Original Article

Android App Antiforensics

Alberto Ceballos Delgado, Bing Zhou

Department of Computer Science, Sam Houston State University, Huntsville, TX 77340, USA.

Correspondence to: Alberto Ceballos Delgado, Department of Computer Science, Sam Houston State University, 1905 University Ave, Huntsville, TX 77340, USA. E-mail: aac088@shsu.edu


Received: 5 Dec 2021 First Decision: 7 Jan 2022 Revised: 24 Jan 2022 Accepted: 4 Mar 2022 Published: 29 Mar 2022

Academic Editors: Zheng Xu, Pedro Peris-Lopez Copy Editor: Xi-Jun Chen Production Editor: Xi-Jun Chen

Abstract

Aim: Android is one of the most popular platforms in the market. This popularity has led the operating system to be a potential tool for criminal activities. Law enforcement has noted this development and started incorporating smartphone evidence into their cases. However, digital evidence is susceptible to data modification, and thus anti-forensic techniques have been developed to counter forensic investigations. This research investigates the possibility of generating false data using automation techniques.

Methods: A rooted Android device was acquired. The device screen coordinates were mapped using screenshots and Gimp. The coordinates were used to develop a Python script to automate common user tasks such as making a phone call, sending a text message, or adding a contact. These tasks were performed manually and using the automation script. A system image was acquired of the device before and after data population. The images were analyzed using Autopsy and Cellebrite’s Inspector. The forensic artifacts retrieved were compared between the manual and automatic data population.

Results: The artifacts show that the data was added successfully and that forensic tools may not detect that the data was automatically generated.

Conclusion: This research shows that it is possible to populate an Android device with false forensic artifacts using automation scripts. Being able to generate forensic artifacts using automation scripts can allow educators to more easily generate datasets to teach forensic techniques. Additionally, it can also be used by malicious actors to generate false forensic artifacts to mislead an investigator. Future work could improve the proposed data generation technique via machine learning to prevent hardcoding the screen coordinates in the automation script or improve the technique to generate data with old timestamps. Another avenue of future work includes the
development of techniques to identify false data.

Keywords: Anti-forensics, Android, data generation, android antiforensics, mobile data generation

INTRODUCTION

There are many different smartphone operating systems used to access social media platforms, but Android is the dominant operating system in the market\cite{1-5}. Statista\cite{1} reports that Android’s worldwide mobile market share is 71.93\% as of February 2021. Similar to other popular operating systems, Android collects user data to improve its services\cite{6,7}. Active data collection encompasses data collected when the user is actively interacting with Google’s application and websites such as Chrome, PlayStore, Maps, Images, Drive, Sheets, and other applications\cite{8}. Whereas passive data collection includes data acquired from Android devices without actively interacting with Google’s products, such as usage information, device configuration, crash reports, backups, and other types of data\cite{8,9}.

Law enforcement and prosecutors have taken advantage of smartphone data collection practices to aid their investigations\cite{10,11}. Google’s location data has been used by state law enforcement to aid their investigation of murder and potential arson cases\cite{12}. The federal government has also taken an interest in Google’s location data as the FBI has used location data to aid in their investigations on several occasions\cite{13}. Since Google’s location data is more accurate than the traditional cell tower triangulation, reverse search warrants are becoming an increasingly common practice for law enforcement agencies\cite{14}. More recently, Google’s location data has been used to help law enforcement identify perpetrators in the capitol riot on January 6, 2021\cite{15}. Criminals have started developing and employing anti-forensic techniques as a response against law enforcement usage of digital evidence\cite{16-18}. Encrypted devices have been used by criminals to aid in drug trafficking\cite{16,18}. Encryption has also been involved in murder cases, illegal firearm sales, and torture\cite{16,19,20}. Steganography is used by malware developers to hide malicious software\cite{17,21,22}. The anti-forensic technique is also used by spies to covertly share information\cite{23}.

Academicians and practitioners have taken an interest in the field of anti-forensics to assist law enforcement in their investigations. Researchers have focused on the development of new anti-forensic techniques, potential countermeasures, and on the investigation of existing anti-forensic methods\cite{24-27}. Android anti-forensics is an area of ongoing research which has resulted in the development of new anti-forensics techniques\cite{24,25}. However, limited research exists on data counterfeit generation on Android applications. This scenario motivates the hypothesis that it is possible to produce false evidence on Android applications. The following research inquiries are proposed to investigate this hypothesis:

- Can you populate an Android device with false information?
- Are there forensic artifacts left on the device that could be used to identify that the data was generated using the proposed data generation technique?

The research contributions of this paper are two-fold. First, this research investigates the generation of false Android application data as an anti-forensic technique. Second, it establishes a foundation to conduct future investigations on the topic. The paper is structured as follows: the next section discusses existing Android anti-forensics research. Section three presents the methodology and experimental design. Section four discusses the implementation and results. Section five discusses the research conclusions and presents future work.
As a response to criminal usage of anti-forensic techniques, researchers and practitioners have investigated this novel field\cite{24-27}. Research on this topic diverges into different concentrations. Some researchers construct novel anti-forensic techniques\cite{24-26,28}, while others postulate solutions to existing methods\cite{27} and develop comprehensive taxonomies\cite{29}.

Research on data generation focuses mostly on Windows and Android platforms\cite{24,26-27}. Windows data generation involves the usage of virtual machines to generate false data\cite{36} or forensic artifact generation via data injection\cite{37}. However, these techniques only work on Windows environments, these techniques fail to generate older data, and they may leave forensic artifacts that could be used to detect them\cite{36-37}. On the other hand, Android data generation focuses on the usage of Android terminal tools such as the “sendevent” and “getevent” commands\cite{24} or relies on other tools such as SQLite or Android’s “touch” and “date” commands\cite{24,35}. However, these techniques may generate forensic artifacts that could be used to identify them; they are limited and may not work on novel Android devices.

Research on Android anti-forensics focuses on hindering forensic processes or compromising the data’s veracity\cite{26,30,31}. Literature on the topic includes the development of data modification techniques on Android devices with the objective of modifying the data partition and deleting evidence\cite{37}. However, this technique may lead to additional forensic artifacts that could be used for identification and may not work on newer Android devices. Other techniques include the modification of the Android content providers to hamper forensic acquisition\cite{31}. However, this technique ignores data generation and may not work on novel devices. Researchers have also examined the anti-forensics potential of application features such as WhatsApp\cite{30} to delete forensic artifacts. However, this technique ignores data generation. Other researchers have examined anti-forensic techniques to develop detection mechanisms\cite{27}. Existing literature has examined the possibility of detecting database manipulation on Android and iOS devices via a series of forensic artifacts\cite{27}. However, their research focuses mostly on detection and not data generation.

Existing Android anti-forensic research concentrates on the development of anti-forensic techniques\cite{24,26,28}, the evaluation of existing techniques\cite{32,33}, and the development of a comprehensive taxonomy\cite{29}. Research on data generation on an Android device as an anti-forensic technique is both limited and minimal\cite{35-37}. To the best of our knowledge, there is little research on false data generation of Android applications as an anti-forensics technique.

**METHODOLOGY**

The study presented in this article implements a controlled experiment methodology as described by\cite{38}. This research design employs experiments to assess a hypothesis. The hypothesis measured in this experiment is that the artificial creation of realistic Android application forensic artifacts is possible. To test this hypothesis, an Android application data population technique is developed. Artificial data generation is achieved employing Python, Android Debug Bridge, and Android’s input command.

The development and testing environment are illustrated in Table 1. The population technique uses Android’s input command to simulate user’s swipe and touch. Different screen element coordinates are used in conjunction with Android’s input utility to interact with various applications.

Figure 1 details the proposed data population technique. First, each target automation activity is performed manually, such as creating a text message or adding a phone contact. This step is performed to identify smaller tasks required to complete the activity, such as touching a text box, typing the desired message, and clicking the “Send” button. Once all smaller tasks are identified, a device’s screenshot is captured at each
Table 1. Instruments used

<table>
<thead>
<tr>
<th>Hardware &amp; Software</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>Android Debug Bridge (ADB)</td>
<td>v1.0.41</td>
</tr>
<tr>
<td>Python</td>
<td>v3.9.2</td>
</tr>
<tr>
<td>Gimp</td>
<td>v2.10.22</td>
</tr>
<tr>
<td>Autopsy</td>
<td>v4.11.0</td>
</tr>
<tr>
<td>Inspector</td>
<td>v10.4</td>
</tr>
<tr>
<td>Rooted Android Device</td>
<td>SM-G935V with Android 8.0.0</td>
</tr>
<tr>
<td>Dell Desktop Computer</td>
<td>Intel Xeon W-2102 CPU @ 290 GHz, 32 GB RAM, 500 GB Hard Drive. Running Windows 10 Pro</td>
</tr>
<tr>
<td>SQLite DB Browser</td>
<td>v3.12.0</td>
</tr>
</tbody>
</table>

Figure 1. Android App population diagram.

Task using Android Debug Bridge and Android’s screencap utility<sup>[44]</sup>. This is accomplished using the command “adb shell screencap -p <download location on device>”<sup>[44]</sup>. The screenshot is extracted using ADB pull command to a remote computer. The command executed at this step is “adb pull <device location> <remote computer location>”<sup>[44]</sup>. The extracted image file is then examined using image manipulation software, namely GIMP<sup>[40]</sup>, to identify the X and Y coordinates of different screen elements used by the subtask. These coordinates are supplied to Android’s input utility to emulate user’s screen touch. The commands used for this task are “adb shell tap <X coordinate> <Y coordinate>”<sup>[45]</sup>.
Android’s input swipe command is used to simulate long touch and swipe actions. This command also uses the previously mentioned X and Y coordinates. The command used to achieve this task is “adb shell input swipe <X1 coordinate> <Y1 coordinate> <X2 coordinate> <Y2 coordinate> <duration in milliseconds>”\[^{[45]}\]. To simulate long screen taps, the X and Y coordinates are kept constant, and the duration is set to a random value between 1 and 3 s. The X coordinate is kept constant, the Y coordinate changes, and the duration is set to value between 1 and 3 s to emulate vertical screen swipes. The same process is used but with Y constant and X varied for horizontal screen swipes.

Android’s input utility is also used to simulate user typing behavior. A screenshot of the different states of the keyboard is acquired. This screenshot is examined using GIMP\[^{[40]}\] to map each key to a set of X and Y coordinates. A small random delay between 0.5 and 1 s is added to emulate human typing behavior. The aforementioned functions are compiled into a Python script for ease of execution.

A set of rooted and non-rooted emulators were configured with Android API level versions ranging from 16 to 28 to identify Android’s input syntax differences between versions. Using ADB, Android’s input utility was executed on each emulator to determine if the command exists on that API version. If the utility is found, the help function is executed to verify command syntax differences. The previously mentioned process is repeated for all emulators. Table 2 details the results of this evaluation. These experiments confirm that Android’s input command exists on the tested Android API level for both rooted and non-rooted devices. This situation suggests that device population using Android’s input command should work for a large variety of Android versions.

The automation technique is developed using a rooted Android device. The device specifications are detailed in Table 1.

A factory reset is performed on the device to delete all previous data. The device is connected to the internet. A new Google account is created for the device. The following activities are performed on the device:
1. Add a phone contact;
2. Send a text message to a phone number;
3. Call a phone number.

The aforementioned tasks were performed manually, and a system image was extracted after the manual population. The tasks were also performed using the automation script, and a system image was extracted after data population. The forensic artifacts were identified and examined using Autopsy and Cellebrite’s Inspector. The forensic artifacts for each system image were compared to verify the effectiveness of the proposed data generation technique.

The technique used in this research relies on screen coordinates. This technique uses the screen layout and icons to interact with different applications and generate data. The environment used to develop and evaluate the technique is described in Table 1. The technique was evaluated on rooted Android emulators with different API versions described in Table 2. This technique was implemented on a rooted Android device running Android 8.0.0. The technique was evaluated using Autopsy and Cellebrite’s Inspector. For the purposes of this research, all other Android versions and population techniques are considered out of scope.
Table 2. Android input evaluation

<table>
<thead>
<tr>
<th>API level</th>
<th>Rooted</th>
<th>Syntax difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>Yes</td>
<td>Input swipe command does not contain duration</td>
</tr>
<tr>
<td>17</td>
<td>Yes</td>
<td>Input swipe and tap command have an optional [touscreen] and [touchpad] parameter. Swipe command does not contain duration</td>
</tr>
<tr>
<td>18</td>
<td>Yes</td>
<td>Input swipe and tap command contain optional [touchscreen</td>
</tr>
<tr>
<td>19</td>
<td>Yes</td>
<td>Input command has optional source argument. Default source is touchscreen for both tap and swipe command</td>
</tr>
<tr>
<td>21</td>
<td>Yes</td>
<td>Same as 19 rooted</td>
</tr>
<tr>
<td>22</td>
<td>Yes</td>
<td>Same as 19 rooted</td>
</tr>
<tr>
<td>23</td>
<td>Yes</td>
<td>Same as 19 rooted</td>
</tr>
<tr>
<td>24</td>
<td>Yes</td>
<td>Same as 19 rooted</td>
</tr>
<tr>
<td>25</td>
<td>Yes</td>
<td>Same as 19 rooted</td>
</tr>
<tr>
<td>26</td>
<td>Yes</td>
<td>Same as 19 rooted</td>
</tr>
<tr>
<td>27</td>
<td>Yes</td>
<td>Same as 19 rooted</td>
</tr>
<tr>
<td>28</td>
<td>Yes</td>
<td>Same as 19 rooted</td>
</tr>
<tr>
<td>16</td>
<td>No</td>
<td>Same as 16 non-rooted</td>
</tr>
<tr>
<td>17</td>
<td>No</td>
<td>Same as 17 non-rooted</td>
</tr>
<tr>
<td>18</td>
<td>No</td>
<td>Same as 18 non-rooted</td>
</tr>
<tr>
<td>19</td>
<td>No</td>
<td>Same as 19 non-rooted</td>
</tr>
<tr>
<td>21</td>
<td>No</td>
<td>Same as 21 non-rooted</td>
</tr>
<tr>
<td>22</td>
<td>No</td>
<td>Same as 22 non-rooted</td>
</tr>
<tr>
<td>23</td>
<td>No</td>
<td>Same as 23 non-rooted</td>
</tr>
<tr>
<td>24</td>
<td>No</td>
<td>Same as 25 non-rooted</td>
</tr>
<tr>
<td>25</td>
<td>No</td>
<td>Same as 25 non-rooted</td>
</tr>
<tr>
<td>28</td>
<td>No</td>
<td>Same as 28 non-rooted</td>
</tr>
</tbody>
</table>

RESULTS
Forensic examination

The forensic analysis was performed using Autopsy and Cellebrite’s Inspector. Autopsy required a forensic image to analyze, so the image was acquired in a separate step. First, Android Studio was installed on the forensic workstation to acquire Android developer tools to extract the forensic image from the device. Next, the device was powered on, logged in, and Android’s developer mode was activated, as shown in Figure 2.

Next, the device was connected to the forensic workstation as a USB device to be able to execute commands on the device. In the forensic workstation, two command lines prompts were started and navigated to the Android developer tools folder in the Android Studio folder. In the first command prompt, the following command was executed: “adb shell”. Next, using the “adb shell” the following command was executed to identify the “data” partition of the device: “mount | grep /data”. The output of the command was examined to identify the partition that contains the “/data” folder. Next, the “adb shell” was exited, and the following command was executed “adb forward tcp:8888 tcp:8888”. This process is shown in Figure 3.

The “Disk Dump” utility (dd) and the “netcat” tool were installed on the forensic workstation to extract the partition of the Android device. The following command were typed on the second command prompt: “nc 172.0.0.1 8888 > image.dd”. This command extracts the information received in the localhost port into the “image.dd” file. In the first command prompt the following command was typed in the “adb shell”: “dd if=<data partition path> | busybox nc -l -p 8888”. This command extracts the content of the data partition and outputs it to the “netcat” port. The commands in both command prompts were executed at the same time to extract the information from the device, as shown in Figure 4.
This image extraction process was performed before and after the execution of the proposed data generation technique to acquire two forensic images for comparison. Both images were loaded into Autopsy by creating a new case and loading each forensic image file. The “hash calculation” module was executed on each image to compute an MD5 hash of each file. Autopsy’s report feature was utilized to create a comma-separated value (CSV) file with the filenames, file paths, and file hashes to a CSV file for easier analysis. This process is summarized in Figures 5 and 6.
The process was repeated for each forensic image. Afterward, the data in the CSV files were compared using a Python script to identify differences between both images. These differences were then further examined in Autopsy to verify their data contents and confirm their forensic relevance.

Next, Cellebrite’s Inspector was started, and a new case was created. The Android device was connected, and data were acquired from the device using Inspector’s data acquisition features. The forensic artifacts identified in the comparison were verified using Inspector’s data analysis feature, and their forensic relevance was confirmed. The identified forensic artifacts are summarized in Table 3.
The forensic image extraction and analysis process was repeated for both manual data generation and using the technique proposed in this research. The relevant forensic artifacts were compared, and no significant difference was found.

Analysis
The “mmssms.db” is an SQLite database which contains text message data. This database shows the text messages that were sent along with their sent date, receiver, and the message body. The database shows the messages that were sent both with the automation script and manually in the “SMS” table. The “contacts2.db” is an SQLite database which contains the contacts data. The contact information is stored across different tables which include the “raw_contacts”, “contacts”, “data”, “phone_lookup”, “name_lookup”, and other tables. This database shows that the contacts that were added manually and with the automation script were added successfully to the Android device. The “calllog.db” is an SQLite database which contains the call log information. The “calls” table in this database stores information about the calls, which includes its state, duration, number, and other data. This database shows that the automation script was successful in placing a call.
The results of this research suggest that an Android device can be populated with false data. The case study presented in this research using contacts, SMS, and calls shows that it is possible to populate an application with fake data using an automation technique. However, this technique is screen-specific since it relies on screen elements X and Y coordinates. The method will be impacted by layout changes that come with application updates or website updates. The technique proposed in this research does not require a device to be rooted to work correctly as Android’s input and Android’s screencap are both available using Android Debug Bridge. Despite its shortcomings, the technique is able to interact with a variety of Android applications. If a user is able to accurately map the X and Y coordinates of their actions in target applications, they could generate an extended amount of false data. This technique can be implemented using a scripting language like Python to insert delays to emulate human interaction with Android applications.

Regarding the second research question proposed in the introduction, the research results suggest that there are no relevant forensic artifacts that could be used to detect that the data was generated using the proposed technique. The rooted android device was populated with false data using the anti-forensic technique proposed in this research and analyzed using two popular forensic tools: Autopsy and Cellebrite’s Inspector. A forensic analysis was performed based on a list of known forensic artifacts and techniques. This analysis showed that multiple desired artifacts were generated, such as database entries. The forensic investigation did not find any signs of evidence tampering or data generation that would raise suspicion towards the forensic artifact’s validity.

Validation
The proposed technique was compared with existing data generation techniques for Android devices. More precisely, the proposed technique was compared with Ceballos Delgado et al. [35] and Albano et al. [24] techniques. Ceballos Delgado et al. [35] technique utilizes Android Debug Bridge (ADB), Android’s “date”, Android’s “time”, ADB’s “push”, and Android’s “SQLite” utility to populate the device with false data. Differential analysis of two images acquired before and after the usage of their technique shows that the device holds forensic artifacts that could be used to verify the veracity of the data. First, the database journal file does not match the database file. There is also a substantial timestamp difference between the journal file and the database, which can indicate that the database was proctored using the technique proposed by Pieterse et al. [25]. Furthermore, the timestamp modification capabilities of their tool allow the user to modify the affected file timestamps to a time older than the filesystem creation time, which could be used to detect the data generation technique. Additionally, this data generation technique fails to generate events related to user activities such as event logs since it modifies the database directly. The technique proposed in this research generates said event logs since it recreates the actions performed by the user via interaction with the user interface. While this technique is unable to generate data with an older timestamp, it maintains the consistency between journal file and database files which reduces detection probability using Pieterse et al. [25] technique.

The data generation technique proposed by Albano et al. [24] uses Android’s “sendevent” and “getevent” command to interact with the user interface and generate false data. Their technique works by mapping the events related to a specific activity in an application using the “getevent” command and repeating them using the “sendevent” command to recreate the activity. The result is that the device generates the events, which emulate user activities. However, their proposed technique requires a significant number of events and parameters to generate a single activity. Furthermore, their tool is limited to a small set of automation activities, and their research lacks replication details on the forensic examination which hinders replication. The technique proposed in this research can generate a diverse set of automation activities depending on the screen coordinates of the activity to replicate. This proposed technique requires less information to
recreate an emulation activity than Albano et al.\textsuperscript{[24]} method. Lastly, this research provides more details on the forensic examination to aid experiment replication.

Other data generation techniques focus on Windows platforms such as Scanlon et al.\textsuperscript{[37]} and Du et al.\textsuperscript{[36]} methods. Their proposed techniques are more versatile and can generate a wide variety of false data. However, their techniques are unable to generate data on Android devices, which the technique proposed in this research can accomplish.

**CONCLUSION AND FUTURE WORKS**

The research presented in this document shows an investigation of Android application anti-forensics. The software developed in this research can populate an Android application with false data indistinguishable from real data. The software uses Android Debug Bridge and Android’s input to simulate screen touches and interact with a target application. The X and Y coordinates used by Android’s input are acquired using Android’s screencap and an image editor. The proposed population technique is evaluated using contacts, call logs, and SMS as a case study. A forensic analysis was performed on a rooted Android device. A variety of general and application-specific artifacts were retrieved and analyzed. The forensic investigation was not able to detect false data generation.

The scope of this research is limited to Android devices, all other operating systems are considered out of scope. This research evaluates data generation of text messages, phone calls, and phone contacts as an anti-forensic technique, other possible techniques are considered out of scope.

A potential improvement in this research would be the automation of screen coordinate identification. Image recognition technology could be applied to detect the different elements’ screen coordinates and update them automatically. Thus, addressing one of the shortcomings of the application. Additionally, this improvement would reduce the time required to automate false data generation and improve the technique’s efficiency. Another possible future avenue includes the automatic detection of artificially generated data. This could be accomplished using machine learning models and event log timestamp differences. Lastly, a potential avenue for future research is the generation of false data with older timestamps.

**DECLARATIONS**

**Authors’ contributions**
Made substantial contributions to conception and design of the study and performed data analysis and interpretation: Ceballos Delgado A
Provided administrative, technical, and material support: Zhou B

**Availability of data and materials**
Not applicable.

**Financial support and sponsorship**
None.

**Conflicts of interest**
Both authors declared that there are no conflicts of interest.
Ethical approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Copyright
© The Author(s) 2022.

REFERENCES
4. NetMarketShare. Operating System Market Share. Mobile 2019-11 to 2021-02. Available from: https://netmarketshare.com/operatingsystem-market-share.aspx?option=%7B%22filter%22%3A%7B%22%22id%22%3A%22Operating+System%22%2C%22%22type%22%3A%22Mobile%22%7D%22%22type%22%3A%22Operating+System%22%7D%22%22dateLabel%22%3A%22Trend%22%7D%22%22attributes%22%3A%22%22%2C%22%22group%22%3A%22%22%2C%22sort%22%3A%22%22%7D [Last accessed on 28 Mar 2022].


