Forensic research on data recovery of android smartphone

CHANG Xu
Department of Information Science and Technology
Evidence Forensic laboratory in Colleges and universities of Shandong Province (Shandong University of Political Science and Law)
Shandong University of Political Science and Law
Jinan, China
e-mail:changxumail@qq.com

TANG Xin-hua
Department of Information Science and Technology
Evidence Forensic laboratory in Colleges and universities of Shandong Province (Shandong University of Political Science and Law)
Shandong University of Political Science and Law
Jinan, China
e-mail:txhwhu@163.com

WU Jian
Department of Information Science and Technology
Evidence Forensic laboratory in Colleges and universities of Shandong Province (Shandong University of Political Science and Law)
Shandong University of Political Science and Law
Jinan, China
e-mail:jinanwujian@163.com

***Abstract*—With the popularity of the Android smart phone, life of people has become more and more convenient, it also bring new ways for criminals to commit a crime. Data recovery is an important part of forensic, because it can mine potential evidences. In this paper ,we find chance to recover the data which is deleted but not erased after analyzing the structure of NAND and YAFFS2 file system, especially the mechanism of deleting data, this paper put forward a method to recover data of Android smart phone with Yaffs2 file system.**

Keywords- Android; forensic; YAFFS2;data recovery;NAND

I. INTRODUCTION

With the development of communication technology, mobile phone has become life necessities for the people, especially the smart phone with independent operating system of intelligent became more and more popular, Smart phone refers to " kind of mobile phone like a personal computer, with independent of the operating system, can install the software, such as game third party service providers program by the users, can expand mobile phone functions through such program constantly , and can realize the wireless network through the mobile communication network [1]. Because of powerful advantages of smart mobile phone and the promotion of smart phone developers, the permeability of smart phone is still in the growth.

Smart phones offer great opportunities but also cause a lot of problems. While the forensic analysis of ordinary cell phones typically results in a well known set of data (e.g. call history, text messages, contacts, photos) an analysis of a smart phone reveals a plethora of information, because each app stores application-related data [2]. Mobile phone and mobile networks is being used for crime increasingly. Mobile phone and mobile phone related crime do great harm to social, led the law-enforcing department attaching great importance to it. in treatment process of many cases, the related evidence from mobile is became more, generally the mobile phone often retained the important information, which can provides clues and basis to clarify the facts of the case, in some cases, it even become major evidence. Currently, the crime of mobile phone can be divided into three types: one is the implementation of the criminal behavior in the process, the mobile phone is used as communication tools, The second is mobile phone is used as a kind of storage media of criminal evidence; The last one is mobile phone is used as criminal activities implementation tool for SMS fraud, SMS harassment and virus software communication and so on. The data stored on smart phones could be extremely useful to analysts through the course of an investigation [3]. So, for mobile phone forensics research become one of the focuses of criminal evidence.

Android system is based on Linux kernel of the open source phone operating system developed by Google Company, which is the first open and complete mobile software for the mobile terminal. The new smart mobile phone system shows a rapid development trend since first come to market in 2008. according to analyze data of" the 2011 third quarter China mobile terminal market quarter monitoring", which is recently released by EnfoDesk, showed that: Android system mobile phone sales accounted for 58% of the sales on this quarter terminal market [4]. So, the forensics research of Android smart phone is imminent.

II. THE IMPORTANCE OF DATA RECOVERY

Mobile phone forensics is also one of the digital forensics. In NIST (National Institute of Standards and Technology) Guidelines on Cell Phone Forensics in the mobile Phone Forensics is defined: Mobile phone forensics is the science of recovering digital evidence from a mobile phone under forensically sound conditions using accepted methods [5]. During forensics process, the first step of work is to acquire valuable electronic evidence from mobile phones. Cell phone SI M card, memory, external memory CARDS and mobile network operators business database constitute the important evidence in the mobile phone forensics source [6].

Data recovery is an important part of the mobile phone forensics; we can explore potential evidence through the recovery of deleted data. The Android file system is Yet Another Flash file System 2 (YAFFS2). Located in directory "fs/yaffs/". We can search and restore deleted data through the analysis of the Yaffs2 file format in memory, protect the original file for making image of mobile phone memory,

Published by Atlantis Press, Paris, France.

then, analysis the image, look for the data which is deleted but not covered by new data. We can mine potential evidence to recover deleted messages, call records, Internet trace, and the picture etc. with this method.

III. ANALYSIS OF NAND AND YAFFS FILE SYSTEM

A. *Analysis of NAND*

Different from the system of desktop and server, mobile equipment mostly use Flash as a storage medium instead of hard disk, Flash Memory is divided into two types: NOR Flash Memory and NAND Flash Memory. NAND Flash, its characteristics is that it can provide extremely high cell density, can undertake high density storage, and the speed of write and erase is very fast. NAND Flash basic unit is block, each block consist of multiple pages, each page data area size mainly are 512 b, 2 KB, 4 KB, At the end of each page is a spare area (1/32 of data area size) used for storage ECC check code information. Because of different physical data pages by techniques, data block of NAND Flash have the following kinds:

1) Consist of 32 pages, each page for 512 bytes + 16 bytes, namely 16 KB;

2) Consist of 64 pages, each page for 2 KB + 64 bytes, namely 128 KB.

B. *Analysis of YAFFS*

YAFFS File System is an embedded File System which is specialized in NAND Flash design, a good support for the NAND - Flash chip. on each page of NAND FLASH, there are spare space used to store additional information, usually, NAND drive only use part of the space, YAFFS is just use the remaining parts to store the tags which is related to the data. YAFFS is the file system with structure of log, provide a loss balance and power-fail protection, can effectively avoid the accident to the file system consistency and completeness of the influence. YAFFS is designed according to the hierarchical structure design, for the file management interface, internal realize layer and NAND, simplify itself and system interface design, can be more easily integrated into the system. At the present stage has developed two versions, YAFFS and YAFFS2, YAFFS2 is an extension of YAFFS designed to fulfill a new set of objectives to work with newer NAND types [7].

YAFFS2 writes data in unit of chunk sequentially. There are two types of chunk:

Object header:

All data(such as normal file, directory, links, equipment files, etc.) of YAFFS file system are unified as a object to deal with, each object has a chunk Object Header, Object Header save the document patterns, the owner id, group id, length, filename, Parent Object id information. Object Header information is needed to rebuild the file and folder structure [9]. Shown in figure 1. Because of the need to storage all the tags in one chunk, the length of the file name, symbol linked object path name and length should be limited.

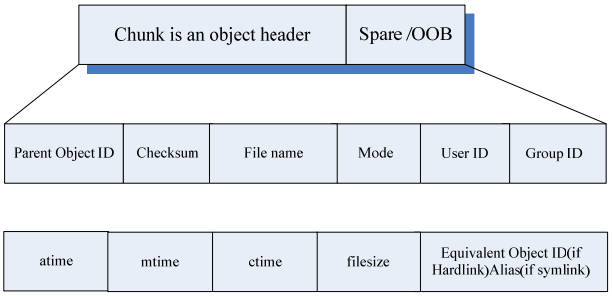

Figure 1. Data Object Header structure expansion of YAFFS

Data chunk: a chunk which contains the normal Data of a file. Structure of YAFFS file system Data and spare space as shown in figure 2:

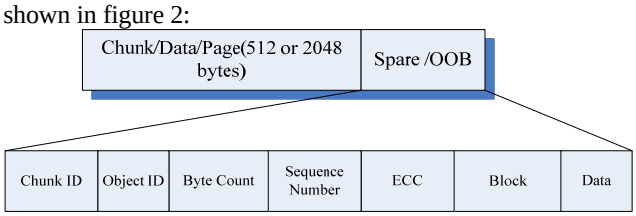

Figure 2. Data and Spare structure of YAFFS

Data structure definition of the YAFFS2 OOB in the file “yaffs_guts”, TABLE I. lists the structure of the YAFFS spare areas for YAFFS2.

TABLE I. YAFFS2 SPARE DATA STRUCTURE

Bytes	Spare of YAFFS2(64 Bytes)consists of:
4	ChunkId(20)(if 0 is a header(directory entry)if>1 is data and position)
4	ObjectId (0 if unused)
2	nBytes,number of bytes used in the chunk,0x0008=0x0800=2048=full
4	Sequence number
3	ECC for tags
24	ECC for data
1	Block status(damaged)
1	Data status(dirty)

After analysis ,we can get each chunk has tags associated with it and those tags contain important information which provided great opportunities for data recovery, several tags which are helpful to recover data listed as follows:

ObjectId: The number used to identify the object [7].

ChunkId: Identifies where in the file this chunk belongs. A ChunkId of zero signifies that this chunk contains an Object Header. ChunkId==1 signifies the first chunk in the file (i.e. At file offset 0), ChunkId==2 is the next chunk and so on, tells where the chunk belongs within the object, there may be several chunks which share the same ObjectId, among them only one is valid and the others hold previous version data.

Sequence Number: As each block is allocated, the file system's sequence number is incremented and each chunk in the block is marked with that sequence number. The sequence number thus provides a way of organizing the log in chronological order. It distribute for each block from start to the end in a sequence way when distributing chunks, until

the all the chunk being distributed. The Yaffs2 sequence number is not the same as the Yaffs1 serial number!

When deleting data, Yaffs2 is just set state on the file structure which was constructed according to tags from OOB from 1 to 0. At this time, chunk on the object is not released or distributed immediately but deleting describes structure on the memory, until the garbage collection mechanism triggered. The specific process is:

1 set the data state tags to 0, demonstrate this chunk has been deleted;

2. Call underlying function write spare space of deleted pages, the data status tag write to FLASH physical media;

3. Modify file page management data bitmap on the memory, set the delete tag to 0.

IV. DATA RECOVERY DESIGN OF YAFFS2

After the analysis of YAFFS2 file system above. At first, we can obtain Android smart phone system image files, then, analyze of the file system format, extraction YAFFS2 file system information, scanning OOB area, set up the index tree and extract system structure information. Extract data delete state tags which is 0. At last, we can recover deleted data according to data analysis, data reorganization which is not covered by new data. The paper presents a method to recover deleted data which consist of three modules; they are Android smart phone mirror, extraction and analysis of data and data reorganization, as shown in Fig.3.respective introduction as follows:

A. Acquire image files

Users must have the root permeations to rooting the device which can obtain when the products (smart phone) leave the factory or installing operating system. Another step prior to collecting data is to enable USB debugging mode in the device. This can be achieved by altering the application development setting in the device to enable USB debugging. then can execute the command “adb shell ls - l” to check file system root directory structure, its similar to the root directory of Linux file system substantially, the cache folder is used for temporary file, data folder is used to store the user’s installed applications and user data of all the application, etc is known as the configuration file storage directory, SD card is SD card of FAT32 file system mount directory, system is very important in the Android file system possession, which is used to store all the tools, library and system application basically. For mobile phone built-in storage, and collect evidence need pay attention to system and the user's all information, through the command “adb shell mount”or “adb shell cat/proc/MTD” to check all of the file system loading condition of phone [8].In this module, we use The Android SDK adb tools to backup memory of smart phone into mirror, use the command “adb pull/dev/MTD/mtd3 Androiddata. Img” to backup to image files of data directory which is named “Android - data. Img” .

B. Extract and Analysis data

Because there is no concentrate index area in YAFFS file system, in order to construct a concentrate index area by scanning the OOB area of chunks during the system startup. When we got the Android mobile image files, first of all ,we need to scan YAFFS file system data storage area, extract data from each page, construct the index tree, system structure information and structure file page management data bitmap.

According to the structure of constructed file management data bitmap, scan and search data bitmap tags whose value is 0, then, calculator corresponding page address respectively, after that, insert into the queue1, and sort Queue1 in the non-decreasing order by ObjectId.

As analysis above, the data in the chunk may be is not erased by new data, we can acquire data which is not covered according to the information we have got in Queue1 and system structure. The specific processes are shown in figure 3 named Extract and Analysis data. Specific steps are as follows:

1) get an address of chunk from Queue1

2) Judge whether extraction is successful or not, if yes, turned to the third step continue to analysis; If no ,we should see the Queue1 is empty or not, if that is not empty jump to step 1,If that is empty Queue1 analysis over, jump to step 10.

3) Jude whether ObjectId is equal to TempObjectId or not. If equal, means analyze the same objectthen jump to step 4 continue; If not ,means previous object analysis is over and start a new one object to carry on analyzing,jump to step step 5.

4) Judge whether count is equal to TotalPages, if equal, demonstrate recovered chunks is exactly equal to the chunk number according to the calculated chunk number by Object Header information,in this cases,we can fully recover the data, jump to step 7 continue processing;If not equal means we had not fully recovered after this document, there may be have more chunks of the current object in Queue1. Turn to step 8 to continue to carry on the analysis.

5) Judge the value of ChunkId of the address which is stored in Queue1,if 0,means this is Object Header.jump to step 6;if not,means this object do not have Object Header,mark this chunk as “undone”,set count and tempId to 0,insert into Queue2,then jump to step 1.

6) Set ObjectId to TempObjectId, and store the related Object header information, named ObjectId and marked as"undone", then,insert into Queue2, then jump to step 1.

7) Extract and store data information, and name block as ObjectId, marked as "perfect", Insert into Queue2, set count and tempid to 0, And Jump to step 1 continue to extract and analyze;

8) If current recovery page number is less than the value which is calculated by Object Header above,at this cases,count++,because current object contains one more rcovered chunk, extract and save data information block, named as ObjectId, marked as "undone", Insert into Queue2; And Jump to step 1 continue to extract and analyze.

9) When extract failure,if Queue1 is empty,means all the free chunk address corresponding data that is extracted above are all analyzed. Jump to step 10.

10) Analyze and restructure the data according to ObjectId which are stored in Queue2,jump to step 11.

11) Process over,exit.

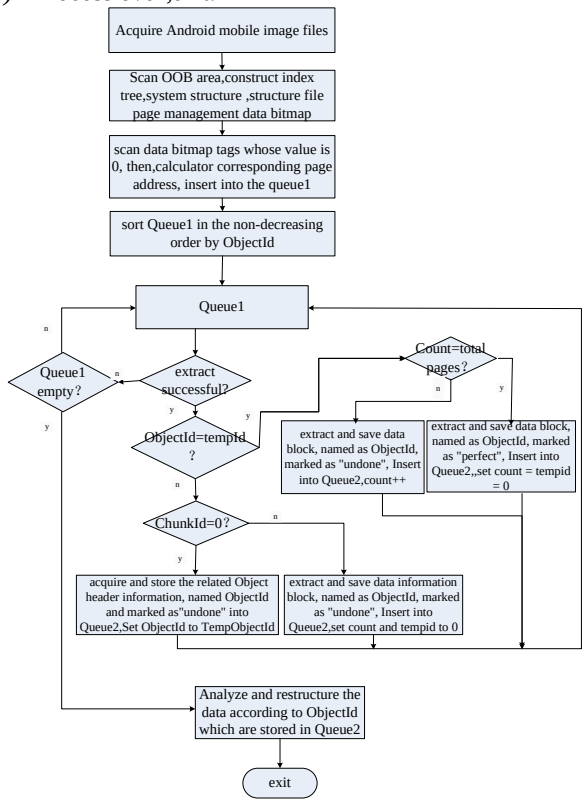

Figure 3. The process of method for recovering data

C. *Data reconstitution*

Extract information of Object Header according to the YAFFS file system principle, we can calculate how many chunks the Object have. Then find the corresponding data, conduct effective combination, add Object Header according to the structure of YAFFS file system, finish delete data recovery. But in reality, some data or head information may be spoiled, left only part of data information or Object Header. In this paper, the recovery method has shown in section 3.2 use Queue2 for storing extracted data, construct data according to the ObjectId as the same object. For the chunk that only has Object Header or Data Information, we marked as “undone”, and we marked “perfect” to demonstrate the object is recovered perfectly.

V. CONCLUSION&FUTURE WORK

After analysis NAND storage structure, YAFFS2 file system structure, especially the mechanism of deleting files. When deleting data, Yaffs2 is just set state on the file structure which was constructed according to tags from OOB from 1 to 0. At this time, chunk on the object is not released or distributed immediately but deleting describes structure on the memory, until the garbage collection mechanism triggered. After analysis, we can get each chunk has tags associated with it and those tags contain important information which provided great opportunities for data recovery. In this paper, we combining work characteristic of ChunkId, Sequence number, ObjectId in OOB area, putting forward a method to recover data of Android smart phone with Yaffs2 file system.

The next step work is extending the method of data recovery by study method to recover incomplete data. Develop a new data recovery software use the method in this paper.

ACKNOWLEDGMENT

This paper is supported by the Scientific Research Project of SDUPSL—“Forensics Research on Android smart phone” (No. 2012Z25B).

REFERENCES

[1] http://baike.baidu.com/view/535.htm

[2] Stefan Maus, Hans Höfken, Marko Schuba ,Forensic Analysis of Geodata in Android Smartphones, Cyberforensics 2011, Glasgow, June 2011; Schuba M.

[3] Jeff Lessard,Gary C. Kessler, Android Forensics: Simplifying Cell Phone Examinations. SMALL SCALE DIGITAL DEVICE FORENSICS JOURNAL VOL. 4, NO.1.

[4] http://www.analysys.com.cn/shujufenxi/119445.htm.

[5] Wayne Jansen,Rick Ayers,Guidelines on cell phone Forensics NIST Special Publication 800-101,May 2007.

[6] DAI Ji-ming, Study on Mobile Phone Forensic and Digital Evidence Acquisition. Computer and modernization.2007.(5).36.

[7] http://www.yaffs.net/documents/how-yaffs-works.

[8] YAO Wei,SHA Jing, Digital Evidence Investigation on Android Smart Phone [J]. Chinese Journal of Forensic Sciences, 2012, 60(1):45-49.

[9] Darren Quick, Mohammed Alzaabi, FORENSIC ANALYSIS OF THE ANDROID FILE SYSTEM YAFFS2 , The Proceedings of the 9th Australian Digital Forensics Conference, December, 2011. 100-108.