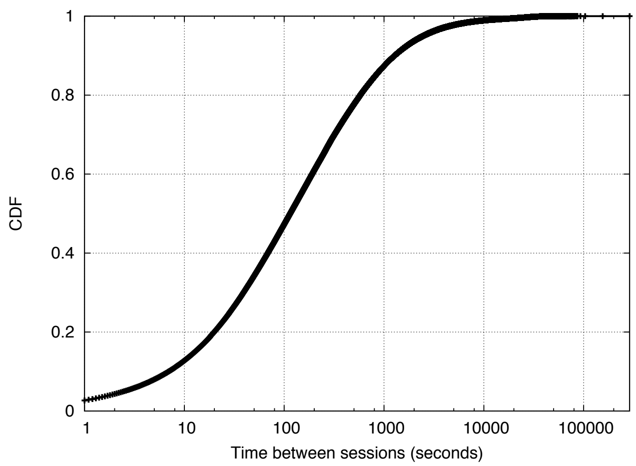
2.0.3 Diurnal Patterns

Studying user interaction throughout the day can provide insight into the lifestyles of smartphone users. Interaction early in the morning and late at night could, for example, indicate an unhealthy level of sleep or an excessively long work day [6]. Our analysis confirms the observation by Falaki that strong diurnal patterns exist [4]. Figure 5 illustrates the daily distribution of interaction for two sample users. Each sample user contributed approximately 3 months in traces. The distribution of sample user 1 follows a typical diurnal pattern. Interaction is highest during the business day and drops to zero during the night while the user is sleeping. The second sample user has an abnormally high inter-day variation. The absence of nightly gaps is also indicative of irregular sleeping patterns.

Figure 6(a) and Figure 6(b) illustrate the distribution of user interaction throughout the day and the week respectively for the sample population. Not surprisingly, we observe a strong correlation between user interaction and the

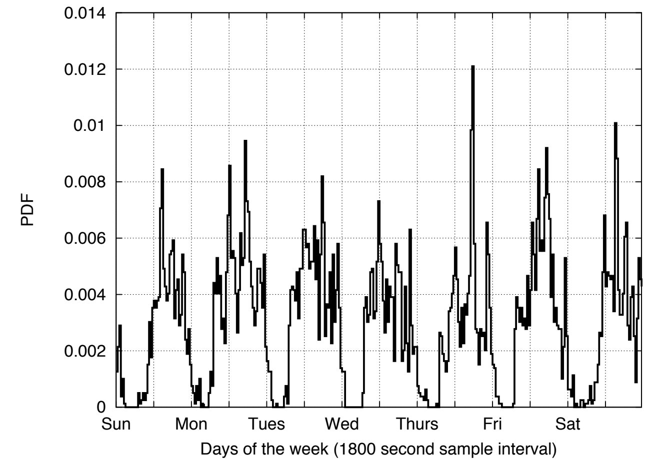


(a) Session length

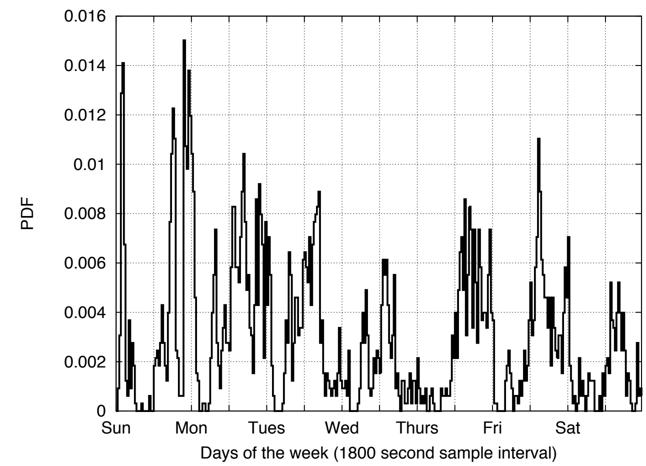


(b) Time between sessions

Figure 4: CDF of session length and the time between sessions for all users.



(a) Sample user 1



(b) Sample user 2

Figure 5: Weekly distribution of interaction sessions for two sample users.

hour of the day. Users are also slightly less likely to interact with their device on the weekend than during a weekday.

3. CHALLENGES

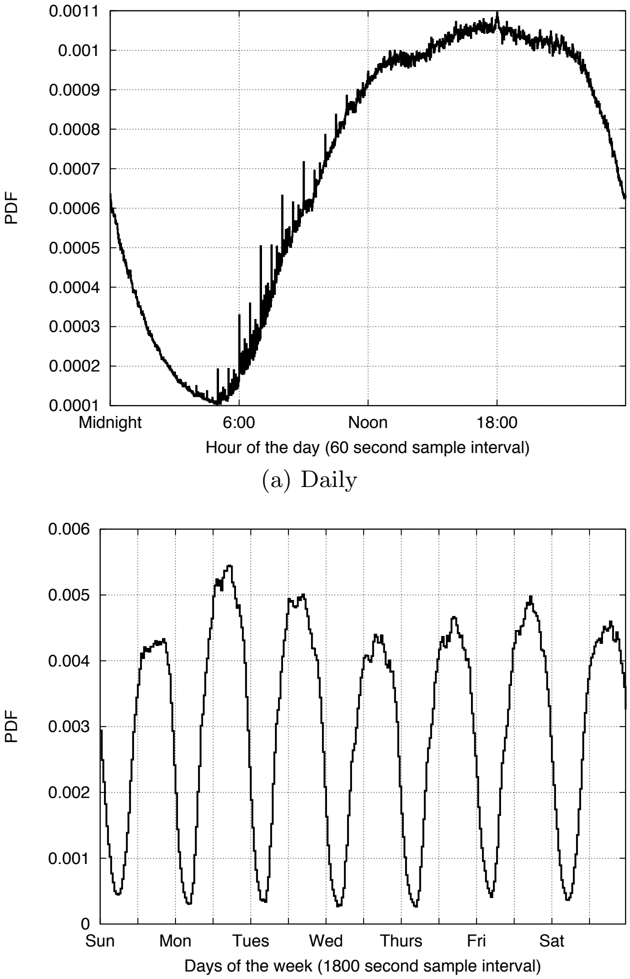
While we are optimistic that our dataset can answer many interesting questions pertaining to smartphone user interaction, there are many research questions that cannot be answered using our data and further studies are needed. In this section we examines the major challenges in conducting a large-scale study on personal mobile devices. Although our experiences [9] are with BlackBerry smartphones, most of the following challenges are isomorphic to the challenges faced on other smartphone platforms [7].

3.1 Volatile file systems

Logging data to a file is a standard technique to ensure persistent storage in the presence of power failures or other interruptions. However, file systems can be unmounted at any time either manually by the user or automatically when connected to a PC. For example, on BlackBerry, of the three file system roots `/system`, `/store`, and `/SDcard`, the SD card root has the greatest capacity and no restrictions on file size, but may be removed at any time. The `/store` path has various capacities and restrictions on file size (depending on the device) and can also be unmounted. Finally, `/system` is a fixed path on some devices, and non-existent on others. An autonomous logging application must be heterogeneous and dynamic file-system configurations. Adapting to this problem by frequently flushing I/O buffers to persistent storage can have a significant impact on battery life. We therefore exploits two trends among BlackBerry users to provide reliable logging while saving energy: users maintain a high battery level and rarely fully power off their devices [9]. We found that buffering log data in volatile memory until uploading it to a collection server each night caused only 2.31% of data to be lost.

3.2 Energy constraints

Logging applications that have a noticeable impact on energy consumption will not be successful. In preliminary experimentation, we found that users were intimately aware of their device's normal battery life; any noticeable or per-



(a) Daily

(b) Weekly

Figure 6: Daily and weekly interaction session distribution.

ceived decrease in battery life would be attributed to the experimental application. As observed in previous studies [10], polling a device's state can have a detrimental impact on battery life. We therefore exploit callbacks from the operating system whenever possible, and found that this approach has a negligible impact on battery life. Unfortunately, a pure event-driven model can produce inconsistent sequences of events. For example, when powering off a BlackBerry, we frequently did not receive an event signalling that the screen was off until after powering the device on. Similar scenarios were observed when sampling charging behaviour. An event-drive logger and subsequent data analysis should therefore be able to adapt to or detect race-conditions from the underlying OS.

3.3 Third-party application intervention

User-centric studies must take into account intervention by third party applications. For example, an application that records a user's battery consumption can be affected by a spyware application that was accidentally installed on the device [3]. For example, on BlackBerry it is possible to programmatically control the LCD backlight and reset the device's idle time, which would adversely impact the measurements presented in this paper. One solution to this problem is to collect a manifest of all active applications on a device. In preliminary experimentation we have found that most BlackBerry users run few applications, and less than 2% of applications run for long durations in background. Devices that contain long-running background applications should be flagged for review and possibly discarded.

3.4 Non-linear time

Time synchronization, a task often taken for granted, is the most problematic challenge to overcome in autonomous logging. A device's clock can be updated through three means. The clock can be changed by plugging a device into a personal computer and synchronizing with a 'device manager' [2]. Users can manually change the time or time zone at any time. A device may also synchronize its clock with the timestamp broadcasted by the cellular network. Changes to the device clock manifest as the appearance of non-linear time. Unfortunately, there is no way to programmatically detect changes to the device clock. We mitigate errors introduced by time changes by detecting when the device is connected to a computer and analyze these events separately from non-connected events. In a large-scale user study, users can also traverse many different time zones. UTC time should be used to provide the absolute time of an event. Experiments must also record the device's time zone and when a time zone change takes place.

3.5 Malicious users

Malicious users are a unfortunate side effect of large-scale studies. If left unchecked, malicious activities can significantly bias the accuracy of an experiment. Over the course of our study, approximately 3.8 years of data from the Standard Logger was discarded due to a variety of malicious activities: file manipulation, programmatic manipulation, and simulated user attacks. File manipulation attacks are the easiest to detect. These attacks involve modifying the contents of files, renaming/copying files, or deleting files, and can be countered by a secret hash of the contents of each file and through a sequential file naming scheme.

Simulated user attacks are the hardest to detect. These attacks involve downloading experimental software into a BlackBerry simulator and then replaying an existing usage log or running a programmatic manipulation attack. These attacks were often detected by recording the IP address of each set of uploaded data. Multiple uploads from a single IP address where then flagged for review.

The types of attacks are highly dependent on the experiment parameters. The best approach to detect malicious users and mitigate negative results is to analyze data early and analyze data often - looking for abnormal patterns in every participant's data.

4. CONCLUSIONS

Through a large-scale user study we have built a dataset that contains over one millennium of the user interaction traces from over 17300 BlackBerry users. A preliminary analysis of this data reveals that BlackBerry users interact with their device on average 1.68 hours per day across 86.9 interactin sessions, that approximately 80% of device interactions are less than 90 seconds in length, and that 50% of interactions occur within 90 seconds of each other.

Although our dataset can be used to answer many interesting questions about smartphone users, there are many questions that it cannot answer. For example, we cannot examine how interaction sessions are sub-divided among multiple applications, nor does can we explore the relationship between interaction and network communication. The set of open research questions is unlimited and additional user studies are inevitable. We therefore share our insight into the major challenges in conducting a large-scale study. We believe that other researchers will benefit from our experiences.

5. REFERENCES

[1] A. Agarwal, F. Deepinder, R.K. Sharma, G. Ranga, and J. Li. Effect of cell phone usage on semen analysis in men attending infertility clinic: an observational study. *Fertility and sterility*, 89(1):124–128, 2008.

[2] BlackBerry Desktop Manager. Last visited: 29/09/2009. *http://na.blackberry.com/eng/services/desktop/*.

[3] BlackBerry Spyware Wasn't Ready for Prime Time. Last visited: 29/09/2009. *http://www.wired.com/threatlevel/2009/07/blackberry-spyware/*.

[4] Hossein Falaki, Ratul Mahajan, Srikanth Kandula, Dimitrios Lymberopoulos, Ramesh Govindan, and Deborah Estrin. Diversity in smartphone usage. In *MobiSys '10: Proceedings of the 8th international conference on Mobile systems, applications and services*, New York, NY, USA, 2010. ACM.

[5] N. Kakabadse, G. Porter, and D. Vance. Addicted to technology. *Business Strategy Review*, 18(4):81, 2007.

[6] N.K. Kakabadse, G. Porter, and D. Vance. The unbalanced high-tech life: are employers liable? *Strategic Change*, 18(1-2):1–13, 2009.

[7] Earl Oliver. A survey of platforms for mobile networks research. *SIGMOBILE Mobile Computing and Communications Review*, 12(4):56–63, 2008.

[8] Earl Oliver and Hossein Falaki. Performance evaluation and analysis of delay tolerant networking. In *MobiEval '07: Proceedings of the 1st international workshop on System evaluation for mobile platforms*, pages 1–6, New York, NY, USA, 2007. ACM.

[9] Earl Oliver and Srinivasan Keshav. Data driven smartphone energy level prediction. Technical Report CS-2010-06, University of Waterloo, April 2010.

[10] Ahmad Rahmati, Angela Qian, and Lin Zhong. Understanding human-battery interaction on mobile phones. In *MobileHCI '07: Proceedings of the 9th international conference on Human computer interaction with mobile devices and services*, pages 265–272, New York, NY, USA, 2007. ACM.

[11] Research in Motion. Last visited: 29/09/2009. *http://www.rim.com*.

[12] C. Rosen. Our cell phones, ourselves. *The New Atlantis*, 6:26–45, 2004.

[13] A. Sleep. Technical Note Automatic Sleep/Wake Identification From Wrist Activity. *Sleep*, 15(5):461–469, 1992.