The Analysis of Visual Forensics in Cloud Computing Environment

Leng Jing
Department of Information Technology
Hubei University of Police
Wuhan, China
e-mail: daleng0127@sina.com

Zhu Bo*
Orthopedics department, Tongji Hospital
Huazhong University of Science and Technology
Wuhan, China
e-mail: Jubal_cn@hotmail.com
(* Corresponding author)

Abstract—New characteristics of cloud computing network environment put forward new requirements of electronic forensics，such as positioning key evidence from big data fast and accurately, analyzing the correlation between evidences, and presenting the results of analysis humanely, which will be the new aspect of research. First, forensics technology new features and challenges of cloud computing network were analyzed. Second, based on the research of integrate framework of cloud forensics, the concept of visual evidence collection and visualization of data object representation were put forward, the data fusion technology was introduced, and visual cloud forensics of evidence was made operable by Web interaction technology. Finally, the forensics process in cloud environment was elaborated in details. The research aims at performing new techniques and applications of electronic evidence collection in a complex cloud environment. Visual evidence collection methods could assist forensic analyzing data and locating the key evidence from huge amounts of data quickly. Visual forensics in cloud environment will greatly improve the work efficiency of forensics, having certain research value.

Keywords-Cloud computing; Electronic forensics; Data fusion; Visualization

I. INTRODUCTION

With the advent of cloud computing and its explosive application in social aspects, computer forensics technology also faces new challenges.

Traditional computer forensics technology is mainly focus on the data recovery and access, rather than on data analysis and interactive presenting.

Existing technology architecture, however, has been difficult to cope with cloud computing forensics. Positioning key evidence from big data fast and accurately, analyzing the correlation between evidences, and presenting the results of analysis humanely, will be the core issue of the future forensics technology.

To solve problems mentioned before, the concept of visualization was introduced to the forensics in cloud computing environment. Visual representation on object's properties and relevance of data facilitates analyzing data and rapid positioning key evidence, and improves the efficiency of the forensics.

II. CHARACTERISTIC AND CHALLENGES OF FORENSICS TECHNOLOGY IN CLOUD COMPUTING ENVIRONMENT

The traditional computer forensics technology is mainly for single and small scene forensics, and data extraction and evidence protection, which data is relatively simple. As a brand new environment, cloud computing environment has many new features beyond traditional network. To find the key information from big data and present it friendly is more important than data acquisition.

A. The definition of cloud computing

The national institute of standards and technology (NIST) defined cloud computing as follows: Cloud computing is internet-based computing in which large groups of remote servers are networked to allow sharing of data-processing tasks, centralized data storage, and online access to computer services or resources. Clouds can be classified as public, private or hybrid.

B. The characteristics of cloud computing

1) Large scale

"Cloud" has a considerable scale, such as one enterprise private cloud typically has hundreds of thousands of servers. This characteristic determines the method of static forensics is no longer applicable in cloud computing environment, only the real-time, high efficient algorithm could adapt to it.

2) Virtualization

Cloud computing allows users to access application services in any place, using a variety of terminals. The requested resources are from the "cloud", rather than a fixed tangible entity, while users don't know where there are. Some application services provided by public cloud of some companies such as gambling are prohibited in our country. That requires our security personnel to embody the "cloud", and shield it outside the domain of our country.

3) High reliability

The cloud computing is more reliable than the local computer, as it has multiple data copies of fault-tolerance, and computing nodes of isomorphism interchange .This characteristic determines the real-time forensics model in cloud computing environment should be extendible.

4) Generality

One cloud computing is not only for a specific application, which can support different applications running at the same time. It increases the complexity of cloud forensics.

5) High scalability

The scale of cloud is dynamical, that meets the needs of the application and the growth of the user scale.

6) The distribution of data storage

As the resource is not local, and data is distributed in many servers of cloud, the data storage is distributed. This

characteristic makes it more difficult for the forensics model in cloud computing environments to monitor the usage of each public cloud, while recording and analyzing the precise location of data storage in the cloud resources timely.

7) Potentially dangerous

Cloud computing services provide not only computing services but also storage services. Cloud computing services are current monopoly in the hands of the private enterprises, and they can only provide business credit. The data in cloud is not secret to those enterprises, which is the biggest potential danger.

C. The challenge of cloud forensics technology

As the characteristics of the cloud computing environment above, it brought some new challenges to forensics, that as follows:

- Due to the regional distribution of cloud computing is wide, jurisdiction become one of the most difficult problem. So it should try to work coordinately, and know the local legal procedures well, to ensure the validity of the evidence.
- Because of the complexity and uncertainty of the data, using visualization methods, pattern matching and other statistical analysis tools to increase reliability of data, in the process of investigation.
- From the Angle of time, space and correlation, analyzing the data acquired could rationalize the data structure.
- Establishing the correct event sequence and clearing data attributes through the public key infrastructure services, the network service discovery mechanism and naming service mechanism.
- Semantic description obtained should be matching with data.
- It is necessary to maintain access to the long-term stability of the evidence, that we need to establish a set of data captured by a maximum extension of information, and describe the mechanism of semantic information, such as using XML.

With the further development of technology, new problems will appear constantly and security personnel should face it seriously in cloud environment.

III. THE INTEGRAL ARCHITECTURE OF CLOUD FORENSICS

A. Integral architecture

The integral cloud forensics architecture can be divided into five parts: the target cloud, forensics agent cloud, other data sources, forensics cloud and terminal. The forensics cloud adopts modular design, mainly including four modules: collecting module, storage module, analysis module and external service module, as shown in Fig. 1. Each module can be run on different computer group, in response to the heterogeneous data processing needs of cloud computing. Modular design can not only simplify cloud forensics system design, but also reduce the control complexity effectively.

B. The working mechanisms of forensics in cloud environment

In cloud computing environment, forensics agent cloud collects the log data .etc from the target cloud, and sends data to collecting module of the forensics cloud. After receiving the data, collecting module sends the data to storage module, at the same time informs analysis module to update data. Then, the analysis module processes data from storage module. External service module receives requests of credit forensics terminal, and informs the analysis module to process. After processing, analysis module returns the results to the external service module. At last, the external service module will sent the results to credit forensics terminals.

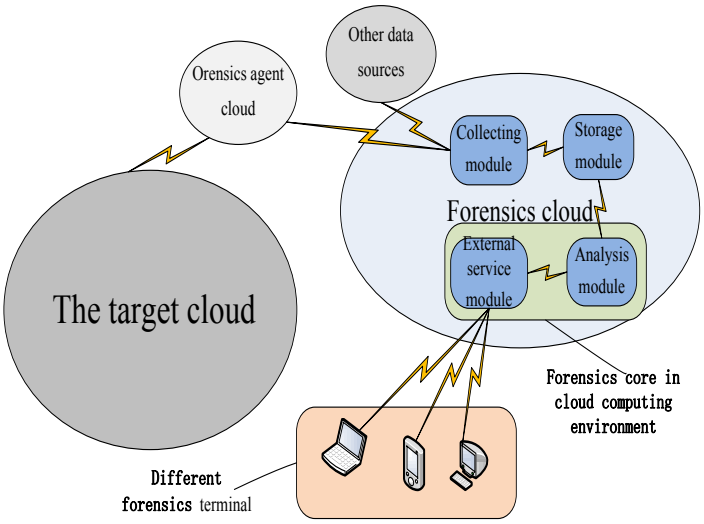

Figure 1. Integral architecture of cloud forensics diagram

Collection module in addition to receive data from forensics agent cloud, can also receive data from his trusted sources.

C. Multiple interfaces design of External service module

External service module adopts the design of Multiple interfaces, in order to satisfy the requirements of C/S and B/S. Forensics terminal can login cloud forensics either in the form of the client terminal, also through the browser.

IV. VISUAL FORENSICS IN CLOUD COMPUTING ENVIRONMENT

Computer forensics technology can be divided into two stages: acquiring the electronic evidence; analyzing and presenting the electronic evidence. Putting forward the concept of visual forensics in cloud computing environment is mainly aims at the second stage.

A. The concept visual forensics in cloud environment

Visual forensics is a technology which can help forensics staff query, analysis data and position key evidence rapidly. It's not concerned about data acquisition methods.

Visual technology applied to the business is mainly to display data in a chart after statistical analyzing, while visualization in cloud environment is very different. The visualization should be associated and interactive.

Association refers to the data presented should be meaningful, associated; interactive means to match the correlation of the data, query, connecting, and so on.

In cloud computing environment, the object of visual forensics is user's behavior pattern. Pattern matching is used to find suspicious behavior and record for forensic analysis. Pattern recognition approach has some limitations that can only identify the known suspicious behavior, and its defects can only be made up through self-learning algorithm slightly. For known behaviors, real-time response according to the predetermined strategy can be used.

B. Visual representation of forensics objects

The data of visual forensics in cloud computing environment is a kind of operable object, which related to each other according to constraint rules of its properties. The correlation between different data may be different. Through this kind of correlation, we can quickly locate from a data object to some other data objects associated, and track to the key evidence from the numerous and complicated data in fast. The rules of visualization of data objects are as follows:

- Using meaningful icons to represent data objects, and view specific data content (only contains useful information for forensics).
- Using the length of line to represent the correlation between objects.
- Using the location and size of the data objects to represent the importance or urgency of the evidence.
- Through Clicking, drawing, dragging and dropping the object icon to locate data and query, etc.
- The associated effect of visualization of data objects is shown in Fig. 2:

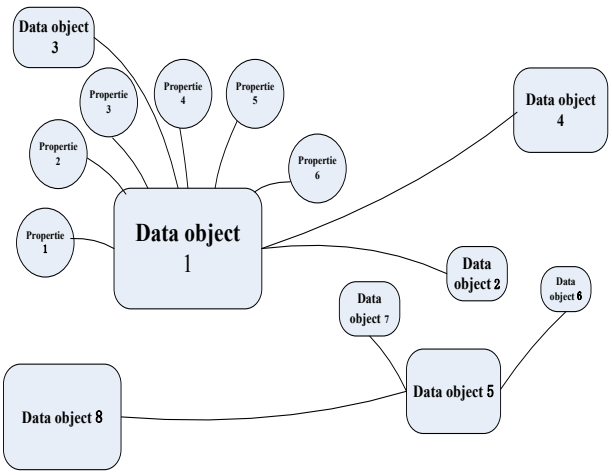

Figure 2. Object associated effect diagram

In the figure above, the data object 1 is associated with a data object 2, 3, 4; data objects 5 is related to object 6, 7, 8; object 1 and object 5 are uncorrelated. Data objects attributes are displayed, indicating that the current focus on the data object 1. It also can be seen from the picture that the suspicious degree of data object 1 is higher than object 5; the relation between data object 3 and 1 is closer than which between object 4 and 1.

Through the correlation of data object, we can quickly locate the relevant evidence, and do correlation analysis.

C. The key technology

1) Data fusion technology

Data fusion technology originated in the military field has developed dramatically over the last decade, and its technology thought gradually applied to other areas. It includes useful information collection, transmission, integration, filtering, correlation and synthesis from the various information sources, in order to assist people in situational or environmental decision, planning, detection, validation, diagnosis.

- The basic principle of work

Data fusion center can fuse information from multiple data sources , also can fuse information from multiple data sources with the observation results of man-machine interface at decision-level. After feature information extracted, and matched with the knowledge from knowledge base under the action of reasoning machine, decision was done and provided to the user. Learning module could be added to the center, analyzing and giving feedbacks to knowledge base, modifying the corresponding confidence factor and updating the knowledge base. Self-learning module can deduce according to the knowledge in the knowledge base and the dynamic result of the questions asked by users to the system, which obtain new knowledge and new experience, expand the knowledge base, and realize the self-learning function of expert system.

- Classification

Data fusion technology has three aspects: data fusion layer, feature fusion layer and the decision-making level fusion, as shown in Fig.3.

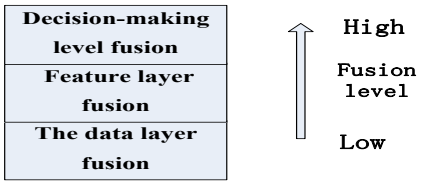

Figure 3. Fusion level diagram

a) The data layer fusion

Data layer fusion generally uses a centralized fusion system process to synthesis and analysis of the raw data. It's a low level fusion. As shown in Fig.4.

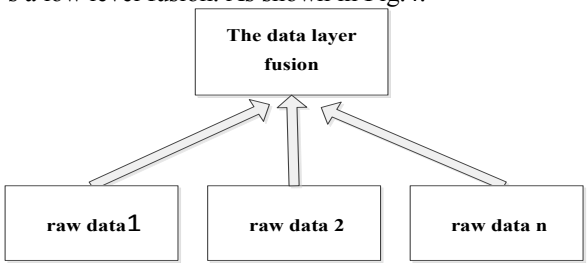

Figure 4. Data layer fusion schematic diagram

b) Feature layer fusion

Feature layer fusion extracts features of original information from the data sources .The feature can be the edge of the goal, direction, speed, etc. Then it makes comprehensive analysis of information and processes the data. The advantages of feature layer fusion is considerable information compression, contributing to real-time

processing, giving maximum the feature information of the decision analysis needs , due to the extracted features is directly related to decision analysis, as shown in Fig.5.

Feature fusion layer can be divided into two categories: target state fusion and target feature fusion, generally adopting the distributed or centralized fusion system. It belongs to middle level layer fusion.

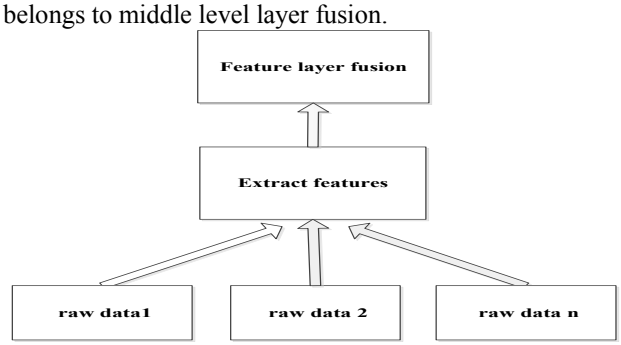

Figure 5. Feature layer fusion diagram

c) Decision-making layer fusion

As shown in Fig.6, decision-making layer fusion observes a target to establish a preliminary conclusion, through different types of data sources, and each data source completes the basic processing locally, including preprocessing, feature extraction, recognition or judgment. After doing decision of decision-making fusion layer through the relevance processing, final result of joint inference is obtained.

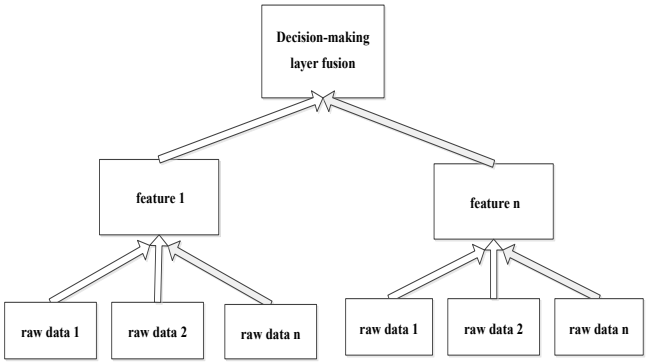

Figure 6. Decision-making layer fusion diagram

2) Evidence of interaction technology

Web front-end technology has the characteristics of "What You See Is What You Get", that can provide perfect interaction experience of visual forensics in the cloud computing environment. At present, the latest Web front-end technology mainly covers HTML 5, CSS 3 and JavaScript.

- HTML5

HTML (Hyper Text Markup Language), is an application of the standard generalized Markup Language. It uses the labels to display object in page. As the latest version of HTML, HTML 5 has many unique features:

a) The semantic features

HTML 5 makes web better meaning and structure. More labels will support micro data and micro formats, to build more valuable data driven Web for the program and users.

b) Equipment compatibility features

HTML 5 provides unprecedented open access to data and application interface. External applications can be directly connected to the data within the browser, such as video audio can be directly associated with microphones and cameras.

c) Connection properties

HTML5 has more effective server push technology, helping us to push data to a client server.

d) Three-dimensional, graphics and special features

Visual effect based on SVG, Canvas, WebGL and 3D function is amazing.

In summary, for visual forensics in cloud computing environment, the semantic features can be used to express the data object; the equipment features can make forensics interactive technology unrelated to platform; the connection properties can make interaction running real-time; the three-dimensional, graphics and special features can support operation of interaction effects and provide rich interface.

- CSS3

CSS (Cascading Style Sheet) can effectively control the effects such as font, color, background and layout of the page, etc. CSS 3 is an upgraded version of the CSS technology, added more new modules.

- JavaScript

JavaScript is a prototype inheritance, object-based, dynamic, case-sensitive client scripting language developed from LiveScript by Netscape, for solving the speed problems of server-side language (e.g. Perl), to provide customers with more smooth browsing effect.

Overall, in visual forensics process of cloud environment, HTML5 is used for description of data objects; CSS 3 is for the layout of the data objects; JavaScript is used for the realization of the interactive logic. JavaScript can manipulate CSS properties of HTML tags, and this feature can be used to achieve to display interactive response results in real-time.

V. FORENSICS PROCESS IN CLOUD ENVIRONMENT

The forensics process in cloud computing environment is shown in Fig.7:

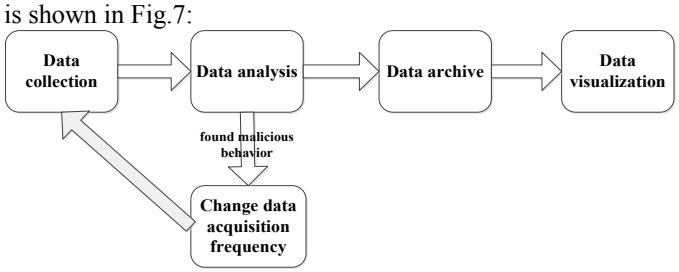

Figure 7. Flow chart of forensics system

A. Data collection

There are two forensics data sources in cloud computing environment: the acquired data and data from the network proxy. The latter will become the main source, with the automation of forensics technology.

B. Data analysis

In cloud computing environment, data analysis mainly works by the ways of correlation analysis and pattern matching in the data fusion technology. Its main job is to establish relationships within recorded data, which is the core of visualization technology.

- Data analysis principle

Operation time, operation permission, user operation and operation properties are the objects of data analysis. Suspicious operation will be found through the data fusion technology and pattern matching. Data analysis process is shown in Fig.8:

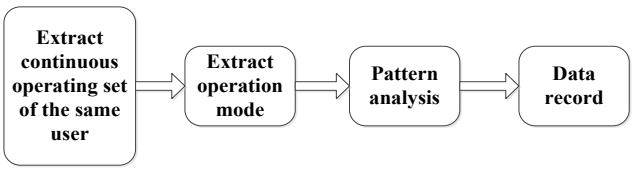

Figure 8. Flow chart of data analysis

Analysis principle: After setting up a time threshold, the continuous logs, whose operating interval is the maximum one of the same user actions that less than the threshold, are set as an analysis object. Then set a similarity threshold, and analyze object permissions, the nature of the operation according to the time change in the pattern matching found its suspected the level of operation, if the layer is less than the similarity threshold, it is not suspicious, conversely as suspicious. It is important to note that the similarity value is greater its fishy is higher.

C. Data archive

In cloud computing environment, data archive mainly is putting the analysis object of data analysis phase into a record, and saving the record in the database according to certain rules, such as in suspicious operation time or suspicious degree. Data archive is convenient for the operation of data presentation. A good archive will greatly improve the efficiency of the system design, and it is the system time bottlenecks.

D. Data visualization

Data presentation in cloud computing environment is realized mainly through Web technology. With the development of computer software technology, Web technology can not only be presented in the browser, but also can be made into client software, running on a PC, mobile phone, tablet and other platforms.

As shown in Fig.9, the data object is shown in the form of icon while HTML tags as the carrier. The operations as clicking on icon, scaling, dragging and dropping and so on, will change the state of a data object in real-time. The correlation can be established between associated data with JavaScript event. At the same time, JavaScript is also the core of the dynamic interaction technology, whose principle is changing the CSS properties of HTML tags to change the state of the data in real time.

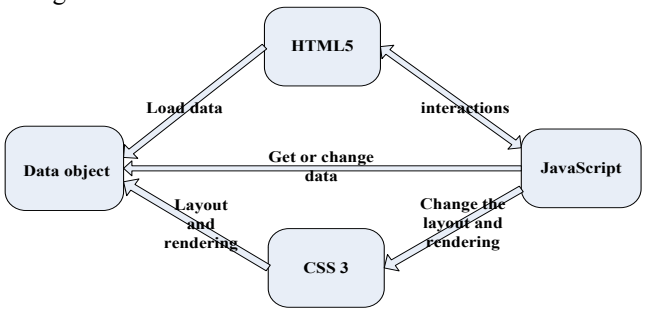

Figure 9. Data presentation

E. Data acquisition frequency

Getting and analyzing log files from different data sources in real time, may bring pressure to network load. In order to reduce the pressure of the network load and load of data acquisition host, and access relevant data timely, the data acquisition time interval will use the principle of dynamic adjustment. The rules are as follows: if the malicious behavior isn't detected, updating data acquisition time for half an hour; if malicious behavior is detected, the time will be adjusted as malicious behavior frequency, until taking the necessary protective measures or confirming not malicious behavior, to restore the initial frequency. It is important to note that when found malicious behavior, data acquisition frequency interval should be slightly smaller than malicious behavior.

VI. CONCLUSION

Cloud computing technology has brought the new change to the Internet, brand-new network application experience for users, and higher requirements on forensics in complicated network environment. In this paper, based on the studies of cloud forensics mechanism, the concept of the visual evidence of cloud computing environment is put forward according to the stages of the electronic evidence analysis and present. The analysis result of forensics process and key techniques such as data fusion, evidence interaction, et al. provides technical support on visual presentation and relevant operation.

The research aims at performing new techniques and applications of electronic evidence collection in a complex cloud environment. Visual forensics in cloud environment will greatly improve the work efficiency of forensic.

REFERENCES

[1] Ding Qufeng, Sun Guozi, "Cloud computing forensics technology," Netinfo security, 2011 (11), pp.36-38 dio:10.3969/j.issn.167-1122.2011.11.010.

[2] Peter Mell ,Timothy Grance, "The NIST Ddfinition of Cloud Cpmputing R National Institute of Standards and Technology," 2011.

[3] Wu Shaobing, "Research of key technologies of Electronic Evidence Forensics based on cloud computing environment," Computer science, Vol.39, No.11A, Nov.2012 ,pp.139-142.

[4] Jiang Zhongyun," The research of intrusion dynamic forensics based on multi-agent network − − access network evidence technology, Jiang nan university, June 1.2006.

[5] " Cloud computing," Wikipedia, http://en.Wikipedia.Org/wiki/Cloud_computing.

[6] Liu Ling, "Discussion on static computer forensics and dynamic forensics ," Network & Computer Security, 2009 (8), p.64-66.

[7] Yin Kang, "Cloud computing concept model and key technology", ZTE technology, 2010 (4), P.18 – 23.

[8] Yin Xiaoming, "Research on business model of cloud computing based on value net," Beijing University of Posts and Telecommunications, 2009.

[9] Wang Peng, "The key technology of cloud computing and applications," Posts & Telecom Press, 2010, P.5-6.

[10] Li Xiaoli, "The electronic evidence recoverable sex research of civil action," Journal of LiaoNing administrators college of police and justice,No.4,2009,p.55-57.