Compression and Reconstruction of Virtual Machine for Digital Forensics in Cloud Computing

Taimour Nazar, Malik Tahir Hassan, Sheraz Naseer
s2016114004@umt.edu.pk, tahir.hassan@umt.edu.pk, sheraz.naseer@umt.edu.pk
School of Systems and Technology
University of Management & Technology, Lahore, Pakistan

Abstract---Cloud computing is an emerging trend these days. It offers computation and storage at a relatively low cost due to its pay per use policy. However, it has created new concerns regarding security, as all the conventional methodologies and tools for investigation fall short for cloud computing investigation. Study of different research papers has shown that no definite strategy exists to cater this issue. Certain methodologies have been purposed by researchers but a major issue, i.e. termination of virtual machines remains unsolved and unaddressed. The main aim of this paper is to address this issue and propose a possible and practical solution for it.

Keywords---Cloud computing, Digital forensics, Digital investigation, Digital crime, Cybercrime

I. Introduction

The famous and well know description of cloud computing given by NIST is that the model of cloud computing provides us at anytime with easy access to a common pool of configurable processing assets that can be immediately made available with least interaction of the cloud service provider. This model is made up of 5 important characteristics, 3 different service models and 4 different deployment models [11]

DFRWS has defined digital forensics as follows:

Those methods are used which are deduced and established from science to protect, gather, approve, distinguish, analyze, interpret, record and demonstration of digital evidences which are from digital sources which can will help in remaking the events which are seen as criminal, or bringing to light the unapproved activities which appear to cause troubles in the path of planned activities. [10]

NIST describes forensics in cloud computing as the utilization of scientific rules, technical practices, and deduced and demonstrated techniques to recreate past events with the help of forensic activities like 1-Identification, 2-Collection, 3-Preservation, 4-Examination, 5-Interpretation, 6-Reporting [11]

There are five essential characteristics of cloud: 1-On-demand self-service, 2-Broad network access, 3-Resource pooling, 4- Measured service 5-Rapid elasticity (Figure1).Three service models of cloud include software as a service (Saas), platform as a service (Paas) and infrastructure as a service (Iaas). Table 1 describes these service models briefly. The four deployment models of cloud are public model, private model, hybrid model and community model. These deployment models are briefly described in Table 2.

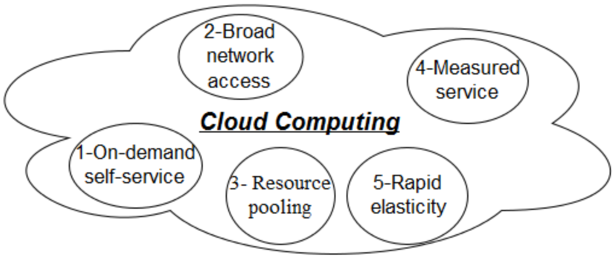

Figure 1: Five essential characteristic

Table 1 Service Models

Service Models	**Description**
Saas	You can use deployed software on cloud.
Paas	You can create & deploy your own software using cloud as a platform.
Iaas	You can use the underlying infrastructure of cloud to install any OS of your choice, create and deploy your own software.

Table 2 Deployment Models

Deployment Models	**Description**
Public	Cloud is available to general public.
Private	Cloud is used by an organization for itself.
Community	Cloud is used by multiple organizations, having common goals.
Hybrid	It is a combination of public, private or community cloud. Combination can be of any two or all three.

Cloud computing can be used for storing and processing data at cheaper cost as compared to conventional services available for storing and processing. It has gained a high level of acceptance in global village and IT world. The main reason of different companies shifting to cloud is that it helps them to reduce their expenses. However, concerns about the security of data in cloud are increasing. The attackers can use cloud to initiate attacks at a very large level e.g. distributed denial of server attack (DDOS). The area of digital forensics concerned with cloud is not very mature and a lot of loop-holes are still present which need to be addressed.

II. Related work

Different people have shown their concerns regarding the lack of forensics tools, and ineffective and inefficient methodologies present today for conducting digital forensics in the cloud.

It is rightly predicted in [1] that in future more organizations will move towards cloud and currently available conventional methodologies of digital forensics will not be sufficient for cloud based systems. We have to make tools, methods and procedures to conduct effective forensics in the cloud.

According to [2] during different process of evidence collection in cloud, different challenges are encountered. For example, during the process of identification following issues are encountered: decentralization of data and logs, physical location not known or out of access and volatile logs. Issues faced during the process of preservation include virtual machine not accessible, dependence on the cloud service provider (CSP), difficulty in protection and preservation of metadata. During acquisition following issues are faced: difficulty in separation of data due to multi tenancy model in multi tenancy model data of different users is present on same server thus, separation of data, logs and network logs becomes difficult, problems in establishment of chain of custody and different laws of different regions.

Digital evidence should fulfill the legal requirements that are compulsory for any conventional evidence, i.e. it should be authentic, reliable, complete, believable and admissible [3]. Furthermore, authors [3] say that digital evidence is present in physical context and logical context. By physical context they mean the data which is present on the hardware and by logical context they mean the file and directory structure which is essential for human beings to view the file, as the humans can't understand the evidence in its physical form so human requires logical context of data to understand it. Thus both physical and logical contexts should be considered during collection and analysis of the evidence and both of them are also important during presentation of digital evidence.

The work [4] claims that clouds are not forensic friendly because several characteristics have made it difficult to conduct forensics in cloud environment and researches are being carried out to protect cloud from both internal and external threats and attackers.

Reference [7] draws our attention towards important issues related to digital forensic. New techniques are evolving to breach security while most of the organizations are not capable to conduct proper forensic activity in case of a security breach. Limited number of forensic experts are available and data is growing in size, which makes efficient forensic more difficult. It is difficult to conduct forensic on volatile data. Further they have proposed that legal experts and experts of cloud computing should work together to find and create digital evidence so that they can be used by experts of digital forensics.

Moreover [9] has presented two hypothetical case studies and from those case studies they have highlighted the following issues of cloud forensics. 1- Cloud environment makes collection of data more difficult. 2- Support of cloud service provider is very important. 3- Conventional digital forensic tools are not sufficient for conducting forensics in cloud. 4- Cloud metadata misses key elements for forensic. 5- Chain of custody gets highly difficult in cloud computing environment. In chain of custody we have to document every person in whose custody any collected evidence was. Simply it involves every person from the point of collection of evidence to its disposal. Disposal occurs when case is closed. In cloud computing environment it is extremely difficult to document, that who has accessed the VM and when someone accessed it. Even we can’t be sure that did the employees of the cloud service provider access it and made any changes to it.

III. Major challenges and solutions for forensics of cloud computing

Challenges for digital forensics in a cloud environment are as follows:

1. Decentralization of data and Jurisdiction of area
2. Segregation of data of different users
3. Chain of custody
4. Virtual machine compromised

5. Termination of Virtual Machine

These five are the main issues/challenges faced by forensic experts in case of clouds as all of these create hindrances in their path. We should keep in mind that cloud forensics is not same as ordinary forensic and we can't utilize same forensic tools for cloud which we have been using for many years for networks, routers, computers, mobiles, laptops and tablets, etc. The main focus of this paper is to address the issue of termination of virtual machine, but we will briefly go through the other four issues as well to put forward the importance and significance of why addressing termination of virtual machine is necessary.

1. Decentralization of data & Jurisdiction of area

A cloud service provider can have multiple data centers which are located in different regions around the world in different countries. When we say different countries we should understand that different countries will have different law, rules and regulations. Laws of one country will not apply on other countries and if we get a warrant to search and investigate a cloud service provider by a court then we can only operate or investigate in the data centers which are present in the country of that court and not outside that country.

Forensics team will require help and deep cooperation from the law enforcement agencies and law regulation authorities of different countries. Once they get help and reach the data center where the virtual machine is located they can carry out their investigation only and only if the virtual machine has not been terminated.

2. Segregation of data of different users

The next problem faced by forensics team will be data segregation or multi-tenancy model (figure 2) which is used by cloud providers for cost effectiveness. In multi-tenancy model although different users are given their own virtual machines but data of different users can be present on the same physical server, which creates difficulty in conducting any forensics activity. As we are not clear which portion of data on the physical server belongs to which user, we will require the help of the cloud service provider to separate the data of the attacker for forensics team. This means we are at the mercy of the cloud service provider to move forward. Suppose cloud service provider agrees to help and now we are able to segregate the data of the attacker from innocent people, we can only do this if the virtual machine has not been terminated.

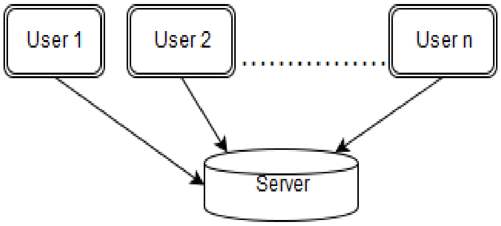

Figure 2: Multi tenancy model

3. Chain of custody

Another problem faced by forensics team is to establish the chain of custody. Keeping under consideration the structure of cloud we can say that it becomes extremely difficult to maintain the chain of custody. The main reasons is that the data is decentralized and spread over different regions and in the end we don't know if even the cloud service provider is providing us correct data or if they have tempered the data before releasing it to the forensics team. It is extremely difficult to prove the chain of custody in the court of law. Chain of custody can only be established if the cloud service provider logs all the access to data from inside and outside of cloud infrastructure, by cloud employees and by cloud users. An example can be Google docs, where Google keeps record of all the people interacting with each document along with its timestamp. Similarly, activity log of Facebook also provides one’s interaction with other users. Termination of virtual machine might result in deletion of all this information.

4. Virtual machine compromised

An attacker pays for a virtual machine, uses it to compromise the virtual machine of an innocent user and launches an attack with it. After a successful attack he terminates his own virtual machine to remove all his traces. When we approach the innocent owner of the virtual machine he can claim that his virtual machine was compromised and he is innocent. When we investigate further we can't find the attacker because he has removed his virtual machine.

Once the virtual machine is terminated all the data relevant to it is removed by the cloud service provider and all data that could be used for forensic purpose is lost. We can't prove the attacker to be a culprit in the court of law, because after termination of virtual machine we have no evidence which we can acquire and use to prove the guilt in the court of law.

5. Termination of Virtual Machine

This is the most critical issue as it erases the entire base on which crime can be proved and any criminal can simply use a few clicks to do it. The main aim of this paper is to address this issue as this issue has not been addressed yet. After going through the four above mentioned challenges one can understand the extent of the significance of this issue. There are three possible approaches to solve this issue.

First approach is simple that the cloud service provider keeps the backup of all the virtual machines that are terminated so they can be used later for forensic purposes if needed. This method seems very easy but it is extremely cost inefficient and will require a huge amount of space and it is not possible for cloud service providers to bear a huge cost to store this huge and massive amount of data. As a result it can't be used.

Second approach will be to compress the data before storing it, again it will still need a lot of space and will require resources to compress and decompress all the virtual machines which are being terminated. As a result, it is not practical as well.

Third approach which is the proposed solution to this issue is described as follows. Let's take an example of a virtual machine. What is the pack of a virtual machine actually containing? it contains an OS and some software installed on it. Now let's take into consideration the information that we need to conduct a successful forensic activity. These include the following:

- OS used and its version.
- List of all the software used to launch the attack and their versions.
- All log files, all dump files, all crash dumps etc.
- In case of windows we need data stored in registry.
- In case of Linux we need data stored in log files under folder /var/log and other important log files.
- Script files e.g. .sh or .py etc.

The OS and the software occupy the most space in the virtual machine and the log files and registry data occupy lesser space. Storing the entire OS and software is not necessary; we can only store the name and version of the OS and the software along with all the log and script files. All this data can be stored in a simple XML file and then compressed; now its size will be in KB or MB. The XML file should be made after command for deletion of VM has been issued and before VM has been deleted (Figure 3).

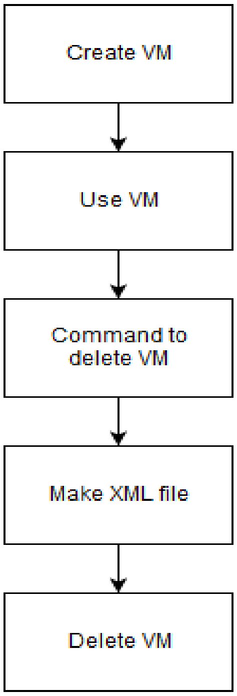

Figure 3 Creation of XML file before deletion of VM

IV. Structure of XML file

The proposed structure of XML file is as follows:

```
<OS>
        <name>Ubuntu</name>
        <bit>64</bit>
</OS>
<software>
        <1>name of software</1>
        <2>name of software</2>
        <3>name of software</3>
    </software>
    <hardware>
        <harddisk>
            <companyname>
                Seagate
            </companyname>
            <space>500GB</space>
        </harddisk>
        <Ram>
            <companyname>
        name of vendor
            </companyname>
        <space>4 GB</space>
    </Ram>
</hardware>
<registry>

</registry>
<logs>

</logs>
```

This XML file above is a sample and it can include all the required and relevant information of the virtual machine

V. Benefit

The main benefit is that in this way we can store the record of huge amount of virtual machines which have been terminated in a very small space. From this minimum information which we have kept we can easily regenerate the complete virtual machine at any time. To regenerate the virtual machine we will simply take information from the XML file and build a virtual machine according to it. Moreover, the log files and registry data which we have stored will provide us with most of the information which we require; they can even tell us about the software which was installed and uninstalled before termination of the virtual machine. This regenerated virtual machine will play a vital role in successful conduction of forensics activity.

If CSP has created such XML files, then any agency can request them to get the XML. After receiving the XML, they can reconstruct the virtual machine and start forensics on it. Upon completion, results can be prepared and presented (Figure 4).

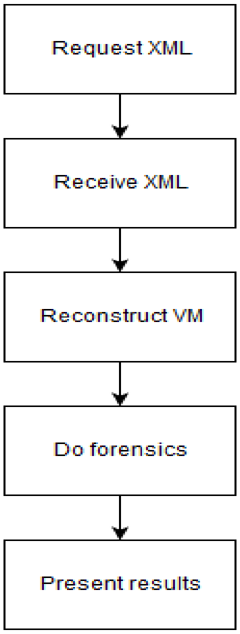

Figure 4 Steps for reconstruction & forensics of VM

VI. Conclusion and Future Work

It can be troublesome if we identify a culprit who is involved in any illegal online activity and before we can catch him, he deletes the entire virtual machine. When he tries to terminate the virtual machine, the cloud service provider should store the minimum necessary and sufficient amount of information about his virtual machine. Forensics team can ask the cloud service provider to give them that virtual machine information, which they can use to regenerate the virtual machine and it will help them in complete timeline reconstruction. This information can be very crucial in proving the guilt of the culprit in the court of law. Many issues related to virtual machine termination can be resolved by storing the necessary data before termination of the virtual machine.

In this work, we have proposed a method to help digital forensic activities in cloud computing. Availability of virtual machine and usage data under investigation is crucial for digital forensics in a cloud but taking all data backup or keeping all virtual machines alive are not feasible solutions. We identify the key information of a virtual machine and propose to store this information in an XML file before the termination of the virtual machine. The compressed XML file information is sufficient to reconstruct the virtual machine after its termination.

We have proposed a framework for compression and reconstruction of virtual machine for digital forensics in cloud computing. Implementation of the complete framework and its comprehensive experimentation on a cloud is the next step in our work.

References

[1] G. Grispos, T. Storer and W. Glisson, "Calm Before the Storm", *International Journal of Digital*

[2] G. Meyer and A. Stander, "Cloud Computing: The Digital Forensics Challenge", in *Proceedings of Informing Science & IT Education Conference*, 2015, pp. 285-299.

[3] D. Reilly, C. Wren and T. Berry, "Cloud Computing: Pros and Cons for Computer Forensic Investigations", *International Journal Multimedia and Image Processing*, vol. 1, no. 1, pp. 26-34, 2011.

[4] S. Zawoad and R. Hasan, "Digital Forensics in the Cloud", *CrossTalk Journal of Defense Software Engineering*, vol. 26, no. 5, pp. 17-20, 2013.

[5] F. Daryabar, A. Dehghantanha, N. Udzir, N. Sani, S. Shamsuddin and F. Norouzizadeh, "A Survey About Impacts of Cloud Computing on Digital Forensics", *International Journal of Cyber-Security and Digital Forensics*, vol. 2, no. 2, pp. 77-94, 2013.

[6] Ericsson, "Cloud security architecture (White paper)", 2015.

[7] L. Slusky, P. Partow-Navid and M. Doshi, "Cloud computing and computer forensics for business applications", *Journal of Technology Research*, vol. 3, 2012.

[8] J. Dykstra, *Digital Forensics for IaaS Cloud Computing*, 1st ed. Austin, 2012.

[9] J. Dykstra and A. Sherman, "Understanding Issues In Cloud Forensics: Two Hypothetical Case Studies", in *Conference on Digital Forensics, Security and Law*, Richmond, 2011, pp. 45-54.

[10] S. Un, "The Future Of Digital Foreniscs", RSA Conference, Asia pacific, 2013.

[11] J. Dykstra, L. Gowen, R. Jackson, O. Scot, E. Rojas, K. Ruan, M. Salim, K. Stavinoha, L. Taylor and K. Zatyko, "NIST Cloud Computing Forensic Science Challenges", NIST, 2014.