Comparing Sources of Location Data from Android Smartphones

Michael Spreitzenbarth, Sven Schmitt, Felix Freiling

Chapter 10

COMPARING SOURCES OF LOCATION DATA FROM ANDROID SMARTPHONES

Michael Spreitzenbarth, Sven Schmitt and Felix Freiling

Abstract It is well-known that, for various reasons, smartphone operating systems persistently store location data in local storage. Less well-known is the fact that various network applications (apps) do this too. This paper considers the issue if location data extracted from mobile phones can replace or complement the location data obtained from network operators. Experiments with Android smartphones reveal that location data stored on the phones is often much more precise than the rather coarse-grained data stored by network operators. However, the availability of location data on smartphones varies considerably compared with the data stored by network operators.

Keywords: Mobile phone forensics, Android phones, location data

1. Introduction

This paper focuses on the location data stored by Android smartphones. In particular, it attempts to answer the question: In what sense can the location data extracted from mobile phones replace or complement the location data obtained from network operators?

Unlike other research in the area, this work also considers location data maintained by mobile phone applications (apps). The ADEL forensic tool [7] was modified to extract all the different sources of location data and merge them to create a more complete picture of the movements of a phone. The extracted data includes location data from popular social networking apps as well as location data from cache files and GPS coordinates associated with pictures.

Experiments were performed using three brand-new Samsung Galaxy S2 (Android) smartphones. These phones were given to three students who used them with the understanding that the smartphones would

be analyzed forensically. The students were encouraged them to take many photographs and make use of social networks and Twitter. After almost four weeks, the data was extracted and analyzed, and the quality of the location data was compared with the data available from Spitz [17]. Overall, the experimental results demonstrate that location data stored on smartphones is often much more precise than the rather coarse-grained data stored by network operators. However, the availability of location data on smartphones varies considerably compared with the data stored by network operators.

2. Background

Following the 2004 terrorist attacks in Madrid, the European Union issued a 2006 directive [6] to harmonize regulations in EU member states concerning the retention of data generated by publicly-available electronic communications services. The directive seeks to enable law enforcement to access traffic data pertaining to suspects, e.g., to discover who the suspects communicated with and the digital services that had been used. In addition to data about individual communications, the directive also requires certain location data to be retained by network operators. Specifically, the directive requires that the following data be retained for at least six months:

- Identity and exact GPS coordinates of the radio cell where the user started a phone call.
- Identity and coordinates of the radio cell that was active at the beginning of a GPRS data transmission.
- Time stamps corresponding to this data.

This information can help investigators create movement profiles of suspects. Also, the information may be used to locate and monitor suspects.

Many EU member countries have implemented this directive in national laws. However, in some countries, there has been an intensive public debate about the laws, especially in relation to their threats to privacy. In Germany, discussions were fueled by a data set provided by the German politician Malte Spitz [17]. The data set contained location data over a period of six months that was preserved by his mobile network operator under the data retention law. A German newspaper created a graphical interface that enabled users to visually replay Spitz's detailed movements [17].

Overall, it is argued that retaining large amounts of data creates new risks of abuse. Also, the requirement to store data pertaining to millions

of innocent people is out of proportion to the small number of cases in which the data is used by law enforcement. As a result, in 2011, the German Constitutional Court dismissed the original legislation requiring data retention. Meanwhile, the search for less invasive techniques to analyze the movements of criminals continues.

2.1 Location Data in Mobile Phones

In recent years, many new types of mobile phones (smartphones) have flooded the market. Since they are essentially small personal computers, they offer much more than the possibility to make phone calls and surf the Internet. Over the last two years, devices based on the Android operating system have become very popular, with a market share of more than 40% and sales of more than 46 million units during the second quarter of 2011 [8]. Increasing numbers of subscribers are using apps (mostly third party applications that are directly installed on their phones) and are communicating with friends and family via social networks such as Facebook, Google+ and Twitter. The ubiquity of these services is overwhelming. Facebook had more than 600 million active users in January 2011 [4]; Google+ attracted 25 million active users in less than one month [9]; Twitter claimed that it topped 100 million monthly users for the first time in August 2011.

For performance and other reasons, mobile devices persistently store location data in local memory. In April 2011, it was reported that Android and iOS store sensitive geographical data [1, 13]. This data, which is maintained in system cache files, is regularly sent to platform developers. But generating geographical data is not restricted to the operating system – many apps that provide location-based services also create and store such data. For example, Benford [2] has shown that pictures taken by an iPhone contain the GPS coordinates of the locations where the pictures were taken. Such data is sensitive because it can be used to create movement profiles. Unlike the location data retained by network operators, location data stored on smartphones can be accessed by law enforcement via an open seizure.

2.2 Forensic Analysis of Mobile Phones

Mobile phones are an increasingly important source of digital evidence. However, until recently, extracting and analyzing data from mobile phones was rather cumbersome because of the diversity of their hardware and software. The situation is gradually changing as the market of mobile platforms consolidates and more tools become available. Hoog [10] discusses the forensic analysis of Android devices, including

details about the file system and information stored in apps such as Facebook and Google Maps.

Several commercial forensic tools have been developed for mobile phones, including EnCase and MOBILedit! [5]. Additionally, hardware solutions for forensic analysis such as XRY [12] are available. The only extraction and analysis tool available as a research prototype is ADEL [7], which incorporates several programming scripts developed for Android 2.x platforms. While some of these tools extract location data stored by the operating system, none of them address location data stored by popular apps.

Up to now, forensic examinations of social networks have primarily focused on computers. A SANS Institute post [3] outlines a method for finding Facebook data in the memory dump of a computer. Wong, *al.* [16] describe Facebook forensics and the reverse engineering of the Facebook API on virtual environments and mobile devices. However, with regard to the Android operating system, only the design of the database associated with the Facebook app has been studied. Twitter has not been investigated on mobile devices as yet, and only a few publications are available for Google+, among them, a discussion of artifacts related to URL forwarding [15].

3. ADEL Forensic Tool

ADEL (Android Data Extractor Lite) [7] is a forensic data extraction and analysis tool for versions 2.x of the Android platform. The tool incorporates multiple scripts (modules) written in Python and can be extended rather easily. It can automatically dump predefined SQLite database files from Android devices and extract the contents stored in the dumped database.

Figure 1 shows the ADEL workflow. In the first step, ADEL establishes a connection to an Android device via the Android debugging bridge (`adb`), dumps the predefined SQLite database files from the phone and stores them on the investigator's machine (dump module). All of the subsequent steps are performed on copies of the database files in the read-only mode to ensure data integrity. In the second step, the contents of the dumped database file copies are analyzed and extracted (analysis module). To accomplish this, we developed a specialized parser module for the SQLite database file format [14], which extracts content by directly parsing database files. After the contents are extracted, an XML-based report is generated to support the further use and presentation of data (report module). The report can be viewed using an ordinary web browser and can be refurbished with the help of an XSL

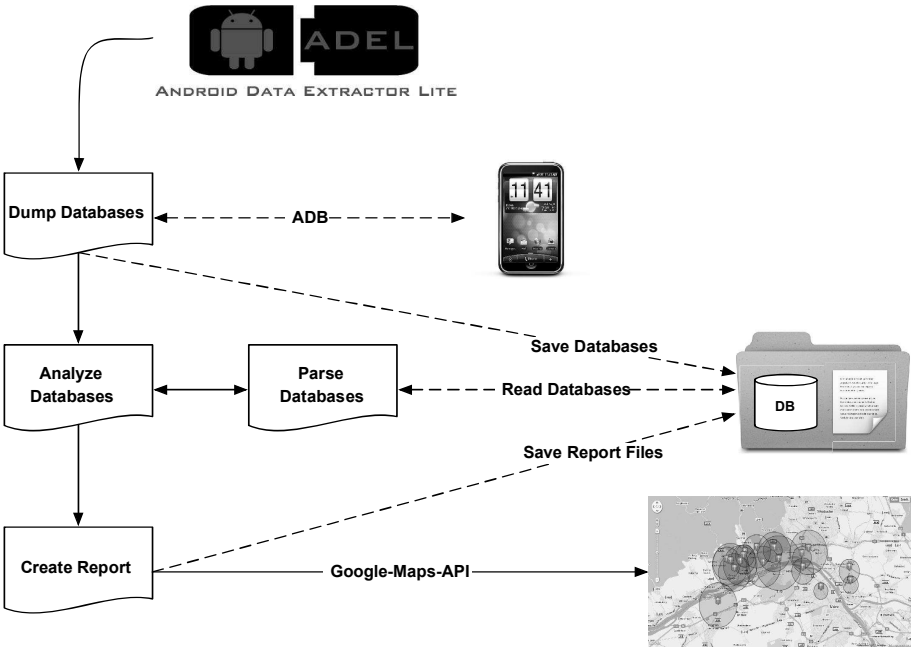

Figure 1. Android Data Extractor Lite (ADEL) workflow.

file. Although it is possible to optionally generate an extensive log file containing entries for all the major steps performed during script execution, the use of ADEL is intended to be as simple as possible for experts as well as non-experts.

In the original development state [7], the following information can be dumped and analyzed:

- Telephone and SIM-card information (e.g., IMSI and serial number)
- Phone book and call lists
- Calendar entries
- Browser history and bookmarks
- SMS messages

ADEL is built in a modular manner and, thus, can be augmented quite easily to provide additional functionality. A disadvantage with ADEL is that it can only be used with mobile phones that provide root access to applications ("rooted phones"). The root access is necessary in order to access the `adb` interface.

4. Accessing Location Data

This section describes a method for using ADEL with a non-rooted smartphone. Also, it describes the additional modules incorporated to extract location data from specific apps on a smartphone and combine the extracted data in an intuitive visual representation.

4.1 Daemon Replacement

Since many manufacturers have publicly released their boot loaders [11], it is no longer necessary to exploit the Android system to gain root access to execute ADEL correctly. Indeed, it is sufficient to modify the original kernel so that a root shell is included. As a result, the amount of modified data is significantly less compared with other approaches (e.g., using the "rage against the cage" root exploit).

After a smartphone is updated with the modified kernel, some changes must be made manually to guarantee the trustworthiness and integrity of the extracted data. In particular, the original, untrusted `adb-daemon` on the smartphone must be replaced by a trusted version. The following commands can be used to copy a trusted `adb-daemon` to the smartphone and, subsequently, remount the system partition to make it writeable:

```
adb push adbd /sdcard/
adb shell
su -
mount -orw,remount /
mv /sbin/adbd /sbin/adbd.old
mv /sdcard/adbd /sbin/adbd
mount -oro,remount /
kill \$(ps $\mid$ grep adbd)
```

Note that the existing daemon is backed up and the new daemon moved to its place. Following this, the partition is mounted to read-only again to prevent further changes. In the final step, the running (original) `adb-daemon` is terminated.

4.2 New Location Data Sources

We investigated the location data stored by several well-known Android apps. Table 1 lists some of these apps, their databases and their content. The upper half of the table contains data originating from system services, the lower half shows data retrieved from third party apps. The first group includes the cache files and pictures taken by the integrated camera. The camera records GPS coordinates and stores the coordinates in the EXIF data of a JPEG image (if the GPS is switched

Table 1. Android applications and stored location information.

App	Storage Location	Content
System	`cache.cell`	Last 50 mobile telecommunications cells
System	`cache.wifi`	Last 200 WiFi routers
Camera	Pictures	Latitude and longitude of picture location
Browser	`CachedGeopositions.db`	Latitude, longitude, accuracy and timestamp
Twitter	`author_id.db` – status	Latitude and longitude of status message
Twitter	`author_id.db` – search_queries	Latitude, longitude and radius of location search queries
Facebook	`fb.db` – user_status	Latitude and longitude of status message
Facebook	`fb.db` – user_values	Latitude, longitude and timestamp of last check-in
Google-Maps	`da_destination_history`	Source and destination of navigation

on). The integrated browser stores location data that is normally used for Google searches.

The Twitter app adds GPS coordinates to every published message. Although this function is deactivated in Android devices by default, it is often enabled by users after installation. Twitter also stores location data related to local search requests.

Google Maps stores all data pertaining to navigations in a separate database. This includes, for example, the current address, destination address and current time. However, since this data is stored when Google Maps computes a navigation, it cannot be assumed that the user actually traveled to the destination.

We also investigated the Facebook app. According to Hoog [10], this app stores location data in a database. However, we could not verify this fact during our investigation. The database and the written status messages were empty although the app had been used extensively. Only the data pertaining to the last location was found in the database.

5. Experimental Results

This sections describes the experiments that were conducted and their results.

5.1 Experimental Set-Up

We purchased three brand new Samsung Galaxy S2 smartphones and gave them to three undergraduate students in computer science who had volunteered to participate in the study. We explicitly selected participants who were active members of multiple social networks. The participants were asked to use the smartphones on a daily basis "according to their normal behavior." The smartphones were provided free-of-charge for a period of approximately four weeks.

After the smartphones were returned, we extracted and analyzed the stored data with the help of ADEL. We modeled each piece of information as a data point, i.e., a tuple (s, g, a, t, d) where s is the specific service or app, g is the GPS measurement (latitude/longitude), a is the measure of accuracy, t is the timestamp, and d is the duration of time for which the measurement holds.

The following assumptions were made regarding the accuracy of location data:

- Location data stored in the EXIF header of pictures has an accuracy of 51 to 100 meters.
- Location data stored by Twitter during the transmission of a message has an accuracy of 51 to 100 meters.
- Google Maps data has an accuracy of less than 50 meters.

The assumptions are made because apps do not always store the exact distance to the broadcasting tower. In the case of GPS sensors, it can be assumed that more precise location data is more accurate. To validate this assumption, we took 100 pictures with a smartphone and compared the stored data with data from a high-precision GPS receiver. We also checked the results inside and outside buildings. Much of the inaccuracy appears to arise in measurements made indoors.

We performed similar experiments with Twitter and Google Maps. In the case of Twitter, it can be expected that a large fraction of messages would be produced and consumed indoors.

Our experiments indicated that Google Maps has by far the best accuracy. Of course, most users would use Google Maps for outdoor navigation as opposed to indoor navigation.

We used a special convention with regard to timestamps when performing the analysis using ADEL. In particular, we assumed that the user remains at a given location for a time period of 15 minutes. This convention was used whenever the interval between several stored location data points was greater than 15 minutes. Otherwise, the interval of time between two timestamps was been chosen as the length of stay.

Table 2. Stored data from smartphones vs. network operators.

Data Source	Phone 1	Phone 2	Phone 3	Operator
Cell-Cache	50	50	50	3,223
WiFi-Cache	67	200	175	0
Twitter	1	0	5	0
Google Maps	2	0	3	0
Pictures	20	0	31	0
Browser	1	1	1	0
Facebook	1	0	1	0

5.2 Data Comparison

We compared the data extracted using ADEL with the Spitz data [17]. The Spitz data was collected by a large German network operator according to regulations enacted as a result of the EU data retention directive [6]. The Spitz data only contains the GPS coordinates of base stations and the rough directions of radio beams. Since cell site locations are smaller in densely populated areas than in the countryside and Spitz had mainly visited larger cities, we assumed that the accuracy was between 501 and 1,000 meters most of the time. The remainder of the time we assumed an accuracy of at least 1,000 meters.

5.3 Collected Data

Table 2 provides an overview of the smartphone data extracted using ADEL and network operator (Spitz) data. The number of points in the Spitz data was scaled down to cover approximately the same time frame as the ADEL data. Clearly, the number of data points retained by the network operator is much greater than those found upon forensically analyzing the three smartphones. The reason is that smartphones only save data associated with the last 50 mobile phone cells. Note, however, that the Spitz data is only associated with mobile phone cells, while the smartphone data comes from various sources.

Table 2 also shows that Phone 2 accessed the largest number of WiFi networks. However, according to the extracted data, it did not use the integrated camera, navigation system and social networking apps. Fortunately, the other two phones used the camera, navigation and social networking apps during the test period.

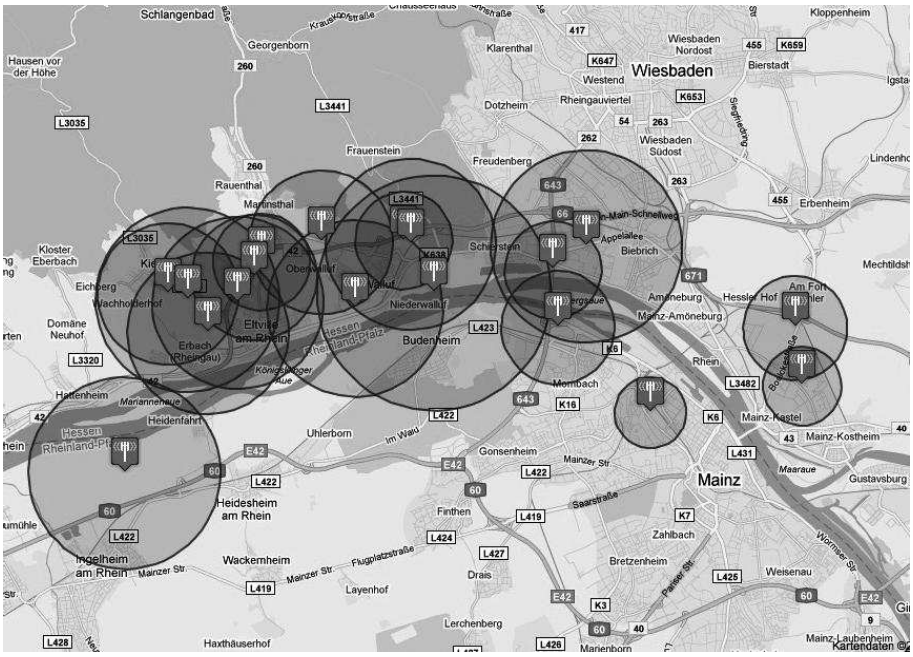

Figure 2. Movement profile generated from Smartphone 1 data.

6. Movement Profiles

Detailed movement profiles can be created using the data extracted with ADEL. Figures 2 and 3 show two examples of movement profiles. Each circle in the figures represents the approximate position of a user when a message was published in a social network or when the user made a phone call. The figures also contain marks corresponding to when Google Maps navigations were initiated. However, the navigation destinations were intentionally ignored because it cannot be guaranteed that the routes were ever taken.

Figures 2 and 3 also present data from the two cache files, `cache.wifi` and `cache.cell`, along with GPS data for pictures taken with the phones. Note that data from the smartphones themselves yields movement profiles that are much more detailed that those generated using network operator data (for which individual cells can have diameters exceeding 1 km). Fusing additional data from the smartphones can yield even more detailed movement profiles.

Figure 2 shows data associated with a trip through the Rhinegau, a region in Germany. All the data points were generated within a time-

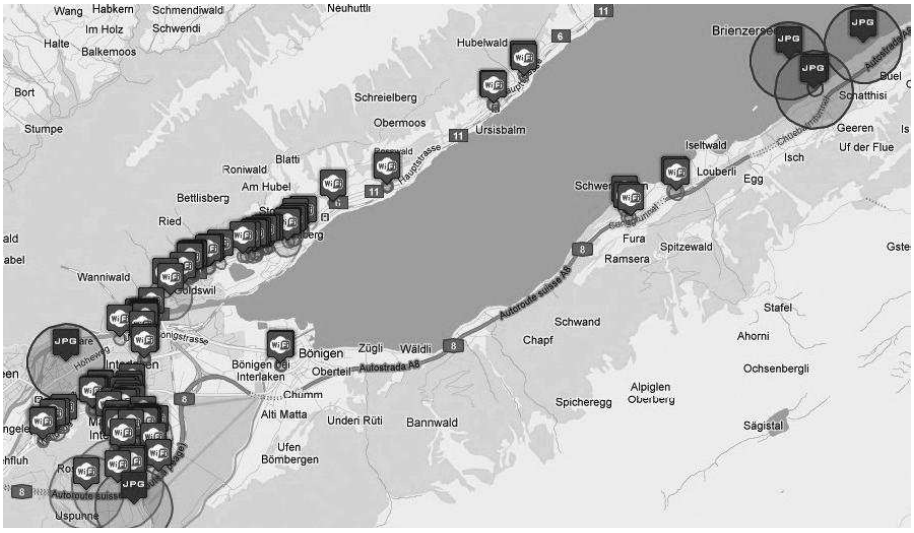

Figure 3. Movement profile generated from Smartphone 3 data.

frame of five hours. According to stored data, the user did not use any other apps during the trip nor did he take any pictures with the smartphone. Based on the overlap of the mobile cells to which the user was connected, it is possible to make assumptions about the streets on which the user traveled.

Figure 3 shows data associated with a trip around Brinzer Lake in Switzerland. No mobile cell data was recorded during this trip. The reason could be that the smartphone was not connected to the network because it had a German SIM card or because the maximum storage capacity of 50 cells was insufficient to store data associated with this portion of the trip. However, the smartphone apparently connected to a large number of WiFi routers. These "cells" are much smaller than mobile cells, which is why the location data has more precision. Also, some locations are recognizable in the pictures taken by the smartphone. Thus, it can be inferred that the user took Federal Road B11 and Highway A8.

The two examples demonstrate that the aggregation of GPS data from different sources can lead to very precise movement profiles.

7. Tracking Coverage

We now address the research question posed in the introduction: In what sense can location data extracted from mobile phones replace or complement the location data obtained from network operators? In

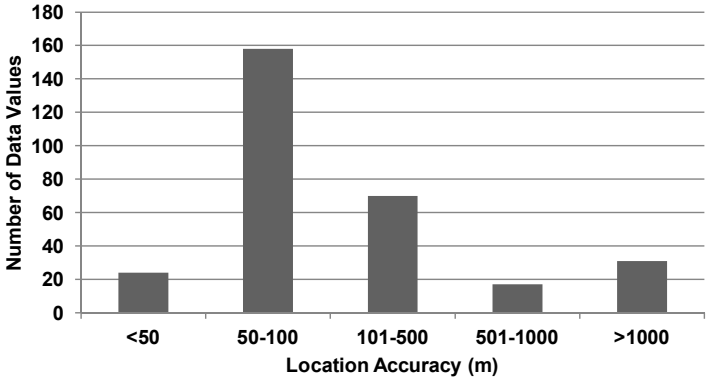

Figure 4. Location data from smartphones.

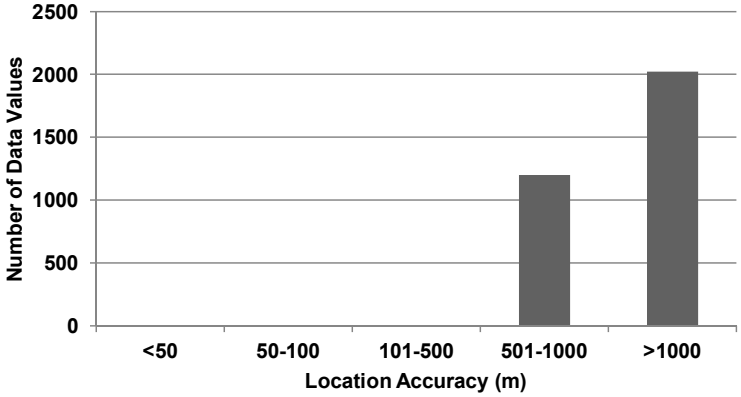

Figure 5. Location data from data retention.

attempting to answer this question, we assume that the location data from the mobile phones is extracted by ADEL.

Figures 2 and 3 show the movement profiles that were obtained from smartphones using ADEL. Thus, the question arises if data retention by network operators is at all necessary. We investigate this issue.

Figures 4 and 5 compare the location data obtained from smartphones with the data retained by network operators. Clearly, the number of data points is much higher in the case of data retention. However, the location data in much more accurate in the case of smartphones (50 to 100 meters) as opposed to data retention (500 meters or more). Clearly, location data extracted from smartphone using ADEL allows for much better positioning of users.

Another question pertains to the effects of the large difference in the number of location data points obtained from smartphones compared

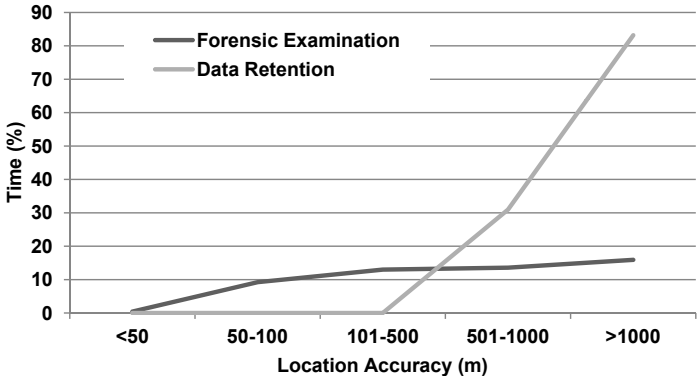

Figure 6. Percentage of time smartphones were traceable.

with data retention. Note that the smartphone experiment set the number of data points (including stored timestamps) in relation to the maximum possible time period (Figure 6). Since the experiment covered a period of two weeks, the maximum time during which the user is trackable amounts to 20.16 minutes. Taking the steeper line in the figure into consideration, the user is trackable about 83% of the time in the case of data retention – this corresponds to about 16.765 minutes during the two-week period. In contrast, based on data extracted from the smartphones using ADEL, the user is trackable about 17% of the time, which corresponds to only 3.428 minutes.

Upon comparing the two sets of results, it is evident that location data extracted from smartphones is much more precise that the data retained by network operators. However, smartphone-based data exhibits more time-related gaps. In a criminal investigation, tracking a user for 17% of the time using smartphone-based data is quite low compared with 83% of the time with data retention. However, if the time period of interest in the criminal investigation is within the trackability period, then a smartphone would yield much more precise location data for the investigation.

8. Conclusions

The extension of the ADEL tool described in this paper allows easier access to evidence residing on mobile phones. This includes evidence related to system services as well as popular web services.

With regard to location data, several services and applications running on Android devices store data about the geographical locations of the devices. An additional source of location data available to investigators is the data retained by network operators as a result of prevailing

laws and regulations. Our experiments comparing the data sources indicate that location data stored on an Android device is more accurate than the location data retained by network operators. However, network operators often store data for long periods of time whereas the data stored on a device is regularly overwritten with newer data.

One avenue for future research is the implementation of privacy enhancing techniques that reduce the types and amount of data stored on phones. For example, disabling the "Use Wireless Networks" option in the "Location and Security" settings menu of a device could result in the deletion of the `cache.wifi` and `cache.cell` files. Other options include turning off "Geotagging" in the camera settings and "Use my location" in the device privacy settings.

Finally, it is important to consider the possibility that the location data retrieved from a mobile device may not be completely reliable. This is true for location data pertaining to WiFi routers because the data is recorded only when a router is encountered for the first time. Since WiFi routers could be moved to new locations, the location data stored in `cache.wifi` may be outdated. Furthermore, in the case of apps such as Facebook and Google+, it is possible to link a user to a certain location although the user may not actually be at that location.

Acknowledgements

This research was supported by the German Federal Ministry of Education and Research under Grant No. 01BY1021 (MobWorm). A portion of this research was completed at Schloss Dagstuhl, Leibniz Center for Informatics, Wadern, Germany.

References

[1] J. Angwin and J. Valentino-Devries, Apple, Google collect user data, *Wall Street Journal*, April 21, 2011.

[2] D. Benford, Geotags: Friend or foe? *Forensic Focus* (`www.forensicfocus.com/geotags-friend-or-foe`).

[3] J. Bryner, Facebook memory forensics (`computer-forensics.sans.org/blog/2009/11/20/facebook-memory-forensics`), 2009.

[4] N. Carlson, Goldman to clients: Facebook has more than 600 million users, *Business Insider*, January 5, 2011.

[5] Compelson Labs, MOBILedit! Forensic Overview, Eugene, Oregon (`www.mobiledit.com/mef-overview.htm`).

[6] European Parliament and Council of the European Union, Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC, *Official Journal of the European Union*, vol. L(105), pp. 54–63, 2006.

[7] F. Freiling, S. Schmitt and M. Spreitzenbarth, Forensic analysis of smartphones: The Android Data Extractor Lite (ADEL), presented at the *ADFSL Conference on Digital Forensics, Security and Law*, 2011.

[8] Gartner, Gartner says sales of mobile devices in second quarter of 2011 grew 16.5 percent year-on-year; smartphone sales grew 74 percent, (`www.gartner.com/it/page.jsp?id=1764714`), August 11, 2011.

[9] P. Gobry, Google+ hits 25 million users, is the fastest growing website ever, *Business Insider*, August 3, 2011.

[10] A. Hoog, *Android Forensics: Investigation, Analysis and Mobile Security for Google Android*, Syngress, Waltham, Massachusetts, 2011.

[11] HTC, HTC Developer Center (`www.htcdev.com/devcenter`).

[12] Micro Systemation, XRY Physical, Stockholm, Sweden (`www.msab.com/xry/xry-physical`).

[13] J. Raphael, Apple vs. Android location tracking: Time for some truth (`blogs.computerworld.com/18190/apple_android_location_tracking`), April 25, 2011.

[14] SQLite, The SQLite Database File Format, Charlotte, North Carolina (`www.sqlite.org/fileformat2.html`).

[15] L. Whitfield, Flashpost: Google+ artifacts – URL forwarding (`www.forensic4cast.com/2011/07/flashpost-google-plus-artefacts-url-forwarding`), July 5, 2011.

[16] K. Wong, A. Lai, J. Yeung, W. Lee and P. Chan, Facebook Forensics (`www.fbiic.gov/public/2011/jul/Facebook\_Forensics-Finalized.pdf`), 2011.

[17] ZEIT Online, Tell-all telephone (`www.zeit.de/datenschutz/malte-spitz-data-retention`), August 31, 2009.