Chapter 20

FORENSIC ANALYSIS OF POPULAR CHINESE INTERNET APPLICATIONS

Ying Yang, Kam-Pui Chow, Lucas Hui, Chunxiao Wang, Lijuan Chen, Zhenya Chen and Jenny Chen

Abstract When the Digital Evidence Search Kit (DESK) was first used in Mainland China, it was found to be inadequate because it did not support criminal investigations involving popular Internet applications such as QQ, MSN and Foxmail. This paper discusses the enhancements made to DESK to conduct forensic analyses of QQ, MSN and Foxmail.

Keywords: Forensic analysis, Internet applications, QQ, MSN, Foxmail

1. Introduction

The Digital Evidence Search Kit (DESK) is a digital forensic tool developed by the Center for Information Security and Cryptography at the University of Hong Kong [1]. DESK was introduced to Mainland China in 2007 and is now being used on a trial basis by several Chinese public security departments. During the performance review phase, it was discovered that DESK was inadequate for investigations in China because it did not support forensic analyses of some Internet applications that are popular in China, but rarely used elsewhere in the world. These applications, which include QQ, MSN and Foxmail, have special data formats and often encrypt important data.

QQ [4] is an instant messaging (IM) application with Chinese characters developed by Tencent Holdings; MSN is a similar application developed by Microsoft. According to the 2007-2008 China Internet Survey, QQ has an overwhelming market share in China. QQ is followed by MSN, which is the most popular instant messaging tool among office workers. Other instant messaging tools have a miniscule market share.

K.-P. Chow, S. Shenoi (Eds.): Advances in Digital Forensics VI, IFIP AICT 337, pp. 285–295, 2010.
© IFIP International Federation for Information Processing 2010

QQ's popularity stems from the fact that it is designed to accommodate Chinese Internet communication habits, including the need to communicate with friends and strangers. QQ is positioned as a comprehensive platform for entertainment, Internet chat and communications. It continuously incorporates peripheral functions that are attractive to children and young adults, causing the numbers of new QQ accounts and new QQ users to increase dramatically. According to an iResearch report, the QQ IM application had 132,740,000 active users in China as of June 2009. Indeed, almost every netizen in China has a QQ account.

Another popular system is Foxmail [8], a client-side email software also developed by Tencent Holdings. Foxmail functions are similar to Microsoft Outlook, but with specialized Chinese character support, which has contributed to its widespread use in China.

Meanwhile, the use of IM and email by criminal entities is growing in China. In particular, criminals use QQ, MSN and Foxmail to disseminate obscene images or links to pornographic websites, to divulge national secrets, and even to discuss, plan and coordinate criminal operations. This paper discusses the enhancements made to DESK to support forensic investigations involving QQ, MSN and Foxmail.

2. QQ Forensic Analysis

This section presents key details of the QQ IM application and outlines a forensic analysis methodology.

2.1 Overview

Every new user must register with the QQ service before using the IM application. Upon successful registration, the user is assigned a QQ number (i.e., a QQ account). The user may then log into the QQ service using the QQ number and start a session. The QQ processing can be summarized as follows:

- When a user logs into the QQ service, the QQ client obtains the latest friends list from the QQ server and establishes a peer-to-peer (P2P) connection between the user and each friend on the list.
- The user and his/her friends communicate using UDP.
- If a P2P connection cannot be established (e.g., due to a network problem), messages between the two friends are transmitted through the QQ server. The server stores all messages that the sender has sent but the receiver has not yet received. The stored messages are passed to the receiver when he/she next logs into the QQ service.

Table 1. QQ packet structure.

Bytes	Content
0	Start of packet (`0x02`)
1 to 2	QQ version number (expressed using network byte order)
3 to 4	Command number (expressed using network byte order)
5 to 6	Sending serial number (receiver should check the number)
7 to n	QQ data (possibly encrypted)
n+1	End of packet (`0x03`)

QQ messages are sent using UDP. Each UDP packet has no more than 64K bytes. Table 1 presents the packet structure.

Table 2. QQ files.

QQ File	Function
`QQApplication.dll`	Friends panel display program
`LoginUinList.dat`	Login history file
`QQZip.dll`	Compression and decompression utilities
`QQMainFrame.dll`	QQ main panel
`Newface`	Directory containing all portrait files
`QQ.exe`	QQ main executable
`QQFileTransfer.dll`	QQ file transfer utility
`QQHook.dll`	QQ keyboard monitor program
`QQPlugin.dll`	QQ friends searching utility
`QQRes.dll`	QQ resource handling function

2.2 Principal Files

In order to perform QQ forensics, it is important to understand the file organization. Table 2 lists the files present in a QQ directory after the successful installation of the application.

User account information is stored in several files in a directory named after the user's QQ number (e.g., 12345678). The principal files are:

- **`MsgEx.db`:** This file is created after the user registers with and logs into the QQ service. The file stores all chat records using structured storage [3]. Local history data, which includes chat records and logs, are encrypted using TEA [7] and stored in `MsgEx.db`.

- **`ewh.db`:** This file stores the MD5 hash [5] of the user's password. When a user attempts to log into the QQ service, the QQ client verifies the submitted password with the password stored in the

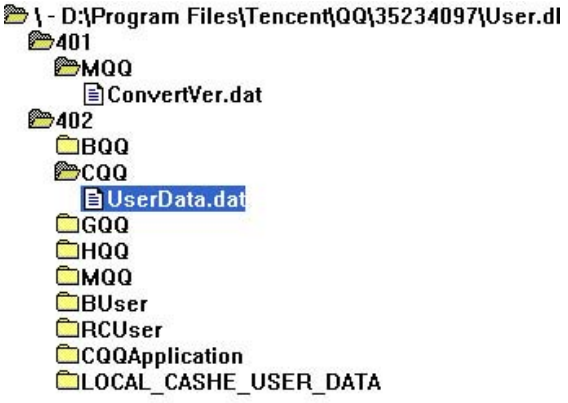

Figure 1. `User.db` data structure.

`MsgEx.db` file. The QQ server then performs a second validation before the user can successfully log in. This involves hashing the user-supplied password and comparing the value with the password hash saved in `ewh.db`.

- `Notes.db`: This file stores the QQ memorandum.
- `User.db`: This file stores the friend records. It uses the same structured storage format and TEA encryption as `MsgEx.db`.
- `QQAVFile`: This directory stores all QQ image files.
- `CustomFace`: This directory stores all self-defining expressions.
- `CustomFaceRecv`: This directory stores all received self-defining expressions.
- `ShareInfo.db`: This file stores the configuration information of the shared directory.

2.3 Key Technologies

Structured Storage Structured storage was developed by Microsoft for storing hierarchical data in the Windows operating system [3]. It improves disk space efficiency and simplifies software distribution by gathering all the data files into one file. In the structured storage paradigm, storage can contain other storage, just like a directory can contain subdirectories. The `MsgEx.db` and `User.db` files store data using structured storage. Figure 1 presents a sample `User.db` structured storage file.

Table 3. Registry information.

Registry Field	Description
HKEY_LOCAL_MACHINE\SOFTWARE\Tencent\QQ	
Install=c:\Program Files\Tencent\qq	QQ installation path
version=1413.192	QQ version number

MD5 Hashing MD5 [5] is a popular cryptographic hash algorithm that converts a variable-length message into a 128-bit hash. The input message is broken up into chunks of 512-bit blocks. The output consists of four sub-groups of 32 bits, which are cascaded to form the 128-bit hash value. QQ uses MD5 to generate the encryption key from the user's QQ number.

TEA Encryption Tiny Encryption Algorithm (TEA) [7] implements a block cipher that uses a 128-bit key and operates on 64-bit blocks. It has a Feistel structure with suggested 64 rounds, typically implemented in pairs called "cycles." It has a very simple key schedule, mixing all the key material in exactly the same way for each cycle. Different multiples of a magic constant are used to prevent simple attacks based on the symmetry of the rounds. The magic constant 2654435769 (`9E3779B916`) is computed as $\left\lfloor \frac{2^{32}}{\phi} \right\rfloor$ where ϕ is the golden ratio. Although 64 rounds are suggested for security reasons, QQ uses only sixteen rounds of TEA to encrypt the `MsgEx.db` and `User.db` files.

2.4 Forensic Analysis

Analyzing the Registry Table 3 lists the Windows registry information maintained about the QQ application after a successful installation. The QQ installation information may be extracted by exploring the Windows registry.

Decrypting Data One of the important tasks in QQ forensics is to extract the encrypted chat records from the `MsgEx.db` file. Figure 2 presents the structured storage scheme used by `MsgEx.db` to store data. `C2CMsg` stores chat record messages, `SysMsg` stores system messages and `GroupMsg` stores group messages. The message contents themselves are stored in `Data.msj` files. Peer-to-peer messages are stored in `Data.msj` files under the QQ number directory within the `C2CMsg` folder and are indexed by the `Index.msj` file. Group messages are stored in `Data.msj`

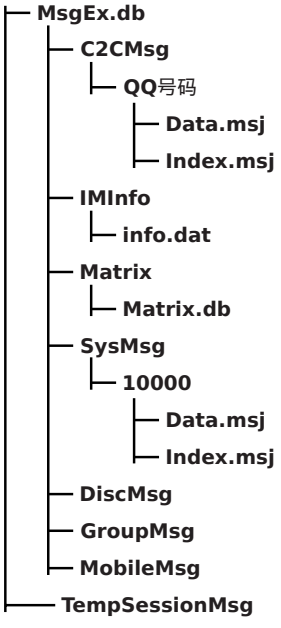

Figure 2. `MsgEx.db` data structure.

files under the directory "SysMsg\10000" and are indexed by `Index.msj` in the same directory. The file is encrypted using TEA.

The primary task in QQ forensics is to extract the decrypted data from `C2CMsg`, `SysMsg` and `GroupMsg`. The following steps are involved in decryption and extraction:

- Obtain the directory and the QQ number and transform the QQ number to the MD5 key using the MD5 algorithm.

- Obtain data from `Matrix.db` for `C2CMsg`, `SysMsg` and `GroupMsg`. Note that QQ often applies padding and permutation. Decrypt the data using sixteen rounds of TEA with the MD5 key.

- Translate the decrypted chat records and friends list into Chinese and display the chat record messages as shown in Figure 3.

3. MSN Forensic Analysis

This section presents key details of the MSN application and outlines a forensic analysis methodology.

Figure 3. QQ chat records.

3.1 Overview

After a user installs MSN and executes the software, several digital traces remain, these include the MSN system configuration, friend's messages and communication messages.

The MSN installation directory is recorded in the registry field:

HKEY_LOCALMACHINE\SOFTWARE\Microsoft
\MSNMessenger\InstallationDirectory.

Each MSN user account has its own configuration settings, which are recorded in the registry field:

HKEY_CURRENTUSER\Software\Microsoft
\MSNMessenger\PerPassportSettings\$\star$

where "$\star$" denotes PassID, which is generated from the user name.

The default path for storing MSN chat records is:

%SysDisk%\Documents and Settings\(Windows login user)
\My Documents\My Received Files\(emailName+PassID)
\history.

The user may change the default path, which is then stored in the registry field:

HKEY_CURRENTUSER\Software\Microsoft
\MSNMessenger\PerPassportSettings\PassID
\MessageLogPath.

MSN chat records are stored in files using the XML format. The filename is of the form `user-nick-name+account-number+.xml` [6], where

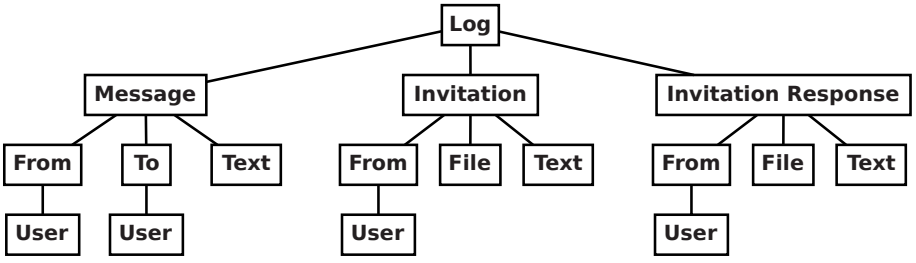

Figure 4. XML file tree structure.

`user-nick-name` is the nickname of conversation counterpart. Each file stores the chat records of the user and file information (name and path) that the user has received. The XML file uses a tree structure to store the chat records (Figure 4).

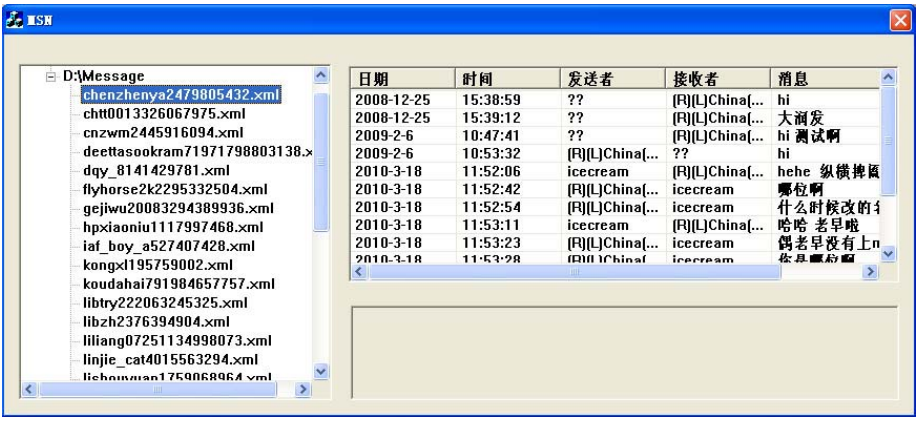

Figure 5. MSN chat records.

3.2 Forensic Analysis

The chat record files may be processed by a standard XML parser (e.g., Microsoft's XML parser) to extract the necessary information. Figure 5 shows example MSN chat records that are obtained using this method.

4. Foxmail Forensic Analysis

This section presents key details of the Foxmail application and outlines a forensic analysis methodology.

4.1 Overview

Foxmail emails are stored in the directory:

foxmail-installed-path/mail/FOXMAIL account number.

Four files are found under the Foxmail account number subdirectory: `in`, `out`, `send` and `trash`. These four files correspond to the email inbox, email outbox, email sent items and email deleted items, respectively.

According to the structured storage paradigm, all the emails in a mail folder are stored in one file. Our analysis revealed that the following byte sequence is used as the header for email in the mail file:

```
10 10 10 10    10 10 10 11 11 11 11 11 11 53 0D 0A
```

4.2 Forensic Analysis

The main task in Foxmail forensic analysis is to search for the email header and extract the mail content that follows the mail header [2]. Next, the extracted mail is exported to the EML format, which can then be processed by any email forensics tool.

5. Case Study

This section describes the use of Enhanced DESK in a case involving the theft of a computer.

A student named Mr. Zhang consigned a logistics company to deliver a computer to his home. The shipment was signed by his father upon arrival at his home. However, upon inspecting the contents of the shipment, Mr. Zhang discovered that his original computer was replaced by a cheaper machine, causing a direct financial loss of 6,500 Yuan.

Mr. Zhang contacted the company about the switch but got no results. He then lodged a complaint with the police, and the case was placed on file for investigation and potential prosecution. The Shangdong Computer Science Center was assigned to examine the computer and obtain digital evidence.

Enhanced DESK was used to clone the computer hard disks and create a backup. The examination of the forensic image revealed more than ten QQ accounts. Enhanced DESK was then used to extract QQ-related information. The recovered chat logs were saved to a `.txt` file for analysis. Photographs were taken of the entire process as subsidiary evidence. Finally, text and photographs from QQ Zone were collected based on information recovered about the owners of the QQ accounts.

Upon reviewing the digital evidence, the police officer assigned to the case confirmed that the suspect was an employee of another logistics

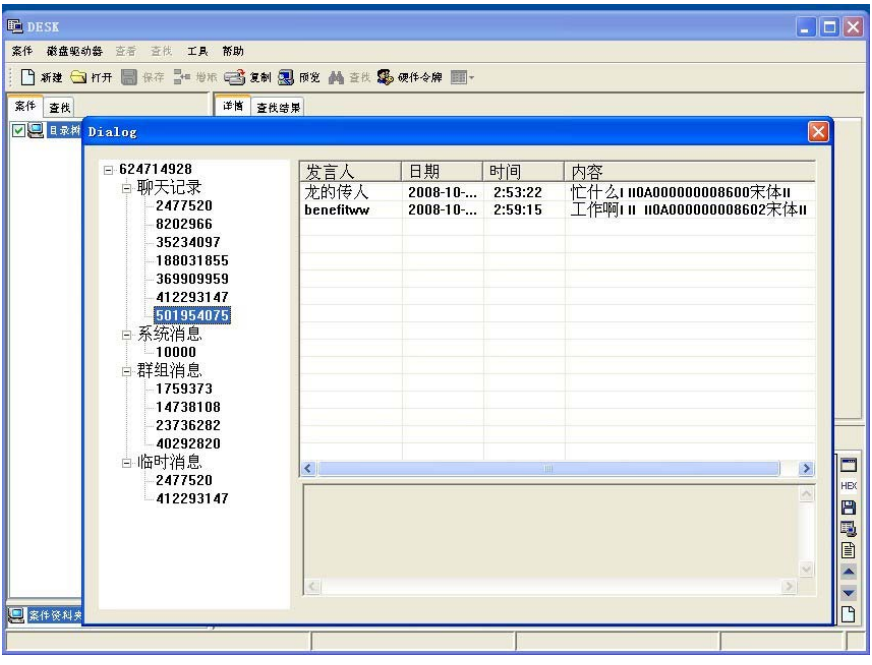

Figure 6. QQ forensic analysis using Enhanced DESK.

company. When confronted with the evidence, the suspect confessed to taking Mr. Zhang's computer and replacing it with a cheap substitute.

Figure 6 shows a sample Enhanced DESK screen dump during a forensic investigation involving the QQ application.

6. Conclusions

The enhanced version of DESK supports forensic investigations of major Chinese Internet communication applications such as QQ, MSN and Foxmail. Our future research will attempt to develop a lightweight version of DESK targeted for forensic investigations in the field.

References

[1] K. Chow, C. Chong, P. Lai, L. Hui, K. Pun, W. Tsang and H. Chan, Digital Evidence Search Kit, *Proceedings of the First International Workshop on Systematic Approaches to Digital Forensic Engineering*, pp. 187–194, 2005.

[2] J. Feng, The implementation of Foxmail email converter by VC platform, *Computer Programming Skills and Maintenance*, vol. 10, pp. 45–46, 2003.

[3] Microsoft Corporation, Structured Storage, Redmond, Washington (msdn.microsoft.com/en-us/library/aa380369(VS.85).aspx).

[4] QQ International, QQ, Shenzhen, China (im.qq.com).

[5] R. Rivest, The MD5 Message-Digest Algorithm, IETF RFC 1321, 1992.

[6] Y. Shi and Y. Zhang, IM system model based on MSNP protocol, *Computer Engineering and Applications*, vol. 41(36), pp. 142-144, 2005

[7] R. Spillman, *Classical and Contemporary Cryptology*, Prentice-Hall, Upper Saddle River, New Jersey, 2004.

[8] Tencent Holdings, Foxmail, Shenzhen, China (fox.foxmail.com.cn).