Forensic Analysis to China's Cloud Storage Services

Chen Long and Zhang Qing

***Abstract*—Nowadays, cloud storage is becoming increasingly popular among individuals and businesses. At the same time, there are an increasing number of illegal cases about preserving illegal information or stealing the company's confidential data through cloud storage service. Therefore, a study on digital forensic investigation of cloud storage services is necessary. Using two china's cloud storage services(360 and Baidu cloud storage service) as case studies, this paper discusses the types of terrestrial artifacts that are likely to remain on a client's machine and analyses the law of terrestrial artifacts after accessing to the cloud storage. At last the paper proposes a method to investigate and analyze the artifacts for reconstructing the event of user's activities.**

***Index Terms*—Cloud computing, cloud storage, digital forensic, user's activities.**

I. Introduction

In cloud environments the common data and processing power can be shared and distributed across single or multiple datacenters that are spread across a specific geographical area or even the entire globe. The structure and characteristics of the complex to computer forensics work bring huge challenges, in order to adapt to these changes, computer forensic in cloud computing has became an important topic, is of very important theoretical and practical value. The current domestic and foreign scholars on cloud forensics research mainly concentrated in two aspects: 1) In the cloud server design scheme to record user information and the customers can obtain network, process, and access logs over a read-only API on the server. The investigator uses the information to analyze the user's activities. 2) Collected suspicious data from client's machine, then analyze the user's activities.

Shams Zawoad [1] *et al.* introduce Secure-Logging-as-a-Service, which stores virtual machines' logs and provides access to forensic investigators ensuring the confidentiality of the cloud users. Shams Zawoad [2] *et al.* also introduce the idea of building proofs of past data possession in the context of a cloud storage service and discuss how this proof of past data possession can be used effectively in cloud forensics. Li-ping Ding [3] *et al.* has proposed a forensics framework under an infrastructure as a service cloud model. Experiments show that the framework can obtain evidence data in cloud platform effectively and efficiently. Ting Sang [4] *et al.* propose a approach which using logs model to building a forensic-friendly system. Using this model we can quickly gather information from cloud computing for some kinds of forensic purpose.

Darren Quick [5] used Microsoft SkyDrive as a case study, they identified the types of terrestrial artifacts that are likely to remain on a client's machine. Fabio Marturana [6] has discussed technical aspects of digital forensics in cloud computing environments and present results of a case study about user-cloud interaction, aimed at assessing whether existing digital forensics techniques are still applicable to cloud investigations. Jason S. Hale [7] discusses the digital artifacts left behind after an Amazon Cloud Drive has been accessed or manipulated from a computer. Kim-Kwang Raymond Choo [8] used three popular public cloud storage providers (Dropbox, Google Drive, and Microsoft SkyDrive) as case studies to explore the process of collecting data from a cloud storage account using a browser and also downloading files using client software. Darren Quick [9] used Dropbox as a case study, research was undertaken to determine the data remnants on a Windows 7 computer and an Apple iPhone when a user undertakes a variety of methods to store, upload, and access data in the cloud.

Above studies, the first study records user information and stores the logs in the cloud storage service, which requires cloud storage provider change the current design framework of cloud storage. The second study get information from client's machine, foreign research only focuses on the specified cloud storage service and just has a simple analysis of the data generated by using cloud storage service. It doesn't propose the process of collecting and analyzing these data, the universal capability is not very strong. In addition, during cloud forensic investigation, there are always a huge number of suspected data generated by using cloud storage service, the forensic investigators have to spend a lot of time to analyze these data manually.

Generally speaking, the forensic analysis of the client has the following several advantages: the devices can be accessed easily, the cost of forensic analysis is relatively low etc. However, as the existing research on the forensic analysis of the client only focuses on analyzing the historical traces, it doesn't have further processing with the historical traces , and there is also no the corresponding tool to help the investigator finish the forensic analysis task very effectively. This paper focuses on the forensic analysis of the client. In this paper, we use 360 and baidu cloud storage service as case studies to discuss the types of terrestrial artifacts that are likely to remain on a client's machine and analyses the law of terrestrial artifacts after accessing to the cloud storage. Then we put forward a series of methods to acquire the terrestrial artifacts on the user's devices after the user's accessing the cloud storage service as fully as possible and reconstruct event of user's activities by combing logs and history data remnants together. At last, we develop an autopsy tool to help the forensic investigators finish some tasks automatically.

Manuscript received February 12; revised July 15, 2015.

The authors are with the Institute of Computer Forensics, Chongqing University of Posts and Telecommunications, Chongqing 400065, China (e-mail: chenlong@cqupt.edu.cn, zhangqing7441@163.com).

This tool can save the forensic analysis time, greatly improve the efficiency of the forensics.

II. Important Factors in an Investigation

Currently, users usually access cloud storage services through browsers or clients, whether it is the browser or client, it will be a lot of evidence on the user device. This section outlines and provides a rationale for the choice of elements that are prioritized for investigation, among the data collected from browsers and clients.

A. Log Files of Web Browsers

Although there is a difference of kernel structure and the method of storing traces of online activities among different browsers, the information can be recorded in different methods, such as history, cookies and cache. In a browser, uploading and downloading files and logging users' behavior are possible The history record of the browser is an important consideration. There will be a large number of URL records generated by the browser. Analyzing the history record of the browser can indicate that the user has ever used the browser to access the cloud service, but it's not enough for us to know the detail information of user's operation. By analyzing the user's browsing cookie, the investigator can get much useful information related with the case, such as access time, login name, access frequency, operation event and the content of relevant file operated *et al.*

The cache of the browser is the crucial information of the forensic investigation. The browser's accessing the cloud storage service is essentially calling the network APIS, namely, after the client send the request information to the cloud server, the cloud server will send the corresponding reply information to the client. The cache file is actually used to store these response information, including the pictures, Flash, JS script, CSS files and some html files from the site visited. Analyzing the cache files can obtain the detail operation information of the user's using browser to access cloud storage service.

On the Windows system, Internet Explore(IE) is the most famous web browser. Therefore, this paper only focuses on log files of Internet Explorer.

B. Artifacts of Client Application in PC

Most cloud service providers provide the user with client application to access the cloud service. Many files will be generated in the disk of the user's PC after the client software is installed, such as log files, database files, configuration files etc. These files may have many suspicious data. Analyzing and mining these data can get user's activities, reconstructing the event of user's activities, determine the possible event sequences and reconstruct the activity scene. It can be helpful for the investigator to know what and how the event is taking place, then provide the foundation of auditing the user behavior.

The log files contain much key information that the user has requested to the cloud service provider. When the user upload, rename, download or delete a file, some information will be recorded in the log files. We can reconstruct the timeline of a user's activities in cloud storage. The database files usually stores information about folders and files on a PC. The information contain the path of file、the filename、the hash of file、the size of file、the create and modify times etc. The configuration files usually contain the account ID, the username、the email etc.

C. Procedure for Digital Investigation of Cloud Storage

The investigator collects and analyzes data from device that a user has used to access a cloud storage service. There are five steps during the forensic investigation of cloud storage, the detail information is as followings:

1) Analyze the registry in the user's device, obtain the information of the user-installed browser, cloud storage client and the corresponding installation directory;
2) Collect the suspicious data (related with the evidence of targets) of each browser and cloud client stored in the user's device. These data include browser cache, history record, download history, web log of cloud client, synchronous log, database file and configuration file;
3) Analyze and mine the suspicious data to extract user's activities, then standardize the user's activities, the corresponding format is as followings: user's activities = <event ID, event type, the filename, the hash of the file, the time of occurrence>;
4) Analyze and process the standard user's activities data. Firstly, store these data in a dataset, group the similar data, delete the repeated data. Secondly, sort these data by time sequence. Lastly, iterate over this dataset, complete the miss information by reasoning forward and rebuild the event of user's activities;
5) Obtain the event of user's activities according to the requirements, analyze the correlations and rules of user's activities among different time, different targets and behavior intention. Then determine the possible event sequences, reconstruct the activity scene. This work can be helpful for the investigator to know what and how the event is taking place, then provide the foundation of auditing the user behavior.

III. Artifacts of Cloud Storage Services

Here we use two popular public cloud storage providers (360, baidu) in China as case studies to describe the artifacts left in the Windows after a customer has used a cloud storage service. In general, the client application of cloud service use two methods to record use's operation: database and log file. The two cloud services are typical representative of the two kinds of storage ways.

A. 360 Cloud

In the domestic various cloud storage services，360 cloud storage service is one of the most famous cloud storage services. It not only provides a larger free storage space, and has a fully functional and better user experience. The 360 cloud storage service records user's actives by log.

1) Web browser

When using Internet Explorer version 8.0 to open a file, a cache file named intf[n].js is created in the local path. The URL attribute of the cache file begin with "http://pXX-X.yunpan.360.cn/ intf.php?method=", and it also

contains extra information. This extra information in the form of key-value pairs to record user activities, it is shown in Fig. 1.

http://p53-3.yunpan.360.cn/intf.php?method=Preview.getHtmlInfo&fhash=d6e31e02121a01fdb47107a8c05e86e7d470fc62&fname=%E6%B5%8B%E8%AF%95%E6%96%87%E4%BB%B6.docx&pub=0&ck=609ada7600e39a68d8e612126cdfde0c&ofmt=jsonp&callback=QWJsonp1406769490989

Fig. 1. The URL attribute of the cache file after the user open a file.

The method field is the user's action type. The fhash field is the hash of the file on which the user acted. The fname field is the name of the file on which the user acted. The callback field is the time at which the user performed the action.

When using Internet Explorer version 8 to upload, rename, download or delete a file, a file named webclick[n].htm is created in local path. The URL attribute of the cache file begin with "http://s.360.cn/yunpan/webclick.html?u=http%3A%2F%2Fyunpan.360.cn%2Fmy". Fig. 2 shows the URL attribute of the cache file after the user upload a file to cloud storage services by IE browser.

http://s.360.cn/yunpan/webclick.html?u=http%3A%2F%2Fyunpan.360.cn%2Fmy&id=3537848.1651113778616214000.1414724501831.406&buttonid=Upload&t=1414724877888

Fig. 2. The URL attribute of the cache file after the user upload a file.

2) Client software

Some folders and files were created when client software is used on a windows system. The observed folder structure is listed in Table I. Among these folders and files, history.dat、filecache.db and sync.log contain important information.

TABLE I: IMPORTANT FILES AND PATHS

Path	Details
%profile%\Roaming\360CloudUI\sync.log	the client log
%profile%\Roaming\360CloudUI\user ID\ filecache.db	the local cache file
%profile%\Roaming\360CloudWin2\user ID\ history.dat	Information the history of upload
%profile%\Roaming\360CloudWin2\sync.log	synchronous log
%profile%\Roaming\360CloudWin2\user ID\ config.ini	the user information
%profile%\Roaming\360CloudWin2\user ID\ filecache.db	the local cache file
%profile%\Roaming\360CloudWin2\user ID\ history.dat	the history of upload

Firstly, history.dat and filecache.db contains the same information. They recorded the history of the users to upload files in a different way. Secondly, config.ini contains the user name, the account ID, and the user email. Thirdly, sync.log contains some key information that the user has uploaded, edited, opened, downloaded, and deleted most recently. This file contains authentication information, the account ID, IP and the times at which the application started and ended.

Some information are recorded in sys.log when users has uploaded a files shown in Fig. 3. The information include the operation type, filename, the hash of file, the operation time, the client's IP *et al.*

[2015-01-18 16:22:39.103] DLL[3.0.0.1500] DEV[UI 3.7.5.2300] os6.1 ie9 206cca6a7afe7048f4666fbda7646a3d
[2015-01-18 16:22:39.103] SetUser 262965246, type 0
[2015-01-18 16:22:39.107] SetDiskRoot D:\360CloudUI\Cache\262965246
[2015-01-18 16:22:39.710] [resp] user detail. ver 18683, node_count 11930, last_login_ip:113.250.159.87
[2014-11-01 10:16:38.558] status 6(ok) -> 5(monitor)
[2014-11-01 10:16:38.558] [db] Transaction Begin
[2014-11-01 10:16:38.558] [out_upload] [Queue:1] [new] \test.docx
[2014-11-01 10:16:38.862] [upload][192810392] begin: \test.docx, size:10258, fhash 6e553062ce4565dc230e6c598288a8036e42658e
[2014-11-01 10:16:38.862] [req] upload filesize=10258, \test.docx
[2014-11-01 10:16:39.016] [upload][192810392] have, new_ver:1, name:\test.docx
[2014-11-01 10:16:39.016] status 5(monitor) -> 6(ok)

Fig. 3. Sync.log.

B. Baidu cloud

In the domestic various cloud storage services，360 cloud storage service is one of the most famous cloud storage services. It not only provides a larger free storage space, and has a fully functional and better user experience. The 360 cloud storage service records user's actives by log.

1) Web browser

When using a browser to open a file, a cache file named A[n].html is created. The URL attribute of the cache file begin with "http://www.baidupcs.com/", and it also contains some extra information. This extra information also in the form of key-value pairs to record user's activities. It is shown in Fig. 4.

The method field is the user's action type. The md5 field is the md5 of the file on which the user acted. The time field is the time at which the user performed the action. We can't get the name of the file on which the user acted. But we can inquiry the file information from the cache_file table of client app by the value of md5.

http://www.baidupcs.com/doc/d0770031ef57bacc4d312dced0256952?fid=621326181-250528-1069966796137668&time=1417955739&rt=pr&sign=FDTAER-DCb740ccc5511e5e8fedcff06b081203-kZhuu5nQxLYVbCgEBpox8JMZOcE%3d&expires=8h&chkbd=0&chkv=0&method=view&md5=d0770031ef57bacc4d312dced0256952&type=swf&pn=0&rn=1

Fig. 4. The URL attribute of the cache file after the user upload a file.

2) Client software

Whenever using client software to upload a file, edit a file, or delete a file, some information will be stored in database files. The database file structure is showed in Fig5. BaiduYunGuanjia.db sqlite includes six important tables. The backup_file records backup file information using the client. The bache_file records all file information on the server. The download_file records current download file information. The download_file records have been downloaded file information. The upload_file records current upload file information. The upload_history_file records have been uploaded file information. These tables contains some key information that the server_path, the filename, the md5 of file,

the File Created time and modified times. We can reconstruct user's activities through the information.

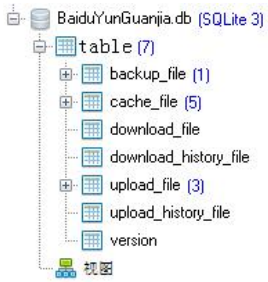

Fig. 5. Database table information.

IV. Case Study of a Cloud Storage Service

A. Case Overview

Suppose one employee of an enterprise disclosed the company's important design documents. According to the investigation, there is likely that the employee used 360 cloud storage service to copy and steal this design document. Except for this, the employee may also change the original file name and delete some documents in order to hide the traces after the crime.

B. Method

The investigator firstly found there is the 360 cloud storage client installed in the employee's PC. Secondly, collect the suspicious data (related with the evidence of a crime) of each browser and cloud client stored in the user's device. The record of accessing 360 cloud storage service from the history record of IE browser was also be found. The investigator obtained the user's activities information and the copied files by analyzing the cache file. Then get user's activities information form sync.log of the client software. At last, standardize the user's activities and used the developed automated tool to rebuild the event of user's activities, analyzed the correlations and rules of user's activities among different time, different targets and behavior intention, extracted the relationship among user's activities. The result provide clues for the forensic analysis.

C. Result

For this case, the forensic investigator can analyze the user's data operation behavior according to the event of user's activities, then determine whether the user disclosed the company's confidential information. The investigator determined the possible event sequences according to the obtained event of user's activities, traced every step of processing each file, then reproduced the crime scene. This work is helpful for the investigator to know what and how the event was taking place, then the investigator can judge whether the user has disclosed this file.

V. Conclusion

This paper analyzes the left traces in the user's device and their storage methods and rules after the user using client application and browser to access cloud storage service in the Windows operating system. These left traces and storage rules are helpful for the investigator to extract completed and reliable evidence information quickly. Then this paper presents a method to reconstruct the use's activities, it can associate different left traces extracted from the user's activities, rebuild the user's accessing cloud storage service, analyze the user's data operation behavior and provide clues for further investigation and analysis. This paper uses Baidu cloud storage service and 360 cloud storage service as case studies, but the method mentioned in this paper can also be applicable to other cloud storage services. This paper mainly focuses on the PC client of using cloud storage service, the similar applications on the mobile device will be our next research work.

Acknowledgment

The research work was supported by National Social Science Foundation of P.R. China under Grant No. 14BFX156 and Natural Science Foundation of CQ CSTC of P.R. China under Grant No. 2011jjA40031.

References

[1] S. Zawoad, A. K. Dutta, and R. Hasan, "SecLaaS: Secure logging-as-a-service for cloud forensics," in *Proc. ASIA CCS'13 Proc. the 8th ACM SIGSAC Symposium on Information, Computer and Communications Security Table of Contents*, New York: ACM, 2013, pp. 219-230.

[2] S. Zawoad and R. Hasan, "I have the proof: providing proofs of past data possession in cloud forensics," *Cyber Security*, Washington, DC, 2012, pp. 75-82.

[3] Y. L. Xie, L. P. Ding, L. Y. Q. Lin *et al.*, "ICFF: A cloud forensics framework under the IaaS model," *Journal on Communications*, vol. 34, no. 5, pp. 200-206, 2013.

[4] T. Sang, "A log based approach to make digital forensics easier on cloud computing," in *Proc. Third International Conference on Intelligent System Design and Engineering Applications (ISDEA)*, Hong Kong, 2013, pp. 91-94.

[5] D. Quick and K. K. R. Choo, "Digital droplets: Microsoft SKYDRIVE forensic data remnants," *Future Generation Computer Systems*, vol. 29, no. 6, pp. 1378-1394, 2013.

[6] F. Marturana, G. Me, and S. Tacconi, "A case study on digital forensics in the cloud," in *Proc. 2012 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, 2012, pp. 111-116.

[7] J. S. Hale, "Amazon cloud drive forensic analysis," *Digital Investigation*, vol. 10, no. 3, pp. 259-265, 2013.

[8] D. Quick and K. K. R. Choo, "Forensic collection of cloud storage data: Does the act of collection result in changes to the data or its metadata," *Digital Investigation*, vol. 10, no. 3, pp. 266-277, 2013.

[9] D. Quick and K. K. R. Choo, "Dropbox analysis: Data remnants on user machines," *Digital Investigation*, vol. 10, no. 1, pp. 3-18, 2013.

Chen Long is a professor and an associate director of the Center for Information Security Technology Engineering, Chongqing University of Post and Communications. He is an intelligent digital security professional committee of the China Association for artificial intelligence, the editor of international journal, the appraiser of computer judicial. His main researched directions include network security, computer forensic, intelligent digital security, published papers in the authoritative journal, such as the Journal of Computer Science, Journal of Software, Electronic Journal. In all, he has published more than 30 papers in the domestic and foreign important journals or conference.