Analyzing Mobile Device Ads to Identify Users

Jayaprakash Govindaraj, Robin Verma, Gaurav Gupta

Chapter 6

ANALYZING MOBILE DEVICE ADS TO IDENTIFY USERS

Jayaprakash Govindaraj, Robin Verma and Gaurav Gupta

Abstract User browsing behavior is tracked by search providers in order to construct activity profiles that are used to fine-tune searches and present user-specific advertisements. When a search input matches a commercial product or service offering, ads based on the previously-saved interests, likes and dislikes are displayed. The number of web searches from mobile devices has exceeded those conducted from desktops. Mobile devices are being used for critical business tasks such as e-commerce, banking transactions, video conferences, email communications and confidential data storage. Companies are moving towards mobile-app-only strategies and advertisers are displaying ads on mobile apps as well. Mobile device ads can often reveal information such as location, gender, age and other valuable data about users. This chapter describes a methodology for extracting and analyzing ads on mobile devices to retrieve user-specific information, reconstruct a user profile and predict user identity. The results show that the methodology can identify a user even if he or she uses the same device, multiple devices, different networks or follows different usage patterns. The methodology can be used to support a digital forensic readiness framework for mobile devices. Additionally, it has applications in context-based security and proactive and reactive digital forensic investigations.

Keywords: Smartphones, advertisements, user behavior, user identification

1. Introduction

A 2014 mobile security survey by Checkpoint [5] reported that 64% of Android devices, 16% of Apple/iOS devices, about 16% of Windows phones and about 36% of BlackBerry devices are vulnerable to security threats. Insecure web browsing accounted for 61% of the total factors impacting the safety of mobile data. Meanwhile, 82% of security professionals expect mobile security incidents to increase, 98% have concerns

about the impact of mobile security incidents and 95% face challenges with bring your own device (BYOD) policies.

Mobile devices can be taken anywhere, increasing the possibility of the devices getting stolen or tampered with. Smartphones enable users to access the Internet from anywhere. Mobile devices are vulnerable to remote attacks through SMS/MMS or via the exploitation of insecure connections. Unlike hard disk drives, it is challenging to forensically-image phones without changing the states of the devices. Since phones use flash memory, every time an extraction is made, a different hash value is obtained.

All forensic images of phones are not equal. Logical extraction only provides a dump of the existing files such as call history, SMS and text messages; it does not acquire a dump of the unused space on the phone. Physical extraction can obtain a complete memory dump, but the process is difficult to perform without invasive techniques that could damage the phone. Most commercial forensic tools cannot bypass passcodes of smartphones [1]. Smartphones have to be jailbroken or rooted to access evidence required in digital forensic investigations.

Most mobile forensic solutions and products are designed for use in post-incident scenarios. At this time, there is no well-defined digital forensic readiness framework for mobile devices. Data collection is a key requirement in readiness scenarios [6, 10]. However, it is not clear what evidence to collect, how to handle situations where the collected evidence has been tampered with [11] and how to monitor and target particular evidence. Without question, there is a great need for new ways to identify users before security incidents occur as well as after the incidents.

Tracking a user's search keywords is one way search providers are gathering user preferences and targeting the most appropriate ads to display to users [13]. It is often the case that potential criminals plan their crimes using Internet searches or make purchases of objects or services needed to perpetrate their crimes. Knowledge of an accused's ad preferences could be useful in attempting to establish the sequence of events involved in the planning and execution of the crime. If the suspect had used mobile devices, the kind of ads that he clicked/viewed could reveal information about himself, including his motives and behavior. Although browser history may reveal more information, the specific ads that the suspect clicked and the websites he visited can reveal valuable information about his interests and behavior [4]. This is the principal motivation for analyzing the ads clicked by a user to identify the user.

This chapter describes a system that can track clicked ads in real time, extract the ads, analyze them to retrieve personal information and use

Table 1. Types of mobile ads.

Ad Type	**Description**
Video Ads	These ads are displayed during a game; a user has the option to skip after certain amount of time or watch the ads
Interactive Ads	These ads are similar to video ads; however, the user also has the opportunity to interact with the ads
Banner Ads	These ads are displayed within an app or site; the ads can be displayed in any location within the display limit based on the developers' intentions
Native Ads	The contents of these ads are aligned with the content of the app; the user does not feel that he/she is viewing ads
Pop-Up and Takeover Ads	These ads are displayed as dedicated full screen pages, requiring users to click to get to the original pages
Lock-Screen Ads	These ads are displayed whenever a device is locked
Notification Ads	These ads are displayed as notifications by apps belonging to a brand or company; the ads notify users of events and sales, and strengthen customer relationships
Rich Media Mobile Ads	These interactive ads provide rich user experiences using the mobile device gyroscope, camera and/or accelerometer
Branded Mobile Ads	These ads are specifically developed by advertisers and uploaded to an app store

this information to construct a user profile. The utility of the system is demonstrated via experiments involving multiple users, multiple devices and the collection and analysis of more than 5,000 ads. It is shown that, if a user operates a mobile phone in an office environment with restricted network access and uses the same device in a home environment with unrestricted network access, the two different usage patterns can still be used to identify the user. The system can also be used as a digital forensic readiness framework for mobile devices. Additionally, it has applications in context-based security and proactive and reactive digital forensic investigations. Finally, the system can be used to identify ads that violate constraints imposed by enterprises or government.

2. Background

This section provides background information on the main aspects of this research, including mobile ads, mobile ad targeting and mobile ad architecture. Ads are displayed to mobile device users via SMS, MMS, phone calls, web browsers and mobile applications. Table 1 presents the

Table 2. Mobile ad targeting methods.

Ad Targeting	Description
Content Targeting	Based on the app or the site where ads are displayed
Behavioral Targeting	Based on user behavior, browsing, recent downloads and interests by analyzing recent device locations
Device- or Carrier-Based Targeting	Based on the type of device or carrier used; ads for iPhone cases would only be displayed on iPhone devices
Demographic Targeting	Based on information such as user age, gender, ethnicity and language preference
Geographic Targeting	Based on the user location obtained from the device GPS or nearest cell tower location
Re-Targeting	Based on users who viewed or clicked the ad in the past
Time-Based Targeting	Based on the particular time of day

various types of mobile ads [14]. Different types of mobile ad targeting [7] are used by advertisers to reach users. Table 2 describes the various mobile ad targeting methods.

2.1 Information Revealed by Ads

An ad may reveal private information about a user that the user would not otherwise disclose. This corresponds to an unintended data leakage vulnerability [2]. The following information about a user or device may be leaked:

- App name and version of the ad that was clicked.
- List of device capabilities.
- Name of network operator.
- User-provided age.
- User-provided gender.
- Ad publisher account ID.
- Type of network used (e.g., 3G, 4G or Wi-Fi).
- User-set system language.
- App-supplied keywords.

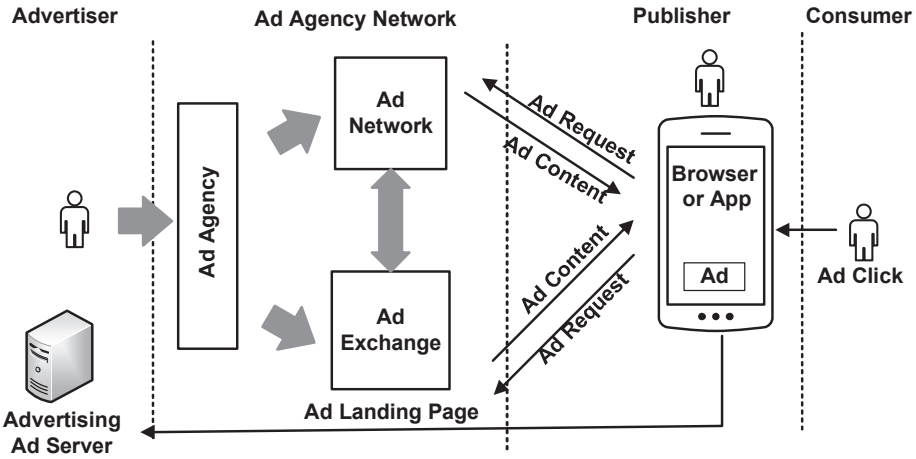


Figure 1. Mobile advertisement architecture.

- User location.
- Time zone.
- User demographic information.
- User emotional state (e.g., anger, fear, sadness, depression or hopelessness).

2.2 Mobile Advertisement Architecture

Figure 1 presents the mobile advertisement architecture. The workflow includes four participants: (i) advertisers; (ii) ad agency network; (iii) publishers; and (iv) consumers [8].

- Advertisers are commercial entities that wish to promote their products and services. An advertiser contracts an ad network for an ad campaign. The ad campaign typically specifies the advertising budget and the target numbers of clicks and impressions over a certain time period. An ad agency manages marketing, advertising and public relations services on behalf of the advertiser.
- An ad agency network consists of three components: (i) ad agency; (ii) ad network; and (iii) ad exchange:
  - An ad agency handles the marketing and branding strategies of advertisers.

  - An ad network collects ad space from publishers and segments the space to offer specialized groups of ads that match advertiser needs. An ad network participates in the exchange of ads and places bids on behalf of advertisers. An ad network receives information such as user profiles, contexts and device types from an ad exchange server. An ad exchange server also collects and shares ad metrics such as the numbers of clicks and impressions.

  - An ad exchange is a marketplace for publishers and advertisers to buy and sell ad space. It connects publishers with multiple ad networks. Buying and selling in an ad exchange occur via real-time auctions. An ad exchange is a neutral party that collects ads from different ad networks. An ad exchange server tracks down the list of displayed and clicked ads and determines the fees that an advertiser has to pay; some of the collected fees are passed to the publishers of the apps where the ads were displayed.

- Publishers develop mobile applications for mobile device users (consumers). A mobile application includes an ad control module that notifies the associated ad exchange server that there is an available slot for an ad on a user's device. The app also sends user information such as the user profile, context and device type to an ad exchange. The ad exchange server decides how to monetize the particular ad slot.

- Consumers are the end users of mobile apps who actually click on the ads.

3. Related Work

Toubiana et al. [12] have demonstrated that Google web search session cookies can expose personal user data. They claim that the cookies capture as much as 80% of a user's search/click history. Castelluccia et al. [3] have also shown how private user information is leaked during the web searches.

Korolova [9] has shown how the micro-targeted ads on Facebook leak private information belonging to users. Castelluccia et al. [4] show that knowledge of only a small number of websites containing Google ads can reveal an individual's interests with an accuracy of 79%. In fact, up to 58% of an individual's Google ad profile can be reconstructed using this information. Castelluccia et al. reportedly accessed the ads (which

are almost always served in the clear) by capturing data packets over a network.

Most recent research has focused on desktop web searches, browsing histories and ads displayed on desktop browsers, in the process demonstrating that private user information is leaked. In contrast, the system described in this chapter captures user URLs from mobile device cache files, cookies and the history database. The system retrieves ad-specific URLs and extracts private information that can be deduced from the ads. The system then reconstructs user behavior and sequences of events with the goal of establishing user identity. The system is capable of identifying users across multiple networks, even when they exhibit different usage patterns on a single device or on multiple devices.

4. Methodology

Ads are of various types, including location-based ads, content-based ads and targeted ads. Ads often reveal some private information about the user that the user would normally not disclose. An experimental setup was created to capture and analyze the ads. The targeted operating systems were iOS and Android. The targeted browsers were Safari and Chrome. Experiments were conducted with apps belonging to popular categories on four different devices and with four different users. More than 5,498 ads were captured.

The methodology involved four steps: (i) simulation of the mobile ad ecosystem; (ii) ad extraction; (iii) ad analysis; and (iv) inferences based on ad analysis (i.e., reconstructing a user identity).

5. Mobile Devices

Mobile devices display ads at two locations: (i) on the app itself; and (ii) on the search browser.

5.1 iOS Ad Architecture

As shown in Figure 2, whenever a user clicks on an ad on an iOS app, the ad information is stored in cookies (Figure 3(a)) and cache files (Figure 3(b)) in the corresponding app folder. Additionally, an entry is logged in the Safari history database.

Whenever a user clicks on an ad on the Safari browser or on an ad on an app itself, an entry is logged in the history database in the format shown in Table 3.

Figure 4 shows that the Safari history database is the common location for ad-related information.

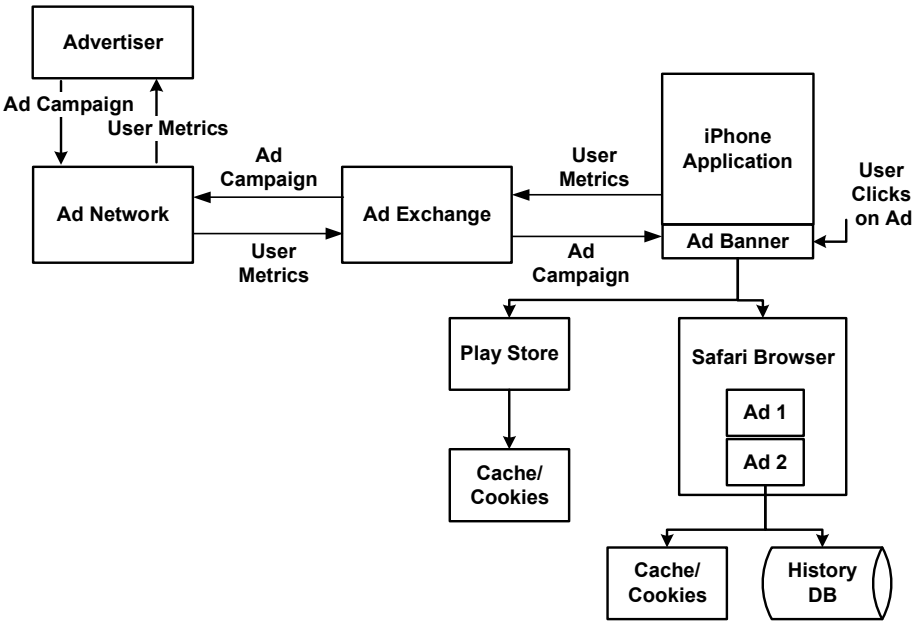


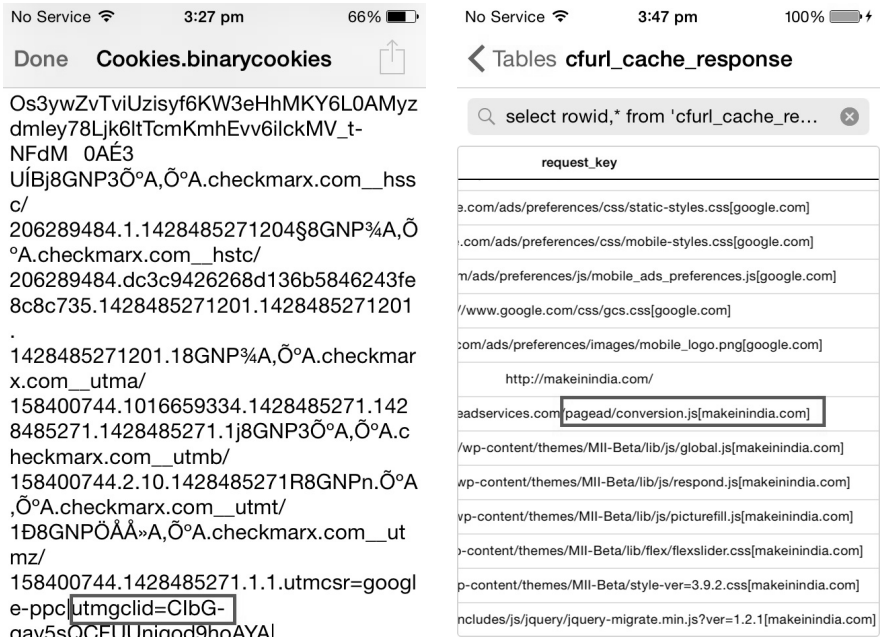
Figure 2. iOS ad architecture.

Table 3. Ad information format.

Timestamp	URL	Ad
10-Jun-2015 20:13:40	www.googleadservices.com/pagead/aclk?sa=L&aiCnX3vFT&adurl=http://99acres.com	99acres

iOS Ad Extraction. The Safari history database was accessed after jailbreaking the iOS device. Ads in the history database were tagged with `googleads.g.doubleclick.net` or `adclick.g.doubleclick.net`. As shown in Table 3, the history database contains the ad timestamp and URL. The ad URLs were extracted from the history database by searches using the keywords "adurl," "googleads" and "doubleclick." The ads sent to the app store had intermediate links that opened in the browser and were then redirected to the app store. These intermediate ad URLs were also stored in the history database, but the actual ad URLs were not stored. The intermediate URLs were extracted from the history database. A custom app was then used to replay the intermediate URLs and capture the actual ad URLs and other ad information.

To capture the ads in real time, the iOS device had to be jailbroken and some browser and app store functions had to be hooked. Figure 5(a) presents the iOS ad extraction process.



(a) Cookies. (b) Cache Files.

Figure 3. Cookies and cache information.

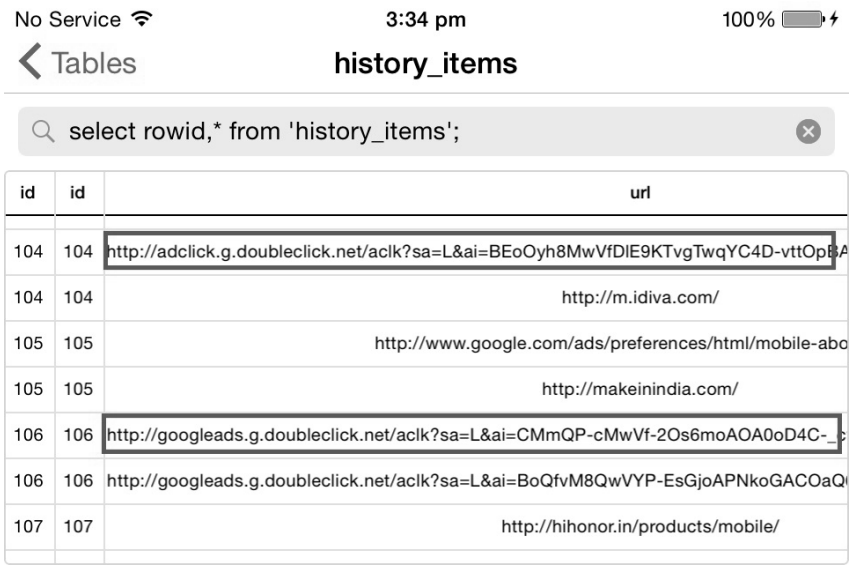
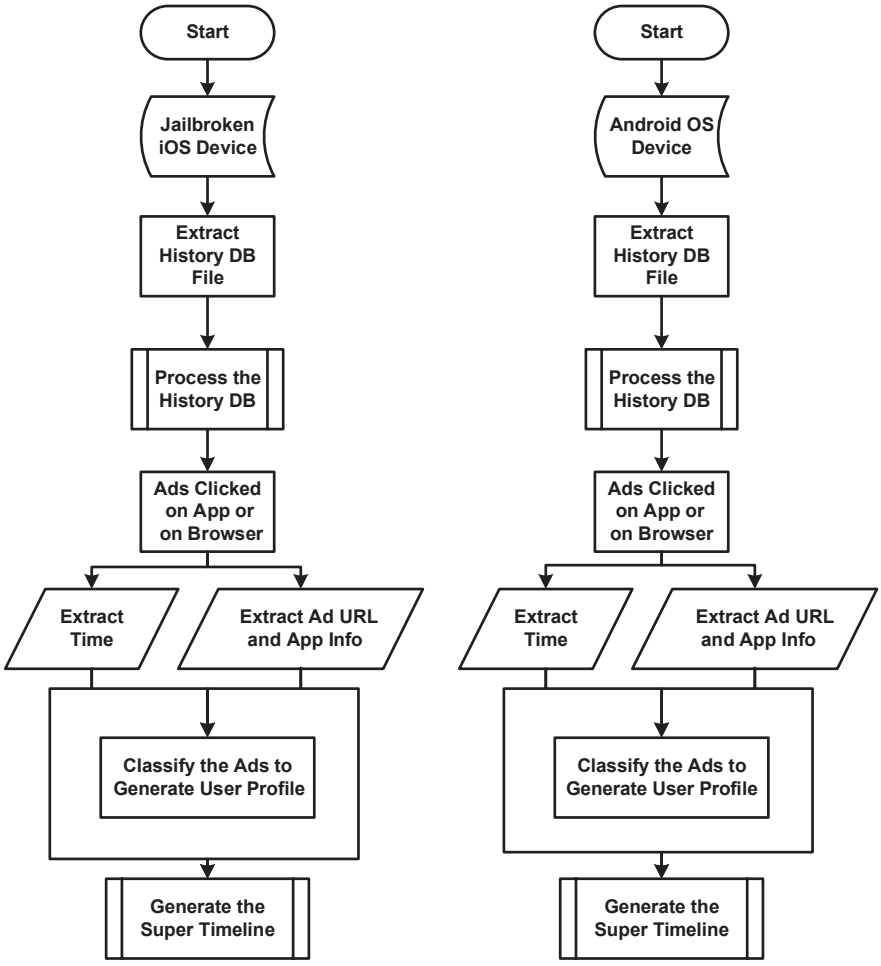


Figure 4. Safari history database information.



(a) iOS process. (b) Android process.

Figure 5. iOS and Android ad extraction processes.

5.2 Android Ad Architecture

As shown in Figure 6, when a user clicks on an ad on an Android app, the ad information is stored in a `logcat` file. Figure 7 presents the information stored in the `logcat` file.

On the other hand, when a user clicks on an ad in the Chrome browser, an entry is logged in the history database as shown in Figure 8. However, user clicks on some ads are redirected to the Google Play Store and the corresponding entries are logged in the `logcat` file.

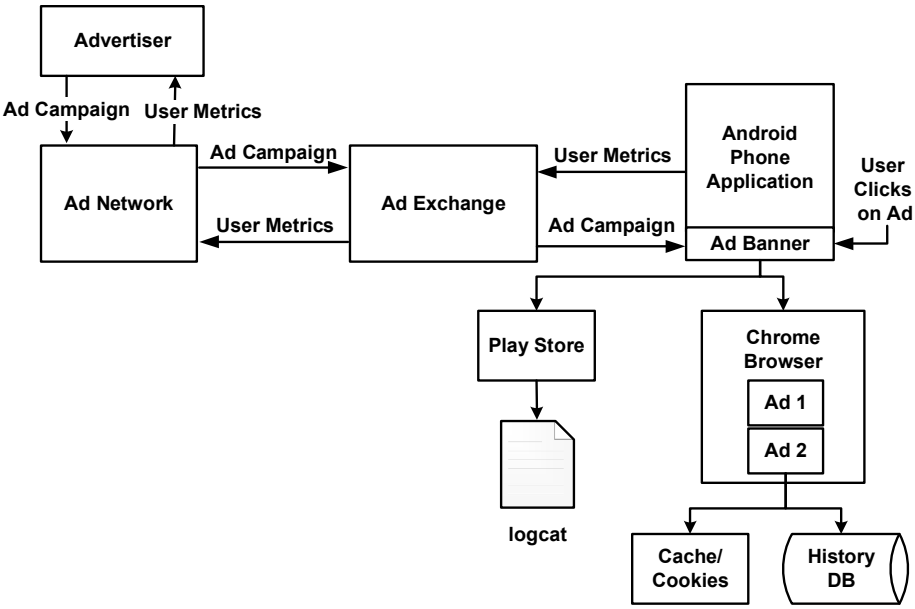


Figure 6. Android ad architecture.

```
Bangalore(UCA)&gclid=CM2IrcXX-sQCFRUVjgodaTQALQ
V/timeIds ( 8368): 16-4-2015 06:22:35
V/titleIds( 8368): Online Shopping India, Search and Buy Product Deals for Cheap
V/urlIds  ( 8368): http://www.askmebazaar.com/product.php?app_data=cHJvZHVjdF9pZ
V/timeIds ( 8368): 17-4-2015 02:08:58
V/titleIds( 8368): New Cars, Used Cars, Car Prices, Reviews & Photos in India -
V/urlIds  ( 8368): http://www.carwale.com/m/?ltsrc=l7831&gclid=COn-tJn4_MQCFYEoj
V/timeIds ( 8368): 17-4-2015 04:08:06
V/titleIds( 8368): Idea Cellular - Cell Phone Services | 3G, Prepaid, Postpaid &
V/urlIds  ( 8368): https://m.ideacellular.com/mobile/default/index.html?pn=insta
V/timeIds ( 8368): 17-4-2015 04:09:38
V/titleIds( 8368): Ford Figo Price in India, Photos & Review - CarWale
V/urlIds  ( 8368): http://www.carwale.com/m/ford-cars/figo/?ltsrc=21545&gclid=CI
```

Figure 7. Android `logcat`.

Android Ad Extraction. Android in-app ads and browser ads either open in the browser or in the Play Store. Information about in-app ads

Figure 8. Android history database.

that open in the browser and information about browser ads are stored in the browser history database. The challenge was to extract the history database and separate the ad URLs from normal web browsing URLs stored in the database. Typically, an Android device must be rooted to access the browser history database. However, it was discovered that the APIs provided by the Android SDK, corresponding to content providers of Chrome and the Android default browser, enable the extraction of history data without having to root an Android device.

After extracting the history database, it was necessary to separate the ad URLs from the normal web browsing URLs. After analyzing the ad URLs, it was discovered that they had some unique keywords such as "googleadsservices," "adclick" and "adurl." These keywords were employed as distinguishing features to separate ad URLs from web browsing URLs. Figure 5(b) presents the Android ad extraction process.

Based on this research, an Ad Extractor Android app was implemented to save the date, time, title and URL of each ad in the browser history database to the `logcat` file. As mentioned above, an algorithm was implemented to separate ad URLs based on specific keywords. The ad URLs were separated from the other URLs and copied to a file using ADB commands. Whenever the Ad Extractor app was launched, it dumped all the ad data to a specified file. Thus, the ads on a browser as well as ads from apps that go to a browser were captured.

The next task was to capture the ads that go to the Play Store from the app or browser. The URLs of these ads are not recorded in the browser history file. It was discovered that around 40% of the ads in Android apps (in-app ads) go to the Play Store. These ad URLs are also not stored in the browser history. However, because no Play Store history database exists, it was not possible to capture the required ad information.

To address this problem, a custom app that uses the Play Store URL schema is required. The idea is to open an app using the URL schema and enable the user to select whether to go to the Play Store or to use the custom-created URL schema app. Two URL schema exist for the

Table 4. Genymotion emulators.

Devices	Description
Nexus (Device 1)	Google Nexus 4
Samsung (Device 1)	Galaxy S3
Nexus (Device 2)	Google Nexus 4
Nexus (Device 2)	Google Nexus 4

Play Store – `market` and `play.google.com`. Thus, a URL schema app was developed with the same URL schema as the Play Store. When this custom app is used, instead of an ad going directly to the Play Store upon being clicked by a user, two options are provided to the user:

- Google Play Store app (built-in app).
- URL schema app (custom-built app).

Upon selecting the URL schema app, an ad is opened in the app, enabling the capture of the ad URL. Two URL schema apps were developed, one for capturing ads with the URL scheme `market:` and the other for ads with the URL scheme `play.google.com`.

All the steps (browser ad extraction and Google Play Store ad extraction) and the individual app functionality were consolidated into a single app that was deployed in the experiments. When a user clicked on ads, the consolidated app captured all the ads, stored them in temporary locations and sent the collected information via email to the researchers based on a predefined schedule.

In order to simulate a mobile ad ecosystem, an experimental system was created using Genymotion emulators. This system automated the ad clicking process for in-app banner ads (i.e., ads that appear in apps). Five apps were selected for the experiments: (i) Cricbuzz; (ii) Reddit Sync; (iii) 4 Pics 1 Word; (iv) Times of India; and (v) Advanced Permission Manager. The five apps were selected based on their popularity and coverage of categories such as sports, news and games; a controller app was also included. All the apps had banner ads at fixed locations that made automated clicking more efficient.

The five apps mentioned above and the two URL schema apps were installed in Genymotion emulators (see Table 4). A shell script was created to launch each app in turn and click on the ad. This process was repeated 1,000 times. In the simulations, ads that go to the browser directly from the history database were captured; in the case of ads that go to the Google Play Store, the ad URLs were captured using the URL

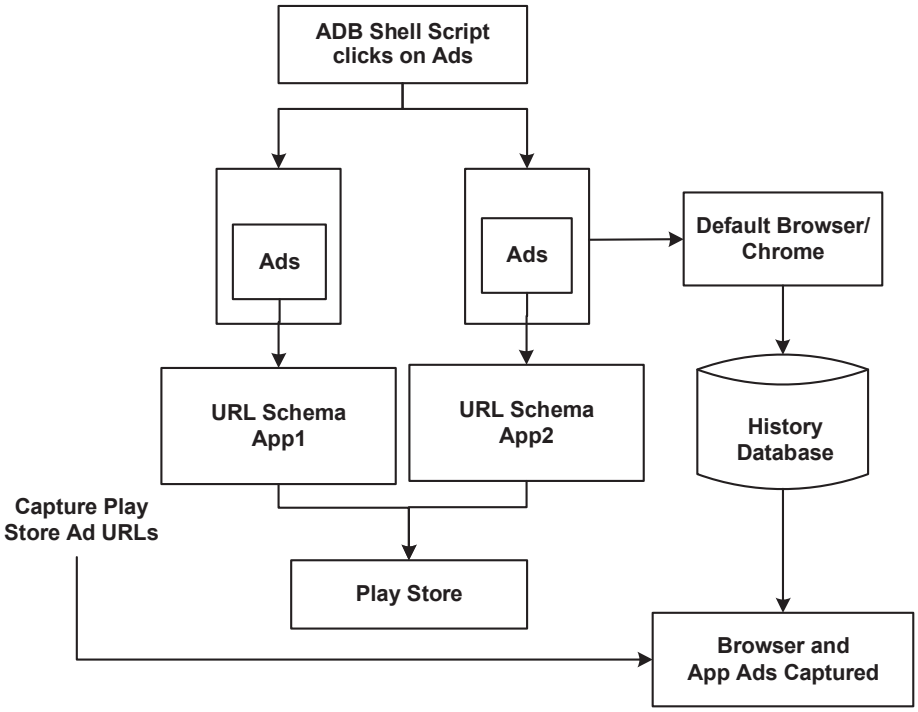


Figure 9. Experimental setup for Android ad extraction.

schema apps that logged them to separate text files along with their timestamps.

The experiments with the five apps were conducted by two different users (running two emulators per user) over a period of eight weeks. The ad data was subsequently collected from the history database and analyzed to construct user behavior profiles. Figure 9 summarizes the ad extraction process.

Android Ad Analysis. The experiments captured more than 5,498 ads from the four emulators (see Table 4). The ads were mined to collect the following data:

- **Ad Category Information:** To understand the different categories of ads.

- **Ad Interest Information:** To understand the different types of user interests that ads reveal.

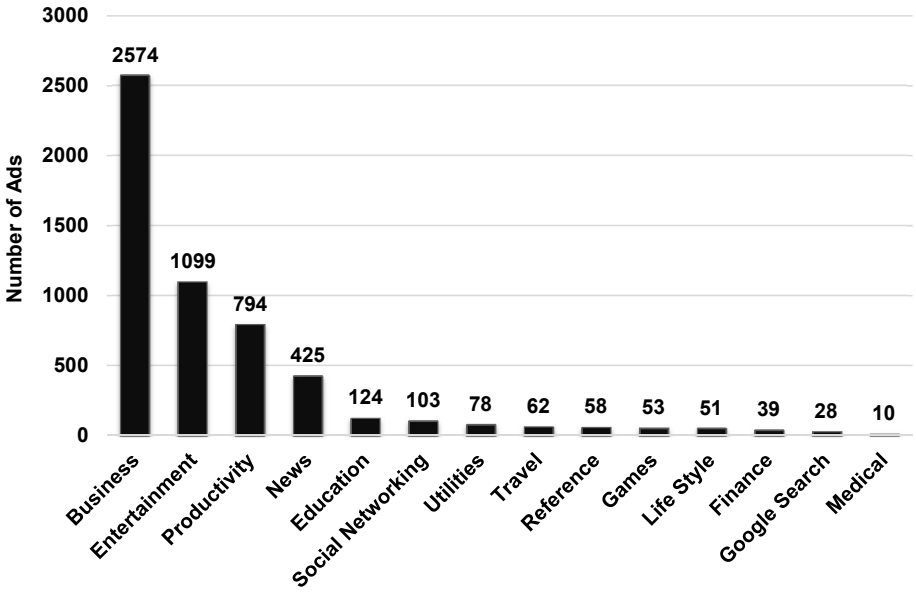


Figure 10. Ad category distribution.

- **User Information:** To obtain user-related information for constructing user profiles and generating user preferences based on the clicked ads.

Two Google accounts (User1 and User2) were set up and an automated program was created to read the URLs and browse the links. In the first experiment, User1 logged in and the automated program was executed to read the URLs and browse the links for one device in an office environment (restricted network). In the second experiment, User2 logged in and the automated program was executed to read the URLs and browse the links for one device in a home environment (unrestricted environment). Two user profiles were created and the results were compared. The next section presents the experimental results and inferences.

6. Results and Discussion

More than 5,499 ads were captured using the four emulators. This section discusses the statistics related to the captured results.

6.1 Ad Category

Figure 10 shows that 47% of the ads deal with business, 20% with entertainment, 14% with productivity and 8% with news.

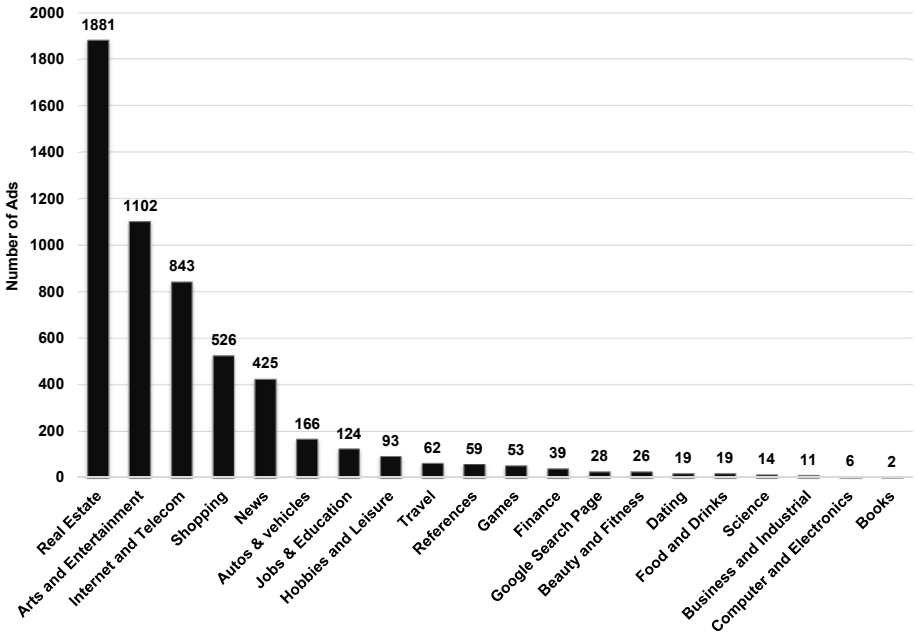


Figure 11. Ad interest distribution.

6.2 Ad Interest

Figure 11 shows that 34% of the ads deal with real estate, 20% with arts and entertainment, 15% with the Internet and telecommunications, and 10% with shopping.

6.3 User Information

Analysis indicated that 36% of the ads contained location information, 33% time information, 13% language information and 5% app information. If the ads are assumed to be clicked by the user more than 100 times, then 38% would contain location information, 34% time information and 14% language information (see Figure 12).

If the ads are assumed to be clicked by the user less than 100 times, then 22% would contain information about financial transactions, 21% about mobile phone account recharges, 10% about offers, 10% about cars and 7% about marital status (see Figure 13).

Figure 14 shows the preferences generated for a user who used two different environments. Based on this information, it was possible to link the user preferences to the same user.

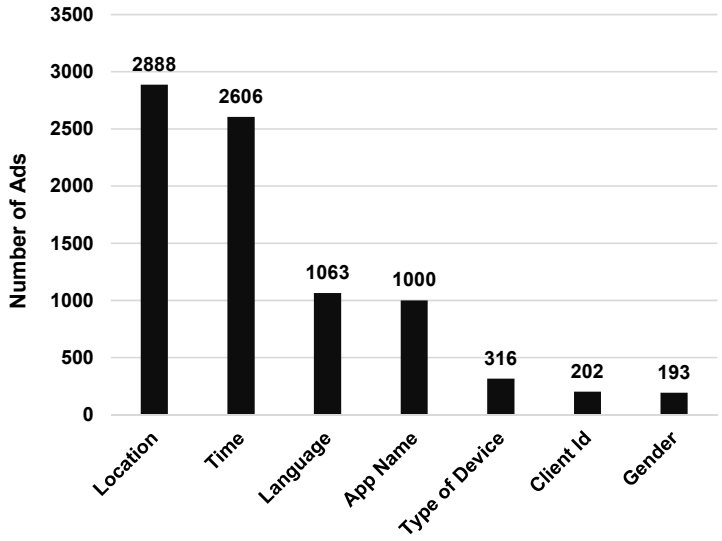


Figure 12. User information based on ads clicked more than 100 times.

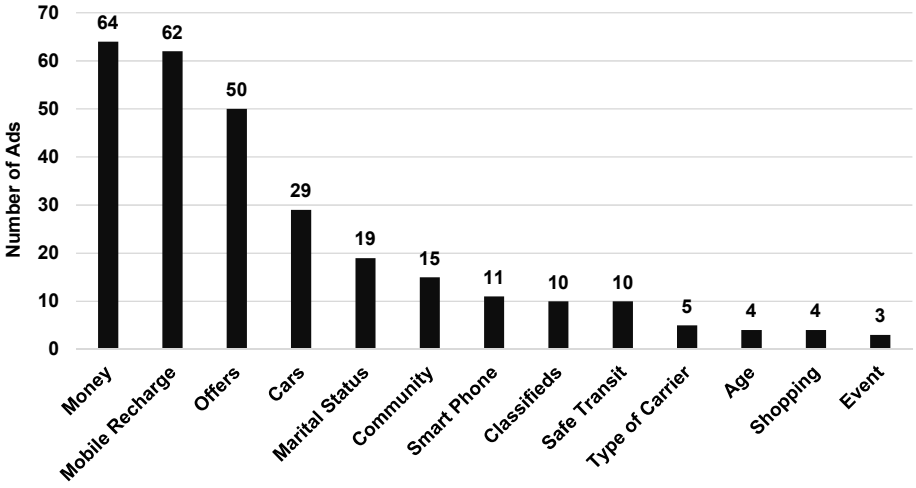


Figure 13. User information based on ads clicked less than 100 times.

6.4 Observations and Inferences

The following observations and inferences were made using the data collected from ads displayed in Android apps:

- Ads provided information relevant to events, offers and events. For example, a real estate offer lasted for five days and it was observed that the ads during the five days were related to the offer.

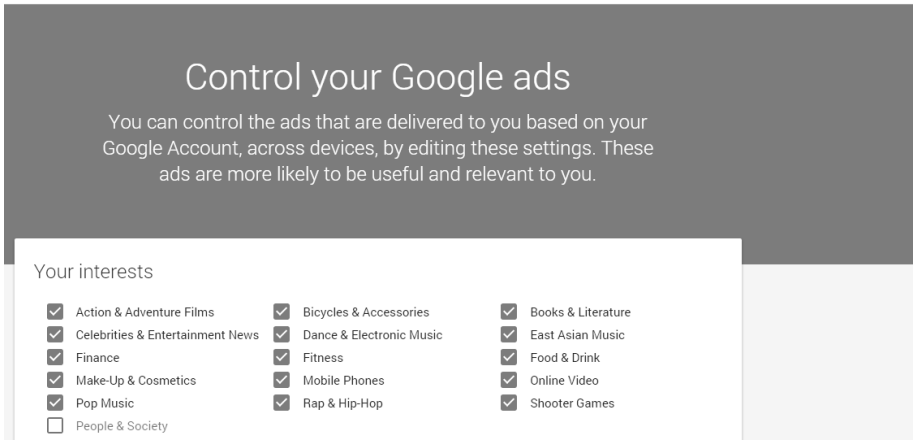


Figure 14. User preferences generated via Google.

- Ads provided information about user locations. For example, an ad for a jeweler showed the store locations and offers in the city of Bangalore; another ad was about a Kannada news app (Kannada is the regional language).

- No significant difference was observed in the displayed ads with respect to the user who was logged in.

The following observations and inferences were made using the data collected from ads displayed in Google search:

- Ads provided information about regional festivals, events and occasions (e.g., Diwali, Christmas and New Year festival offers, and sports events).

- Ads customized for geolocation provided information about user locations (e.g., `taxiwala.com`, a taxi service located in Delhi).

- The average database size of ads corresponding to a keyword was between five and ten; ads were almost always getting picked up from this database of ads.

- The percentage of ads displayed from the database ranged from 30% to 100%

- In the case of some search keywords, more ads seemed to appear late in the evening, thereby disclosing information about time of the day.

- No significant difference was observed in the displayed ads with respect to the user who was logged in.

From these results, it was possible to construct user profiles, predict user locations, days, dates and times, regional festivals and other events.

7. Conclusions

This research has demonstrated how information gleaned from ads on a mobile device can be used to identify users. The methodology can be used to identify a user even if he or she uses the same device, multiple devices, different networks or follows different usage patterns. Also, the methodology can support a digital forensic readiness framework for mobile devices. Additionally, it has applications in context-based security and proactive and reactive digital forensic investigations.

The current research has focused extensively on the Android environment. Future research will examine iOS devices as well as desktop systems, laptops and tablets. Furthermore, the methodology will be extended to link a user across multiple devices and systems. Finally, future research will attempt to track and capture live ads at the network and firewall levels, which could support proactive crime prevention efforts.

References

[1] D. Abalenkovs, P. Bondarenko, V. Pathapati, A. Nordbo, D. Piatkivskyi, J. Rekdal and P. Ruthven, Mobile Forensics: Comparison of Extraction and Analyzing Methods of iOS and Android, Gjovik University College, Gjovik, Norway, 2012.

[2] T. Book and D. Wallach, An Empirical Study of Mobile Ad Targeting, Department of Computer Science, Rice University, Houston, Texas, 2015.

[3] C. Castelluccia, E. De Cristofaro and D. Perito, Private information disclosure from web searches, in *Privacy Enhancing Technologies*, M. Atallah and N. Hopper (Eds.), Springer-Verlag, Berlin Heidelberg, Germany, pp. 38–55, 2010.

[4] C. Castelluccia, M. Kaafar and M. Tran, Betrayed by your ads! Reconstructing user profiles from targeted ads, in *Privacy Enhancing Technologies*, S. Fischer-Hubner and M. Wright (Eds.), Springer-Verlag, Berlin Heidelberg, Germany, pp. 1–17, 2012.

[5] Dimensional Research, The Impact of Mobile Devices on Information Security: A Survey of IT and IT Professionals, San Francisco, California, 2014.

[6] A. Guarino, Digital forensics as a big data challenge, in *ISSE 2013 Securing Electronic Business Processes*, H. Reimer, N. Pohlmann and W. Schneider (Eds.), Springer Fachmedien, Wiesbaden, Germany, pp. 197–203, 2013.

[7] O. Hamoui, Targeting an Ad to a Mobile Device, U.S. Patent Application US20080059300 A1, 2007.

[8] Interactive Advertising Bureau, IAB Platform Status Report: A Mobile Advertising Overview, New York, 2008.

[9] A. Korolova, Privacy violations using microtargeted ads: A case study, *Proceedings of the IEEE International Conference on Data Mining Workshops*, pp. 474–482, 2010.

[10] G. Mohay, Technical challenges and directions for digital forensics, *Proceedings of the First IEEE International Workshop on Systematic Approaches to Digital Forensic Engineering*, pp. 155–161, 2005.

[11] R. Rowlingson, A ten-step process for forensic readiness, *International Journal of Digital Evidence*, vol. 2(3), 2004.

[12] V. Toubiana, V. Verdot and B. Christophe, Cookie-based privacy issues on Google services, *Proceedings of the Second ACM Conference on Data and Application Security and Privacy*, pp. 141–148, 2012.

[13] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang and Z. Chen, How much can behavioral targeting help online advertising? *Proceedings of the Eighteenth International Conference on the World Wide Web*, pp. 261–270, 2009.

[14] J. Yu, You've got mobile ads! Young consumers' responses to mobile ads with different types, *International Journal of Mobile Marketing*, vol. 8(1), pp. 5–22, 2013.