Research on Data Analysis Cloud Service based on Digital Forensic Cloud Platform

Weiping Li *

School of Management, Hebei University, Baoding0 71002, China

*43499622@qq.com

Abstract

The appearance of cloud computing has promoted the development of different application fields and presented new chances and challenges for digital forensics. At first, this paper would discuss the structure and design of digital forensic cloud platform for digital data analysis service. Secondly, according to the application requests of multiple users on the platform, this paper proposed the request scheduling strategy based on queuing model. The experimental results would testify the efficiency of the method based on the structure and design.

Keywords

Cloud computing; digital forensics; analysis cloud service; request scheduling

1. Introduction

With remarkable advantages of low cost, scalability and elasticity, high computing performance, resources integration, multi-tenancy, reliability and security, cloud computing provides users with services via network according to their requirements. Nowadays cloud computing products and solutions become more and more matured and cloud computing became the domain information application mode in many fields. However, in digital forensics, the gradually constructed digital forensics lab system servers are independent and dispersive with low information sharing and high maintenance cost, which causes the imbalance of the distribution of forensics analysis resources. Especially in grassroots departments who deal with basic mirroring and copy of basic digital data extraction and cases input and flow, the forensic equipments are outdated and insufficiently intelligentized, which causes a low forensic capacity[1]. With the implementation of 2013 New Criminal Law, the legislation of digital evidence became more optimized. More and more digital data will turn up due to the increasingly rampant of cyber crimes. The analysis and securing of the abundant digital data would benefit a lot in fighting against network crimes.

Hence, it is a practical and significant research to design a digital forensic cloud platform with a combination of digital data collection, management and analysis, so that to take full advantage of cloud computing technology to solve related digital forensic problems and improve the forensic equipment level of grassroots departments.

As one of the basic properties of cloud computing, multi-tenant technology can realize a high sharing of resources, improve the utilization of the resources and reduce the cost of grassroots departments. Under the multi-tenant environment, different tenants have different requirements of the service property, which requires a standard of measuring service property[2]. Quality of Service (QoS) is a standard for users to measure whether they are satisfied with the service or not. Generally, the increase of tenant number will reduce the property of each tenant, but will reduce the system cost of the service provider. As a result, the service provider is required to provide specific property assurance for each tenant in the case of satisfying their own utilization of resources.

The main tasks of this research include:

1). Design of the all-in-one digital forensic cloud platform with digital collection, management and analysis;

2). On the basis of the platform, propose the scheduling algorithm based on QoS queuing model request to satisfy the QoS of multiple tenants, and testify the efficiency of AQRS algorithm by simulation experiments.

2. Design and realization of data analysis cloud service based on digital forensic cloud platform

As shown in Figure 1, the research target provides digital data analysis cloud service based on the digital data management cloud platform. Digital data collected from different channels are stored in the digital data management cloud platform and formed up the digital data centre. Users can visit the digital data analysis cloud platform with good data visual display to obtain the data information that needs to be analyzed. Network crime evidences can be obtained from the uploading, processing, reappearing and analysis of the digital data, and can be used for the detection of cases or handling of the cases for judicial organs and administrative departments. In addition, digital data forensic departments can also optimize their forensic resources and service configuration.

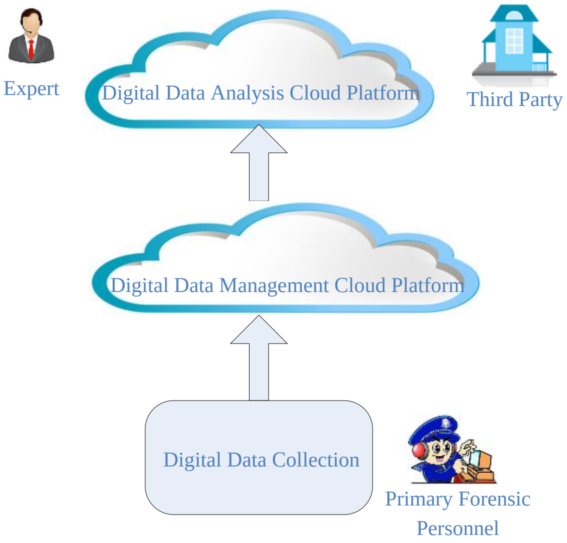
Expert
Digital Data Analysis Cloud Platform
Third Party
Digital Data Management Cloud Platform
Digital Data Collection
Primary Forensic Personnel

Figure 1 Application Environment of Digital Data Analysis Cloud Service

2.1 Design of data analysis cloud service based on digital forensic cloud platform

BI (Business Intelligence) uses DW (Data Warehouse), OPLA (On-line Analytical Processing) and DM (Data Mining) technologies to obtain useful information from the data to support business operation and management decisions. It is used to analyze the digital data in this research. However, the old BI cannot satisfy the analysis of current digital data due to the complexity of the data and rapid increase of data volume. Hence, we should establish a cloud-based BI that can support the digital data analysis by taking advantage of the high computing ability, extendibility and low cost of cloud computing[3]. The structure of the data analysis cloud service platform based on digital forensic cloud platform is shown in Figure 2.

The analysis service cloud platform consists of 4 parts:

Data layer: store and management of digital data from each grassroots forensic departments, including organization management information, case information, case digital data, as well as data warehouse or data mart constructed by data formatting and data classification, providing data source for OPLA analysis and data mining[4].

Analysis layer: Provide all data processing and analysis capability in a safe and configurable way. Analysis mainly consists of analysis engine and operation engine. Analysis engine contains all the software and requirements for the extraction of digital data analysis, including related engines that support the analysis, to assure the flexibility of the structure and satisfaction of the platform users.

Visualization layer: Obtain digital data through OPLA, data mining, simple statistical computing and current forensic software, and present to cloud platform users in a simple, visual and easy way.

Transport layer: end users can obtain related access through Apps and download corresponding functions in the transport layer.

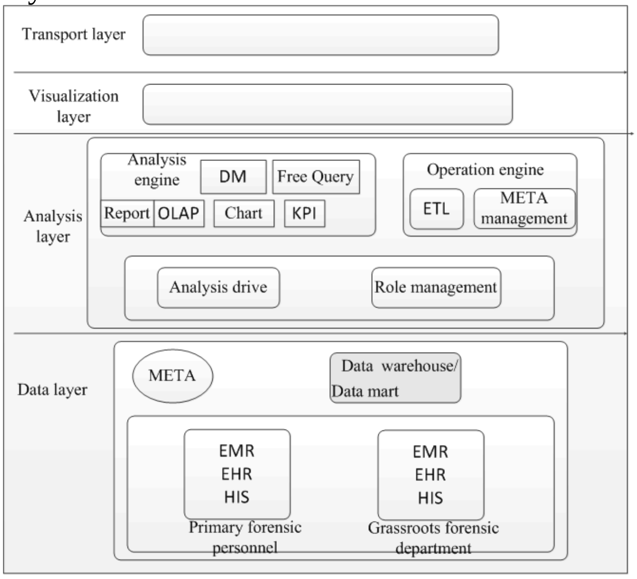

Figure 2 The structure of the data analysis cloud service platform

2.2 Realization of data analysis cloud service based on digital forensic cloud platform

As a new service mode in IT industry, the service forms of cloud computing include Saas (Software as a Service), PaaS (Platform as a service) and Iaas (Infrastructure as a Service). For users, the 3 service modes are independent. As they provide completely different services to different users[5]. Technically, however, there exists a dependent relationship between each of the service mode of the cloud service. For example, except their own technology, the product or service of SaaS also need PaaS to provide development and deployment platform for them or deploy directly on IaaS. The product or service of PaaS is also constructed on the basis of IaaS, and this is just the service deployment of this paper.

Virtualization and monitoring of digital data analysis platform

Resource integration and utilization is one of the advantages of cloud computing. While virtualization is the key of cloud computing. Compared with traditional computing mode, the important property of cloud computing is to use virtualization technology to take the computation, storage, network and applications in the cloud computing as a resource and provide to users. Virtualization technology mainly includes server virtualization, storage virtualization and network virtualization. The research in this paper is a server virtualization[6]. On account of Linux users, open source management programs that have realized virtualization include KVM and Xen. KVM is a Kernel-based Virtual Machine which is a completely native and fully virtualized solution of Linux. It is X86 hardware based on virtualization extension. KVM has succeeded the strong memory management function of Linux and can store any VM image supported by Linux and it also takes the property and scalability of Linux. Hence, considering the high security requirement of digital forensic applications, KVM can be used in IaaS to realize the virtualization of computing resources. In order to improve the operation and service quality of cloud computing platform, we also need an overall and efficient monitoring on the use and operation of all the resources on the cloud platform[7]. An efficient monitoring can realize effective scheduling strategy, property prediction and bottleneck analysis, etc. In this paper, we use Zabbix as the platform monitoring software. As an open source solution, Zabbix has a distributed system monitoring function and provides monitoring on CPU load, memory usage, disk usage, network status, port use and log, etc.

Multi-tenancy of digital data analysis platform

Multi-tenancy is one of the key technologies of cloud computing. The key point to realize multi-tenancy is to isolate the application environment and data of unused tenants. The isolation of application environment can be realized by the isolation of progress or hosting environment that allows multiple applications operate simultaneously (e.g. Webserver). There are 3 solutions for the isolation of multi-tenancy in data layer:

1) Data base isolated. Each tenant possesses their own data base. With this solution, the isolation level is high, but the cost is also very high.

2) Data structure isolated but data base shared. Each tenant has their own mode, but they share the same data base. This solution can improve the sharing of data base but the isolation level is reduced.

3) Data structure and data base shared. Each tenant shares the same data base with same mode. As shown in the chart, tenant ID is added to isolate the data. This solution has the highest sharing level but with lowest isolation.

Considering the requirements of digital data on isolation and safety, we adopted the independent data base solution. The structure of the mixed multi-tenancy model based on resource sharing is shown in Figure 3.

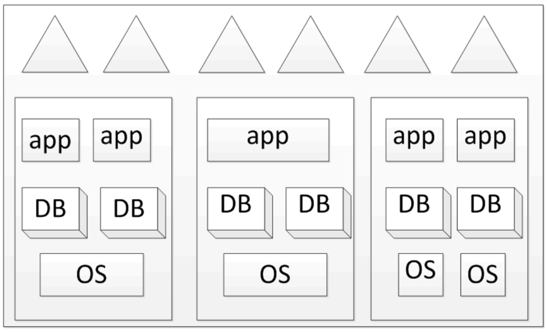

Figure 3 Structure of mixed multi-tenancy model

The mixed multi-tenancy structure in this paper is designed by combining multi-tenancy application layer and data layer isolation, which supports 3 kinds of multi-tenancy implementation models.

1) Application examples isolated; data base isolated; address and space shared;

2) Application examples shared; data base isolated; address and space shared;

3) Application examples isolated; data base isolated; address and space isolated;

3. Request scheduling strategy based on queuing model

In order to satisfy the QoS of each tenant on the multi-application and multi-tenancy platform and improve the utilization of the cloud resource, this paper proposed a request scheduling strategy based on queuing model. The chart of the strategy model is shown in Figure 4.

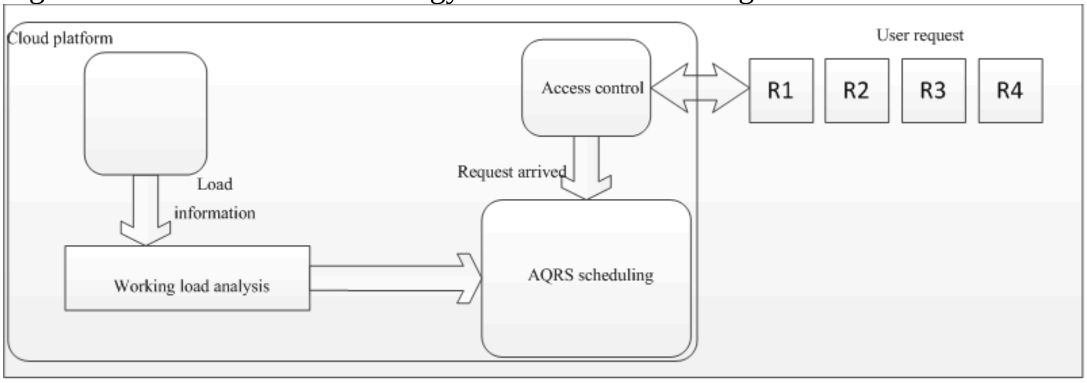

Figure 4 Request Scheduling Strategy Based on Queuing Model.

Working load analysis is usually applied in predicting the request arrival rate to calculate the key parameters used in AQRS scheduling method to satisfy QoS of tenants and assure the efficient utilization of cloud resources. [8] Different data can be used in the prediction. For example, all kinds of historical data collected by platform monitoring software can be also used for a comprehensive analysis. In the application layer, the access control based on user request can satisfy M/M/m queuing model for the whole ASHC platform. But for a single VM, M/M/1/L queuing model is adopted. In M/M/1/L queuing model, when there are L requests in the queue for each application case, tenant's new request will be rejected so that to assure the QoS of each tenant. The accepted request will be implemented by the AQRS scheduling algorithm proposed by this paper. L is the queue size, which can be calculated from Formula (1). Ts refers to the negotiated service time and Tr means the real working time of a single request. If the request in the VM request queue is more than L, new request will be rejected by access control and not implemented. So each request is rather rejected or implemented.

$L=|Ts/Tr|$(1)

The scheduling strategy of M/M/1/L queuing model uses AQRS （Application-level QoS-based Request Scheduling）algorithm as proposed by this paper, where Tc refers to request processing time, Td stands for request blocking time.

AQRS algorithm is shown as follow.

Algorithm 1 ： AQRS

Input：TC,Td,Count

```
While(true){
 Count=0;
 Requestr=retriveRequestFrom Queue();
 While(true){
  executeRequest(r,Tc);
  If(r.isFinished()){
   break:
}
Count++;
While(true){
 Sleep(Td);；
 If(oktocontinue()){
  break；
 }
}
}
}
BooleanokToContinue(){
 Set〈Thread〉 threadPool=getAllThrwads();
intminDelay=MAX_INT;
For(Thread t:threadPool){
 If(t.isRunning()){
  If(t.Count{minDelay){
   minDelay=t.Count;
   }
  }
 }
 If(Count==minDelay){
```

```
Return true;
}
Return false;
}
```

According to the AQRS algorithm, the blocking count of each request is taken as the relative priority of the request. The request priority is the lowest when the count is 0. For request with higher service demand, the count will increase with the processing of request and the priority will go up. When the count reaches a specific value, that is when the request processing is hold on, preemptive method will be used to obtain processing authorization so that to assure the QoS of tenants of task request submitted. Among them, the value of Tc and Td is the key to influence the property of the algorithm. Previous study [9] showed that Tc=100ms is a proper option and better property and lower cost can be obtained when Td=5ms. In AQRS algorithm, count value is the key for the request to be processed preferentially and satisfy the QoS. It is recommended to set the count between 1 to 7 to gain better QoS.

4. Results and discussion

This paper uses the cloud computing environment simulation software Cloudsim to design the experiment and testify the proposed algorithm in the paper. At first, we expanded and recompiled Cloudsim and simulated the AQRS algorithm. Then we analyzed the experimental results.

4.1 Introduction of experimental environment

Experimental environment mainly consists of experimental hardware environment and hardware configuration of simulation cloud platform. Experimental hardware environment configuration is shown as Table 1.

Table 1 Experimental Hardware Environment Configuration

Operating System	RAM	Hard Drive	Development Platform	Development Work
Windows	4G	500G	Eclipse	Cloudsim3.0.3

Simulation cloud platform resource configurations are shown in Table 2 and Table 3 respectively.

Table 2 Host Configuration

Host Number	RAM	Hard Drive	Operating System	Bandwidth
2	89G	5T	Linux	4M/s

Table 3 VM Configuration

VM Number	RAM	Hard Drive	Operating System	VT Type
6	6G	500G	Linux	KVM

4.2 Discussion on the experiment results

In the simulation experiment, one cloud task represents one tenant's request. A certain number of requests will be produced randomly. Repeat the submitting 10 times to gain an average value of the experiment results. In this paper, we use the default scheduling algorithm used in the scheduling model named FCFS (First Come Fist Service) as the comparing algorithm.

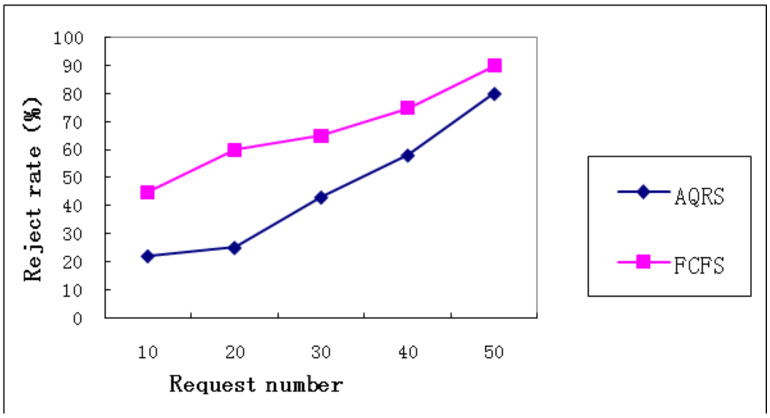

Figure 5 Request reject number under AQRS and FCFS algorithm

In the experiment, the request reject rate, average request response time and CPU utilization of AQRS and FCFS algorithm were measured by changing the number of requests. The result of the experiment showed that, compared with FCFS algorithm, AQRS algorithm has improved the satisfaction degree of tenants and the response time is also assured. But the improvement in the utilization is not apparent as some parts of the resources were used to promote the response time and satisfy the QoS of tenants.

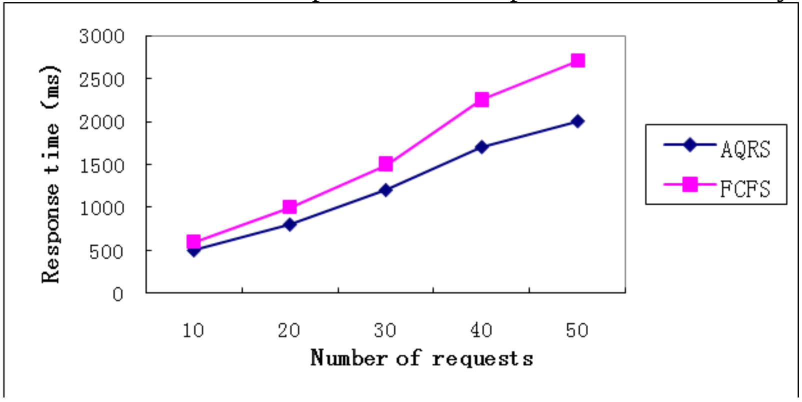

Figure 6 Average request response time under AQRS and FCFS algorithm

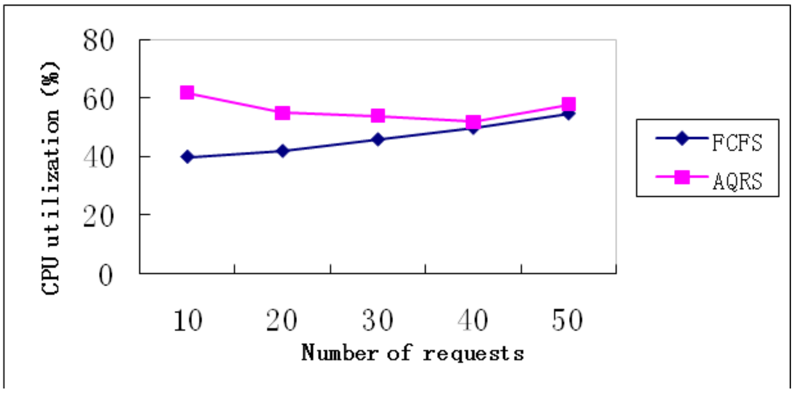

Figure 7 CPU utilization under AQRS and FCFS algorithm

5. Related research

Cloud computing technology has been actively developed and utilized by MPS (Ministry of Public Security) and institutions for academic research in the applications of digital forensics. [10]Many researches were done on how to use cloud computing to establish digital data information cloud platform, analyze the digital data and request for scheduling algorithm on the platform.

analyzed the challenges that cloud forensics were facing and proposed a forensic frame ICFF under IaaS cloud model. It was realized on the open resource IaaS cloud platform of Eucalyptus. At last, ICFF was analyzed and testified through experiment.

stated that, under cloud computing environment, the cloud computing center provides direct services consisted of infrastructure, platform and applications, and users share the whole cloud infrastructure. However, this method has a direct impact on the safety and availability of cloud computing environment and may bring huge hidden danger to cloud computing. It proposed a method with digital forensic technology to solve the safety problem of cloud computing environment and satisfy the demand of traditional forensics on high property computing by using cloud computing technology and super computing technology.

combined the cloud computing with BI (Business Intelligence) and raised the BI system structure based on cloud. This paper will conduct the medical data analysis with cloud-based BI. The task scheduling of system layer and user layer is one of the hot points of cloud computing research. presented the cost-based task scheduling algorithm which divided groups according to task cost before implementing. On account of SaaS application, [15] proposed SLA-based priority scheduling algorithm according to different SLA levels. It has two objectives: higher priority tenants are prior to lower priority tenants on the premise of QoS; make less influence on lower priority tenants while QoS of higher priority tenants is assured. [16] On the basis of the HTTP request property, the AQSR algorithm raised by this paper considered the user experience and no response was activated for long task request.

6. Conclusion

This paper proposed a data analysis cloud service platform based on digital forensic cloud platform to solve the imbalance of digital forensic equipment and technology by taking advantage of the advanced cloud computing technology. Then a remote analysis was conducted on the digital data in the way of cloud service and the forensic problems caused by the imbalance of digital data forensic personnel levels were solved, which promoted the development of digital data forensics. Consequently, we made a deployment of the platform and designed a multi-tenant structure which can take full advantage of the cloud resources. Then we conducted a study on the multi-tenant request scheduling in the applications under such cloud platform structure, proposed a scheduling strategy based on queuing model and testified the efficiency of the AQRS algorithm on account of the platform.

Acknowledgement

This work was supported by the natural science foundation of Hebei province (F2014201117), the natural science for youth foundation of Hebei province (F2014201165).

REFERENCES

[1]Grispos G, Storer T, Glisson W B. Calm before the storm: the challenges of cloud[J]. Emerging Digital Forensics Applications for Crime Detection, Prevention, and Security, 2013, 4: 28-48.

[2] Zareen M S, Waqar A, Aslam B. Digital forensics: Latest challenges and response [C]//Information Assurance (NCIA), 2013 2nd National Conference on. IEEE, 2013: 21-29.

[3] Dykstra J, Sherman A T. Design and implementation of FROST: Digital forensic tools for the OpenStack cloud computing platform[J]. Digital Investigation, 2013, 10: S87-S95.

[4] Dahal S. Security Architecture for Cloud Computing Platform[J]. 2012.

[5] Marrington A, Branagan M, Smith J. Forensic challenges in service oriented architectures[J]. 2007.

[6] Mishin D, Medvedev D, Szalay A S, et al. Data Sharing and Publication Using the SciDrive Service[C]//Astronomical Data Analysis Software and Systems XXIII. 2014, 485: 465.

[7] Simou S, Kalloniatis C, Mouratidis H, et al. Towards the Development of a Cloud Forensics Methodology: A Conceptual Model[C]//Advanced Information Systems Engineering Workshops. Springer International Publishing, 2015: 470-481.

[8] Marturana F, Me G, Tacconi S. A Case Study on Digital Forensics in the Cloud[C]//Proceedings of the 2012 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery. IEEE Computer Society, 2012: 111-116.

[9] Mishra A K, Matta P, Pilli E S, et al. Cloud Forensics: State-of-the-Art and Research Challenges[C]//2012 International Symposium on Cloud and Services Computing.

[10] Sun J R, Shih M L, Hwang M S. Cases study and analysis of the court judgement of cybercrimes in Taiwan[J]. International Journal of Law, Crime and Justice, 2014.

[11] Zargari S, Benford D. Cloud Forensics: Concepts, Issues, and Challenges[C]//2012 Third International Conference on Emerging Intelligent Data and Web Technologies.

[12] Chen G, Du Y, Qin P, et al. Suggestions to digital forensics in Cloud computing ERA[C] //Network Infrastructure and Digital Content (IC-NIDC), 2012 3rd IEEE International Conference on. IEEE, 2012: 540-544.

[13] Ab Rahman N H, Choo K K R. A survey of information security incident handling in the cloud[J]. Computers & Security, 2015, 49: 45-69.

[14] Patrascu A, Patriciu V V. Beyond digital forensics. A cloud computing perspective over incident response and reporting[C]//Applied Computational Intelligence and Informatics (SACI), 2013 IEEE 8th International Symposium on. IEEE, 2013: 455-460.

[15] Marangos N, Rizomiliotis P, Mitrou L. Digital Forensics in the Cloud Computing Era[J].

[16] Khan S, Ahmad E, Shiraz M, et al. Forensic challenges in mobile cloud computing[C]/ /Computer, Communications, and Control Technology (I4CT), 2014 International Conference on. IEEE, 2014: 343-347.