International Journal of Grid Distribution Computing
Vol. 8, No. 3, (2015), pp.127-134
http://dx.doi.org/10.14257/ijgdc.2015.8.3.13

The Research of Digital Forensics Technologies under Cloud Computing Environment

Huan Ma[1], Gaofeng Shen[2], Ming Chen[1] and Jianwei Zhang[1]

[1]*Software Engineering College, Zhengzhou University of Light Industry, Zhengzhou 450002, China*
[2]*School of Computer and Communication Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China*
songge19840416@163.com

Abstract

As the time of cloud computing comes, Cloud computing is estimated to be one of the most transformative technologies in the history of computing. Cloud organizations, including the providers and customers of cloud services, have yet to establish a well-defined forensic capability. Computer cybercrime has become very prominent forms of crime, but the traditional forensic methods have not met the requirements of the new digital forensics environment. In this paper, upon the use of cloud computing resource sharing, storage distribution and other characteristics and based on the analysis of cloud computing environments difficulties on digital forensics, new digital forensics methods and new digital forensics architecture in the cloud-based platform are proposes to meet rapid forensics needs in the era of cloud computing and to deal with the effectiveness, usefulness, depth issues and real-time and reliability problems.

Keywords: *Digital forensics, Cloud computing, Forensics architecture*

1. Introduction

In our current digital age, it is startling to see the ease with which digital media can and is being manipulated to alter our sense of reality. Whether it is a Hollywood studio, a national news organization, or an average computer user, the images and sounds that are being created can no longer be unquestionably believed. The courts, in particular, are wholly unprepared to contend with the sophisticated digital technology that allows even the most novices of users to alter our sense of reality[1].

Cloud computing technology revolution gives birth to the era of big data, and how to quickly obtain digital evidence needed in the vast amounts of data has become an important issue for electronic forensics experts to deal with [2-4]. If there is no evidence of an effective framework, forensics cycle will be greatly extended and digital evidence will be damages and destructed, thereby affecting the successful investigation of cases. This paper proposes a new method of digital forensics and forensic measure to solve the problem of evidence under the new cloud computing environment. In this paper, a great cloud computing platform resource sharing and scalable on-demand computing power advantages are made use and cloud computing architecture based on forensic evidence is proposed to solve the problem of timeliness.

2. Cloud Computing and Digital Forensics

2.1. Conception of Cloud Computing

Currently, cloud computing technology is one of the most concerned new computer technologies, often referred to as "cloud"[5]. It is an infrastructure of on-

ISSN: 2005-4262 IJGDC

demand delivery resources and charge by using. It makes the goal of computing services clearer, at the same time promoting high efficiency and low cost of such services. Its openness has attracted many developers and researchers, and recognized by the market.

Firstly, we identify cloud forensics as a cross-discipline between cloud computing and digital forensics. There are various definitions for both cloud computing and digital forensics to this date, and in this paper we adopt the current definitions for both cloud computing and digital forensics from NIST: Digital forensics is the application of science to the identification, collection, examination, and analysis of data while preserving the integrity of the information and maintaining a strict chain of custody for the data. Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable resources (*e.g.*, networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. Cloud computing has five essential characteristics, *i.e.*, on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service.

Depending on the division of deployment model, cloud computing has three models: public cloud, private cloud and hybrid cloud, in which hybrid cloud is a special kind of model built up based on private cloud. NIST (National Institute of Standards and Technology) with some of the characteristics of a public cloud authoritative believes that cloud computing can be divided into three levels according to the service form: IAAS, PAAS and SAASO.As shown in Figure 1.

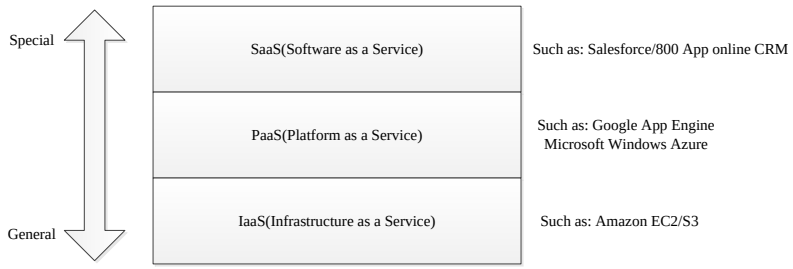

Figure 1. Special To General Process

Secondly, we recognize cloud forensics as a subset of network forensics, as network forensics deals with forensic investigations in any kind of public or private networks, and cloud computing is based on broad network access, thus technically, cloud forensics should follow the main phases of network forensic process with extended or novel techniques. The definition of cloud forensics as the application of digital forensics in cloud computing as a subset of network forensics, as shown in Figure 2.

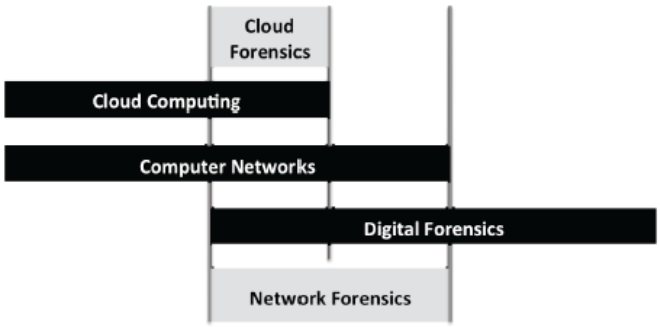

Figure 2. The Definition of Digital Forensics Applied in Cloud Computing

2.2. Digital Forensics

Digital forensics is a branch of forensic disciplines, which includes all of the work process of obtaining evidence from electronic devices and analysis of them. Digital forensics is a relatively new science. Derived as a synonym for computer forensics, its definition has expanded to include the forensics of all digital technology. Whereas computer forensics is defined as "the collection of techniques and tools used to find evidence in a computer", digital forensics has been defined as "the use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and presentation of digital evidence derived from digital sources for the purpose of facilitation or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations".

Digital forensics have been derived from the term of computer forensics, and its oriented object is electronic devices but not just the computer, which is due to law enforcement authorities find in today's environment that people use a variety of electronic devices [6-8]. So the crime is not just computer-related equipment, particularly in the current popularity of intelligent terminals, the source object and criminal offenses are increasing. During the study of digital forensics, many experts has described its concept, but what now widely recognized by everyone, is still DFRWS's definition for digital forensics: "Digital forensics is a process of saving, mobile phones, validation, identification, analysis, interpretation, archiving and reasoning to the data of digital devices with the derived and proven methods, which ultimately facilitates that the investigators reconstruct the chain of evidence and criminal procedure.

Digital forensics has become prevalent because law enforcement recognizes that modern day life includes a variety of digital devices that can be exploited for criminal activity, not just computer system. While computer forensics tends to focus on specific methods for extracting evidence from a particular platform, digital forensics must be modeled such that it can encompass all types of digital devices, including future digital technologies. Unfortunately, there does not exists a standard or consistent digital forensic methodology, but rather a set of procedures and tools built from the experiences of law enforcement, system administrators, and hackers. Palmer suggests that the evolution of digital forensics has proceeded from ad hoc tools and techniques, rather than from the scientific community, where many of the other traditional forensic sciences have originated. This is problematic because evidence must be obtained using methods that are proven to reliably extract and analyze evidence without bias or modification.

3. The Key Technology of Digital Forensics

Digital replication technology. Digital replication can make completely different physical database be shared, including data backup, mirror, taking photos, making videos. Such as the audio and video evidence can use the methods of taking pictures, photography to make evidence of the whole process, so as to increase the credibility. When it needs evidence, through data mirroring method, it can make the backup be quickly copied to another host computer, because it will make the operation more safety, which can ensure the operation safer, by using this kind of method, it can provide the true and effective information. By using this kind of technology, it can provide consumers or businesses with strong evidence in case of the complaint in the future.

Information encryption technology. Information encryption is based on the mathematical algorithm procedure as well as the information of the key to encode information, which can produce the technology of the character that is so difficult to be understood [9]. It can make the information with the feature of confidentiality, integrity and authenticity when it is transmitting, exchanging and storing the electronic commerce information. From the plaintext to the cipher text, the conversion is accomplished by the

key, according to the different keys; it can be divided into symmetric key encryption algorithm and asymmetric key encryption algorithm. Symmetric key encryption refers to the use of the encryption and the use of the decryption keys are the same, so the key is very important, which cannot be lost, such as DES algorithm and IDEA algorithm. While asymmetric key encryption refers to the use of the encryption key and the decryption key are different, the public key is used for encryption, while the private key is used to decryption, such as RSA algorithm. Using encryption technology can almost guarantee the information not to be leaked, not to be tampered during the process of transactions, which is difficult to be attacked by hacking.

The technology of data recovery. Data recovery technology can make the damaged data in different degrees as well as the data in the invisible regions recovered to the normal data, namely, it can be restored to its original figure. The modes of data recovery can be divided into hardware recovery and software recovery. When the computer is suffered with crimes, evidence has been modified and damaged; it can use this technology to obtained the required evidence of the case, which can guarantee the authenticity, integrality of the evidence.

Data interception technology. Data interception refers to the technology of capturing the evidence of a crime as forensics when the crime is occurring. Data interception is generally captured through the medium of transmission, data transmission can be divided into wired transmission and wireless transmission. Cable transmission generally uses the mode of network monitor, using monitor tools (such as Sniffer, TCP, Dump, *etc*.) to obtain information; wireless transmission generally is through electromagnetic wave to capture information. Through the analysis of technology to obtain evidence as the evidence of a crime, to ensure the reliability of the evidence.

Data spoofing. This technology is more used for the evidence forensics among the network attack, in the premise of the offenders who are without awareness, it tries to get the strong evidence of the recorded forensics system during the complete attacking process, path and so on, so as to be the strong evidence to prove the attacking or intrusion behavior. Generally, it uses trap tools such as honeypot . This technology can ensure the evidence to be real, effective, integrity, non-repudiation.

Digital time-stamping technology. This technology can provide complete time prove for the electronic files or online transactions, which also can test whether the electronic files or transaction content with time-stamp is tampered. This technology can provide the safe protection for time, which also can prove the validity of data content, making the data not be tampered and denied, ensuring the certainty of the evidence.

There are many current network forensics methods, and following is a brief introduction of three commonly used network forensics methods:

3.1. IDS Forensics Methods

In 2000, Peter Stephenson proposed that intrusion detection systems can be used in forensics environment. Of course, the idea is not only respected by him, other scholars also very keen on this attempt: the evidence collection and system protection are combined to get more information about the evidence. At the same time this approach can solve real-time and continuity problems of acquirement of evidence. Meanwhile, undo Payer completes real-time forensics model in his research work based on intrusion detection and proposed the ideas that" make sure to save the data in its original format," and "reconstruct of events scene, provide an analysis of the cases of a good start".

3.2. Network Monitoring Forensics Method

Network monitoring method refers to the deployment of a network monitoring system in the network, typically formed by several modules, such as monitoring node, data storage centers and data analysis center. When the abnormal situation happens in the network, the monitoring node will report these abnormalities to data center and stores these abnormal flows, and gradually the original evidence base is formed. Chen Zhen gives a unified threat-based and traffic detection methods network forensics approach in the paper. Meanwhile, the approach has higher efficiency and better results after verification of experiment in terms of fighting the distributed attacks.

3.3. Support Vector Machine Method

Support vector machine is one of the classic feature classification methods. It is mainly used in the feature selection of data packet in the intrusion detection, independent feature filtration and identification to abnormal behavior. Because of its efficiency and accuracy in the nonlinear classification issues, it is the one of the best methods in the intrusion detection and network forensics.

The following is digital forensics process under the cloud computing environment:

(1) Determine the aim and scope of forensics

Determining the aim of forensics means to seize the objectives and significance of this evidence, to ensure compliance with the evidence ultimately achieved with authenticity, legitimacy, relevance so that the evidence can be adopted by the court in the review stage in the proceedings and admitted in the course of cross-examination, and negativity and passivity in the forensics process are prevented. Determining the forensic scope is to determine this evidence taken in the course of forensics is associated with the case, and to avoid taking the irrelevant electronic evidence so that the detection time of cases are wasted.

(2) Determine the data source of forensics

Because data sources of forensics in the cloud computing environment for electronic evidence is multifaceted, some from large-scale cloud computing data centers, some from the cloud service providers, also some from the client's data. Forensics objects of different data sources involved are different. Forensics object in large cloud computing data centre are cloud storage, and forensics object of cloud service providers is relatively small storage and forensics objects of target client is the client's memory, cache, file and so on. Earlier determining forensic source can narrow the range of data sources, so as to accelerate the speed of forensic purposes.

(3) The actual forensic stage

In the cloud computing environment, using forensics software ensures the reliability, completeness and adequacy of the evidence obtained to avoid forensics defect of traditional electronic forensics in a cloud computing environment.

(4) Evidence information processing phase

Due to there are massive electronic data in the cloud computing environment, inevitably it will obtain a lot of information in the evidence collection phase. There are a considerable part of redundant information in the evidence information, so if the information are not processed, the depth and usefulness of the evidence cannot be guaranteed. At this stage, the paper proposes Mahout in Hadoop architecture for mining and processing of evidence in order to obtain useful information on electronic evidence.

(5) Evidence information analysis phase

Traces and illegal evidence obtained from step (V) are selected so that the cases can be figured out fluently and successfully.

The design and implementation of digital forensics analysis based on cloud’s network traffic

Cloud computing platform itself has advantages of high-speed calculation and strong reliability, so it will solve the problems of non real-time, slow speed and week reliability for electronic forensics in big data era if cloud computing technology (including virtualization technologies, Hadoop distributed cloud computing platform) is put into electronic forensics architecture. The benefits of putting cloud computing technology into electronic forensics architecture are following:(1) It can use resources provided by cloud computing platform dynamically, flexibly, and satisfy the needs of forensics. (2) It reduces the demand to forensic equipment and just a browser and installed terminals will be enough, which will greatly reduce the cost of evidence. (3) It can use public cloud forensic evidence to build a private cloud to ensure the confidentiality of the evidence, but when needed, private forensics clouds can be assist and collaborate from the remote among each other.

This paper presents a distributed network traffic forensics system based on Hadoop. According to the function, the system is divided into traffic acquisition layer, traffic analysis layer and storage layer of evidence. The following gives a detailed description for each layer, and proposes an improved algorithm based on past distributions algorithm of type SVM, and verifies and evaluates its effectiveness.

Throughout the model, the first layer is the traffic acquisition layer, whose main goal is to detect the target network using network traffic monitoring tool, and to collect traffic data in real-time which is stored as formatted data so that its convenience of sampling, analysis and archive are improved. In this layer, how we can get real steady flow is very important, because the target traffic to obtain evidence is the first step to obtain evidence sources, which requires to obtain and store in standard way, and also to get legal recognition. The flow got by traffic acquisition layer is directly uploaded to the Hadoop HDFS (Hadoop Distributed File system (HDFS) for storage. As shown in Figure 3.

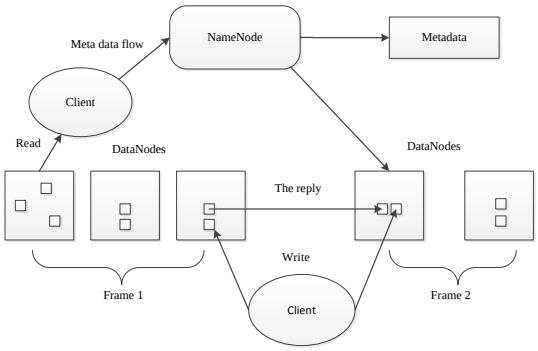

Figure 3. Hpfs Framework

Experimental results and analysis

Experiment on a platform carried out, and the number of training data set were 10,000, 20,000.Compare the result with the stand-alone operation SVM training, 30000 and 20000. The experimental data are shown in Table 1 below:

Table 1. Experimental Dates

The number of training data	10000	20000	30000	40000
The time on the experimental platform	9s	15s	22s	39s
Single SVM time-consuming	14s	44s	99s	176s
Platform accuracy rate	79.6 %	81.1 %	82.5 %	84.6 %
Stand-alone operation accuracy rate	83.2 %	83.5 %	85.4 %	87.3 %

After the understanding of the general process and the core idea of digital forensics and conduction of a more in-depth study, this article will use cloud computing technology in the digital forensics model. And from Hadoop version, distributed SVM algorithm is improved and each layer of cloud computing technology models has this good combination with the cloud computing technology. As an emerging technology, the ability of cloud computing to solve performance bottlenecks cannot be ignored. Meanwhile, the popularity of computer and its lower prices makes the cost of establishing a Hadoop cluster reduced. The algorithms with a way of improving the performance to make up traditional machine learning is insufficient in the face of massive amounts of data, which is an effective way to renew vitality making these classic algorithms.

4. Conclusion

Digital forensics is an application-oriented topic, whose research has always revolved the needs of forensic work, so a good performance, high accuracy and stability are the three basic tasks needed attention. Combining cloud computing with digital forensics solves to find a new solution to many questions. The study results of this paper will provide new research methods and research perspective for electronic forensics researchers.

Acknowledgements

The National Natural Science Foundation of China (U1204609); The Education Department of Henan Province Science and Technology Key Project (14A510011); The Youth Science Foundation of Henan Normal University (2012QK21)

References

[1] M. Reith and G Gunsch, “An examination of digital forensic models. International”, Journal of Digital Evidence, **(2010)**.
[2] M. Sue and A. Glenda, “Describing records in context in the continuum”, The Australian recordkeeping Metadata Schema, **(2000)**.
[3] M. Wen-ping, “Efficient without certificate signature scheme based on ID”, Journal of communications, vol. 29, no. 2, **(2008)**, pp. 87-94.
[4] L. Zhang, “A class without a certificate of the structure of the signature scheme method”, Journal of computers, vol. 32, no. 5, **(2009)**, pp. 940-945.
[5] L. Fengyin and Z. Zhu, “Efficient without certificate signature scheme”, Computer engineering and application, **(2011)**, vol. 47, no. 10, **(2011)**, pp. 23-26.
[6] M. Reim, “CalT&Gregg gunsch an examination of digital forensic models”, international journal of digital evidence, **(2002)**.
[7] M. Reith and G. Gunsch, “An examination of digital forensic models”, International Journal of Digital Evidence, **(2010)**.
[8] S. Liles, M. Rogers and M. Hoebich, “A survey of the legal issues facing digital forensic expert”, Advances in Digital Forensics, **(2009)**, pp. 267-277.
[9] M. Meyers and M. Rogers, “Computer forensics: The need for standardization and certification”, International Journal of Digital Evidence, vol. 3, no. 2, **(2004)**.
[10] N. Beebe, “Digital forensic research”, the good, the bad and the unaddressed, Advances in Digital Forensics, (**2009)**, pp. 17-37.

Authors

Huan Ma, born in June, 1981, Henan, P R china Current position, grades: Lecturer at Zhengzhou University of Light Industry, China University studies: Master degree in computer application technology from Huazhong University of Science and Technology in China Scientific interest: Information processing, algorithm design and analysis

Gaofeng Shen, born in August, 1978, Henan, P R china Current position, grades: Lecturer at Zhengzhou University of Light Industry, China University studies: Master degree in computer application technology from Huazhong University of Science and Technology in China Scientific interest: computer technology

Ming Chen, born in April, 1983, Henan, P R china Current position, grades: Lecturer at Zhengzhou University of Light Industry, China University studies: PhD degree from Beijing University of Posts and Telecommunications in China Scientific interest: big data

Jian-Wei Zhang, born in April，1971，Henan，P R china Current position, grades: Professor at Zhengzhou University of Light Industry, China University studies: PhD degree from The PLA Information Engineering University in China Scientific interest: broadband information network and network security