FORENSIC ANALYSIS OF WECHAT ON ANDROID SMARTPHONES

Songyang Wu, Yong Zhang, Xupeng Wang, Xiong Xiong*, Lin Du

Presented: Negamiye Arlene
2017 Nov 6th
Content

- Introduction
- Related Works
- WeChat Forensics
- Experiments and evaluation
- Conclusion
- Acknowledgements
- References
I. Introduction

• WeChat is one of the most popular instant-messaging smartphone applications in the world. Through the internet, users of WeChat can communicate with each other using the multimedia messages including texts, images, voices and videos.

• The study of WeChat forensics has become increasingly important. Specifically, WeChat can be used as means of communication for criminal activities, and gangs may even use its abundant social functions to organize and coordinate their criminal acts like selling illegal items, defrauding, disseminating pornographic material to children.
WeChat's rise

Data: Tencent
Chart: TECHINASIA
The following pictures (Fig. 12) show the two basic features of WeChat. Fig. 1(a) is the chat screen and Fig. 1(b) is the Moments where users share their life with friends. We use the term “scene” to represent the contexts of the chat conversation displayed as that of Fig. 1.
This paper explores the common questions that arise during investigating WeChat on Android devices including:

1) how to acquire the data of WeChat and how to decode the encrypted database;

2) who did the user communicate with and what was said,

3) what the user was sharing with the Moments.

We also provide useful methods to address common challenges of WeChat forensics including conveniently backing up user data from unrooted Android devices and sufficiently recovering the scene of conversations.
2. Related works

• Most studies of Android forensic technology focus on the acquisition and analysis of smartphone data obtained from flash memory and RAM.

• The necessary background knowledge, technology and operational method were expounded by Hoog (2011), who provided excellent guidelines for forensic workers. (Android Forensics: Investigation, Analysis and Mobile Security for Google Android. Syngress Publishing.)

• The forensic analysis of applications is not their focus. The forensics of applications still requires further study.

• Recently, an increasing amount of literature has considered the forensics of applications on Android smartphones, different studies has been done on a specific SNS application but each application requires its own unique forensic.
Zhou et al. (2015) and Silla (2015) studied the forensic technology of WeChat specifically.

They managed to extract an encrypted and deleted chat history on WeChat tool was carefully checked by extracting logical image ten times.

In their conclusion: out of the two tested tools, ADB recovered all the shared media and downloaded documents files with time stamps.

More importantly, this paper proposes a method for recovering the entire chat scene, and a backup acquisition method for conveniently extracting WeChat data in unrooted cases, which are issues that the above-mentioned schemes did not address well.

According to recent product release notes published on the official website, Cellebrite UFED v5.0 supports decoding and parsing the communications of Android WeChat better than previous versions.
3. WeChat forensics

- **Installation paths and data acquisition**
- **Decrypting the messages database**
- **Communication records**
- **Moments**
- **Conversion of audio file format**
3. WeChat forensics

- **Tools**: Apktool, dex2jar, JD-GUI, etc.
- The focuses of inverse analysing of WeChat including:
  - Extracting the information about required permissions
  - The components and entry points of the application
  - Identifying the key implementation including processes of encryption
  - The generation of encryption keys and the data-storage schemes under the help of decompile tools such as BakSmali or JD-GUI

To protect the privacy of users, WeChat encrypts the database of chat messages with SQLCipher. In this section, we start with:
- the data acquisition of WeChat data from Android devices,
- The next subsection addresses decoding of the encrypted chat messages database.

The remainder of this section address the most concerned problems of the investigators who conduct a digital forensic examination on an instant messaging application:
1) who did the user communicate with and what was said, and
2) what the user was sharing with the Moments.
3. a) Installation paths and data acquisition

- WeChat places the application paths "/data/data/com.tencent.mm/" and "/sdcard/Tencent/MicroMsg" on Android device.

- The data such as chat records, configurations generated during the running of WeChat is stored in three subdirectories they are "databases", "shared_prefs" and "MicroMsg".

- The **databases** and **shared_prefs** directories cache data such as user authentication.

- The **MicroMsg** directory store important users' data of activities such as received images, audio files, etc.
3. a) Installation paths and data acquisition (cont)

- The most critical evidence sources under certain user folder are listed as follows:
  - /data/.../<udir>/EnMicroMsg.db. The encrypted SQLite database of chat messages.
  - /data/.../<udir>/SnsMicroMsg.db. SQLite database of Moments.
  - /sdcard/.../<udir>/image2/. Raw pictures relating to the image messages.
  - /sdcard/.../<udir>/voice2/. Raw audio files relating to the voice messages.
  - /sdcard/.../<udir>/video/. Raw videos relating to the video messages.
  - /sdcard/.../<udir>/WeiXin/. Multimedia files (including images and videos) that are downloaded from Moments (through the command “Saved to phone”).
- The acquisition of digital evidence from Android smartphones is different for rooted or unrooted devices. For rooted devices we can use the Android Debug Bridge (adb) command and for unrooted device unrooted backup method need to be used.
3. b) Decrypting the messages database

- EnMicroMsg.db is the SQLite database of the user's chat messages and is encrypted using the SQLCipher.
- We can identify, through analysing the decompiled code of WeChat APP, that the decryption key is calculated from the International Mobile Equipment Identity (IMEI) of the smartphone and the uin of the current WeChat user as follows:
  \[ \text{dec_key} = \text{Left7(MD5(IMEI)+uin)} \]
  The data of IMEI and UIN can be extracted from CompatibilityInfo.cfg and system_config_prefs.xml.
- To decrypt the database file, we just need to compute plaintext of each 4 KB size block of the encrypted file using the dec_key.
- The uin of the user is a critical element for computing the decryption key.
- In the case of multi-account on the same smartphone, it can take more than 48h and more than 100 GB of storage to acquire the date of other users.
3. c) Communication records

- WeChat scene often contains multimedia information, and the messages of images, emojis or voices during a chat session often convey concrete meanings like the text messages.
- All conversation records of the user are stored in the data table “message” of the database EnMicroMsg.db.
- The data fields of each message record that are interest in forensics include “talker”, “create time”, “type”, “content”, “imgPath” and “isSend”. The “talker” field stores the WeChat account whom the user communicates with, his or her detailed information is stored in the data tables of “userinfo” and “rcontact”.

It is obviously that a complete recovery of chat scene gives digital investigators a better understanding in meanings of communications.

Here we explain the detail processes of retrieving the multi-media resources embed in the conversation as follows.

<table>
<thead>
<tr>
<th>RecNo</th>
<th>msgId</th>
<th>talker</th>
<th>content</th>
<th>type</th>
<th>creatTime</th>
<th>imgPath</th>
<th>isSend</th>
</tr>
</thead>
<tbody>
<tr>
<td>39</td>
<td>40</td>
<td>wxid_i17s4u02q3ka12</td>
<td>oh my God, have you eaten this?</td>
<td>1</td>
<td>1462518990000</td>
<td>(null)</td>
<td>1</td>
</tr>
<tr>
<td>40</td>
<td>41</td>
<td>wxid_i17s4u02q3ka12</td>
<td>Where did you buy it?</td>
<td>1</td>
<td>1462519043000</td>
<td>(null)</td>
<td>1</td>
</tr>
<tr>
<td>41</td>
<td>42</td>
<td>wxid_i17s4u02q3ka12</td>
<td>Yes</td>
<td>2</td>
<td>1462519055000</td>
<td>(null)</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>43</td>
<td>wxid_i17s4u02q3ka12</td>
<td>&lt;xml version=&quot;1.0&quot;/&gt;</td>
<td>3</td>
<td>1462519066000</td>
<td>THUMBائن_DIRPATH://th_dbb5e4622e87f85226c8da6893698fc0</td>
<td>2</td>
</tr>
<tr>
<td>43</td>
<td>44</td>
<td>wxid_i17s4u02q3ka12</td>
<td>Nightclub</td>
<td>2</td>
<td>1462519184000</td>
<td>(null)</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td>45</td>
<td>wxid_i17s4u02q3ka12</td>
<td>wxid_zjvs90v9xyat12:522:3515:0</td>
<td>1</td>
<td>34 1462519228892</td>
<td>171511050616bf5274d86bf101</td>
<td></td>
</tr>
<tr>
<td>45</td>
<td>46</td>
<td>wxid_i17s4u02q3ka12</td>
<td>You know, this is illegal.</td>
<td>2</td>
<td>1462519294000</td>
<td>(null)</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>47</td>
<td>wxid_i17s4u02q3ka12</td>
<td>I know! I will not tell others! [囧]</td>
<td>1</td>
<td>1462519434000</td>
<td>(null)</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td>48</td>
<td>wxid_i17s4u02q3ka12</td>
<td>hello?</td>
<td>1</td>
<td>1462519489000</td>
<td>(null)</td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>49</td>
<td>wxid_i17s4u02q3ka12</td>
<td>You can come to my house.</td>
<td>2</td>
<td>1462519595000</td>
<td>(null)</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>51</td>
<td>wxid_zjvs90v9xyat12</td>
<td>wxid_zjvs90v9xyat12:212188:1</td>
<td>2</td>
<td>34 1462519604000</td>
<td>4515260506160df30cada1101</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>52</td>
<td>wxid_i17s4u02q3ka12</td>
<td>You are now at home?</td>
<td>1</td>
<td>1462519622000</td>
<td>(null)</td>
<td></td>
</tr>
<tr>
<td>51</td>
<td>53</td>
<td>wxid_zjvs90v9xyat12</td>
<td>&lt;xml version=&quot;1.0&quot;/&gt;</td>
<td>2</td>
<td>3 1462519623000</td>
<td>THUMBائن_DIRPATH://th_47a728be5e0c2a1e8905eb80b42b7ce9</td>
<td></td>
</tr>
<tr>
<td>52</td>
<td>54</td>
<td>wxid_zjvs90v9xyat12</td>
<td>wxid_zjvs90v9xyat12:16:0</td>
<td>2</td>
<td>43 1462519660000</td>
<td>1527410605160df30c19261</td>
<td></td>
</tr>
</tbody>
</table>
• For an image:
  file_path= <uDir> + "/image2/" + substr(S1,2,3) + "/" + substr(S1,6,7) + "/th_"+s1)
• For an audio:
  file_path= <uDir>+ "/voice2/" + substr(S1,0,1)+"/"+ substr(S1,3,4) + "/msg_" + st "".amr";
• For a video:
  file_path= <uDir>+"/video1/"+S+".pm4."

After determining the specific storage paths of the multimedia files, the chat scene can be recovered successfully. Fig. 4 represents the investigation result of the communication with the talker “wxid_i17s4u02q3ka12” (described in Fig. 3) using our developed forensics tool.
3. d) Moments

• The Moments is a community where users share their life with friends.
• Messages of Moments are stored in the SnsMicroMsg.db database without encryption.
• The major data tables focused in forensics include “SnsInfo” and “SnsComment”.
• The SnsInfo table stores the Moments messages, including texts and the link of multimedia files (images or videos).
• The SnsComment table stores the associated comments of the sharing message.
The major focused data fields include “userName”, “createTime” and “content”. The “user-Name” field indicates the owner of the sharing message, and “content” is the data of sharing message stored as a binary large object (BLOB), as shown in this picture.
In a Type Length Data structure, see the fig up. the first byte specifies the **type of data** content, the second byte indicates the **data length**, and the third part stores the **data content** itself. The detailed format of the content field of WeChat Moments can be depicted after analysing the BLOB data object.

As shown in this fig, key elements of the content field are “msgOwner”, “msgContent”, “msgResID”, “msgImagePath” and “msgImagePath2”, most of them are stored in TLD structure. The msgContent is the text of Moments message, msgOwner is the user who post the message. msgResID is an identity of the multimedia resource with 20 bytes length.
The “msgImagePath” value of the content.

The URL path of the uploaded multimedia file is stored in the msgImagePath field (Next Fig), msgImagePath2 contains the URL of the thumbnail of the uploaded multimedia file.
The multimedia resources can be acquired from the WeChat server after extracting the URL of the multimedia file from the msgImagePath field.
3. e) Conversion of audio file format

- Audio files of WeChat with the “.aud” extension use a customized format slightly modified from the standard formats of AMR or SILK_v3 (free and provided by skype).
- Audio files can be decoded and played through the common decoder such as FFmpeg package.
- In order to play an audio the file need to be converted.
- Since PCM audio is similar to WAV audio, the PCM audio file can be decoded and played by common audio players by adding a WAV file header.
4. Experiments and evaluation

- Experiment setup

- Test data come from artificial production.
- Major forensic functions including data acquisition, decryption, and communications investigation were tested.
- Data acquisition depends mainly on the specific Android device and the version of WeChat, whereas decryption and communications investigation depend only on the WeChat versions.
- WeChat versions 5.0 to 6.3 were installed on different brands Smartphone.
- Workstation with an Intel Core i7 CPU 2.4 GHz and 16 GB RAM.
## Experimental results

### Table 1
Test results of data acquisition.

<table>
<thead>
<tr>
<th>Smartphone</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.0</td>
</tr>
<tr>
<td>Xiaomi 4C</td>
<td>✓</td>
</tr>
<tr>
<td>Samsung Galaxy Note4</td>
<td>✓</td>
</tr>
<tr>
<td>LG G4</td>
<td>✓</td>
</tr>
<tr>
<td>HTC One</td>
<td>✓</td>
</tr>
<tr>
<td>Moto X</td>
<td>✓</td>
</tr>
<tr>
<td>Samsung Galaxy S6 Edge</td>
<td>✓</td>
</tr>
</tbody>
</table>

### Table 2
Test results of decryption and forensics functions.

<table>
<thead>
<tr>
<th>Function</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5.0</td>
</tr>
<tr>
<td>Decryption</td>
<td>✓</td>
</tr>
<tr>
<td>Text message</td>
<td>✓</td>
</tr>
<tr>
<td>Image message</td>
<td>✓</td>
</tr>
<tr>
<td>Voice message</td>
<td>✓</td>
</tr>
<tr>
<td>Video message</td>
<td>✓</td>
</tr>
<tr>
<td>Moments</td>
<td>✓</td>
</tr>
</tbody>
</table>
Experimental results (Cont)

Table 3
Inconsistencies after degrading WeChat v6.3.27 to v6.0

<table>
<thead>
<tr>
<th>Files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified</td>
</tr>
<tr>
<td>Removed</td>
</tr>
<tr>
<td>Emerged</td>
</tr>
</tbody>
</table>

- Data acquisition approach was tested on six devices, and the data of different WeChat v5.0 e v6.3.27 were acquired successfully.
- The test results of forensic functions including decrypting messages database, parsing communication records and investigating Moments are shown in Table 2, which indicate that our method discussed in this paper are practical for investigating Android WeChat from v5.0 to v6.3.27.
Experimental results (Cont)

- We also tested the WeChat (v6.0.0.54 and v6.3.27) forensics on two emulators, Genymotion and BlueStacks.
- A naturally arisen question is that whether or not the operation downgrading the WeChat to version 6.0 causes any loss of data via the adb pull command with root privilege to that acquired through the “unrooted backup method”
- There are 9 files were modified and 3 files were removed. Fortunately most of extracted files including the important EnMicroMsg.db, SnsMicroMsg.db, etc. were still intact.
- In the future work, more tests on Android v6.0 and v7.0 are required and the “unrooted backup method” is also need to be improved to meet the coming changes of WeChat and Android system
5. Conclusion

• In this paper, we explored several common questions that arise in forensic examinations of Android WeChat including:
  1. Acquisition of the user data and decoding the encrypted messages database;
  2. Investigating the communication
  3. What the user was sharing with the Moments.
• Provide corresponding technical methods that answer to the above questions.
• As far as we known, few literature that analyses forensic of WeChat in detail like this paper.
5. Conclusion (cont)

• This study can provide significant references for investigators and researchers of digital forensics.
• WeChat is updated occasionally, means possible changes in the data storage structure and the data protection measures. Therefore the reverse analysis of Android applications, and the data-protection mechanisms, should be studied continuously to satisfy the new requirements of digital investigation.
• Future works will include more careful tests of “unrooted backup method” using Android v6.0 or later and the coming new versions of potential data volatile
• The unrooted backup method is also need to be improved to meet the new coming changes of WeChat and Android system.
References

Forensic analysis of WeChat on Android smartphones
By: Songyang Wu, Yong Zhang, Xupeng Wang, Xiong Xiong*, Lin Du


WeChat’s rise: Monthly active users
https://www.techinasia.com/wechat-near-billion-users
QUESTIONS?
THANK YOU