Extraction and Analysis of WhatsApp Database

Anish Dassi[1] Viola Gupta[2] Varun Goel[3]
[1,2,3]MAIT, GGSIPU

Abstract— This paper seems to explore the safety and security of our WhatsApp, with over 1.4 Billon people using WhatsApp to stay connected with each other, we try to look at how secure WhatsApp really is along with how you can take some extra steps to ensure the safety of your data. WhatsApp stores conversations of the user in database files as a backup in the internal storage of the mobile device. Security experts at WhatsApp implemented AES, using which the stored backup files prevented or restricted unauthorized access. Using the results discussed in this paper, any person will be able to reconstruct the list of contacts and the chronology of the messages that have been exchanged by users. Furthermore, thanks to the multiple research papers, studies and forum that helped me complete this.

Key words: WhatsApp, Database

I. INTRODUCTION

With the swift evolution in smartphone technologies, the industry has seen an abrupt surge in the number of applications available to a smartphone user. While there are a lot of applications that cater to every one's needs, social networking applications occupy a massive share in the category of users who use the app on a regular basis. Instant messengers available in the application stores are capable of sending a wide range of messages with no limits on message lengths and also allow you to share multimedia like images, videos, location etc. All these features are provided at virtually no cost and any person can avail this by connecting their smartphone to the internet using mobile data plans or Wi-Fi connection.

WhatsApp is one such application that is the leader amongst social networking and instant messenger applications. WhatsApp is cross-platform instant messenger service that has over 1.4 billion users and continues to grow exponentially. It was acquired by Facebook in 2014 which has made many new changes to the set of features that WhatsApp already provides.

WhatsApp stores conversations of the user in database files as a backup in the internal storage of the mobile device. WhatsApp automatically takes backups every day and stores it in the WhatsApp folder, in internal storage of the Android Smartphone. The backups are maintained for a week or so after which the oldest backup gets overwritten by the new backup. The chat backups are stored in a SQLite Format Database named msgstore.db in sdcard/WhatsApp/Databases

Earlier the databases were stored in a non-encrypted plain text format making it considerably easy for hackers to exploit database files and extract personal conversations from them. Considering this a serious vulnerability, security professionals at WhatsApp implemented a security mechanism by encrypting the backup files using AES encryption algorithm to prevent unauthorized access. From a forensic investigation perspective, WhatsApp may contain huge amount of important data that could be used in the court as evidence. This project proposes a step by step process to acquire the encrypted backup files stored on the Android devices and decrypt them to obtain stored backup conversations from a suspect Android device and parse them in human readable format.

II. RELATED WORK

The forensic analysis of messaging applications on smartphones has been the subject of various works published in the literature. However, our contribution

1) Has a much wider scope, as it considers all the different parameters generated by WhatsApp (such as, contacts database, the log files, the WhatsApp status, and the even deleted messages)
2) (b) Presents a more complete and thorough analysis of these artifacts
3) (c) Detailed explanation on how these artifacts can be used in combination to deduce various type of information having an evidentiary value, such as whether a message has been actually delivered to its destination after having been sent, if a user joined or left a group chat before or after a given time, etc.

Furthermore, the information stored into the chat database is analyzed part by part.

The works of [1] and [2] are similar to ours, since they both focus on the forensic analysis of WhatsApp Messenger on Android. However, these studies focus mainly on the acquisition of the details left by WhatsApp Messenger, and deal with their analysis only in part (they limit their study to the chat database, and analyze it only partially). Similar considerations apply to the WhatsApp Xtract tool [3], that extracts some of the information stored into the chat database (and, possibly, in the contacts database), without however providing any description of how these databases are parsed.

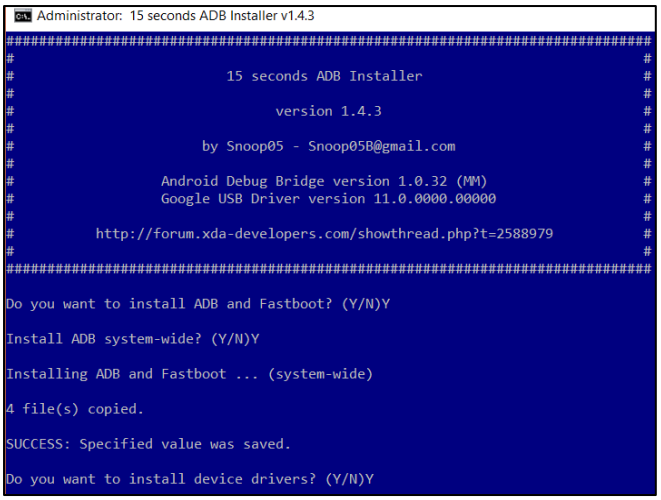
III. SOFTWARE USED ANALYSIS

Let's go one by one, analyzing why we need these different software.

Since we will be using a rooted Android phone, let's go through the software needed to root the phone we have at hand.

A. ADB (Android Debug Bridge):

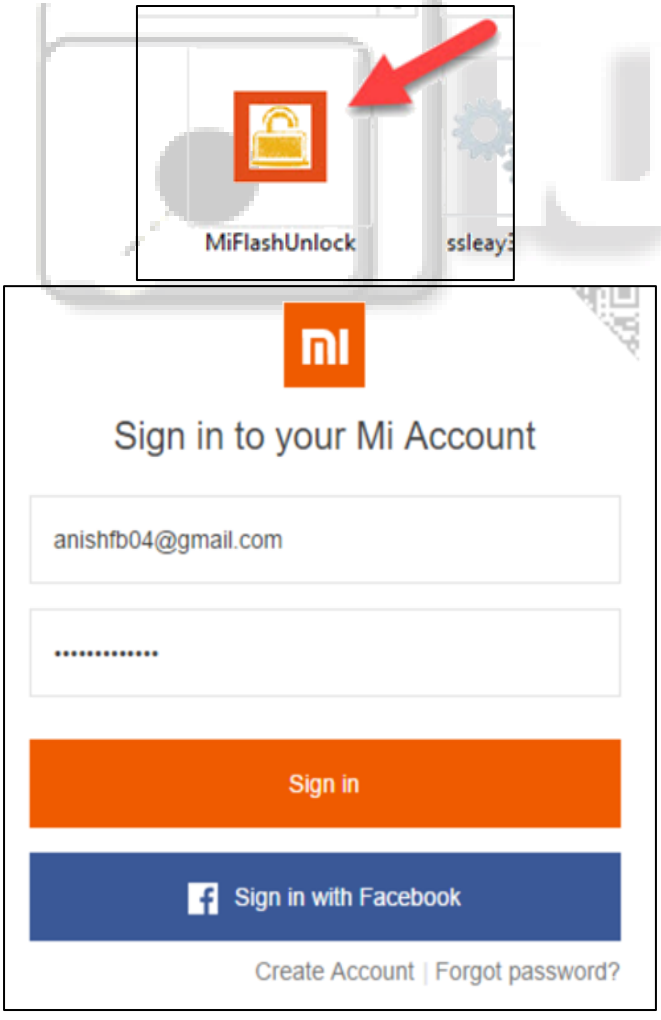
It is a command-line utility included with Google's Android SDK. ADB can control your device over USB from a computer, copy files back and forth, install and uninstall apps, run shell commands, and more.



B. MiFlash Unlock (This unlocker will differ from one manufacturer to other):

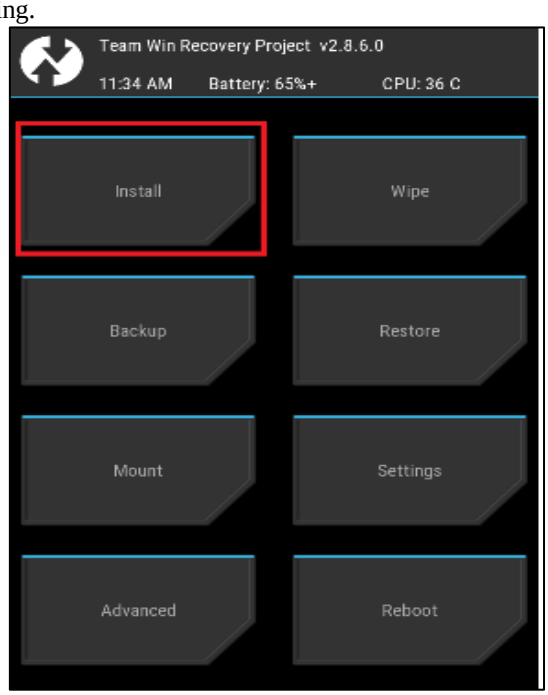
Official Xiaomi Software to unlock Android Bootloader. It is available on the Xiaomi Website.

This is the first step in the rooting process and without unlocking bootloader, we cannot flash twrp recovery & we need twrp recovery for Flashing SuperSU file. Each smartphone manufacturer has their own set of steps that one needs to follow in order to unlock the phone's bootloader.



C. twrp.img & SuperSu Zip File:

Once twrp gets installed on phone, it helps us in flashing our smartphone with the SuperSu zip file which gives us full administrator access to our phone, otherwise called as rooting.



IV. IMPLEMENTATION

The studies shown in this paper have been performed by being carried out in a controlled environment, each one of them referring to a specific usage scenario.

Most of the important files generated by WhatsApp are stored in the area of the internal device memory that is normally inaccessible to users.

We start to explore this area by first unlocking the android phones' bootloader. Bootloader, is like BOIS to windows. It is the first thing that runs when you boot up your Android device. It packages the instructions to boot operating system kernel. Most bootloaders are locked, as well as encrypted, including the developer-friendly Nexus devices. Manufacturers usually lock the bootloader to make sure people stick to their Android OS version specifically designed for the device.

Once we have unlocked our bootloader, we move on to installing Team Win Recovery Project (TWRP) which is an open-source software custom recovery image for Android based devices, providing a touchscreen-enabled interface that allows users to install 3rd party firmware , and also functions that are often unsupported by stock recovery images. It is, therefore, often installed when rooting Android devices

Using the SuperSU root file, specific to the device we're trying to root, we gain root access to the phone. Root access basically means gaining admin privileges. Android uses permissions (Linux-based permissions, to be exact) in the file structure. Every file, every folder and every partition has a set of permissions. These permissions decide who can read a file (look at or access the contents without changing them), write to a file (be able to change the contents of that file, or create a new file inside a folder or partition) and execute a file (run the file if it's a type that can run, like an

app). Once a user has root access, he gains access to do almost everything on a smartphone.

After this, we connect our smartphone to our remote computer on which we have adb (Android Debug Bridge) installed. Using adb, we use its command line to copy the WhatsApp database and its key to our SD card and then download it onto our computer.

The encryption method being used in the encryption of WhatsApp database is AES with a key (K) length of 256 bits and an initialisation vector (IV) size of 128 bits.

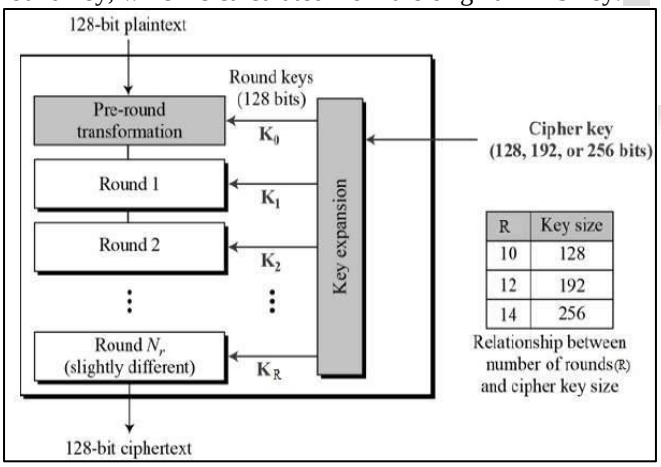
The IV or the initialization vector is saved from offset 0x33 till 0x42 in the crypt12 file (WhatsApp stores its local database in this format). The IV value will be different for every crypt12 file.

AES is an iterative rather than Feistel cipher. It is based on 'substitution–permutation network'.

It comprises of a series of linked operations, some of which involve replacing inputs by specific outputs (substitutions) and others involve shuffling bits around (permutations).

AES algorithm performs all its computations on bytes rather than bits. Hence, AES treats the 128 bits of a plaintext block as 16 bytes. These 16 bytes are arranged in four columns and four rows for processing in the form of a matrix.

The number of rounds in AES is variable and depends on the length of the key. AES uses 10 rounds for 128-bit keys, 12 rounds for 192-bit keys and 14 rounds for 256-bit keys. Each of these rounds uses a different 128-bit round key, which is calculated from the original AES key.



Next we used a code, written using python that took in the WhatsApp key and encrypted database as input and then gave us the output in decrypted .db file.

The way our code worked is simple, we used the help of inbuilt packages of Python, mainly Pycrypto and PyCryptodome, which have inbuilt methods to handle AES encryption and decryption.

This file is still not in user readable form so we will use the SQLite Browser to view and analyze our database.

Now for traversing this huge amount of data and be able to selectively use this data for our analysis, we will sql to extract data search results.

V. CONCLUSION

Using internet and network are increasing rapidly. Everyday a lot of digital data have been exchanging among users. Some of data is sensitive that need to protect from intruders. Encryption algorithms play vital roles to protect original data from unauthorized access. Various kind of algorithms are exist to encrypt data. Advanced encryption standard (AES) algorithm is one of the efficient algorithm and it is widely supported and adopted on hardware and software.

This paper looks to make people understand about how the security in the most used mobile app in the world functions, the ways around it and also how can one protect themselves to such attacks.

From a forensic perspective, WhatsApp contains volumes of evidentiary data that could be used as evidence in a court trial. Therefore it is very crucial to have a methodology to be able to parse these encrypted databases in human readable format.

REFERENCES

[1] A. Mahajan, M.S. Dahiya, and H.P. Sanghvi. Forensic Analysis of Instant Messenger Applications on Android Devices. International Journal of Computer Applications, 68(8), April 2013.
[2] N.S. Thakur. Forensic Analysis of WhatsApp on Android Smartphones. Master's thesis, University of New Orleans, 2013. Paper 1706.
[3] WhatsApp Xtract (v. 2.1) by Fabio Sangiacomo and Martina Weidner., May 2012. Available at https://code.google.com/p/hotoloti/downloads/list.

