Digital forensic analysis of encrypted database files in instant messaging applications on Windows operating systems

Jusop Choi[a], Jaegwan Yu[b], Sangwon Hyun[c], Hyoungshick Kim[a]

[a] Sungkyunkwan University, Korea

[b] LIGNEX1 Co., Ltd, Korea

[c] Chosun University, Korea

Agenda

1. Introduction
2. Our analysis framework
3. Analysis of example IMs
4. Conclusion

Agenda

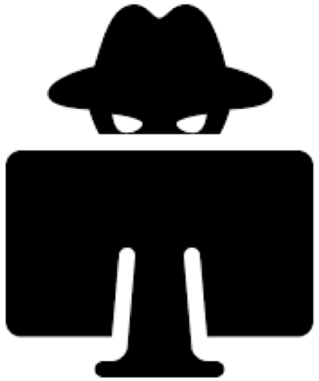
1. Introduction
2. Our analysis framework
3. Analysis of example IMs
4. Conclusion

Instant messengers (IM) everyday


- We are using instant messengers every day.
- There are a lot of users (WeChat: 1.1 billion, Telegram: 200 million, and Facebook Messenger: 1.3 billion users by 2019).

Who can see our messages?

- Government agencies can obtain messages to use as criminal evidence.
- Government agencies can censor messages.
- Hackers can steal messages from us.

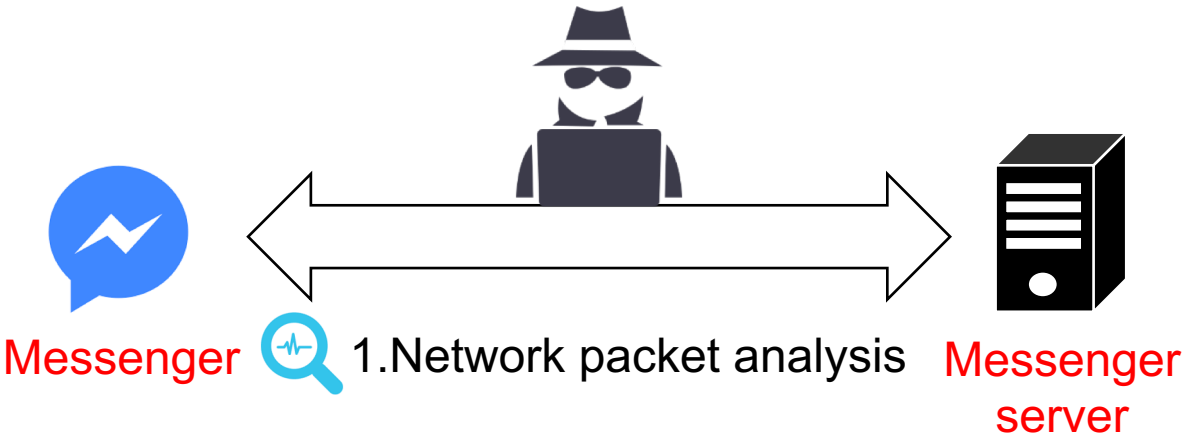


How can messages be obtained?

1. Network packet analysis 2. Server database analysis 3. Client database analysis

How can messages be obtained?

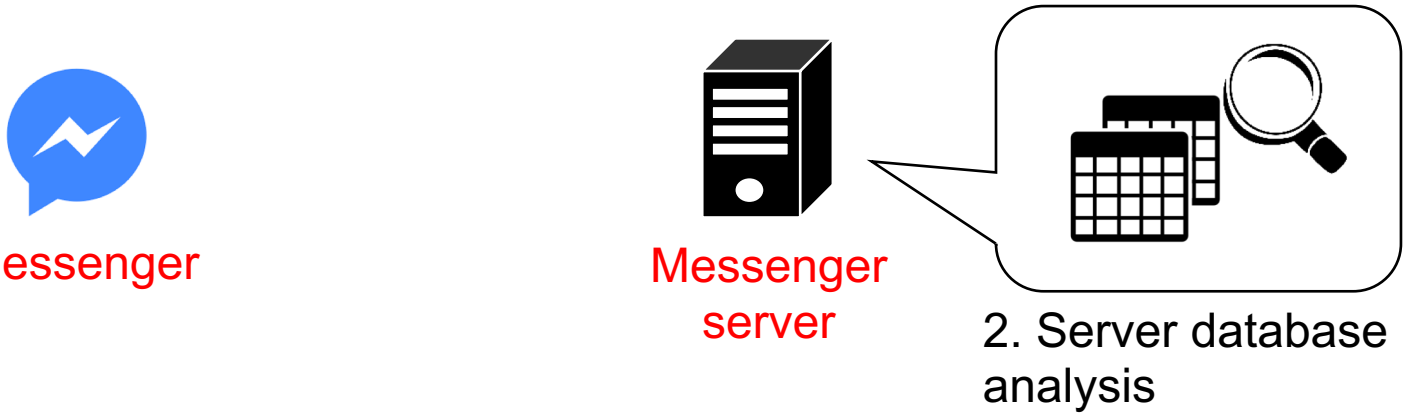
1. Network packet analysis 2. Server database analysis 3. Client database analysis



- Network traffic is securely protected by well designed security protocols.
- Overall, it is **not easy to analyze** the securely protected network traffic.

How can messages be obtained?

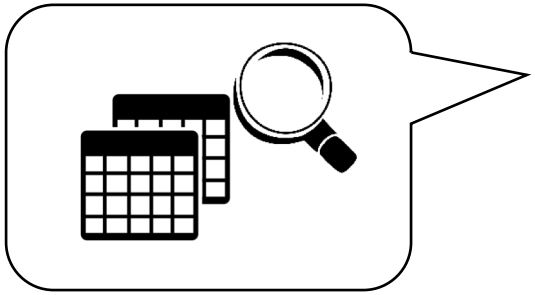
1. Network packet analysis 2. Server database analysis 3. Client database analysis



- It is hard for the attacker to access the server.
- Because of the privacy issue, many messengers **do not store the messages** in the server over a certain time period (e.g., 3 days for KakaoTalk).

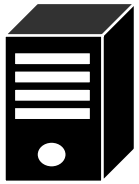
How can messages be obtained?

1. Network packet analysis 2. Server database analysis 3. Client database analysis



3. Client database analysis

Messenger



Messenger server

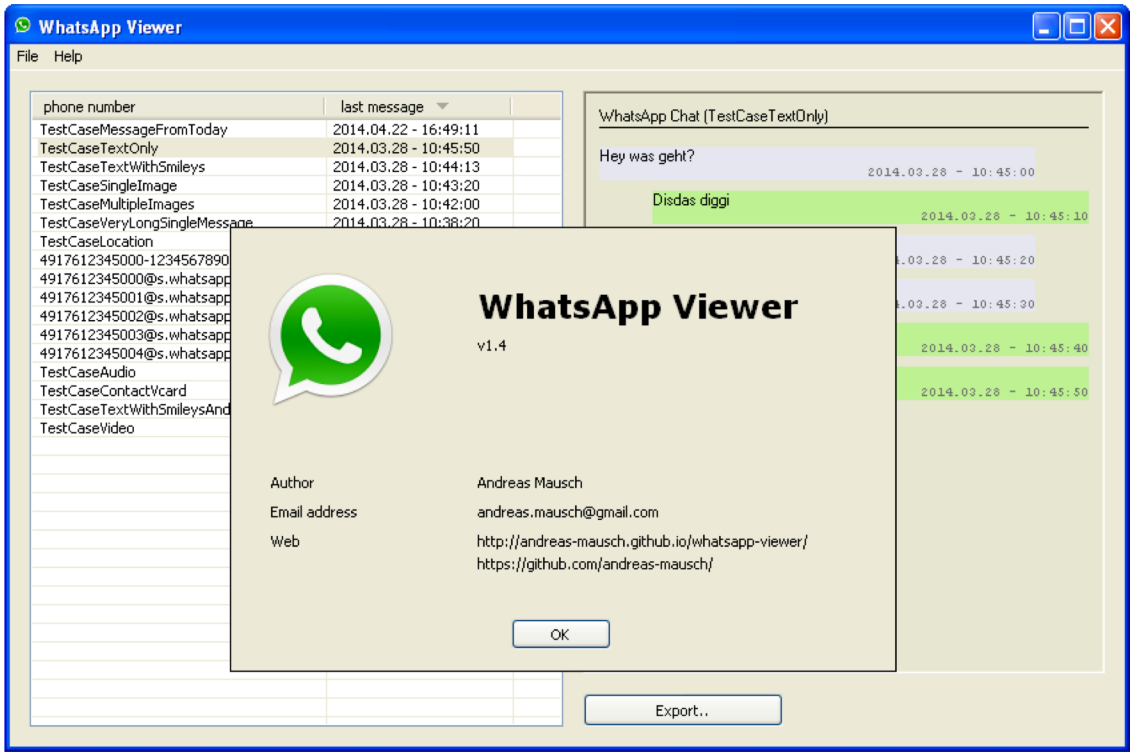
- All messages are stored in database files in the client device.
- In our research, we focused on analyzing the client chat database.

Agenda

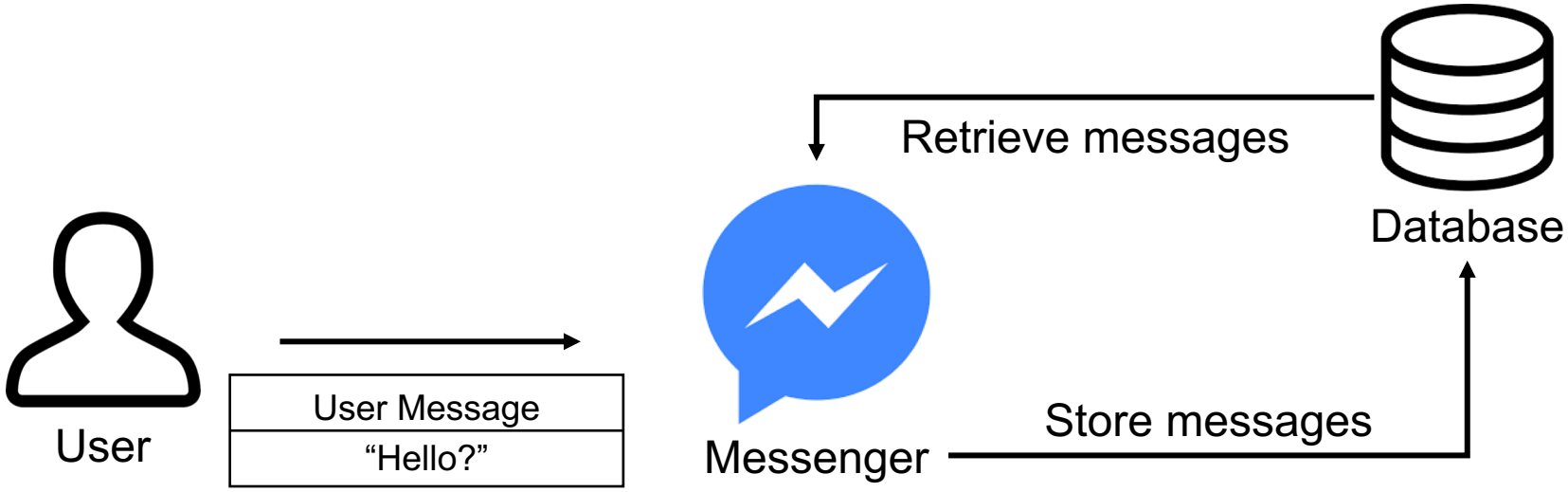
1. Introduction
2. Our analysis framework
3. Analysis of example IMs
4. Conclusion

What databases are used in IM

- There are various database types in instant messengers.
  - Contacts
  - Chats
  - Calls
  - Photos
  - Video

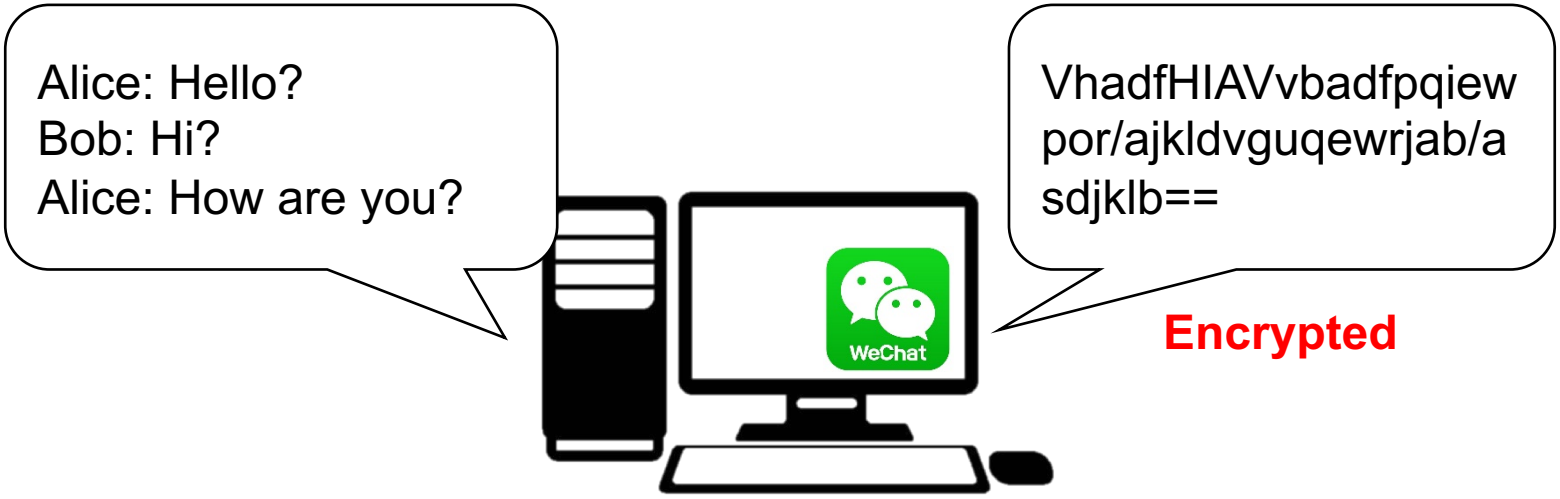


Interaction with database in IM



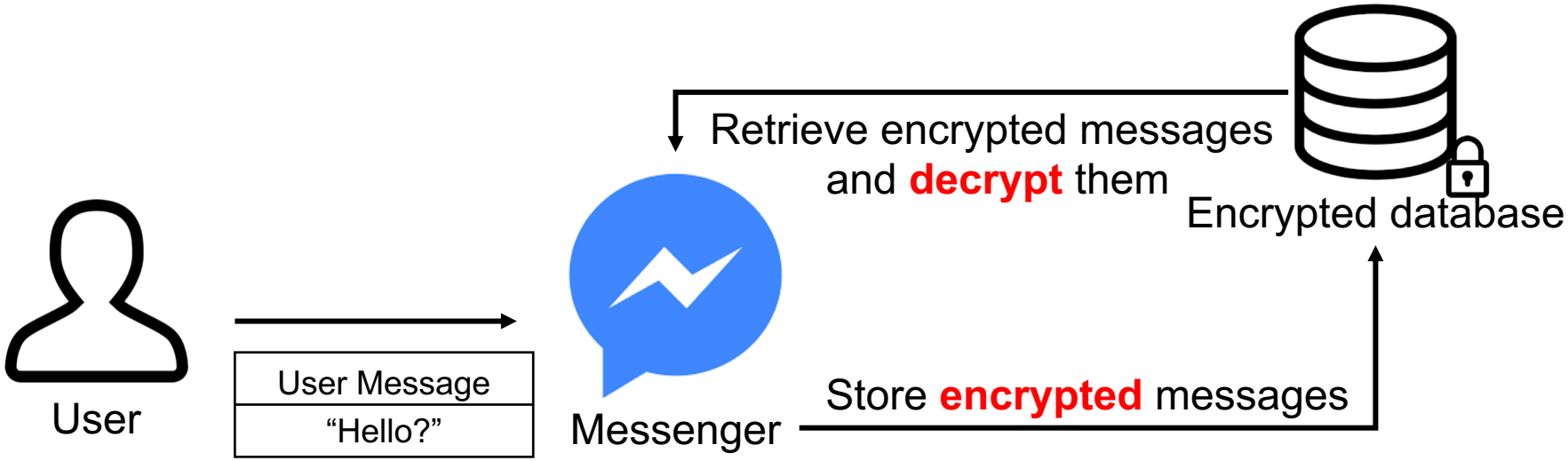
- To minimize network overhead, chat databases are stored in local files in the client device.

Chat database file on client side



- If chat database files are not encrypted (e.g., in WhatsApp), it is easier to analyze!
- However, in some messengers, the database files are **encrypted**.
- The number of IMs using the encryption feature is increasing over time.

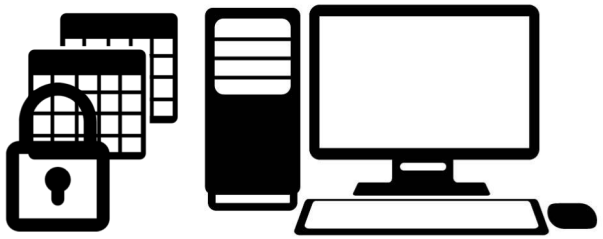
Use of encryption key



- After login, encryption key can often be (temporarily) created (in memory).
- After logout, the encryption key can be deleted from memory.
- In such cases, it is hard to obtain the encryption key.

Adversary model

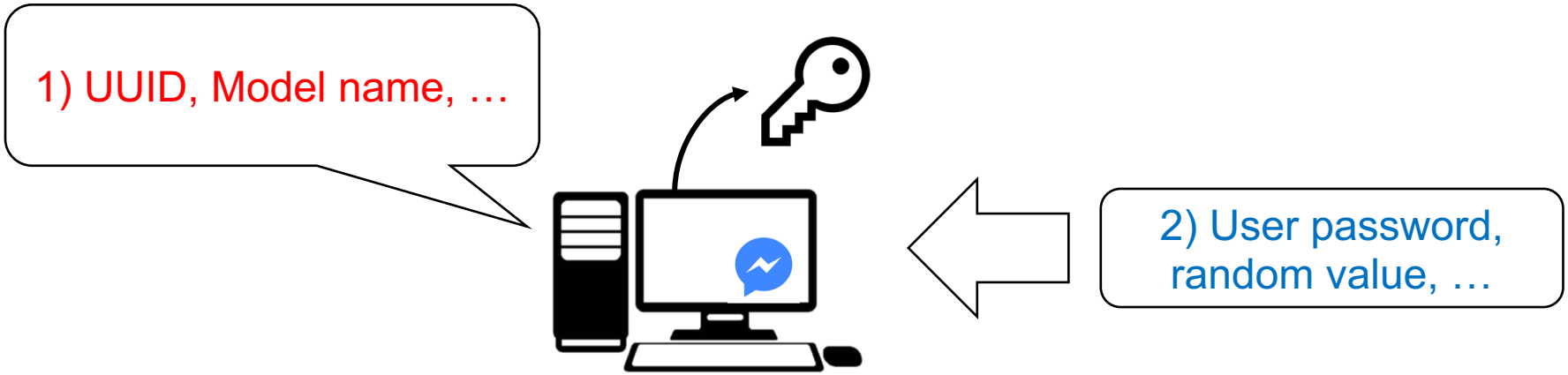
Physical Attacker



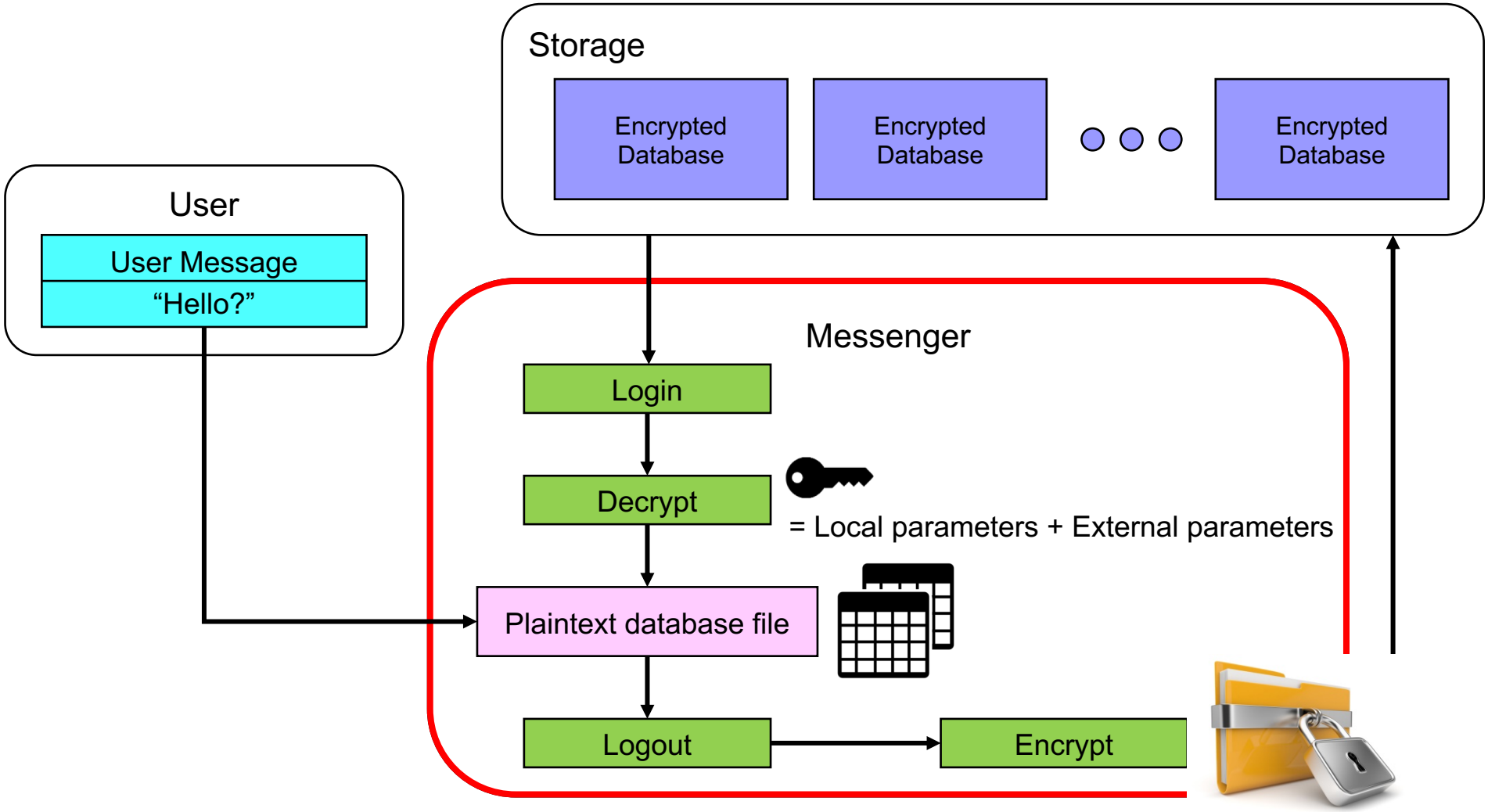
- The attacker can obtain the **encrypted** database file.
- The attacker's goal is to obtain the database **encryption key**.

Key observations in the encryption key generation

- To securely generate the dynamic encryption key, most IMs used the following parameters for key generation.
  1) Device specific local parameters: UUID, HDD serial number, …
  2) Externally obtained parameters after login: user password, user key stored at the server, …



Overall procedure



How to obtain the encryption key

1. Find the (encrypted) chat database files in the target IM.
2. Check whether the chat database files are encrypted.
3. Find the code locations to encrypt/decrypt the chat database files during the execution of the IM.
4. Identify the cryptographic algorithms (e.g., AES, SHA256) used in encrypting chat database files by examining the assembly codes and the dynamic behaviors of the IM.
5. Analyze the procedure of generating an encryption key.

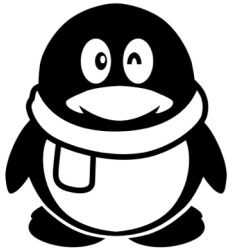
Agenda

1. Introduction
2. Our analysis framework
3. Analysis of example IMs
4. Conclusion

Target messengers

- KakaoTalk is the most popular messenger in South Korea with 49.8 million users.

- NateOn is also popular in South Korea with 15 million users.

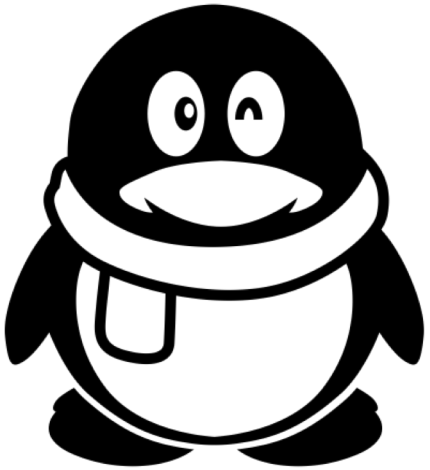


- QQ is the most popular messenger in China with 899 million users.

Example 1. KakaoTalk

KakaoTalk

NateOn

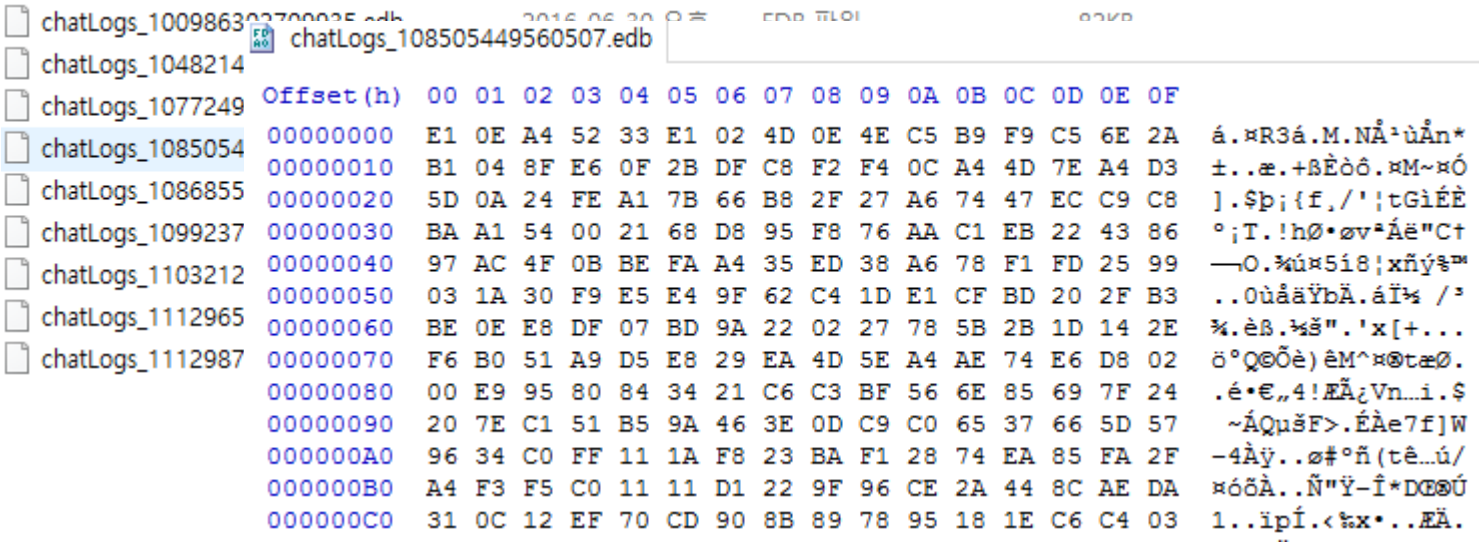


QQ

Experiment environment

- Environment and tools
  - VMware workstation 12
  - Windows 7
  - OllyDbg v1.10
  - KakaoTalk (version 2.0.8.990 ~ 2.5.6.1545)
    → Current version (3.0.0.2110)
  - Packed by *Themida*

Database file path



- The databases are stored in a pre-configured file system path, and the user cannot change it.
  - chatLogs_[random value].db: chat database file
  - TalkUserDB.db: contact database file

Database encryption process

```
0073B4DB   . 8B4D 10        MO
0073B4DE   . 8B71 04        MO
0073B4E1   . 8B51 08        MO
0073B4E4   . 3350 08        XO
0073B4E7   . 3370 04        XO
0073B4EA   . 57             PU
0073B4EB   . 8B39           MO
0073B4ED   . 3338           XO
0073B4EF   . 894D E4        MOV DWORD PTR S
0073B4F2   . 8B49 0C        MOV ECX,DWORD P
0073B4F5   . 3348 0C        XOR ECX,DWORD PTR DS:[EAX+0xC]
```

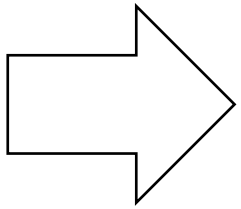
Assembly codes of AES-128 encryption

address	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	CA	CB	CC	CD	CE	CF
0040D8C0	63	7C	77	7B	F2	6B	6F	C5	30	01	67	2B	FE	D7	AB	76
0040D8D0	CA	82	C9	7D	FA	59	47	F0	AD	D4	A2	AF	9C	A4	72	C0
0040D8E0	B7	FD	93	26	36	3F	F7	CC	34	A5	E5	F1	71	D8	31	15
0040D8F0	04	C7	23	C3	18	96	05	9A	07	12	80	E2	EB	27	B2	75
0040D900	09	83	2C	1A	1B	6E	5A	A0	52	3B	D6	B3	29	E3	2F	84
0040D910	53	D1	00	ED	20	FC	B1	5B	6A	CB	BE	39	4A	4C	58	CF

S-Box of AES-128 encryption

- With the assembly codes and S-Box for AES-128, we can guess that AES-128-CBC algorithm was used.
- We also figured out the chat database file format (SQLite format 3).

Extraction of encryption key and IV

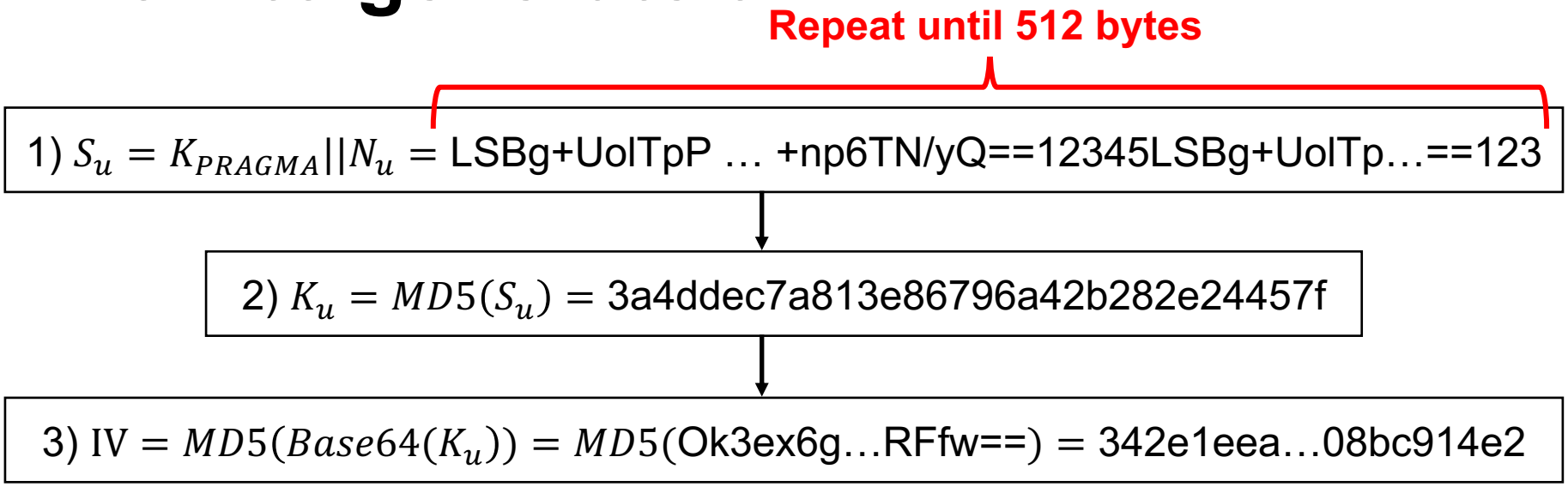


Key: 3a4ddec7a813e86796a42b282e24457f

IV: 342e1eea1209562340012340 8bc914e2

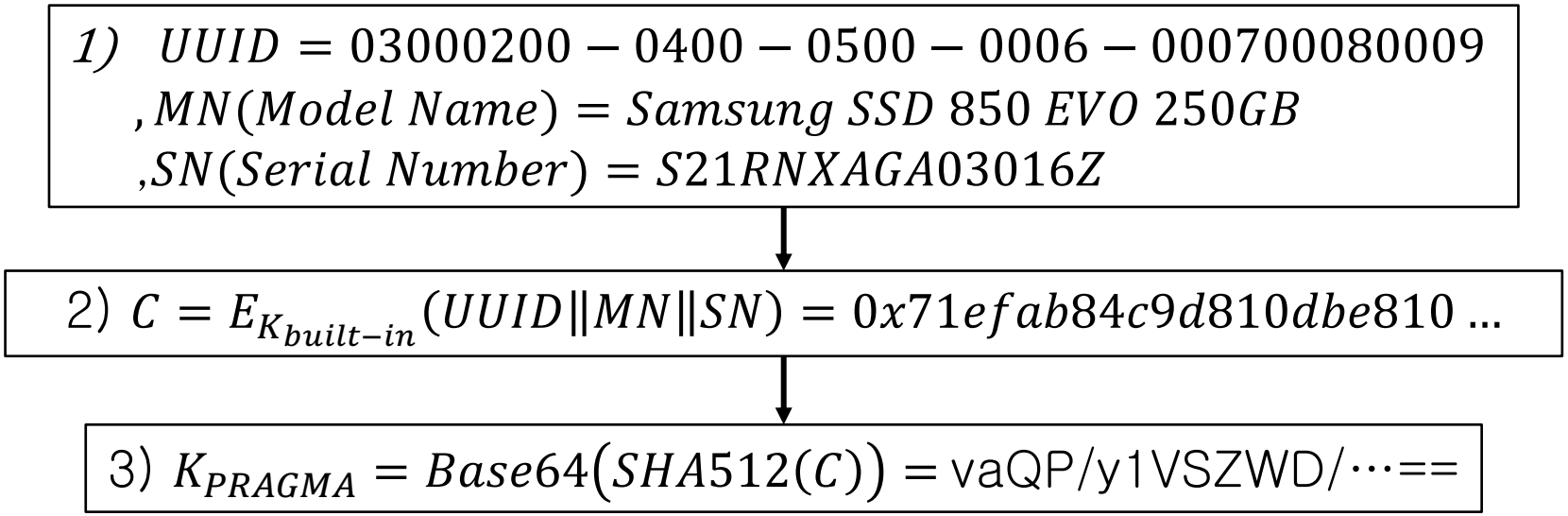
- Through the dynamic analysis, we found which values are used for the encryption key and IV.
- With those values, we tried to trace back the key generation procedure.

How to generate the key



- K_u is used as the encryption key.
- The input of the hash function (S_u) is the PRAGMA key combined with the user sequence number (N_u) given by the server.

How to generate the PRAGMA key



- PRAGMA key generation codes **temporarily appear** in memory to protect the key generation procedure.
  - Dynamic analysis is necessary.

Summary of encryption process

1. Collect the device specific information.
2. Generate PRAGMA key from the collected information.
3. Get the **user sequence number** (the only external parameter for KakaoTalk) from the server after the user logins.
4. Generate encryption key and IV.

How to securely generate the encryption key

- KakaoTalk might use the following strategies.
  - ① Security by obscurity
    - (AES → SHA512 → Base64 → MD5 → Base64 → MD5)
  - ② Moving target
    - Use of device specific parameter
  - ③ Updatable key generation
    - Updatable built-in key
  - ④ External user data
    - Best security practice for disk encryption

How to securely generate the encryption key

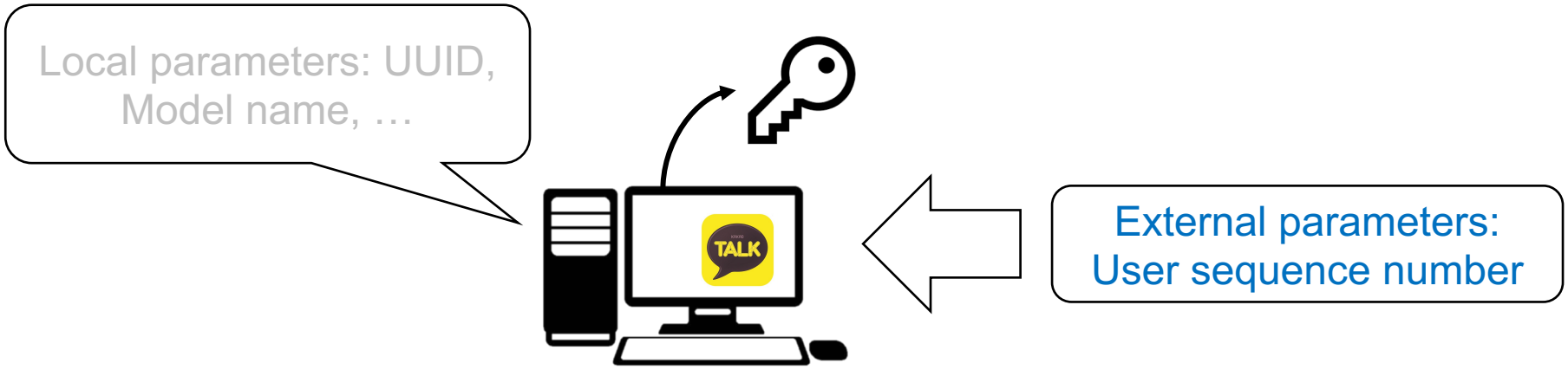
- KakaoTalk might use the following strategies.
  - ① Security by obscurity

However, we might easily obtain external user data.

  - ④ External user data
    - ➢ Best security practice for disk encryption

How to obtain user sequence number

1. Brute-force search to obtain the user sequence number of the target user.
2. Smarter way to obtain the user sequence number using automatic friend addition with the target user's phone number.



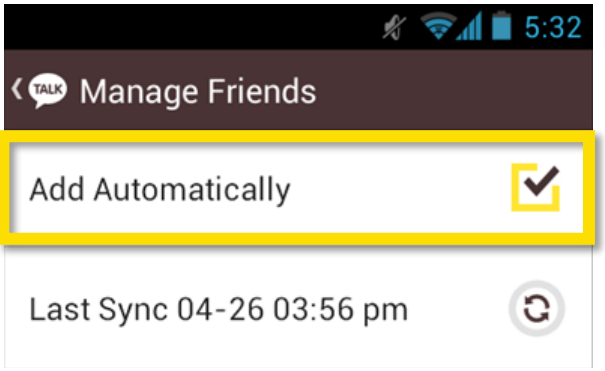
Brute-force search

- User sequence number is smaller than 300 million.
  - When we signed up the KakaoTalk at Jun 21st, the user sequence number is given 251,582,414.
- Through the brute-force attack, the user sequence number can be obtained within **8.4 hours** (10,000 trial per second).
- The user sequence number is related to the date when the user signed-up.
  - If you figure out the victim's registration date of KakaoTalk, guessing range of the user sequence number can be narrowed.

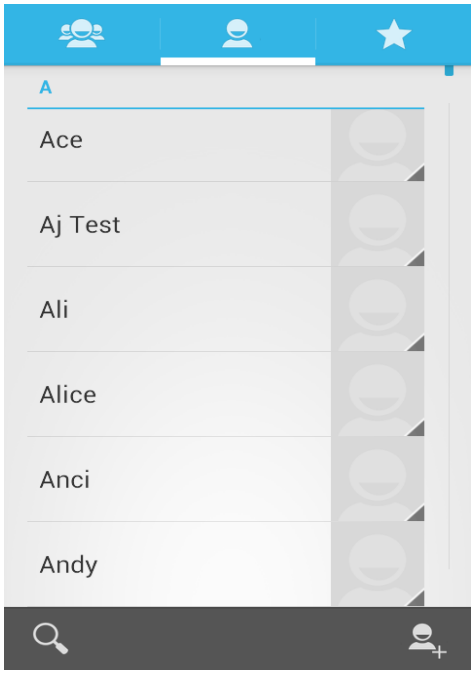
Smarter way to get the user sequence number

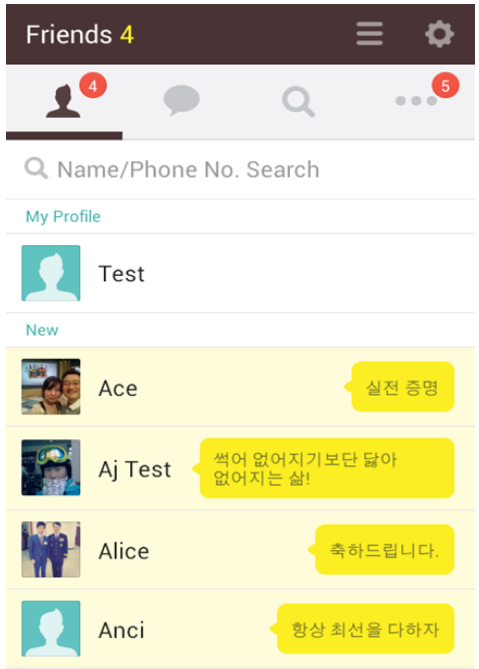
Assumption: Target user's phone number is given.

1. Add the target user as friend in KakaoTalk with the target user's phone number.



Checked by default





Smarter way to get the user sequence number

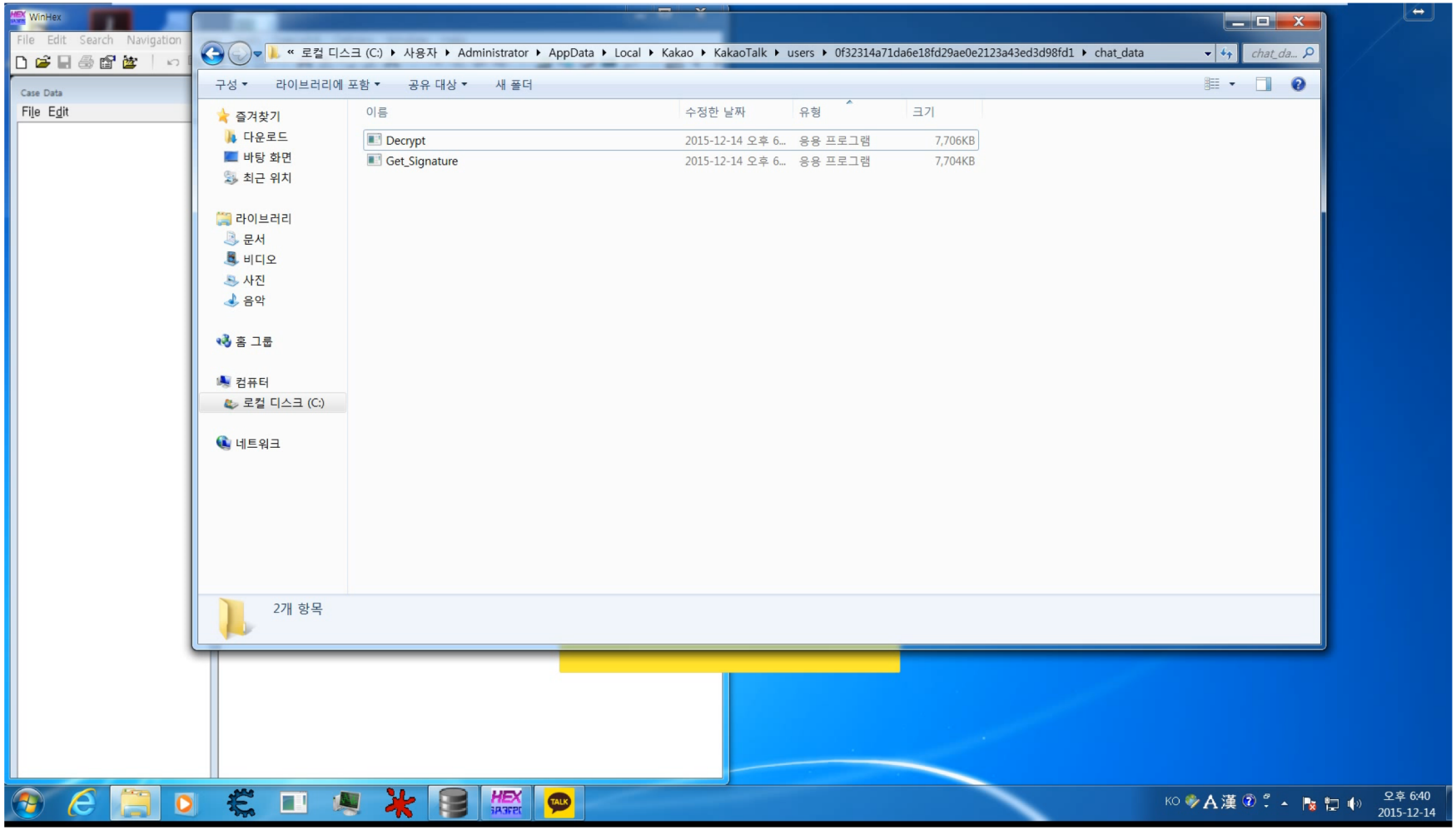
2. The target user's information is added to our contact database file that we can freely access.
3. We can extract the target user's user sequence number from the user's record in the contact database file.

User sequence number

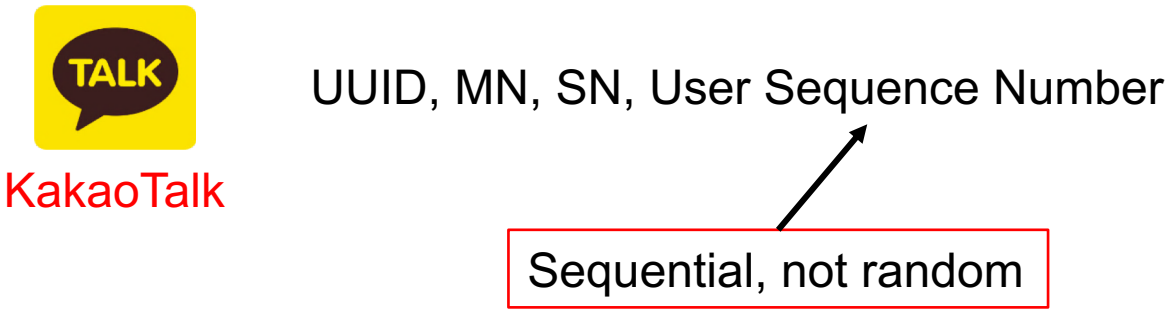
New Recor

	userId	type	nickName	profileImageUrl	llProfileImageU	inalProfileImag	statusMessage	linkedServices
	Filter	Filter	Filter	Filter	Filter	Filter	Filter	Filter
7	[redacted]	2	102					
8	**151741**	2	:)	http://th-p.talk.kakao.…	http://th-p.talk.kakao.…	http://p.talk.kakao.…	▮(ᵔᵕᵔ) / 번…	
9	[redacted]	2	AJ	http://th-p.talk.kakao.…	http://th-p.talk.kakao.…	http://p.talk.kakao.…		
10	[redacted]	2	BuJy	http://th-p.talk.kakao.…	http://th-p.talk.kakao.…	http://p.talk.kakao.…		
11	[redacted]	2	Chan	http://th-p.talk.kakao.…	http://th-p.talk.kakao.…	http://p.talk.kakao.…	빨리읽기를 바라면 문자로	
12	[redacted]	2	David	http://th-p.talk.kakao.…	http://th-p.talk.kakao.…	http://p.talk.kakao.…		
13	[redacted]	2	Geumhwan	http://th-p.talk.kakao.…	http://th-p.talk.kakao.…	http://p.talk.kakao.…	The reason I live	story

Demo – KakaoTalk



How can we securely protect the external parameters?



1. User sequence number is not random, but sequential.
2. User sequence number is used as the primary key for contact database.

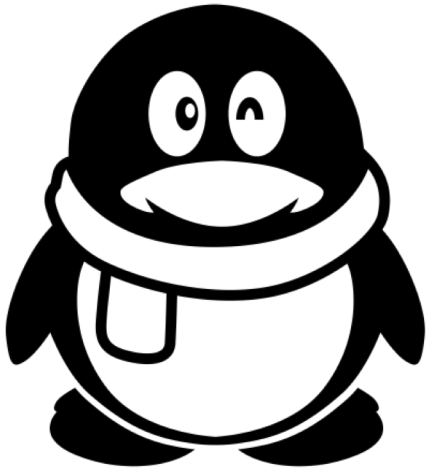
Use of secure external parameter

1. User sequence number is not random, but sequential.
   - ➢ Use a large space of random numbers for external parameters.
2. User sequence number is used as the primary key for contact database.
   - ➢ Use the external parameter for the encryption key purpose only.

Example 2. NateOn

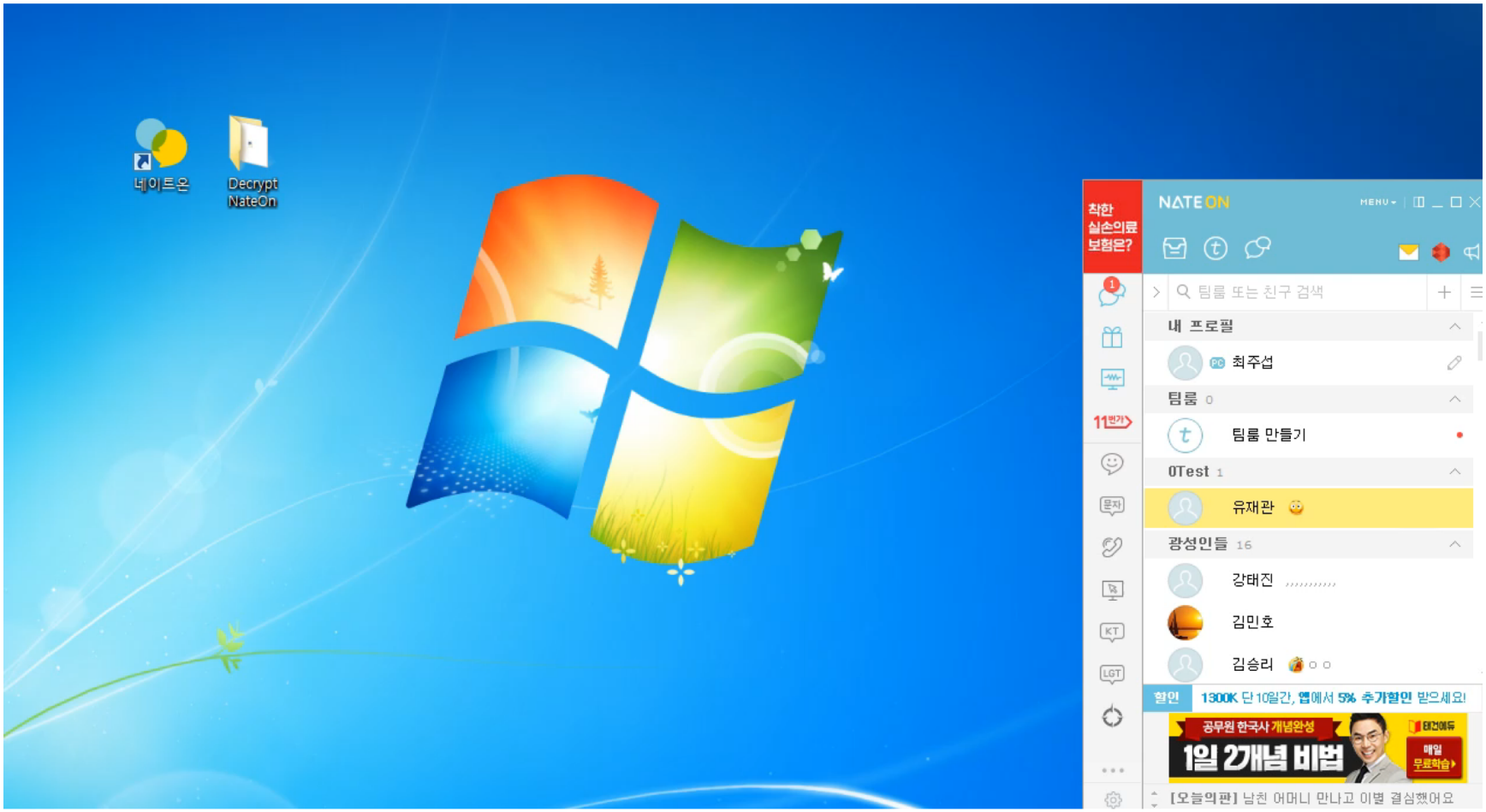
KakaoTalk

NateOn



QQ

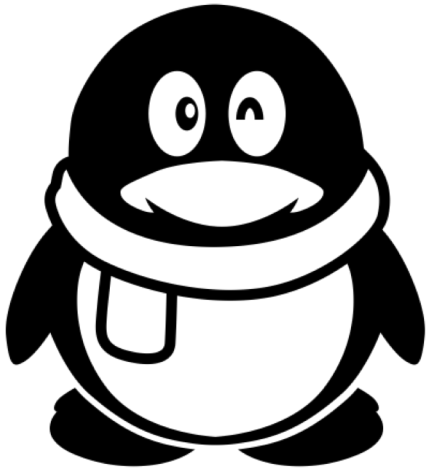
Demo – NateOn


네이트온
Decrypt NateOn
착한 실손의료 보험은?
NATE ON
MENU
팀룸 또는 친구 검색
내 프로필
최주섭
팀룸 0
팀룸 만들기
0Test 1
유재관
광성인들 16
강태진
김민호
김승리
할인
1300K 단 10일간, 앱에서 5% 추가할인 받으세요!
공무원 한국사 개념완성
1일 2개념 비법
매일 무료학습
[오늘의판] 남친 어머니 만나고 이별 결심했어요

Example 3. QQ

KakaoTalk

NateOn

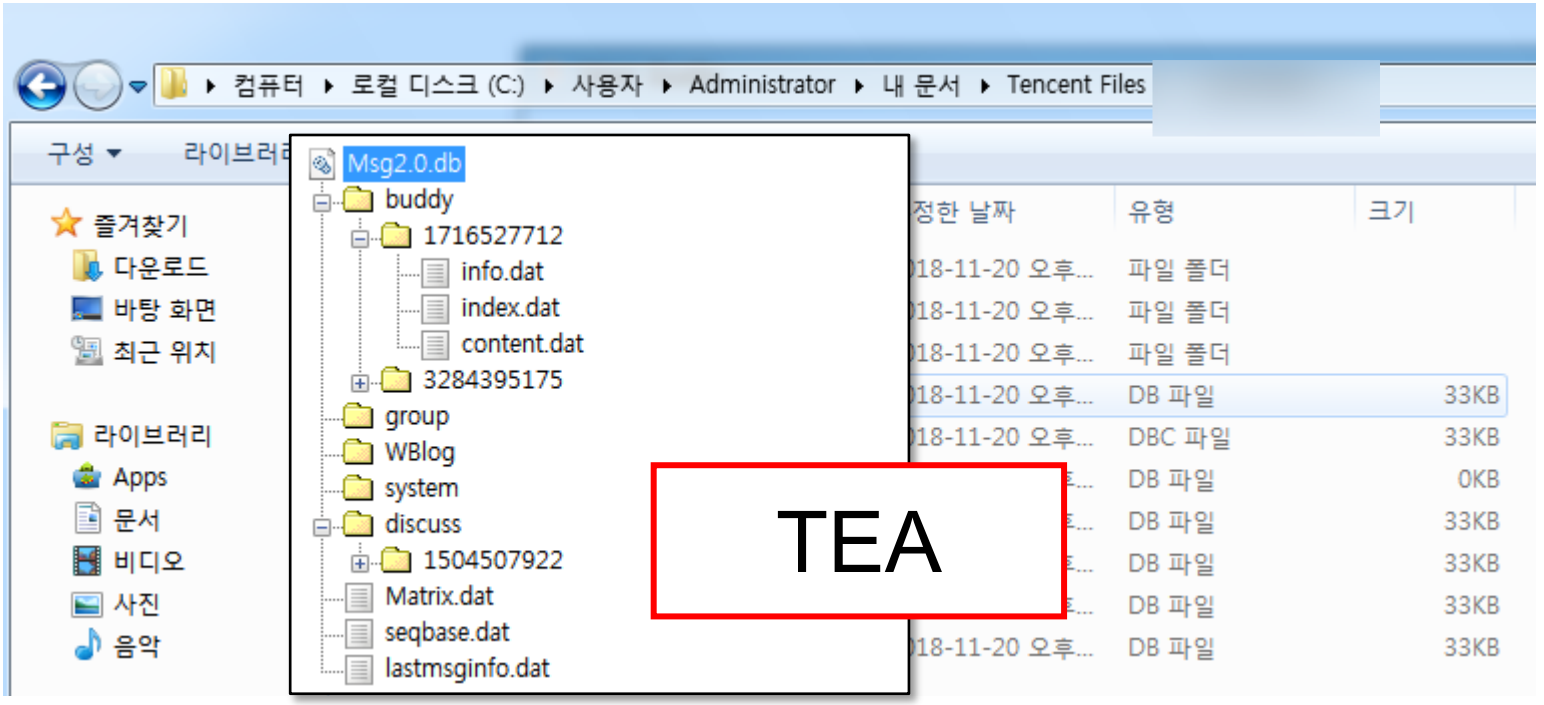


QQ

Experiment environment

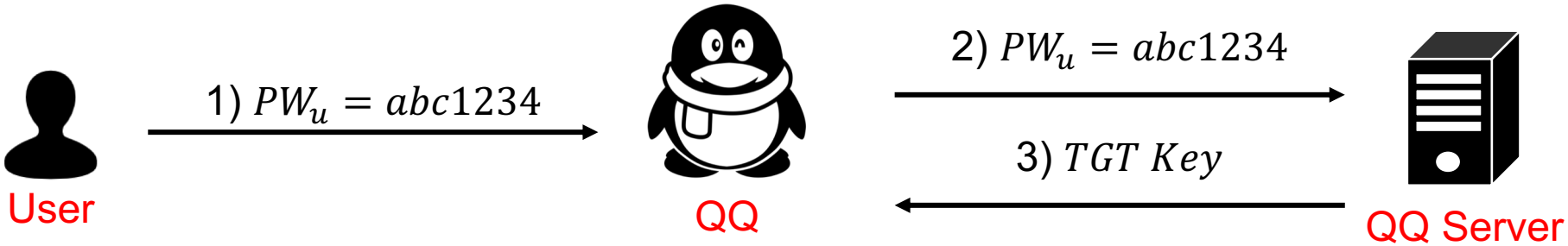
- Environment and tools
  - VMware workstation 12
  - Windows 7
  - OllyDbg v1.10
  - QQ (version 2.1.1369.0)
- Unfortunately, the encrypted database of QQ cannot be decrypted without the server

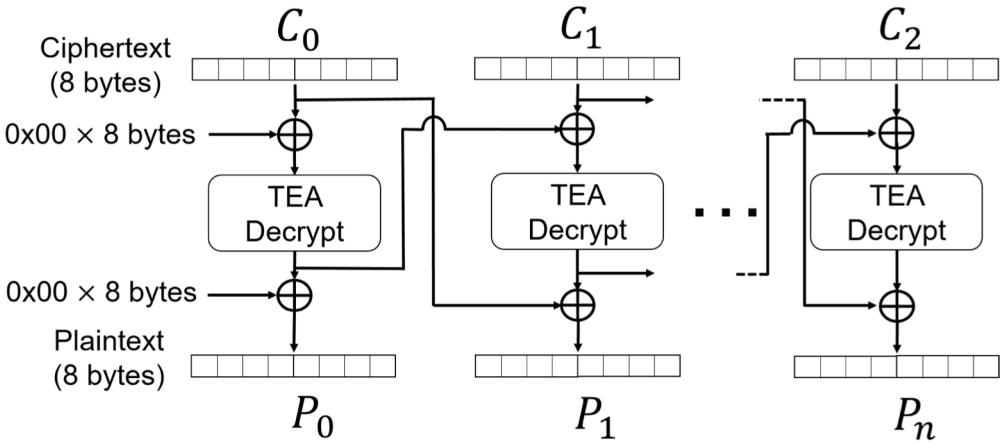
DB format and encryption algorithm



OLE Structure

Encryption key generation algorithm

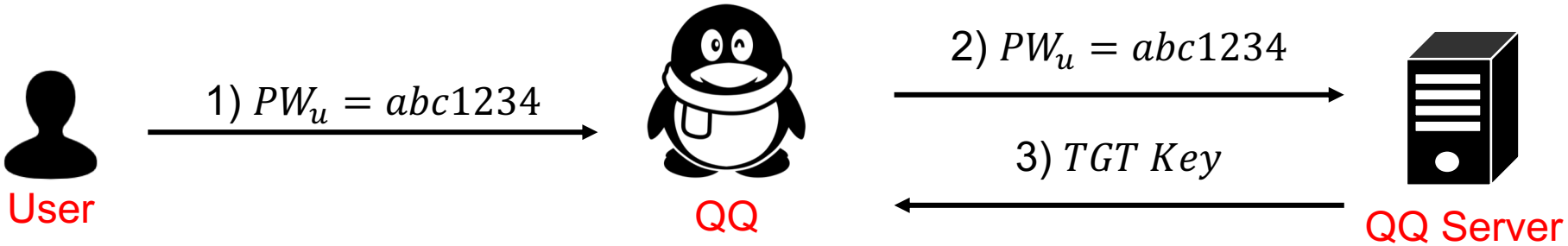




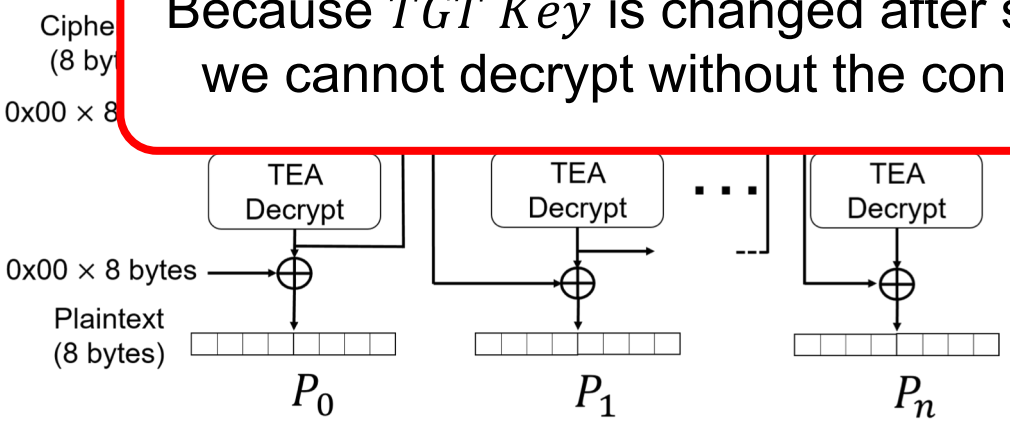
With $TGT\ Key$, QQ Messenger decrypts the encrypted database using TEA with modified block cipher mode.

* **Ticket Granting Ticket** or **Ticket to Get Tickets** (**TGT**) is a small, encrypted identification file with a limited validity period. (https://en.wikipedia.org/wiki/Ticket_Granting_Ticket)

Encryption key generation algorithm



Because $TGT\ Key$ is changed after some period (8 hours), we cannot decrypt without the connection to the server.



nger decrypts the encrypted database using TEA with modified block cipher mode.

* **Ticket Granting Ticket** or **Ticket to Get Tickets** (**TGT**) is a small, encrypted identification file with a limited validity period. (https://en.wikipedia.org/wiki/Ticket_Granting_Ticket)

Agenda

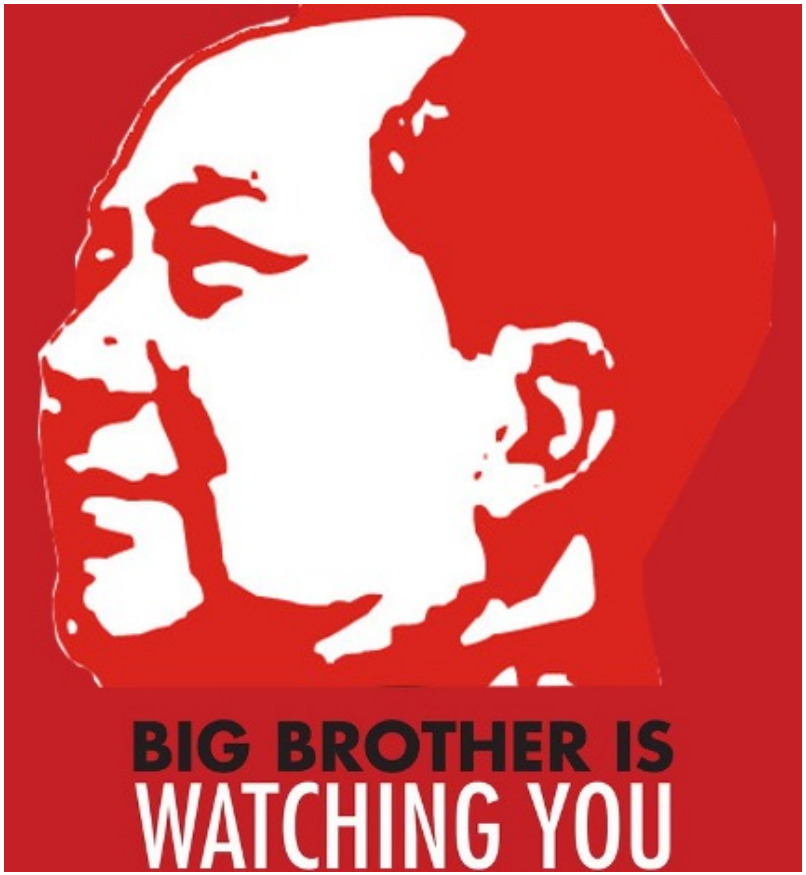
1. Introduction
2. Our analysis framework
3. Analysis of example IMs
4. Conclusion

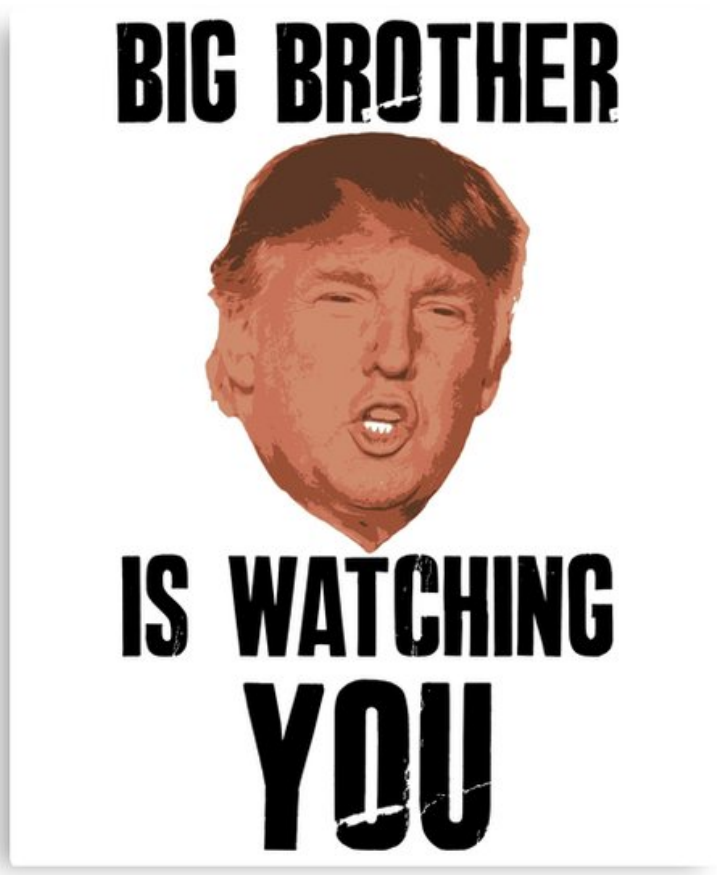
Summary of analysis results

Messenger	Local parameters	External parameters	Decryption without user auth.
KakaoTalk	Device specific	User seq. number	O
NateOn	Fixed string	User seq. number	O
QQ	X	TGT key	X

- In KakaoTalk and NateOn, we can obtain messages without the user's password.
- In QQ, the encryption key for chat database files is stored at the server.

Big Brother is always watching you.





Thank you for listening our talk.