FORENSIC ANALYSIS OF WECHAT ON ANDROID SMARTPHONES

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

Presented: Negamiye Arlene

2017 Nov 6th

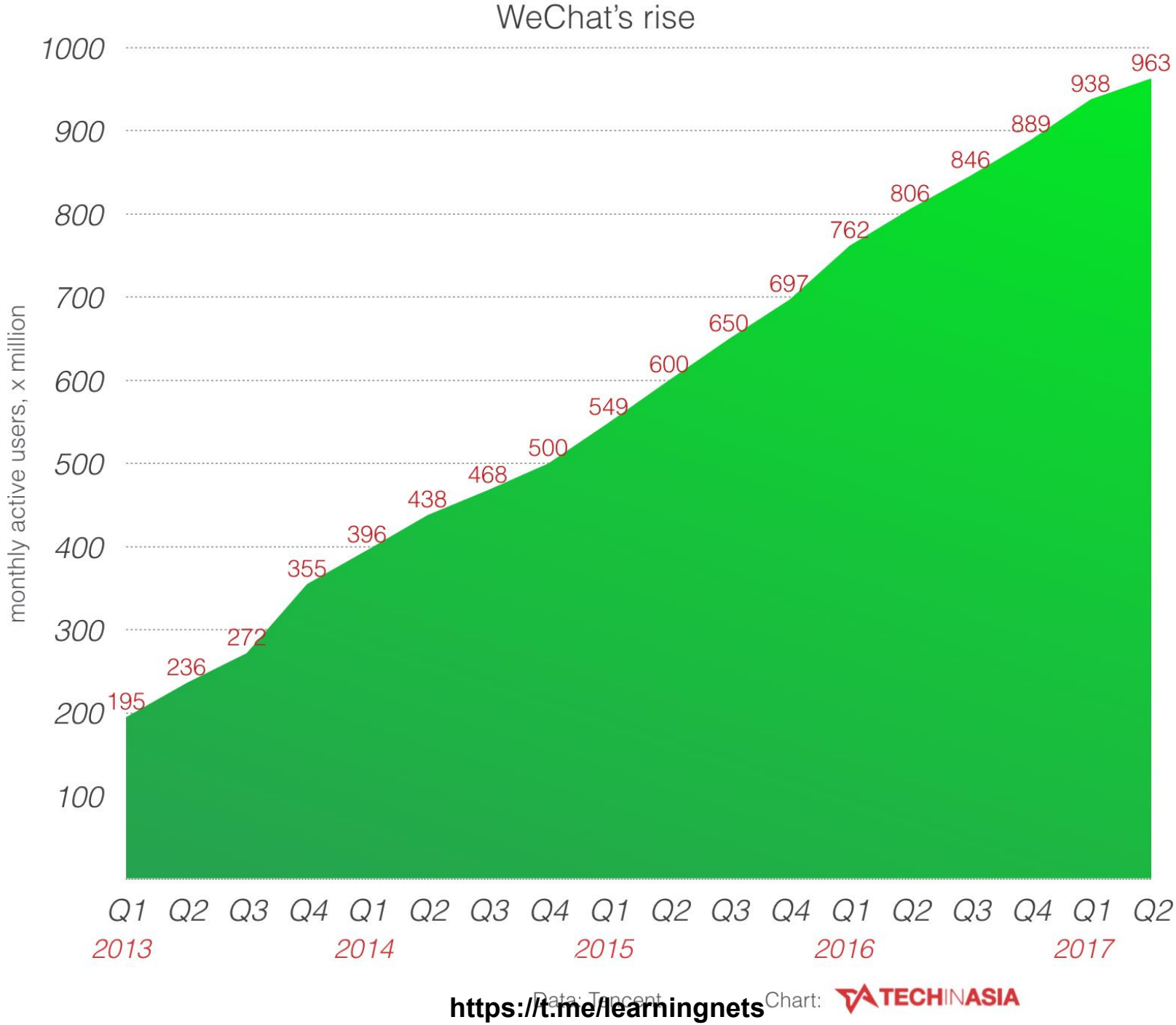
Content

Social Media logo

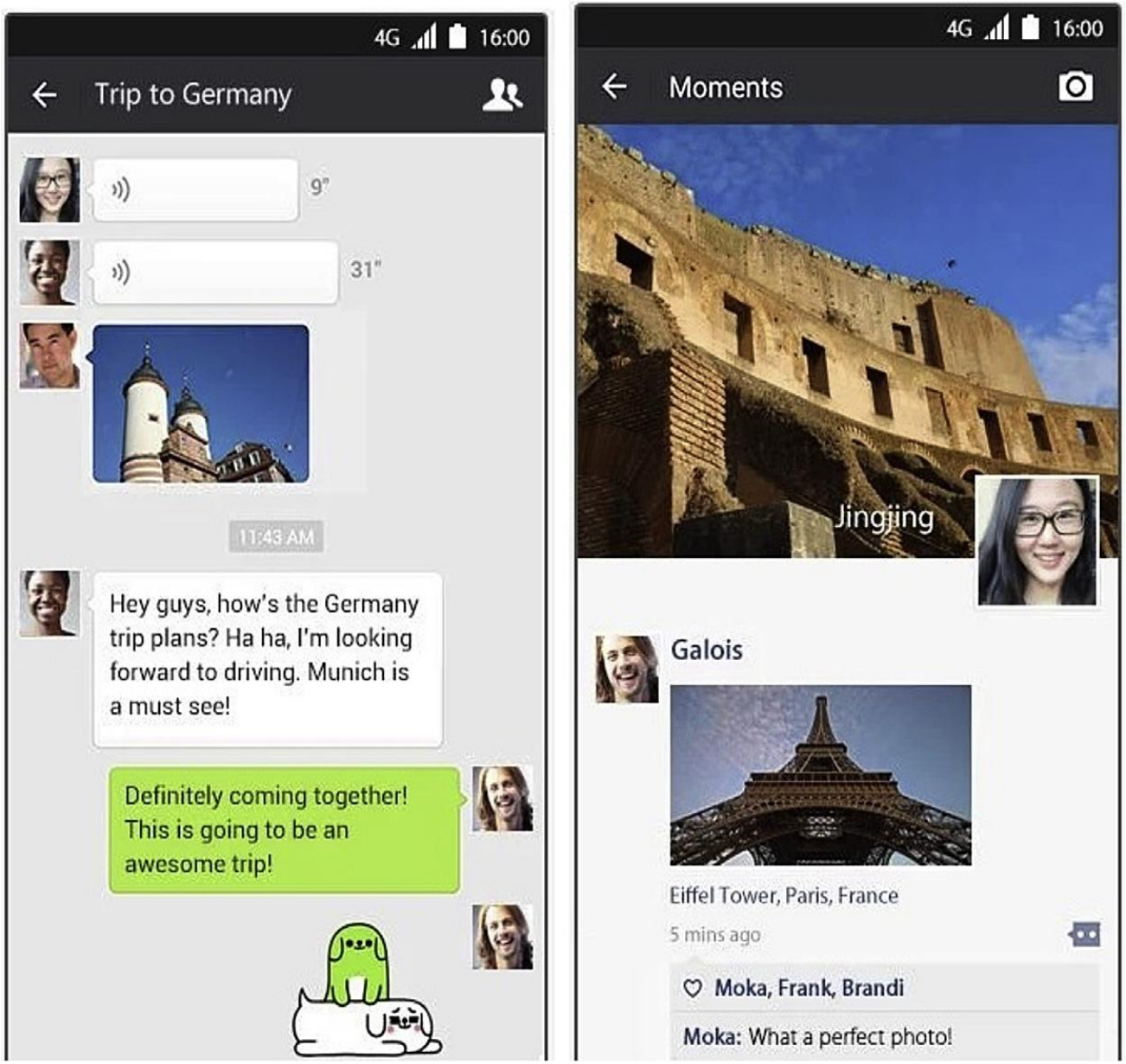
I. Introduction

- 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.



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.



(a) The chat screen (b) Moments

1. Introduction (cont)

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 F 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.

2. Related works(cont)

- 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

a	• Installation paths and data acquisition
b	• Decrypting the messages database
c	• Communication records
d	• Moments
e	• 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/…/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.

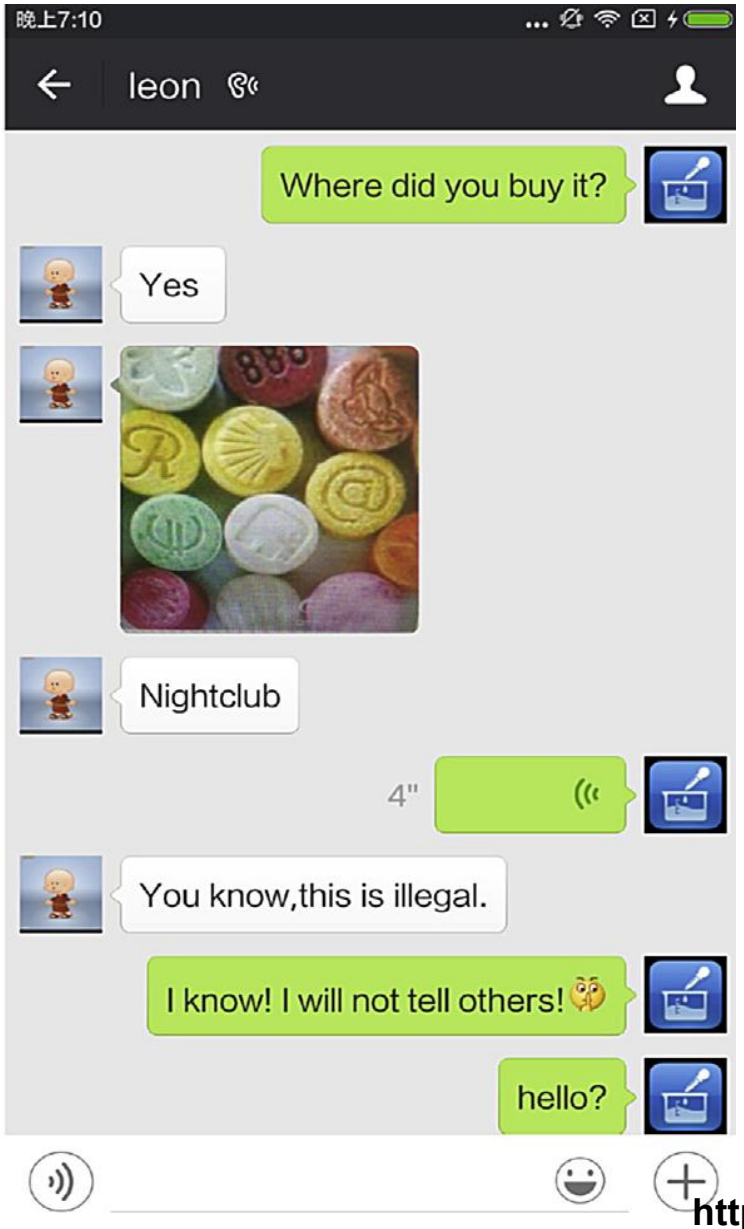
3. b)Decrypting the messages database

- EnMicroMsg.db is the SQLite database of the user's chat messages and is encrypted using the SQLCipher
- We can identity, 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 current WeChat user as follows:

  dec_key = Left7(MD5 (IME)+ uin))

  The data of IME and UIN can be extract from CompatibilityInfo.cfg and system_config_prefs.xml.
- To decrypt the database file, we just need to computer plaintext of each 4 KB size block of the encrypted file using the dec_key.
- The uin of the user is 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”.

Data table of a WeChat chat history.

RecNo	msgId	talker	content	type	creatTime	imgPath	isSend
Click here to define a filter							
39	40	wxid_j17s4u02q3ka12	oh my God,have you eaten this?	1	1462518990000	(null)	1
40	41	wxid_j17s4u02q3ka12	Where did you buy it?	1	1462519043000	(null)	1
41	42	wxid_j17s4u02q3ka12	Yes	1	1462519055000	(null)	2
42	43	wxid_j17s4u02q3ka12	<?xml version="1.0"?>	3	1462519066000	THUMBNAIL_DIRPATH://th_dbb5e4622e87f85226c8da6893698fc0	2
43	44	wxid_j17s4u02q3ka12	Nightclub	1	1462519184000	(null)	2
44	45	wxid_j17s4u02q3ka12	wxid_nxjk2ny7xjc522:3515:0	34	1462519228892	171511050616bf5274d86bf101	1
45	46	wxid_j17s4u02q3ka12	You know,this is illegal.	1	1462519294000	(null)	2
46	47	wxid_j17s4u02q3ka12	I know! I will not tell others![嘘]	1	1462519434000	(null)	1
47	48	wxid_j17s4u02q3ka12	hello?	1	1462519489000	(null)	1
48	49	wxid_j17s4u02q3ka12	You can come to my house.	1	1462519595000	(null)	2
49	51	wxid_zjvs90v9xyat12	wxid_zjvs90v9xyat12:212188:1	34	1462519604000	4515260506160dff30cada1101	2
50	52	wxid_j17s4u02q3ka12	You are now at home?	1	1462519622000	(null)	1
51	53	wxid_zjvs90v9xyat12	<?xml version="1.0"?>	3	1462519623000	THUMBNAIL_DIRPATH://th_47a728be5e0c2a1e8905eb80b42b7ce9	2
52	54	wxid_zjvs90v9xyat12	wxid_zjvs90v9xyat12:16:0	43	1462519660000	1527410605160dff30c19261	2

- 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.

- **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”;

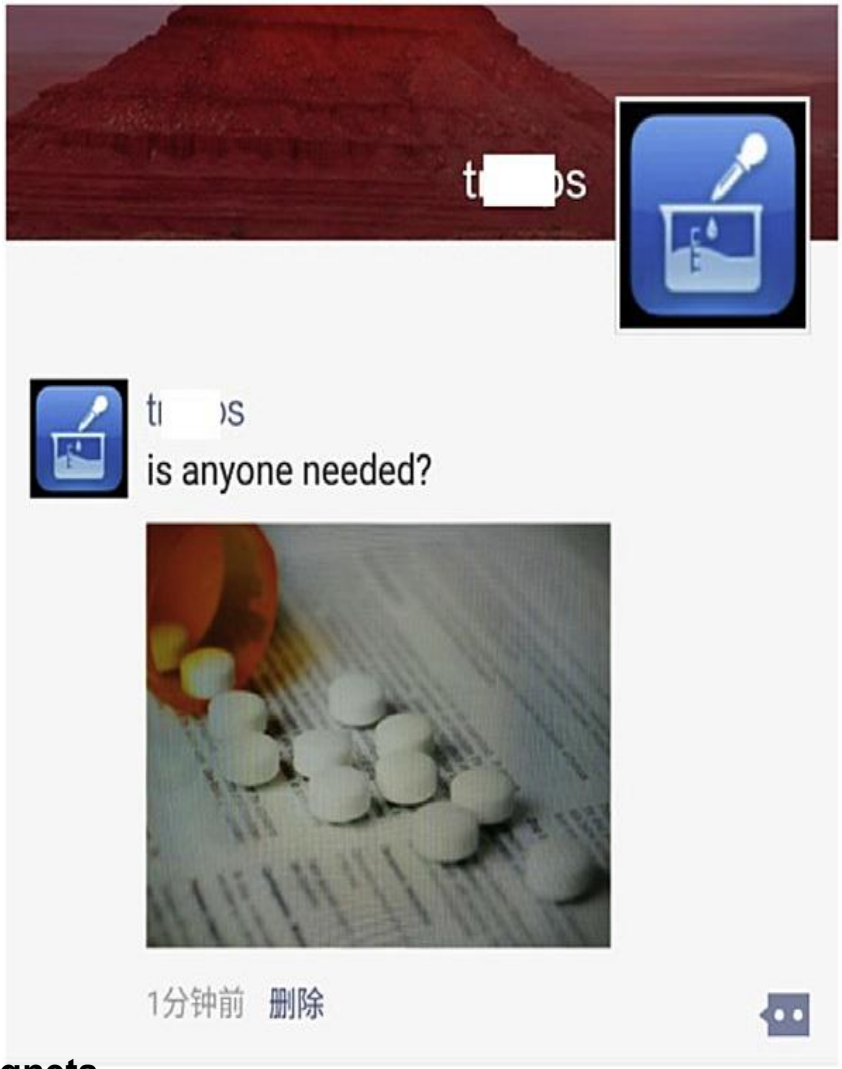
- **For a video**:

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



Storage of Moment using SQLite Viewer

RecNo	snsId	userName	localFlag	creatTime	head	localPrivate	type	sourceType	likeFlag	pravited	stringSeq	content	attrBuf	postBuf
	Click here to define a filter													
1	-6137801086392848296	wxid_nxjk2ny7xjc522	2	1467340348	20160701	0	1	2	0	0	00000123089429873167033 20			(null)
2	-6137805566153838501	wxid_i17s4u02q3ka12	2	1467339814	20160701	0	1	2	0	0	00000123089385075557131 15			(null)
3	-6137816413807431599	wxid_i17s4u02q3ka12	2	1467338521	20160701	0	1	2	0	0	00000123089276599021200 17			(null)
4	-6137858402894663461	wxid_jl3bzxqbghlj12	2	1467333515	20160701	0	3	2	0	0	00000123088856708148881 55			(null)
5	-6138356509528879017	wxid_kr088wz8knfq41	2	1467274137	20160630	0	1	2	0	0	00000123083875641806725 99			(null)
6	-6138872120280207155	wxid_jl3bzxqbghli12	2	1467212671	20160629	0	3	2	0	0	00000123078719534294444 61			(null)
7	-6139915482756935593	wxid_kr088wz8knfq41	2	1467088292	20160628	0	1	2	0	0	00000123068285909526160 23			(null)
8	-6141102535825084182	wxid_jl3bzxqbghlj12	2	1466946785	20160626	0	1	2	0	0	00000123056415378844674 34			(null)
9	-6141429987435204364	wxid_jl3bzxqbghlj12	2	1466907749	20160626	0	4	2	0	0	00000123053140862743472 52			(null)
10	-6142574697127661330	wxid_jl3bzxqbghlj12	2	1466771289	20160624	0	1	2	0	0	00000123041693765811902 86			(null)

Offset(h)	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F	10	11	12	13	14	15	16	17	
00000000	0A	14	31	32	32	36	39	33	36	32	36	33	35	39	39	31	33	36	33	38	39	37	12	13	..1226936263599136389 7..
00000018	77	78	69	64	5F	6E	78	6A	6B	32	6E	79	37	78	6A	63	35	32	32	18	00	20	B3	AE	wxid_nxjk2ny7xjc522.. ³®
00000030	B7	B9	05	2A	11	69	73	20	61	6E	79	6F	6E	65	20	6E	65	65	64	65	64	3F	32	1F	·¹.*.is anyone needed?2.
00000048	0D	00	00	00	00	15	00	00	00	00	1A	00	22	00	2A	00	32	00	38	00	48	00	50	00	..........".*.2.8.H.P.
00000060	58	00	65	00	00	00	00	3A	0A	0A	00	12	00	1A	00	22	00	2A	00	42	EC	02	0A	00	X.e....:.....".*.Bì...
00000078	10	01	1A	00	22	00	2A	E1	02	0A	14	31	32	32	36	39	33	36	32	36	33	36	34	31	....".*á...1226936263641
00000090	30	32	37	30	30	30	35	10	02	1A	00	22	70	68	74	74	70	3A	2F	2F	6D	6D	73	6E	0270005...."phttp://mmsn
000000A8	73	2E	71	70	69	63	2E	63	6E	2F	6D	6D	73	6E	73	2F	64	69	62	43	76	71	48	67	s.qpic.cn/mmsns/dibCvqHg
000000C0	34	57	6E	66	43	47	30	69	61	69	63	6F	73	52	69	63	73	45	39	44	73	64	42	4C	4WnfCG0iaicosRicsE9DsdBL
000000D8	62	44	32	63	64	68	68	6A	74	58	69	61	43	69	63	38	5A	6E	4C	59	30	71	5A	58	bD2cdhhjtXiaCic8ZnLY0qZX

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.

- Storage format of the content field

Type (1 byte)	Length (1 byte)	Data (n bytes)

(a) Definition of TLD format

(TLD) Header	(TLD) MsgOwner	8 bytes	(TLD) msgContent	(TLD) msgProperty	26 bytes .....
20bytes msgResID of the first multimedia file	4 bytes	(TLD) msgImagePath, URL of the first multimedia file	2 bytes	(TLD) msgImagePath2, URL of the thumbnail of the first multimedia file	
. . . . .					
20bytes msgResID of the *Nth* multimedia file	4 bytes	(TLD) msgImagePath, URL of the *Nth* multimedia file	2 bytes	(TLD) msgImagePath2, URL of the thumbnail of the *Nth* multimedia file	
. . . . . (Other data)					

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

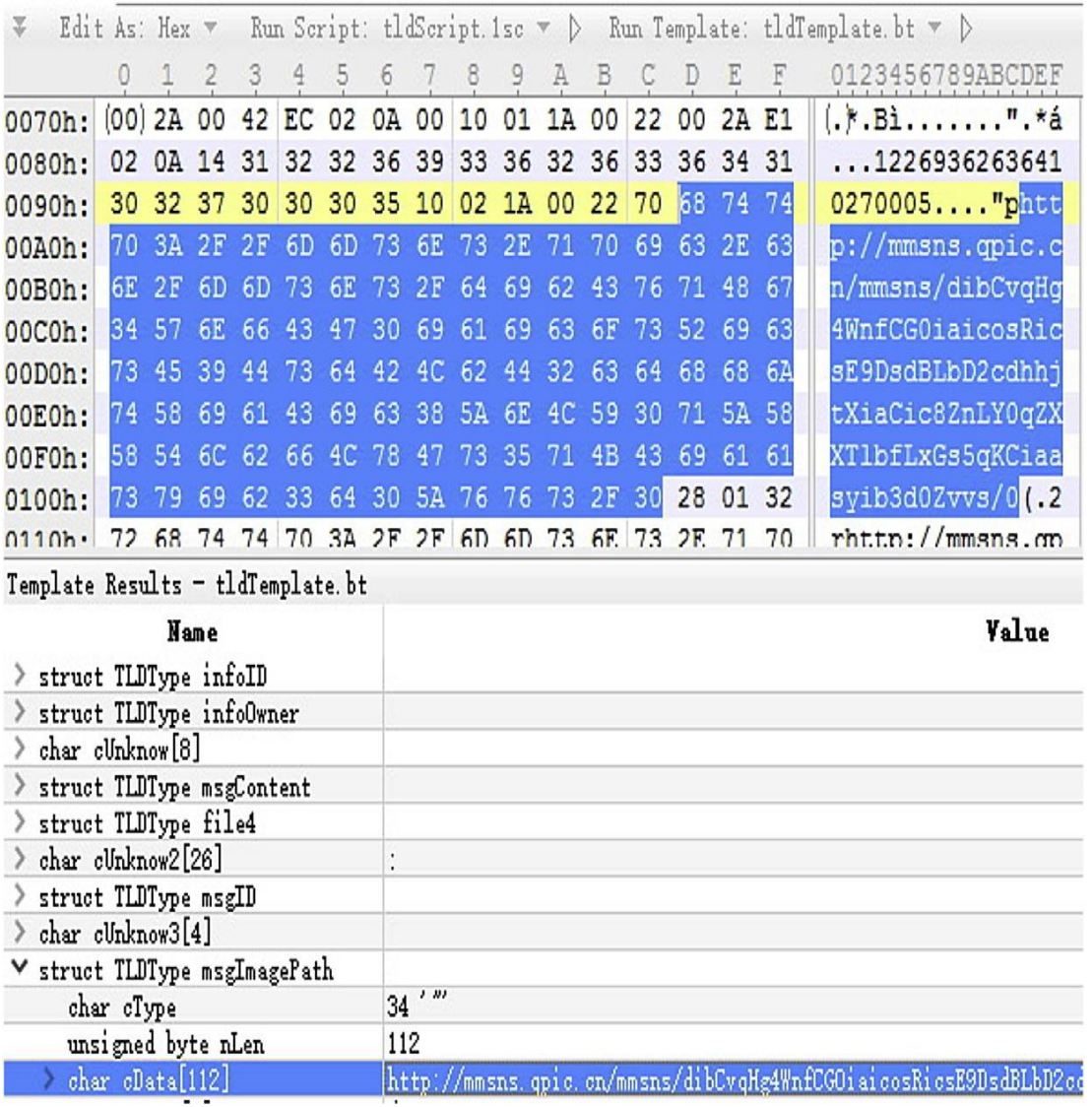
multimedia file is stored in the msgImagePath field (Next Fig), msgImagePath2

contains the URL of the thumbnail of the uploaded

multimedia file.

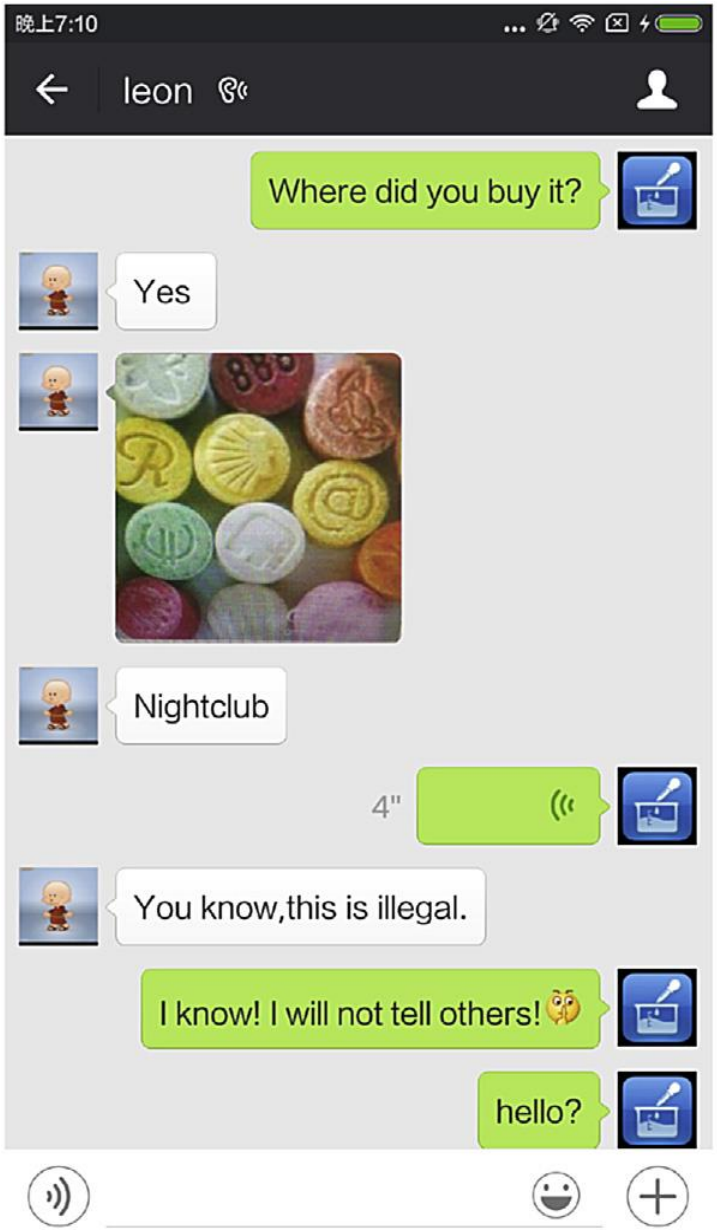
The multimedia resources can be acquired from the WeChat server after extracting the URL of the multimedia file from the

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.

Smartphone	Version						
	5.0	5.2.1	5.4.0.51	6.0.0.54	6.1.0.66	6.2.0	6.3.27
Xiaomi 4C	✓	✓	✓	✓	✓	✓	✓
Samsung Galaxy Note4	✓	✓	✓	✓	✓	✓	✓
LG G4	✓	✓	✓	✓	✓	✓	✓
HTC One	✓	✓	✓	✓	✓	✓	✓
Moto X	✓	✓	✓	✓	✓	✓	✓
Samsung Galaxy S6 Edge	✓	✓	✓	✓	✓	✓	✓

Table 2
Test results of decryption and forensics functions

Function	Version						
	5.0	5.2.1	5.4.0.51	6.0.0.54	6.1.0.66	6.2.0	6.3.27
Decryption	✓	✓	✓	✓	✓	✓	✓
Text message	✓	✓	✓	✓	✓	✓	✓
Image message	✓	✓	✓	✓	✓	✓	✓
Voice message	✓	✓	✓	✓	✓	✓	✓
Video message	✓	✓	✓	✓	✓	✓	✓
Moments	✓	✓	✓	✓	✓	✓	✓

❖ Experimental results (Cont)

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

	Files
Modified	getdns.ini, push_proc_startup.xml, com.tencent.mm_preferences.xml, dcfff5a88f266e972a7fd1cf80a760ec_26031732.getdns2, MM.mmap2, crash_status_file.xml, notify_key_prefxiongx2004.xml, staytime.cfg, systemInfo.cfg
Removed	wakelock_status.bin, com.android.opengl.shaders_cache, psk.key
Emerged	config.dat, md5, manifest and input.monitor

- 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, Genymotion12 and BlueStacks.1
- 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

http://www.sciencedirect.com/science/article/pii/S1742287616301220

WeChat's rise: Monthly active users

https://www.techinasia.com/wechat-near-billion-users

QUESTIONS?

THANK YOU