Research on Digital Forensics Framework for Malicious Behavior in Cloud

Guangxuan Chen[1], Di Wu[2], Guangxiao Chen[3,4], Panke Qin[5], Lei Zhang[6]， Qiang Liu[1]

1. Key Laboratory of Public Security Information Application Based on Big Data Architecture, MPS, Zhejiang Police College, Hangzhou China

2. The People’s Procuratorate of Hangzhou, Hangzhou China

3. Wenzhou Public Security Bureau, Wenzhou China

4. Universidad Carlos III de Madrid, Madrid Spain

5. Henan Polytechnic University, Jiaozuo China

6. Joint Services Academy, National Defence University, Beijing China

chenguangxuan@zjjcxy.cn, ericcgx@163.com, qinpanke@gmail.com,

***Abstract*—The difficult of detecting, response, tracing the malicious behavior in cloud has brought great challenges to the law enforcement in combating cybercrimes. This paper presents a malicious behavior oriented framework of detection, emergency response, traceability, and digital forensics in cloud environment. A cloud-based malicious behavior detection mechanism based on SDN is constructed, which implements full-traffic flow detection technology and malicious virtual machine detection based on memory analysis. The emergency response and traceability module can clarify the types of the malicious behavior and the impacts of the events, and locate the source of the event. The key nodes and paths of the infection topology or propagation path of the malicious behavior will be located security measure will be dispatched timely. The proposed IaaS service based forensics module realized the virtualization facility memory evidence extraction and analysis techniques, which can solve volatile data loss problems that often happened in traditional forensic methods.**

Keywords—malicious behavior; digital forensics; cyber crime; privacy leakage

I. Introduction

There is no doubt that cloud computing technology has become one of the hottest topics of discussion in the last decade. It has won the favor of entrepreneurs at a low price and flexible payment method, and it has also brought new development opportunities to more entrepreneurial companies.

Although cloud computing technology has brought many conveniences to people, there are still some problems that need to be solved that hinder its rapid development, such as cloud security issues and digital forensics issues in the cloud environment. Before the security and forensics technologies and standards in the cloud environment were introduced, the use of cloud computing technology to commit crimes and evade legal investigations may become the first choice for some criminals. The well-known Internet security company CA Technologies reports that the Malware as a Service (MaaS) model is becoming a new trend in malware, and almost all (about 96%) Trojan horse programs will be developed in accordance with the MaaS model [1]. The report also said that cybercriminals are gradually using some cloud applications and cloud services to carry out criminal activities, such as Google Apps, Flickr, Microsoft Office Live and so on [2].

Some people have even developed a web application-based infiltration tool, Incognito, which is a typical representative of the MaaS model. It is in the cloud and provides communication services for “underground organizations” [3]. Kaspersky Security Researcher Dmitry Bestuzhev said that cloud services provide a large amount of computing resources for cybercrime, and with these resources, criminals will have a strong cyber attack capability [4]. Subsequently, Dmitry Bestuzhev discovered a way to spread malware using Amazon Cloud Services and claimed that someone had used it to steal user data from nine banks [5]. In the traditional computer environment, if a "hacker" wants to launch a large-scale network attack, it needs to control a certain size of the zombie computers before using them to initiate a network attack at the same time. Thus, "hacker" may take a long time from preparing for an attack to launching an attack, and it is likely to leave clues to digital forensics experts in the process of finding and controlling the zombie computers. However, in a cloud environment, “hackers” can rent computing resources provided by cloud service providers at a low price and then use these legitimate resources to commit crimes. Therefore, for “hackers”, using cloud services for network attacks can shorten their preparation time, reduce the cost of attacks and making it more difficult for forensics experts to conduct investigations.

As can be seen from the above cases, the use of cloud computing technology for crime has become a new cybercrime method, and we have proposed a new topic of how to quickly discover and dispose of malicious behavior and extract criminal evidence in the cloud computing environment. Propose a method of malicious behavior detection and forensics in cloud computing environment, timely discover malicious behaviors in cloud computing environment, and obtain criminal evidence in a timely and effective manner after the criminal behavior occurs, to find out the time of the invasion, the location and the methods of attack, etc., so as to prosecute and investigate the intruders, can reveal the essence of cybercrime in the cloud computing environment, combat the arrogance of such criminals and protect the legitimate rights and interests of the people.

978-1-7281-1907-6/19/$31.00 ©2019 IEEE

In the following sections, we analyzed the challenges of detecting the malicious behaviors and digital forensics issues in cloud environment, and proposed a forensics framework of judging, detecting, analyzing and forensics for malicious behavior in cloud. This provides support for the further research on malicious behavior detection technology, traceability technology, virtual machine introspection technology and forensics technology in cloud environment, and provides reference solutions for related digital forensics research.

II. Current research of digital forensics in cloud

As early as a decade ago, Keyun et al. listed the key issues in the field of digital forensics according to the interviews of more than 150 forensics experts. Birk et al. analyzed the difficulties faced by cloud forensics from a technical perspective [6]; Professor Elie Bursztein of Stanford University believes that the access point of network services and cloud services should contain a large amount of evidence information, and proposed OWADE forensic tool in 2011 [7]. The tool is mainly used for user-side evidence extraction. Due to the limitations of operating system, the practical application of the tool is not widely used in practice; Hay et al. proposed a Xen-based VI tool that can be used to conduct forensics on many of the volatile data such as the list of currently running processes, open network ports, loaded kernel modules, in-memory data, etc. [8]; Chung et al. proposed a method to collect residual data for different access methods from cloud storage applications such as Amazon S3 and Google Docs and believes that a lot of residual evidence data can be discovered through the analysis on logs, catch and database [9]; Martini proposed a cloud forensic framework and Delport proposed an isolated environment construction and forensics method for virtual machine; in the field of big data forensics, Roussev proposed an improved Map Reduce algorithm, MMR. Therdphapiyanak proposed using Hadoop to improve the efficiency of log analysis [10]. In addition, Reilly et al. analyzed the legal obstacles faced by cloud forensics from a legal perspective [11].

Academically, the institute of software Chinese Academy of Sciences proposed a basic framework of digital forensics in the cloud environment [12]; Dr. Zhou Gang from Huazhong University of Science and Technology proposed a cloud forensics method based on live migration technology on the Xen platform. This method regards the virtual machine instance as forensics objects [13]; Li Hui et al. proposed an intrusion event correlation analysis method based on interactive knowledge discovery [14], focusing on the content of the evidence analysis stage in digital forensics; Fu xiao et al. proposed a hierarchical intrusion detection method, Han Zhengping proposed a network intrusion detection technology based on time series correlation, and Liu Zaiqiang et al. proposed a fuzzy decision tree reasoning method or evidence analysis [15~17]. These methods perform correlate time series analysis on web logs or files to discover and capture clues of malicious behavior; Professor Li Jianhua from Shanghai Jiaotong University put forward the research and application of key technologies for computer network information analysis and evidence collection, and provided an effective means for clue discovery and analysis and evidence collection of cybercrime.

Generally speaking, the current forensic research focuses on single-point technology, lacking systematic and structural research and the field of digital forensics research in the cloud computing big data environment is still in the exploration stage. There is a gap between the basic theory, key technologies, product research and development with forensics practice.

III. Definition of malicious behavior in the cloud

Here, we define the malicious behavior in the cloud environment as: any behavior such as stealing cloud data, abusing cloud resources, or illegally acquiring cloud service resources in a cloud environment [18~20]. According to this definition, we classify the malicious behavior of the cloud environment:

(1) Malicious behavior from the outside the cloud (such as malicious attacks from external hackers, etc.)

(2) Malicious behavior within the cloud service provider (such as malicious stealing of user data by cloud service providers or omission of internal management mechanisms to allow user data to be lost, etc.)

(3) The malicious behavior initiated by the user (such as the user's exploitation of his own authority to maliciously abuse other users' resources and data, etc.). In the complex environment of the cloud, it's difficult to differentiate the malicious behaviors between other abnormal user behaviors. For example, in the data sharing behavior, the boundary of leakage of personal is difficult to be determined. The magnitude of the leak is also difficult to indicate whether it is a malicious act. There is no relevant way to determine the boundaries of behavior in the current field of cloud environmental security research.

Based on the above factors, here we define the malicious behavior in the cloud environment in a simple way:

- Any act of acquiring user data from the outside by means of destruction, infiltration, invasion, etc.
- The behavior that causing user data leakage and privacy leakage due to cloud service providers' own reason.
- The cloud service provider's cloud resources are abused by malicious users that violate relevant laws and regulations.
- The failure of providing service within a certain period of time causing loss of user resources and data due to cloud providers' own special reasons as bankruptcy and so on.
- Users use the authority to abuse cloud resources and data that violates relevant laws and regulation.
- The behavior of users maliciously increasing the authority to steal cloud resources.
- The stolen of user account causing user data loss or leakage.

IV. PROPOSED FORENSICS FRAMEWORK FOR MALICIOUS BEHAVIOR IN CLOUD

Based on the understanding the dilemmas of combating malicious behaviors and forensics, and the in-depth analysis of its essential causes, this paper takes the “evidence chain” problem of malicious behavior in digital forensics as starting point, and proposes the “collection-analysis-presentation” framework of digital forensics oriented to malicious behavior in cloud. This framework will raise some common difficult issues for further research.

The framework is shown in Fig.1. It mainly includes three modules: malicious behavior detection module, emergency response module, traceability and forensics module.

A. Malicious behavior detection mechanism

The malicious behavior detection module in the cloud environment is shown in Fig. 2. The solid lines represent the data and the dashed lines represent the scheduling. First, identify the objects that need to be monitored and detected in the cloud environment. These objects include: network devices, metadata information, flow information, messages, file information, and virtual machines. Then, the corresponding processing is carried out on each detection object.

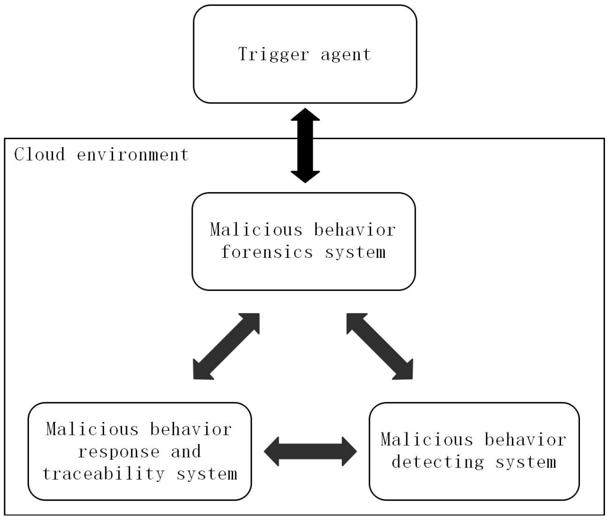

Trigger agent
Cloud environment
Malicious behavior forensics system
Malicious behavior response and traceability system
Malicious behavior detecting system

Fig. 1. The malicious behavior oriented forensics framework.

B. Emergency response and traceability mechanism

As shown in Fig. 3, firstly, we summarize and analyze the relevant detection data obtained, clarify the types of the malicious behavior and the impacts of the events, and locate the source of the event.

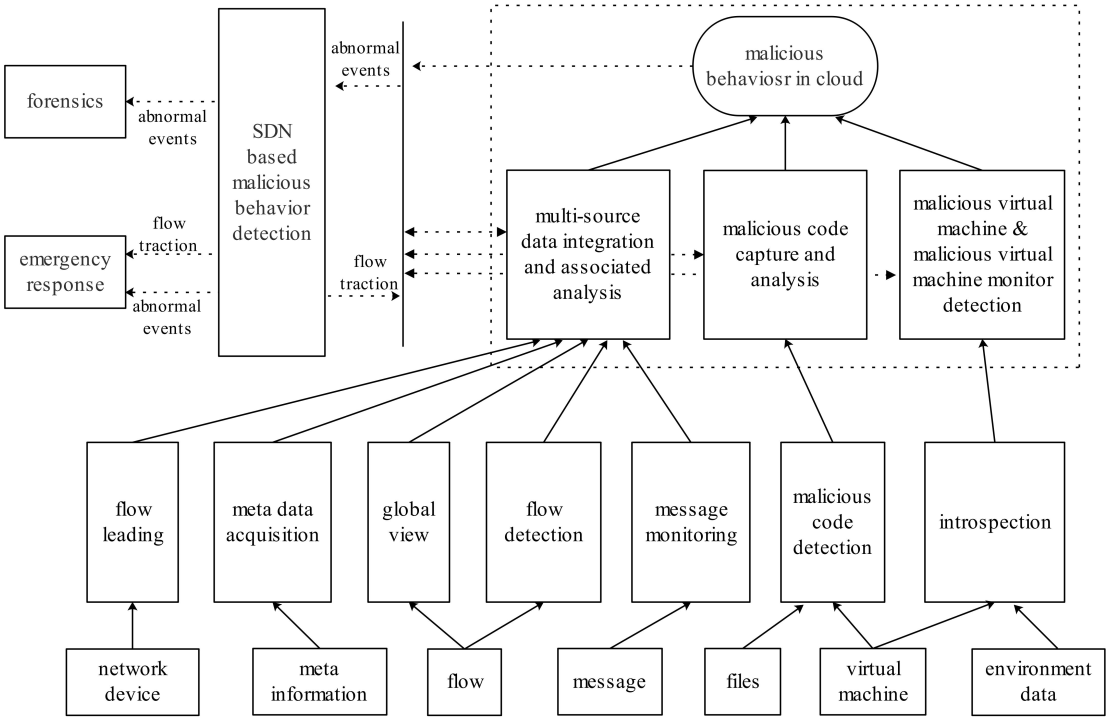


Fig. 2. The malicious behavior detection module.

Secondly, locate the key nodes and paths of the infection topology or propagation path of the malicious behavior, and dispatch measure according to each malicious behavior’s characters. Finally, compare the current security policy with the historical security policy and establish the security policy database which can provide functions such as policy storage, update, and display.

C. Digital forensics module

As for the forensics module, we modified the general forensics model according to the characteristics of the cloud and the requirements of law enforcement norms. Then, we build two forensics systems for cloud infrastructure platform and the cloud storage application respectively.

For the cloud storage applications, firstly, we took DropBox, SkyDrive, 115 network disk, Baidu cloud disk, etc. as examples to study the characteristics of residual data in various operating systems, i.e. the attributes of residual data in

FAT32, NTFS, HFS+, YAFFS2 and EXT4. Then, the evidence discovery method which can automatically identify and collect the local user terminals' evidence data is proposed using machine learning algorithms and data mining techniques. At the same time, the module carry out automatic matching for the digital evidence from user's terminal with the data in the servers, make Hash verification file and produce evidence analysis files in chronological order.

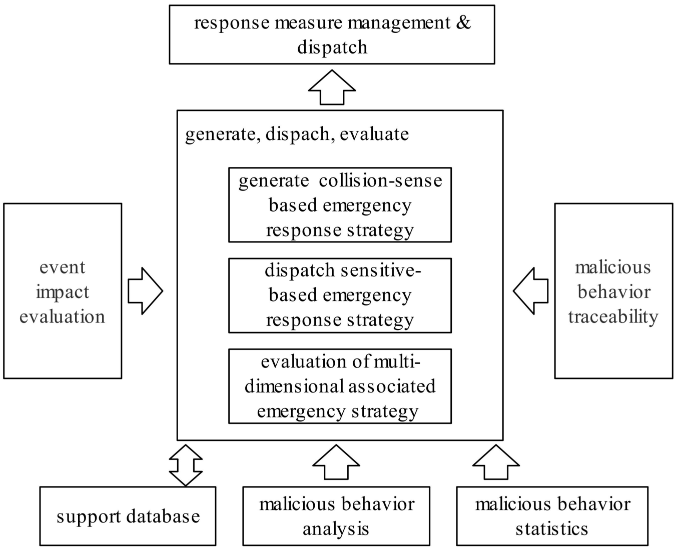

response measure management & dispatch
generate, dispach, evaluate
generate collision-sense based emergency response strategy
dispatch sensitive-based emergency response strategy
evaluation of multi-dimensional associated emergency strategy
event impact evaluation
malicious behavior traceability
support database
malicious behavior analysis
malicious behavior statistics

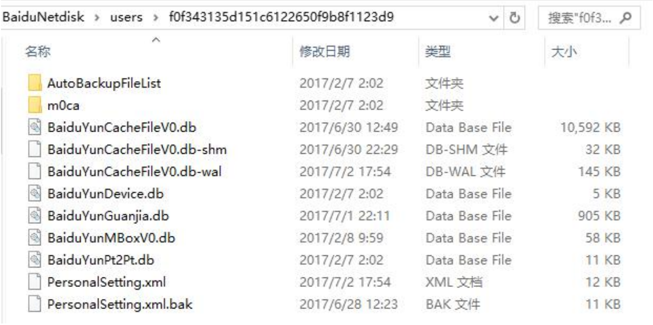
Fig. 3. The emergency response and traceability module

V. EXPERIMENTAL ANALYSIS

Here, prototype system is designed according to this forensics framework. Taking the baidu cloud storage application as example, we conduct a test forensics using the prototype system, i.e., capturing the malicious trace both in the user terminal and the cloud storage, and then analysis the structure and details of the residual data, and match the two samples.

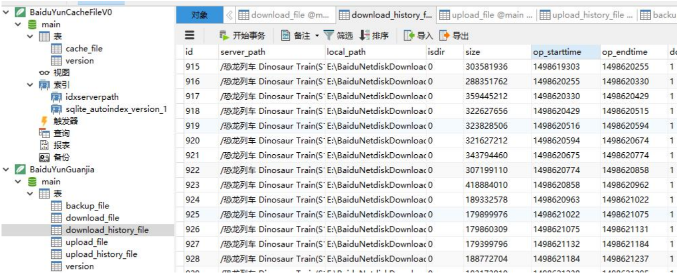
We have omitted some of the previous steps for short. There are two database files, that are BaiduYunCacheFileV0.db and BaiduYunGuanjia.db in the cloud user's folder, show in Fig. 4. BaiduYunCacheFileV0.db records user behavior. Manual analysis showed that there are two tables in BaiduYunCacheFileV0.db, namely cache_file and version. The cache_file contains a list of folder information in the account that is logged in with this computer. The storage path of the related file in Baidu cloud storage, the unique identifier, the file name in the Baidu cloud storage server, the md5 value, and the time of accessing the server can all be saved and exported using Navicat Premium.

The md5 value of the file can be used to verify the consistency of the file in the local terminal and the file in the Baidu cloud storage server. The access time can be used to prove when the computer accessed the file using Baidu cloud storage. There are six tables in the BaiduYunGuanjia.db database, namely backup_file, download_file, download_history_file, upload_file, upload_hisotry_file, and verson. In the table download_history_file, the list of files downloaded in Baidu cloud storage using the account name is displayed, including the file name, size, download starting time and completion time, as shown in the following Fig. 5.



名称	修改日期	类型	大小
AutoBackupFileList	2017/2/7 2:02	文件夹	
m0ca	2017/2/7 2:02	文件夹	
BaiduYunCacheFileV0.db	2017/6/30 12:49	Data Base File	10,592 KB
BaiduYunCacheFileV0.db-shm	2017/6/30 22:29	DB-SHM 文件	32 KB
BaiduYunCacheFileV0.db-wal	2017/7/2 17:54	DB-WAL 文件	145 KB
BaiduYunDevice.db	2017/2/7 2:02	Data Base File	5 KB
BaiduYunGuanjia.db	2017/7/1 22:11	Data Base File	905 KB
BaiduYunMBoxV0.db	2017/2/8 9:59	Data Base File	58 KB
BaiduYunPt2Pt.db	2017/2/7 2:02	Data Base File	11 KB
PersonalSetting.xml	2017/7/2 17:54	XML 文档	12 KB
PersonalSetting.xml.bak	2017/6/28 12:23	BAK 文件	11 KB

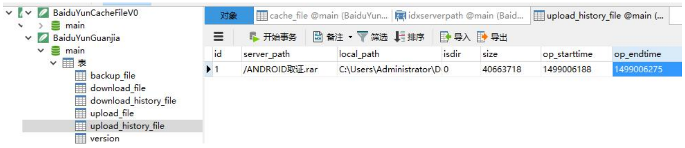
Fig. 4. The structure of the residual data



id	server_path	local_path	isdir	size	op_starttime	op_endtime	d
915	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	303581936	1498619303	1498620255	1
916	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	288351762	1498620255	1498620330	1
917	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	359445212	1498620330	1498620429	1
918	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	322627656	1498620429	1498620515	1
919	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	323828506	1498620516	1498620594	1
920	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	321627212	1498620594	1498620674	1
921	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	343794460	1498620675	1498620774	1
922	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	307199110	1498620774	1498620858	1
923	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	418884010	1498620858	1498620962	1
924	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	189332578	1498620963	1498621022	1
925	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	179899976	1498621022	1498621075	1
926	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	179860309	1498621075	1498621131	1
927	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	179399796	1498621132	1498621184	1
928	/恐龙列车 Dinosaur Train(S	E:\BaiduNetdiskDownloac	0	188772704	1498621184	1498621237	1

Fig. 5. Table download_history_file

The table upload_history_file shows the list that record the file uploaded to Baidu cloud storage by the account. The list displays the file name, size, upload start time and completion time, as shown in Fig. 6.



id	server_path	local_path	isdir	size	op_starttime	op_endtime
1	/ANDROID取证.rar	C:\Users\Administrator\D	0	40663718	1499006188	1499006275

Fig. 6. Table upload_history_file

After the matching is succeed, the evidence information in the residual data can be fixed. Use Navicat Premium to export the above information to an excel table, save the data file, and calculate its hash value. From this, data with evidence value in the residual data is obtained.

VI. CONCLUSIONS

In this paper, we analyzed the difficulties of detecting, response, tracing the malicious behavior in cloud. then, a malicious behavior oriented framework of detection, emergency response, traceability, and digital forensics in cloud environment is proposed. The SDN based malicious behavior detection can implement full-traffic flow detection and malicious virtual machine detection according to memory analysis. The types of the malicious behavior and the impacts of the events can be evaluated and the source of the event can be located through the tracebility module. And the key nodes and paths of the infection topology or propagation path of the malicious behavior will be located security measure will be dispatched timely. The IaaS service based forensics module can finally realized the virtualization facility memory evidence extraction and analysis, which can solve volatile data loss problems that often happened in traditional forensic methods. Further research will focus on the semantic relevance analysis, trusted presentation, and reconstruction of the scene.

ACKNOWLEDGMENT

This work was supported by the Basic Public Welfare Research Program of Zhejiang Province under Grant No. LGF19F020006, the National High Technology Research and Development Program under Grant No.2015AA016003, and the College-Bureau Collaborative Research Program of Zhejiang Police College under Grant No.2018XJY003.

REFERENCES

[1] M. Oh, Y.-G. Kim, S. Hong, and S. Cha, “ASA: Agent-based secure ARP cache management,” IET Commun, vol. 6, no. 7, pp. 685–693, 2012.

[2] Keyun R, Ibrahim B, Joe C, “Survey on cloud forensics and critical criteria for cloud forensic capability: A preliminary analysis,” Proceedings of the ADFSL Conference on Digital Forensics, Security and Law, pp. 105-121, 2011.

[3] Brooks M, Yang B, “A Man-in-the-Middle attack against OpenDayLight SDN controller,” ACM Conference on Research in Information Technology, pp. 45-49, 2015.

[4] Guangxuan Chen, Liping Ding, Jin Du, “Trust Evaluation Strategy for Single Sign-on Solution in Cloud,” International Journal of Digital Crime and Forensics, vol. 10, no. 1, pp. 1-11, 2018.

[5] Guangxuan Chen, Liping Ding, Guangxiao Chen, “Reliable Security Strategy for Message-oriented Middleware,” International Journal of Digital Crime and Forensics, vol. 10, no. 1, pp. 12-23, 2018.

[6] Birk D, Wegener C, “Technical Issues of Forensic Investigations in Cloud Computing Environments,” IEEE 6th International Workshop on Systematic Approaches to Digital Forensic Engineering. Oakland, pp. 1-10, 2011.

[7] D. D. Dinu and M. Togan, “DHCP server authentication using digital certificates,” in Proc. 10th Int. Conf. Commun. (COMM), pp. 1–6, 2014.

[8] Hay B, Nance K, “Forensics Examination of Volatile System Data Using Virtual Introspection,” ACM SIGOPS Operating Systems Review, vol. 42, no.3, pp. 74-82, 2008.

[9] Chung H, Park J, Lee S, Kang C, “Digital forensic investigation of cloud storage services.” Digital Investigation, vol. 9, no. 2, pp. 81–95, 2012.

[10] Guangxuan Chen, Yanhui Du, Panke Qin and Jin Du, “Suggestions to digital forensics in Cloud computing ERA,” 2012 3rd IEEE International Conference on Network Infrastructure and Digital Content (IC-NIDC), pp. 540-544, 2012.

[11] Reilly D, Wren C, Berry T, “Cloud computing: Forensic challenges for law enforcement,” International Conference for Internet Technology and Secured Transactions, pp. 1-7, 2010.

[12] Guangxuan Chen, Yanhui Du, Jin Du, Na Li, Research on digtal forensics in cloud,” NetInfo Security, no.8, pp. 87-90, 2013.

[13] Gang Zhou, “Research on forensic oriented field migration technology in cloud computing environment,” Dissertation of Huazhong University of Science and Technology, 2011.

[14] Hui Li, Chongzhao Han, Qinghua Zheng, “Research on Intrusion Event Correlation Method Based on Interactive Knowledge Discovery,” Computer Research and Development, vol. 41, no. 11, pp. 1911-1918, 2004.

[15] Xiao Fu, Jin Shi, Li Xie, “Hierarchical Intrusion Scene Reconstruction Method for Automatic Evidence Analysis,” Journal of Software, vol. 22, no. 5, pp. 996-1008, 2011.

[16] Zhengping Han, Yan Jin, Taiwei, Chen, “Research on Network Intrusion Detection Technology Based on Time Series Correlation,” Nuclear Electronics and Detection Technology, vol. 27, no. 4, pp. 706-710, 2007.

[17] Zaiqiang Liu, Daidong Lin, Dengguo Feng, “A Fuzzy Decision Tree Reasoning Method for Network Forensics Analysis,” Journal of Software, vol. 18, no. 10, pp. 2635-2644, 2007.

[18] 2015 IEEE International Conference on Communication Problem-Solving (ICCP), 2015.

[19] H. Shulman and M. Waidner, “Towards forensic analysis of attacks with DNSSEC,” IEEE Secur. Privacy Workshops (SPW), pp. 69–76, 2014.

[20] Qiu zheng Ren, Xiaofeng Qiu, Pengcheng Chen, XiaoDong Liang, “The Global Flow Table Based on The Software-Defined Networking,”

[21] Liping Ding, et al, “Research on malicious behavior detection, response and forensics technology in cloud computing environment”, Research report of National High Technology Research and Development Program, 2019.