Device Analyzer: Understanding smartphone usage

Daniel T. Wagner, Andrew Rice, and Alastair R. Beresford

Computer Laboratory, University of Cambridge, UK
{dtw30,acr31,arb33}@cam.ac.uk

Abstract. We describe Device Analyzer, a robust data collection tool which is able to reliably collect information on Android smartphone usage from an open community of contributors. We collected the largest, most detailed dataset of Android phone use publicly available to date. In this paper we systematically evaluate smartphones as a platform for mobile ubiquitous computing by quantifying access to critical resources in the wild. Our analysis of the dataset demonstrates considerable diversity in behaviour between users but also over time. We further demonstrate the value of handset-centric data collection by presenting case-study analyses of human mobility, interaction patterns, and energy management and identify notable differences between our results and those found by other studies.

1 Introduction

Smartphones are highly capable computing platforms containing a wide range of sensors and communications interfaces. They have been widely used in mobile and ubiquitous computing, including location systems [27], measurement [18] and context-sensing [28]. Additional efforts have focussed on understanding and optimising the platform itself, considering the measurement of energy use [23], computation offload [7], or resource sharing between devices [25]. However, understanding the importance or effectiveness of these contributions is difficult due to the lack of generally available, detailed data about how smartphones are actually used.

A large dataset of smartphone use can help direct research efforts, confirm whether local observations hold in a large and diverse population, uncover human behaviour, and show the prevalence of particular software or hardware in the wild.

Collecting usage information on smartphones is difficult: The collection mechanism itself needs to be built and deployed to a diverse group of participants running a multitude of devices in the wild; data must be collected for extended periods of time to overcome novelty effects and find long-term trends. Consequently, researchers are often forced to rely on their intuition or on the results from limited small-scale studies.

We have built and deployed Device Analyzer, which has collected 1,900 years of phone-usage information from 1277 different types of devices used by over 16,000 contributors in 175 countries over the course of two years. Table 1 describes the dataset in more detail. We observe extreme variation in usage not only between individuals but also for particular individuals over time. In this paper:

– We describe and make publicly available the largest dataset of smartphone use in terms of users, duration of study and granularity of collected data available to date.

– We highlight considerable diversity in behaviour not only between users but also over time, which was only possible due to the extended duration of our study. This emphasises the importance of participant selection and extended experiment durations. It also means that summative statistics (such as averages) should be considered and reported carefully to ensure that the true variation in behaviour has been captured.
– We evaluate smartphones as a platform for mobile ubiquitous computing in the wild, presenting for the first time expected device uptime and access to critical resources like power, network connectivity and location context from a large, diverse user base.
– We demonstrate the value of handset-centric data collection by presenting case-study analyses of human mobility, interaction and communication patterns, and energy management. We compare and contrast our results against previous studies.
– We show how limiting data collection to the phone's interaction with the cellular network (e.g. when using Call Data Records from network operators) can produce a notable effect on results like extracted movement patterns.
– We present our mechanism to provide researchers with access our data archive, and to run their own studies using Device Analyzer.

2 Related Work

Some studies reporting on mobile phone usage gather insights from mobile phone providers' Call Data Records (CDRs) that are generated when a phone interacts with the mobile network. By their nature, CDR-based studies have the potential to gather data from vast amounts of users, often 100,000 or more [15] or even capture data about every mobile subscriber of a country [10]. At these scales, CDRs can be used to track migration patterns between rural and urban areas, or build an entire country's social network graph. However, due to their proprietary nature, obtaining the data can be difficult. CDR-based studies lack fine-grained data, which on-device collection provides at the expense of study size. In section 5 we demonstrate that a handset-centric view can come to significantly different conclusions.

Installing software directly on participants' devices allows researchers to access sensor and application data, and record actions that occur offline. Examples of such studies include the MIT Reality Mining dataset which collected data from 100 mobile phones given to undergraduates [11], a study of application usage of 250 Windows Mobile and Android smartphones [12] and a study of application usage of 4,000 Android smartphones [4]. Studies have successfully collected data from up to 20,000 users by providing useful functionality to the user [14] or by bundling the logging software as a library with other applications [20].

A third type of study uses surveys or diary studies to capture usage behaviour. Examples include a 4-week diary study with 20 participants about mobile information needs [6] and an online survey of 350 people, followed by 20 structured interviews, about mobile phone power consumption [21].

Diary studies typically have few participants due to the large amount of work involved. Device Analyzer has captured fine-grained data from several thousand participants in longitudinal collections over the course of months.

3 The Data Collection Tool

Device Analyzer is a free application for Android version 2.1 or higher. This represents more than 99% of the devices connecting to Google Play.[1] The application collects data continuously in the background after initial setup, even when the user is not actively using the device. Device Analyzer was built with a focus on low resource use in order to impact the user experience as little as possible. We collect data asynchronously through event-driven notifications where possible and poll data only where necessary. This allows the device to spend more time in energy-saving sleep states. On most devices the battery drain from Device Analyzer is reported as less than 2-3%. We measured the additional power draw of Device Analyzer on a Samsung Galaxy S III and found the amortized power draw over longer periods of time to be an additional 0.048 mA.

On average, the Device Analyzer mobile application collects of the order of 100,000 data points per day on a given device. Storage files are periodically compressed and later uploaded over an encrypted connection while the device is charging. Local data on the device is deleted after uploading finishes. The server uses a custom-built distributed analysis framework to extract insights from the raw data stream [26].

In contrast to other data collection tools, Device Analyzer is a stand-alone application that must be installed from the Google Play store. We have never embedded our tool as a library into third-party applications. The purpose of the application is clearly stated on the Google Play store, in the confirmation dialog when data collection is activated, and on the project website. We also remind users of on-going data collection once a month. We take care to enumerate the data collected, how sensitive data is processed, and what is transmitted to our servers. We provide a "quick feedback" feature inside the application to allow participants to send feedback without revealing their email address.

One user commented on the Google Play store "I wouldn't normally participate in such a thing, but the attention to detail in disclosure and handling of data is quite refreshing and deserves to be rewarded."

3.1 Design decisions and trade-offs

Device Analyzer was designed from the ground up to be minimally invasive, both in terms of privacy and user experience. Data collection was designed with a public release of the dataset in mind. The following design decisions result from these paradigms.

No demographic information is collected. Instead, we provide a mechanism that allows researchers to recruit, tag, and later re-identify participants (see Section 3.3).

The dataset is inhomogeneous in that we have so far encountered 1277 different types of devices, whereas other studies use a single type of device that facilitates comparisons between people. We see this diversity of devices in the wild as a strength as we avoid novelty effects from handing out handsets for the duration of the study.

Polling of not event-driven data occurs only every 5 minutes; this includes information about the 10 most recently started applications and network traffic. However, for 10% of screen-on sessions application starts are collected with a frequency of 2 Hz, a rate we adopted from a previous study of application usage on Android [4].

---

[1] `http://developer.android.com/about/dashboards/index.html`

Device Analyzer does not collect GPS location, but instead relies on network and WiFi location. Apart from saving energy this means that the set of requested permissions is somewhat less invasive. We also traded participant privacy for reduced utility for researchers by hashing all personally identifying information like phone numbers, network names, and SSIDs with a salted hash (see Section 3.2).

Lastly, we do not provide the application's source code. We believe that a multitude of slightly different, small-scale studies generating separate data silos is not in the interest of the research community. We encourage other researchers to feed data back to the community by deploying Device Analyzer for their own studies, taking advantage of a working, maintained platform where functionality can be augmented and users re-identified as described in Section 3.3.

3.2 Privacy

We strive to protect our participants' privacy as best we can by transferring data only over encrypted connections and removing direct personal identifiers and other sensitive information before they reach our servers. Device Analyzer was reviewed and approved by the University of Cambridge ethics committee. Our approach is compatible with an earlier set of recommendations made to researchers in ubiquitous computing [17].

We use a salted hash function derived from a hardware identifier, which allows us to correlate entries after Device Analyzer was re-installed, but prevents correlating these data between devices. Phone call records for example can only be used to compare whether two calls were made to the same number on the same device. It is not possible to determine whether two separate devices called the same number.

Device Analyzer requires user consent before any data is uploaded to our servers. Users can inspect the data before making that decision. We observe that roughly 40% of our 26,800 installations were never activated. Data collection can be paused and users can explicitly withdraw from further collection at any point in time (4% of users did so) and optionally also request all of their historic data to be deleted (2.5%).

Participants can download all of their collected raw data from the project website. So far, 2100 participants have chosen to do so. Finally, participants can choose the extent to which they want to share data with other researchers. While some participants chose not to share their recorded data with third parties, we have observed that a small subset of participants elected to opt-in to share more sensitive information.

3.3 Access to the dataset and deployment for custom studies

Despite the care we have put into addressing privacy concerns, there remain ways in which the privacy of our participants could be compromised. We invite the community to request summary records from us; we will do our best to accommodate all reasonable requests. Alternatively, researchers can sign a contract that grants them access to the dataset itself as a basis for both further research and commercial work, but prohibits re-identification of individuals.

When a specific group of people should be surveyed, or additional data is required, accessing the existing dataset is not sufficient. We provide researchers with a way to deploy the Device Analyzer platform and receive raw data from re-identifiable users.

Interested parties can contact us to recruit their own users and are given unique participation codes that users enter in the stock Device Analyzer application. This code then allows re-identification of individuals as well as access to the collected raw data. In summary, researchers can leverage the expertise and engineering effort that went into building Device Analyzer without having to implement their own logging application.

4 Description of the Dataset

Device Analyzer captures a rich, highly detailed time-series log of approximately 300 different events. As much detail as possible is captured. For example, Device Analyzer not only records when a device connects to a WiFi network; it records all the details captured whenever a WiFi scan occurs, including AP MAC address, SSID, signal strength, frequency and capabilities. Table 1 shows the categories of data collected. Data is preprocessed on the device to remove direct personal identifiers (see Section 3.2).

			Category	Event types collected
Number of data points	75 billion		Device settings	33
Aggregate trace duration	1,900 years		Installed Applications	17
Countries covered	175		System Properties	29
Unique phone types	1277		Bluetooth devices	21
Unique OS versions	884		WiFi networks	11
			Disk storage	6
Installed copies of DA	26,800		Energy & Charging	5
Consented to collection	16,000		Telephony	20
Users requested deletion	426		Data usage	38
Participation > 1 day	12,300		CPU & memory	11
Participation > 1 month	4,700		Alarms	10
Participation > 3 months	2250		Media & Contacts	8
Participation > 6 months	960		Sensors	15
Participation > 1 year	321			

Table 1: Overview of data collected. A complete list is available online.[1]

While previous work has examined some of the topics present in this paper, we want to stress that we are not aware of any study that rivals this dataset in detail, duration of data collection and size. We believe it is important to optimise all three dimensions in order to derive sound insights into user behaviour. We present evidence that many interesting, abrupt changes in behaviour are visible only in longitudinal datasets.

5 Analysis

To highlight the value of this dataset, we investigate some of the areas we believe are of particular interest for mobile and ubiquitous applications: communication, context, and the capacity to run long-lived applications. We also investigate interaction with the devices as well as communication behaviour.

---

[1] http://deviceanalyzer.cl.cam.ac.uk/keyValuePairs.htm

5.1 Movement patterns

Human mobility patterns are a valuable resource in fields like urban planning and when characterising environmental impact. Previous work looked mainly at Call Data Records (CDRs) that are generated when a phone interacts with the mobile network. Location accuracy of CDRs has been reported as 3 km^2 on average [15].

CDR-based movement traces will be inaccurate if users do not use their phone everywhere they travel. We set out to quantify this error. Device Analyzer collects location data every 5 minutes using the Google location API which fuses signal maps and WiFi fingerprints to improve accuracy. We simulate CDR-based movement traces by selecting only locations where text messages were exchanged or calls started or ended.

A previous study, using data from AT&T, reported users' *daily range*, which is the maximum distance that a phone has travelled in one day [16]. We adopt this measure for the purposes of comparison. Median/90th percentile daily ranges across our dataset were 5.8/51 km but only 0.9/28 km for simulated CDR-based movement traces. We note that a significant number of days record no movement at all for lack of calls and text messages. While CDR-based studies can choose to measure the most active users only, this inherent error is likely present in much previous work dealing with CDRs.

The AT&T data reports human mobility in Los Angeles and New York City. Their dataset partially mitigated the above effect by including locations where data transfers occurred. The authors state that they collected an average of 21 locations per day from 5% of AT&T subscribers in each region, while Device Analyzer collects on average 208 locations per day, but included only about 140 participants in these regions with a total of 8500 phone-days of contributed data. Despite the large difference in scale both datasets show the same trends, namely shorter weekend daily ranges and overall larger numbers for LA residents: Median weekday/weekend daily ranges in our dataset are 7.0/3.3 km for NYC residents and 8.0/4.8 km for LA residents, compared with 7.2/5.6 km for NYC and 9.5/8.5 km for LA residents in the previous study.

We suspect the notably larger weekend movement distances in the AT&T dataset are due to their much larger dataset which is bound to include more variation of use than we were able to observe. At the same time, the reported median figures ignore outliers. Indeed, our 25th and 75th percentiles are much wider than the AT&T figures.

In particular, the reported 98th percentile maximum daily ranges per user are just below 4000 km, covering the continental US. However, the maximum daily ranges in our dataset are over 11500 km from LA and 16000 km from NYC, as we do have data from international travels that are missing from the AT&T dataset. González et al. note that their data does not include movement larger than 1000km for the same reason [15].

The lack of extra-territorial data is a fundamental issue of network-centric datasets which—by virtue of their sheer volume—suggest near-perfect coverage and accuracy. User behaviour abroad can change abruptly due to unfamiliar surroundings and lack of mobile data connectivity. Journeys that exit country boundaries are more common in Europe than in the US, owing to the denser packing of countries and more frequent travel between them. Modelling human mobility from data that ignores these international journeys will result in incomplete and potentially misleading models.

Our dataset captures movement patterns in a global context (see Figure 1). Device Analyzer contains data collected from 175 countries, rather than just a single country
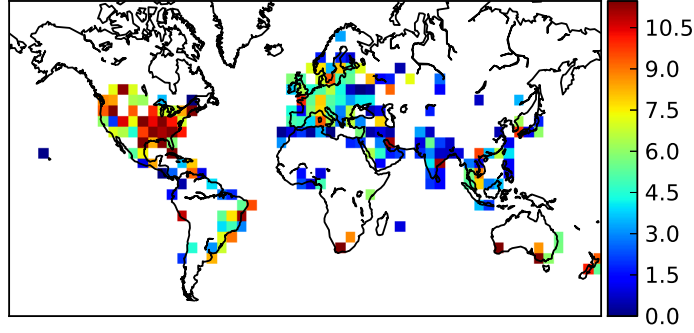
Fig. 1: Median daily ranges (km) of users living in different locations

as in most network-centric datasets. The AT&T 50th percentile maximum daily ranges for LA and NYC residents of approximately 58/43km km are considerably larger than results from a study by González et al. that reports almost 50% of their European participants staying within a 10 km radius over the duration of the 6 month study [15]. We observe the 50th percentile maximum daily ranges to be 89/65 km for LA/NYC residents, 188 km for the US overall, and 144 km for Europe. We also computed the radius within which all recorded locations of a given participant fall and observe the 50th percentile to be 47/40 km for LA/NYC residents, 129 km for the US overall, and 90 km for European participants. Indian participants travel less: The 50th percentile maximum daily range and total travel radius are 57 km and 25 km, respectively.

These figures suggest that at least in the US many city dwellers stay within a smaller area than their rural counterparts. The reported numbers also stand in direct contrast the previous findings of movement radii by González et al. Indeed, we note that for large parts of Europe the median daily range is of similar size as the previously reported gyration radius over multiple months. Device Analyzer was also able to replicate trends that the large-scale study of human mobility using AT&T data found, uncovering between-city effects in an untargeted, global dataset. We believe this showcases the suitability of on-device data collection for large-scale human movements.

5.2 Connectivity

Network communication is a central aspect of many mobile applications. As Device Analyzer collects data on the handset, we have the unique ability to measure connectivity as experienced by the user. We observe that 10% of participants have no network connection for at least 40% of the time, while half of our population spends less than 5% of their time without a connection. When connectivity is available, we observe that 10% of devices communicate over a 2G connection at least 54% of time. However, faster connectivity is often available: 50% (or 80%) of our users are connected to 3G, LTE or WIFI for at least 80% (or 45%) of the time. The top 10% of users spend as much as 98% of their time on these fast connections.

Users experience large temporal changes: Figure 2 shows individual traces for users with good, average and poor connectivity. Each column of an individual trace is a histogram of time spent using various technologies during one week. The second trace
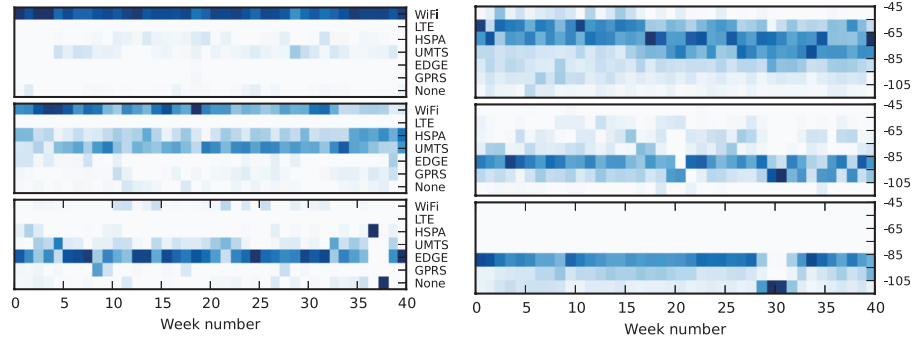
Fig. 2: Technology availability (left), GSM signal strength (right) for three devices

in Figure 2 (left) suffers from low WiFi availability in the last seven weeks. The bottom trace with poor connectivity typically has 2G connectivity but in week 37 sees good HSPA connectivity followed by a week of no connectivity. We find such abrupt changes in many long traces across our dataset.

Energy costs per byte transmitted can be as much as six times higher for a weak connection over a strong one [24]. This can have a noticeable effect on the battery performance of a communication-heavy application [9]. The majority of devices in our dataset saw signal levels of -85 dBm or lower most of the time, with occasional spikes of good signal. Only 15% of devices spending the majority of their time at signal levels of -75 dBm or higher. Figure 2 (right) highlights changes over time and between individuals: The top trace shows mostly signal levels around -65 dBm, which are very rare in the other two traces. The bottom trace shows no signal levels above -85 dBm and includes a period of several weeks with very poor signal.

CDMA and WiFi signal levels followed a similar pattern where less than 10% of devices spend time predominantly in very good signal conditions whereas 80% of devices spend the vast majority of time in medium to poor signal conditions.

Changes in mobile network connectivity present opportunities for applications to improve performance and reduce resource consumption through adaptation. Changes in connectivity manifest themselves either gradually over time or abruptly, which suggests that forecasting and planning cannot be done using signal strength alone and will require additional contextual information. Researchers should take these qualitative differences into consideration when interpreting results or planning studies.

5.3 Location

Location is a primary source of context in ubiquitous computing [1]. We investigate how well mobile ubiquitous systems can establish user location. Smartphones spend 88% of their time in close proximity to the user, making them a good proxy for the user's location [8]. WiFi fingerprinting uses readings of WiFi signal strengths to locate a device indoors [3], producing more accurate results as the density of visible base stations increases [5]. We investigate the number of base stations that participants in our

study were able to see at any given time, thereby giving an indication for the accuracy of positioning that can be expected from WiFi fingerprinting techniques.

Prevalence of WiFi: We observe that the majority of users keep WiFi off most of the time. Only 2% of users enable WiFi over 80% of the time while 56% enable WiFi less than 20% of the time. When WiFi is enabled, 50% (or 10%) of users spend up to half (or 90%) of their time with WiFi enabled but not connected. This suggests that while many people disable their WiFi connection when it is not needed, there is still much potential for further energy savings. Half of our users see on average 17 unique base stations on days when WiFi is enabled for at least 1 hour (top 10%: 82.6 base stations).

Number of visible APs: When WiFi is enabled, the majority of devices in our dataset see 3 or fewer access points most of time, and 1 or 2 access points for 29% of time, which limits the usefulness of WiFi fingerprinting systems. For 4% of the time, the majority of participants see no networks. On the other end of the spectrum we observe that roughly 1 in 10 devices spends most of its time in the presence of at least 8 base stations, providing ample information for WiFi location systems. Individual traces were omitted for reasons of brevity, but many users show significant deviations from their usual WiFi environment for weeks at a time.

Researchers have also considered using Bluetooth for location estimation. We find that over half of our participants activate Bluetooth at most 1 in 14 days. On those days Bluetooth is active for 52 minutes for a typical user, which would make parasitic location estimation highly unreliable.

5.4 Energy management

Battery lifetime depends highly on idiosyncratic usage behaviour and can severely limit the utility of modern smartphones. Power draw varies drastically over time, with one participant draining their battery by 298% one day (charging 3 times) and discharging only 38% on another day. We observe that 17% of participants fully charge their phone on at least 9 out of 10 days and 40% fully charge their phone on at least 7 out of 10 days. Overall, we observe that while 24% of all charges are over USB, the vast majority of these are short charges that may result from transferring data to or from the device.

Emptying the battery can be interpreted as a failure of the user or phone to adequately manage energy consumption. This happens at least every 11 days for 50% of our participants. Some participants exhaust their device's battery nearly every day, and several participants have distinct periods where their device dies frequently. On average, depleted batteries are charged within one hour, which suggests that the device is near a charger when it runs out of battery or that the user is on their way to a location with a charger. 90% of down times due to low battery are resolved within 12 hours.

Ubiquitous applications on mobile devices can expect to see a median uptime of 92.4% (90% of devices have an uptime of at least 63%). Furthermore, while numbers vary a lot across our dataset, 50 (or 80%) of devices can be expected to spend 10.3% (or 3.1%) of their powered-on time plugged into a charger and fully charged. Across devices, a long-running application can expect to see at least 1 hour of plugged-in, fully-charged time on 50% of days (6 hours 35 minutes on 20% of days). Applications could use this time to perform long-running and energy-intensive tasks.

A previous study of charging behaviour that collected data from a battery monitoring application over 4 weeks from 4,000 participants reported a large fraction of charges that lasted 14 hours or longer [13]. We observe only a very small fraction of charges to last 14 hours or longer. Instead, our dataset shows the majority of long charges to be between 6 and 10 hours in duration, which correspond to overnight charges.

Furthermore, this previous study reports mean charging duration as 3.9 hours, based on 1525 charge cycles. Our dataset shows a mean charging duration of 2 hours 21 minutes (median: 43 minutes) based on 1.7 million charge cycles. 95% of datasets with 1525 charge cycles would be expected to have a mean charge duration between 2 hours 9 minutes and 2 hours 33 minutes (two standard deviations). The results presented by this previous study are 16 standard deviations away from the mean found in our dataset, which indicates that a different population was sampled. We suspect that users installing a battery monitoring app may be susceptible to unusual charging behaviour.

Overnight charges A recent study investigating of battery charges of 15 participants over 3 weeks [2] reported that charges during the day and over night vary dramatically in length, with charges between 10pm and 5am having a median duration of 7 hours while other charges had a median duration of 30 minutes. Our dataset supports their findings, with a median AC charge duration of 5 hours 10 minutes (4 hours for AC and USB combined) for the above definition of overnight charges and a 33 minute median for charges during the day, irrespective of charge type. We believe that the difference in median charge duration stems mainly from the small sample size of 15 participants.

5.5 Interactions

The frequency and duration with which a user interacts with their device, as well as the time that passes between interactions, provides a context within which ubiquitous applications need to perform their activities. Modulating energy consumption in reaction to available resources is a key aspect of mobile ubiquitous applications.

We observe that some interactions are purely status checks where the device is never unlocked, but the screen is turned on to check the time or whether any notifications are present. The mobile platform can save energy by not waking up some core-components of the device in such situations, thereby extending battery life. Half of our participants perform these status checks on average 9 times or more per day, while 10% of participants check their device on average 52 times per day.

A 2010 study of 17,000 BlackBerry users by Oliver [19] looked at a number of metrics of interaction length and frequency. We observe mean daily device use across our dataset to be longer (2.05h vs 1.68h on BlackBerry), but median daily device use to be shorter (1.22h vs 1.31h). The presence of days with very heavy use hints at the diversity of the underlying population that we captured and highlights the variability of environments and modes of interaction that mobile ubiquitous systems must cope with. Interestingly, we find that while Android users interact less frequently with their device than their BlackBerry counterparts (on average 57 vs 87 interactions per day), the average duration of each interaction is much longer (115.8s vs 68.4s).

5.6 Calls and texts

Users in our dataset place or receive on average 7.5 calls per day (median: 3) and send or receive on average 11 text messages per day (median: 2). The average length of a text message is 55 characters. Using text messages and phone calls, our participants communicate with 4.9 unique numbers per day on average (median: 3). Per week, our participants communicate on average with 21.3 unique numbers (median: 16).

Across our dataset, 36% of calls end up unanswered. We observe that callers wait on average 21 seconds for the other side to pick up before cancelling the call. This is the same amount of time it takes the callee to pick up 95% of calls.

An analysis of the Reality Mining dataset reported [22] that 71% of calls are shorter than 1 minute, and 90% of calls are shorter than 5 minutes. Our dataset shows slightly longer calls, with 50.4% of calls shorter than 1 minute and 88.5% of calls shorter than 5 minutes. This may be an artefact of participant selection, as participants in the Reality Mining study were all undergraduate students at MIT.

6 Conclusions

We presented a novel collection platform that enabled us to collect the largest dataset of mobile smartphone use to date. Specifically, the dataset improves on previous studies in terms of duration of data collection, amount and diversity of devices covered, and amount of detail collected. In total the dataset contains 75 billion data points covering over 1,900 phone-years of active usage data from 16,000 participants.

To illustrate both the depth and versatility of the dataset, we extracted a multitude of different aspects about human movement patterns, interaction and communication patterns, connectivity and bandwidth, WiFi network availability in the wild, battery use, and reliability of the smartphone platform itself form the broad range of data available.

We contrasted and compared our results with previous work and found important differences in our new dataset. Crucially, we highlighted the importance of long-term data collection: many users show abrupt changes in usage behaviour or resource availability over long time periods to which ubiquitous applications need to adapt. These drastic changes are an important characteristic of real-world usage data and are typically not present in datasets collected over shorter time frames by previous studies.

We provide access to the dataset as well as a mechanism for other researchers to use Device Analyzer as a proven, reliable platform to run their own experiments on. As our dataset grows further we hope to create a sustainable, rich resource that will provide the community with invaluable data for years to come.

References

[1] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. E. Smith, and P. Steggles. Towards a better understanding of context and context-awareness. *CHI*, 2000.

[2] M. Y. Arslan, I. Singh, S. Singh, H. V. Madhyastha, K. Sundaresan, and S. V. Krishnamurthy. Computing while charging: Building a Distributed Computing Infrastructure Using Smartphones. *CoNEXT*, 2012.

[3] P. Bahl and V. Padmanabhan. RADAR: an in-building RF-based user location and tracking system. *IEEE INFOCOM*, 2000.

[4] M. Böhmer, B. Hecht, J. Schöning, A. Krüger, and G. Bauer. Falling Asleep with Angry Birds, Facebook and Kindle-A Large Scale Study on Mobile Application Usage. *MobileHCI*, 2011.

[5] Y.-C. Cheng, Y. Chawathe, A. LaMarca, and J. Krumm. Accuracy characterization for metropolitan-scale Wi-Fi localization. *MobiSys*, 2005.

[6] K. Church and B. Smyth. Understanding mobile information needs. *MobileHCI*, 2008.

[7] E. Cuervo, A. Balasubramanian, D.-k. Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl. MAUI: making smartphones last longer with code offload. *MobiSys*, 2010.

[8] A. K. Dey, K. Wac, D. Ferreira, K. Tassini, J.-H. Hong, and J. Ramos. Getting closer: an empirical investigation of the proximity of user to their smart phones. *UbiComp*, 2011.

[9] N. Ding, D. Wagner, X. Chen, A. Pathak, Y. C. Hu, and A. Rice. Characterizing and Modeling the Impact of Wireless Signal Strength on Smartphone Battery Drain. *SIGMETRICS*, 2013.

[10] N. Eagle, Y.-A. de Montjoye, and L. M. Bettencourt. Community Computing: Comparisons between Rural and Urban Societies Using Mobile Phone Data. *CSE*, 2009.

[11] N. Eagle and A. S. Pentland. Reality mining: sensing complex social systems. *Personal and Ubiquitous Computing*, 10(4), 2005.

[12] H. Falaki, R. Mahajan, S. Kandula, D. Lymberopoulos, R. Govindan, and D. Estrin. Diversity in smartphone usage. *MobiSys*, 2010.

[13] D. Ferreira, A. K. Dey, and V. Kostakos. Understanding Human-Smartphone Concerns : A Study of Battery Life. *Pervasive*, 2011.

[14] A. Girardello and F. Michahelles. AppAware: which mobile applications are hot? *MobileHCI*, 2010.

[15] M. C. González, C. A. Hidalgo, and A.-L. Barabási. Understanding individual human mobility patterns. *Nature*, 453(7196), 2008.

[16] S. Isaacman, R. Becker, R. Cceres, S. Kobourov, J. Rowland, and A. Varshavsky. A Tale of Two Cities. *HotMobile*, 2010.

[17] M. Langheinrich. Privacy by Design - Principles of Privacy-Aware Ubiquitous Systems. *UbiComp*, 2001.

[18] N. Maisonneuve, M. Stevens, M. E. Niessen, and L. Steels. NoiseTube: Measuring and mapping noise pollution with mobile phones. *ITEE*, 2009.

[19] E. Oliver. The challenges in large-scale smartphone user studies. *HotPlanet*, 2010.

[20] E. A. Oliver and S. Keshav. An empirical approach to smartphone energy level prediction. *UbiComp*, 2011.

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

[22] N. Ravi, J. Scott, L. Han, and L. Iftode. Context-aware Battery Management for Mobile Phones. *PerCom*, 2008.

[23] A. Rice and S. Hay. Decomposing power measurements for mobile devices. *PerCom*, 2010.

[24] A. Schulman, V. Navda, R. Ramjee, N. Spring, P. Deshpande, C. Grunewald, K. Jain, and V. N. Padmanabhan. Bartendr: a practical approach to energy-aware cellular data scheduling. *MobiCom*, 2010.

[25] N. Vallina-Rodriguez and J. Crowcroft. ErdOS : Achieving Energy Savings in Mobile OS. *MobiArch*, 2011.

[26] D. T. Wagner, A. Rice, and A. R. Beresford. Device Analyzer: Large-scale mobile data collection. *To appear in ACM SIGMETRICS Performance Evaluation Review, March 2014*.

[27] H. Ye, T. Gu, X. Zhu, J. Xu, X. Tao, J. Lu, and N. Jin. FTrack: Infrastructure-free floor localization via mobile phone sensing. *PerCom*, 2012.

[28] J. Ye, S. Dobson, and S. McKeever. Situation identification techniques in pervasive computing: A review. *Pervasive and Mobile Computing*, 8(1), 2012.