A Log-based Approach to Make Digital Forensics Easier on Cloud Computing

Ting Sang
Shanghai Jiao Tong University, Shanghai, 200240, China.
saintogod@gmail.com

***Abstract*-Cloud computing is getting more and more attention from the information and communication technologies industry recently. Almost all the leading companies of the information area show their interesting and efforts on cloud computing and release services about cloud computing in succession. But if want to make it go further, we should pay more effort on security issues. Especially, the Internet environment now has become more and more unsecure. With the popularization of computers and intelligent devices, the number of crime on them has increased rapidly in last decades, and will be quicker on the cloud computing environment in future. No wall is wall in the world. We should enhance the cloud computing not only at the aspect of precaution, but also at the aspect of dealing with the security events to defend it from crime activities. In this paper, I propose a approach which using logs model to building a forensic-friendly system. Using this model we can quickly gather information from cloud computing for some kinds of forensic purpose. And this will decrease the complexity of those kinds of forensics.**

Keywords- cloud computing; digital forensic; log; security

I. INTRODUCTION

Cloud computing is a hot topic in recent years. It exhibits the following key characteristics: agility, low cost in using, device and location independence, virtualization, reliability, scalability and elasticity, performance and etc. All those features show fascinating benefit to companies. As they can get free from the worry about the investment on hardware and can setup up their business easily. There are three major types of cloud services delivery model: Infrastructure as Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS). Different type of delivery model provides different convenience. IaaS just likes server-hosting services, but consumers do not need pay for hardware and maintain them anymore. They benefit much from the scalability and elasticity. PaaS is like service-hosting, but consumers do not need to worry about the servers out of working or not able to response to large number of requests. They benefit much from the performance and reliability. SaaS looks like the Representational State Transfer (REST) very much, and makes consumers benefit from performance, multi-tenancy architecture and many other features.

But both of the three models share a weakness from the characteristics of the cloud computing. As consumers put their logical procedurals on the cloud, which means that they do not own the control of the hardware especially for PaaS and SaaS. This is not friendly to digital forensic. Because traditional digital forensic is deeply depending on the media seized from the crime scene. At this point, there should be changes or enhancement for cloud computing to be friendlier with digital forensics.

In this paper I will focus on analyzing challenge for forensic investigation in cloud environments. The rest paper is organized as following. In section II, I will show what cloud computing and digital forensics is. And the next section is about the challenges that introduced by cloud computing to digital forensic. I will show a model that can make the cloud computing friendlier to digital forensics in section IV. At last, in section V, a conclusion is given.

II. TECHNOLOGY BACKGROUND

A. *Cloud Computing*

In this era of globalization, concentrating is the only way to success for the small companies or those just on their beginning, especially for those information technology (IT) companies. At this situation, outsourcing is a perfect way to survival, as they can more concentrate on the core business. At first level, they can outsource the noncritical software or modules to other companies to reduce the risk and human resources. The next one to be outsourced should be the services or the hardware for computing and storage. And to my opinion, this is just the original driver to the popular of the cloud computing. Cloud computing is a virtual computing environment which can provide applications, platform and software support as remote services. And those services can be accessed by Internet all over the world. Cloud computing is not a glossary in information technology. It is more like an alias for the new kind of business operation mode named “pay-per-use”. But it involves many new technologies, for example distributed computing, virtual computing, parallel computing, massive data processing and etc. The Cloud Service Provider (CSP) maintains large scale computer systems in clusters, and usually has large storage capacity on Data Center. The CSPs provide high availability system to meeting the customers’ requirements on software, service, computing and storage in a scalable mode. So we can simply think that the customer companies outsource some kinds of requirement to those CSPs.

The cloud computing can provide many benefit to both the small and large companies. To the small companies, they can save the money for physical devices, at the mean times, keep the performance of their “devices” up to date. Another benefit for them is that they can free from the maintaining the physical machines, that may be a lot of money for them. At this view, cloud computing looks like a kind of hardware outsourcing.

To the larger ones, they can focus on their own business logical, and leave the high technology like massive data process to those professional companies. So that, they can

978-0-7695-4923-1/12 $26.00 © 2012 IEEE
DOI 10.1109/ISDEA.2012.29

provide high quality services and have no worrying about the high investment in research and the high risk. At this point, cloud computing is a kind of technology outsourcing.

And to those huge companies, there are benefits too. Firstly, they can more effectively use the computing power and storage ability. Secondly, they can build their own e-ecological environment on cloud easily.

B. *Digital Forensic*

Generally speaking, cloud security should contain two aspects. First is how to protect cloud and applications running inside from attack. And then is how to deal with the happened security events. Precaution is the major topic on cloud security, but we should know that "no wall is wall in the world". Criminal always can find a way to overcome the security blocks to achieve their goals illegally. In the digital world, the security dept. had started a new battlefield with criminals. They use digital forensic technology to reveal the crime events and the criminals.

Digital forensics (also referred to at times as computer forensics) encompasses approaches and techniques for gathering and analyzing traces of human and computer-generated activity in such a way that it is suitable in a court of law. The objective of digital forensics is hence to perform a structured investigation into past and ongoing occurrences of data processing and transmission whilst maintaining a documented chain of evidence, which can be reproduced unambiguously and validated by competent third parties. According to RFC3227 we can conclude the structured investigation into five main steps, collection, extraction, analyzing, reporting and documentation. The figure 1 shows the concise procedural of digital forensics.

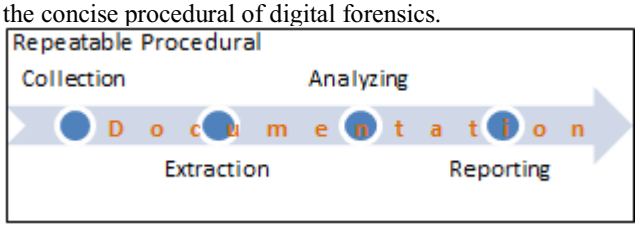

Figure 1. Main steps of digital forensics.

III. CHALLENGES FOR CLOUD FORENSIC

The identification of evidence in the cloud computing environment can be very complex. To different deployment model, which knows as public cloud, private cloud and hybrid, has deep affection on forensics procedural. If the evidence resides within a public cloud, it will be much more difficult to identify. There are different computer forensic challenges related to the different services models, PaaS, IaaS and SaaS. These models present subtly different challenges to the forensic investigator.

While trying to process the forensics procedural in cloud, we will meet grate obstruction at the very beginning. We cannot seize the hardware containing or processing the target applications from the cloud, as they can be everywhere in the world or even no real hardware such as Virtual Machine. And the nature of dynamic scaling up and down makes the possibility of losing information higher. The potential source for evidential data in cloud forensics is very limited. We can have a look at the order of volatility in traditional forensics event.

- Registers, cache
- Routing table, ARP cache, process table, kernel statistics, memory
- Temporary file systems
- Disk
- Remote logging and monitoring data that is relevant to the system in question
- Physical configuration, network topology
- Archival media

To IaaS, if the CSPs provide the suspected VMs, we can collection evidences exactly as the list, even better than in traditional. And if the CSPs don't provide the VMs, we can apply the live forensics too. But to SaaS and PaaS, most of the items shown in the list are not available. Because applications that running on the cloud can be regarded as many individual processes, but not many integral machines. It is impracticable to plan to analyze the VMs directly, even if the CSPs cooperate with investigators. As the VMs for SaaS and PaaS may be have a huge storage system, and contain many other applications in the same VMs. In this situation, the massive data is the biggest challenge for cloud forensics.

There are many other challenges for those models. For SaaS, we cannot obtain the any volatile data and the data that has been deleted logically. And logging files may be collocated or spread across multiple and changing devices. However, multi-tenant environment has may contaminate the acquisition.

Compare with SaaS, the biggest difference for PaaS is the source of Software. So PaaS has all the problems SaaS has, and it has some new ones. As totally losing the control of runtime environment, the information we can get is logging, which is in the CSP's style.

And as mentioned in front of the paper, the biggest challenge for IaaS is that if the CSPs provide the suspected VMs or not. If the CSPs provide, we can do the forensics like traditional one very much. And if not, we just can do live forensics, and should considerate the network band width. There is lack of logging about operations inside the VMs that CSPs can provide.

IV. LOG-BASED APPROACH FOR CLOUD FORENSICS

In digital world, log is a regular or systematic record of actions that object has taken or statuses that object have been. It is the most common component that be used in digital forensics.

Gartner has warned that "Investigating inappropriate or illegal activity may be impossible in cloud computing. Cloud services are especially difficult to investigate, because logging and data for multiple customers may be co-located and may also be spread across an ever-changing set of hosts and data centers. If you cannot get a contractual commitment to support specific forms of investigation - along with evidence that the vendor has already successfully supported such activities - then your only safe assumption is that investigation and discovery requests will be impossible."

In section III, we can know that logging is a challenge for all of the three models. So if we can improve the ability of logging, we may make the some kinds of forensics a little easier on cloud.

On SaaS and PaaS, an application is always for only one single purpose. While what we can get from IaaS are VMs, and we can use them to many different purposes. Logging outside the VMs is not so useful, so the log model proposed later may just suit for SaaS and PaaS.

1) SaaS

The way using SaaS is much like to using software locally. The difference is we send command to server via network in using SaaS. The figure 2 shows the way how to communicate with SaaS.

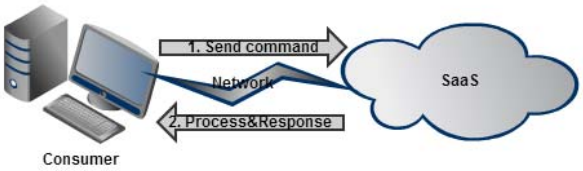

Figure 2. Communication with SaaS

In this model, consumer uses agent to send commands to SaaS. And after SaaS processes the commands and creates logs for that, it will send back response. While consumer gets the response, the agent may make its own logs or just processes the response to user.

However, how can we prove that the consumer had used the SaaS or the nonrepudiation of activity? The simplest answer is asking the SaaS CSP to provide the logs of the software or services traditionally. But in cloud environment, we should not expect the CSPs to supply as much help and quick as the local servers can. It means that we should keep another log locally and synchronously, so we can use it to check the activities on SaaS cloud while without the help of the CSPs. The content that would be recorded in the log files (the log files can be files or database) should be decided by the CSPs, but not the agent itself. That is to say the log files should be operated by a module created by the CSP. This is to make sure that the log files stored in local and in cloud are comparable. While the application on SaaS sends back the response, there will be a summary of the log record stored in SaaS, such as unique id and timestamp. The local log module will use that information on the log record locally. Additionally, for the consideration of protecting personal information, those files should be readable only to particular tools or softwares that made by the CSP. The figure 3 illustrates the whole process of the communication concisely.

Whilst there is another question that is how to guarantee the authenticity of them, if the log files are stored locally. HASH code is known as the simplest kind of fingerprint for digital data. So we can use it to detected modification on the log files. Considering the increasing speed of the log files in size, we can use an incremental HASH algorithm to improve the efficiency and reduce the time spending to verify. It will be very useful, because the investigators may be more interested in the recently events.

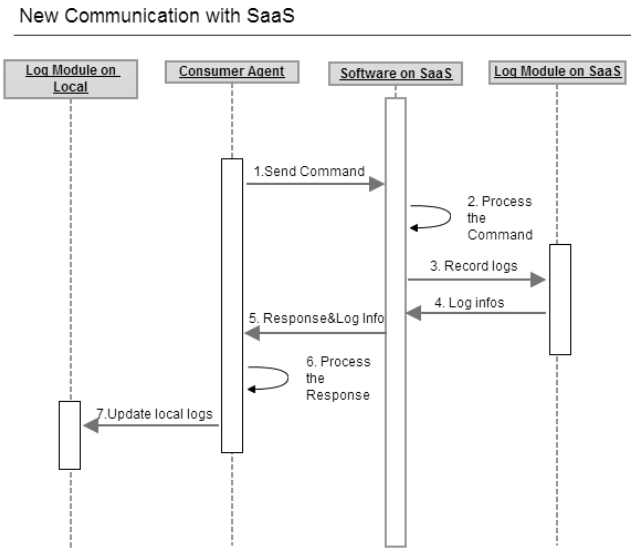

Figure 3. New communication model with SaaS

2) PaaS

In PaaS cloud, CSPs can provide a kind of runtime environment and other useful libraries or tools, for example Microsoft Azure supplies .NET framework and Google App Engine supplies Java and python runtime environment. The biggest difference between PaaS and SaaS is that who develop the softwares, the CSP itself or the third-party.

The CSP can use the resources on the cloud such as file system and storage system easier than the third-party. To apply the new model to PaaS, the CSPs should supply a log module on PaaS to the third-party. So that they can use it to store their own log on the cloud. Even if the CSP supply the lower level APIs, the third-party can create a customized log module, and of course, for both of the consumer side and cloud side. So they can define the best granularity and frequency for their business. And in this manner they can do the forensics totally without the participation of CSPs. Figure 4 and 5 illustrate the little difference of the two types.

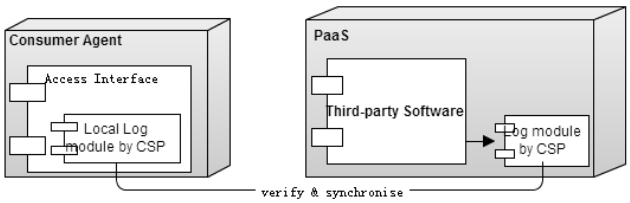

Figure 4. Using the log module of PaaS

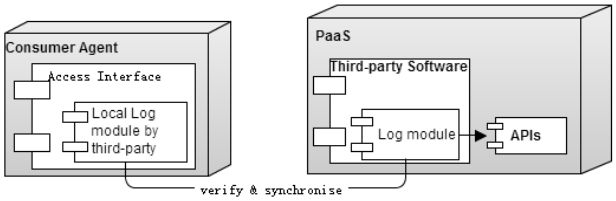

Figure 5. Using the log module of its own

However, if using the log module made by self, there will be many other questions that should be taken care of, for example, how to keep the logs in security and synchronous, and the effectiveness to be evidence. The compromise scheme for this situation is that using an authenticated log module of other party.

By using this model, we can obviously decrease the complexity of verifying if someone or some device has used the cloud services or not. The procedural can be described as the follow figure.

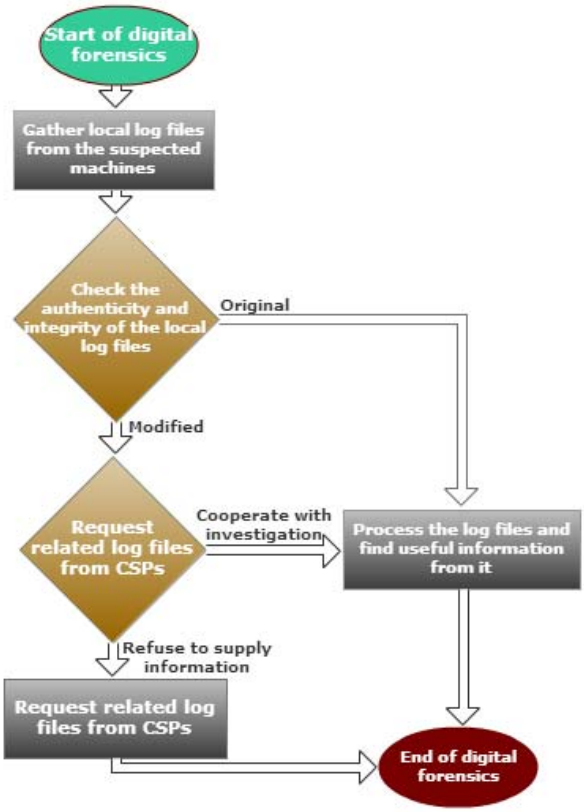

Figure 6. Digital forensics process for log-based model

V. CONCLUSIONS

There is no doubt that cloud computing will be the most popular operation mode for business. Whilst there will be more and more crimes against it too. For all the participator of cloud computing, they should prepare for that change. In this paper we have proposed a log-based model for. The log-based model can help to reduce the complexity of forensic for nonrepudiation of behaviors on cloud. However, it is totally no enough for the other kinds of digital forensics. What makes matters worse is that, till now, there are still no guidelines or standards for the cloud security. Most of times, we modified the guidelines of traditional digital forensics to suit for cloud computing environment independently.

REFERENCES

[1] D. Brezinski and T. Killalea. Guidelines for evidence collection and archiving. RFC 3227, IETF, 2002.

[2] Birk, D.; Wegener, C. Technical Issues of Forensic Investigations in Cloud Computing Environments. 6th International Workshop on Systematic Approaches to Digital Forensic Engineering (SADFE), pages: 1-10, 2011.

[3] Ahmed, S.; Raja, M.Y.A. Tackling cloud security issues and forensics model. High-Capacity Optical Networks and Enabling Technologies (HONET), pages: 190-195, 2010.

[4] Stephen D. Wolthusen, Overcast: Forensic Discovery in Cloud Environments, 5th International Conference on IT Security Incident Management and IT Forensics, pages: 3-9, 2009.

[5] Cheng Yan, Cybercrime forensic system in cloud computing, Image Analysis and Signal Processing (IASP), pages: 612-615, 2011.

[6] Stephen Biggs, Stilianos Vidalis. Cloud Computing: The impact on digital forensic investigations. International Conference for Internet Technology and Secured Transactions, pages: 1-6, 2009.

[7] Hong Guo; Bo Jing. Forensic investigations in Cloud environments. International Conference on Computer Science and Information Processing (CSIP), pages: 248-251, 2012.