Cloud forensics:Volatile data preservation

Prasad Purnaye
Department of Computer Engineering & Information Technology
V.J.T.I., Mumbai, India
prasadpurnaye@gmail.com

Varshapriya Jyotinagar
Department of Computer Engineering & Information Technology
V.J.T.I., Mumbai, India
varshapriyajn@vjti.org.in

Abstract

This paper begins with cloud forensics background, describing steps involved in cloud forensics investigation. When it comes to cloud forensics volatile data plays crucial role. Many of Cloud Service Providers (CSP) do not have proper mechanism for preserving volatile memory data. At most they could do is storing the Virtual Machine (VM) instance. But what if the client (tenant) performs some malicious activity and then ends his subscription? In such case all of his data will be lost. So we propose a mechanism to store volatile data in a dedicated common persistent storage which will provide help during Cloud forensic investigation.

Keywords—cloud forensics, digital investigation, evidence, Random Access Memory (RAM), tenant, volatile data;

I. INTRODUCTION

There's an arms race going on in the business tech world. Cloud computing technology is becoming popular in industry for its many advantages. Reports have indicated that, the popularity of cloud among hackers is also increasing as it gives the perfect platform for high potential cyber-crime. Security is a major concern of every tenant (client) when he is migrating to cloud. We have tools to analyze the data and to get finding of the evidence, but before that it is crucial that all the data is collected properly. Cloud forensics follows generic procedure for any digital investigation in which all the steps from gathering the evidence to presentation before jury are carried out.

The remainder of this paper is structured as follows: In section II we have discussed digital forensics procedure in detail. Section III enlightens the importance of volatile data from a forensics perspective. In the end we have proposed an approach to preserve the volatile data with context to cloud computing in section IV.

II. BACKGROUND

A. Cloud Computing

Cloud computing is cost-effective and efficient computing paradigm. Unfortunately, today's cloud computing architectures are not designed for high security and forensics. Cloud forensics is slightly different than traditional forensics. Many factors complicate forensic investigations in a cloud environment. For example, during acquisition of data in cloud forensics, data is no longer local, we need to subpoena the data center to access the data. Even if we get the access, we cannot confiscate the server as it may contain data of other tenants. Not to mention the integrity of Cloud Service Provider (CSP) can be obscure.

There is difference in degree of control over data in each of the service model in cloud computing. Table 1 depicts that a tenant administrator has more control over Infrastructure as a Service (IaaS) model and level of control drops as we go towards SaaS model.

TABLE 1. DEGREE OF CONTROL IN EACH CLOUD MODE

Parameter	Software as a Service (SaaS)	Platform as a Service (PaaS)	Infrastructure as a Service (IaaS)
Control of Application	No	Yes	Yes
Control of Operating System	No	No	Yes
Networking Control	No	No	Yes
Control of Hardware	No	No	No

B. Cloud Forensics

NIST defines digital forensics as an applied science for "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" [1]. Cloud computing forensic science is the application of scientific principles, technological practices and derived and proven methods to reconstruct past cloud computing events through identification, collection, organization and reporting of digital evidence. In each step there are tools and techniques available. Fig.1 shows different steps of cloud forensics. Traditional methods and tools of forensics cannot cope up with the cloud forensics because of the fact that the retrieval of the information, the major lead of any case, is diversely located and hence difficult to reach. Cloud computing is based on extensive network access, and network forensics handles forensic investigation in private and public networks. However, cloud forensics also includes investigating file systems, process, cash, and registry history. Every data is important for the investigation. So in the collection phase, goal is to collect as much as data which is relevant to the investigation.

Phase 1 is identification of crime. To ensure that actual crime has happened. Evidences involved in crime are identified in this step.

Phase 2 is known as Collection. Tasks performed under this phase related to the acquiring, collecting, transporting, storing and preserving of data from all possible electronic devices. In general, this phase is where all relevant data are captured, stored and be made available for the next phase.

FIGURE 1. STEPS OF CLOUD FORENSICS

Phase 3 is organization. This is the main and the center of the computer forensic investigation processes. It has the most number of phases in its group thus reflecting the focus of most models reviewed are indeed on the analysis phase various types of analysis are performed on the acquired data to identify the source of crime and ultimately discovering the person responsible of the crime.

Phase 4 is known as Presentation. The finding from analysis phase are documented and presented to the authority. Obviously, this phase is crucial as the case must not only be presented in a manner well understood by the party presented to, it must also be supported with adequate and acceptable evidence. The main output of this phase is either to prove or disprove the alleged criminal acts.

Chain of custody is one of the most vital issues throughout the process. Chain of custody should clearly depict how the evidence was collected, analyzed, and preserved in order to be presented as admissible evidence in court. Cloud computing is a multi-tenant system, while traditional computing is a single owner system. An alleged suspect may claim that the evidence contains information of other users, not her. In this case, the investigator needs to prove to the court that the provided evidence actually belongs to the suspect. Conversely, in traditional computing systems, a suspect is solely responsible for all the digital evidence located in her computing system. Moreover, in the cloud, we need to preserve the privacy of other tenants.

III. SIGNIFICANCE OF VOLATILE DATA

Volatile data is a data which cannot sustain without power. Data residing in a VM are volatile, as after terminating a VM, all the data will be lost. There is a wealth of data available in volatile memory. Processes, information about open files and registry handles, network information, passwords on disk, hidden data, and worm and rootkits written to run solely in memory are all potentially stored there. Unfortunately this data is volatile data. In order to provide the on demand computational and storage service, CSPs do not provide persistent storage to VM instances. There is, though, a way to preserve VM data by storing an image of the VM instance. An attacker can exploit this vulnerability in the following way: after doing some malicious activity (e.g., launch DoS attack, send Spam mail), an adversary can terminate her VM that will lead to a complete loss of the evidence and make the forensic investigation almost impossible.

Experts believe that for a good forensic practice actual crime scene, which is the cloud in our case, is a best place to start with any investigation, as you can get most of the evidences in a form of data from a crime scene. But if the electronic devices involved in crime scene are too much compromised to get data from then some extra efforts has to be made for the reconstruction of crime scene. Even if we fetch the data from victim's computer and data from the adversary's devices, we should be able to establish motive and should be able to connect the dots. In such cases data in volatile memory can provide with helpful details.

There are two methods of acquiring volatile memory: hardware-based acquisition, and software-based acquisition. Both methods have their pros and cons. In general, from a forensics perspective, it is better to use hardware-based acquisition because it is more reliable and difficult for an attacker to corrupt, but currently software-based acquisition is the far more popular method due to its cost-effectiveness and ease of availability.

There can be two possible ways of continuous synchronization of volatile data, suggested by D. Birk[2] ,but their implementation is not specified.

• Cloud Service Providers can provide a continuous synchronization API to customers. Using this API, customers can preserve the synchronized data to any cloud storage e.g., Amazon S3, or to their local storage. Implementing this mechanism will be helpful to get the evidence from a compromised VM, even though the adversary shutdown the VM after launching any malicious activity.

• However, if the adversary is the owner of a VM, the above-mentioned mechanism will not work. Trivially, she will not be interested to synchronize her malicious VM. To overcome this issue, Cloud Service Providers by themselves can integrate the synchronization mechanism with every VM and preserve the data within their infrastructure.

IV. PROPOSED THEORY

For better security in cloud computing, and to improve the forensics in cloud, we propose a cloud model which will store volatile data of each tenant in a shared persistent storage. For implementation we are using KVM hypervisor to create para-virtual environment. And a module will collect the volatile data, which then will be stored in persistent storage dedicated to volatile data storage and retrieval.

As we have more degree of control over IaaS, we have considered an IaaS model for experiment. Virtual Machines are deployed on a private cloud. Depending on the Operating System tenant is using, we can decide which scripts to run on virtual machines. The scripts will provide the data which can be stored at server. All tenants share a persistent storage on to which their volatile data is fetched as soon as the VM is booted and that data is stored. Data is synchronized and checked for any new data available in its volatile memory. Data is stored in log format to get clear idea of time and to which user the data belongs.

V. CONCLUSION

We came to know that the main challenge in cloud forensic is of data acquisition. It is important to know exactly where the data is located and actually acquiring it. If Cloud Service Providers practice to preserve volatile data, the loss of important artifacts, which could be potentially crucial evidence, can be made avoided. Every data plays crucial part in forensic investigation. Volatile data issue can be resolved by the proposed project. Using this approach Cloud provider can get hold of the crucial data which will be helpful in forensic investigation. We can have this data, without saving the whole VM instance image which will take a lot more space. This forensic friendly approach is definitely a step forward to more secure cloud architecture.

REFERENCES

[1] Zawoad, Shams, and Ragib Hasan. "Cloud Forensics: A Meta-Study of Challenges, Approaches, and Open Problems." arXiv preprint arXiv:1302.6312(2013).

[2] D. Birk, “Technical challenges of forensic investigations in cloud computing environments,” in Workshop on Cryptography and Security in Clouds, January 2011.

[3] D. Reilly, C. Wren, and T. Berry, “Cloud computing: Pros and cons for Computer forensic investigations,” International Journal Multimedia and Image Processing (IJMIP), vol. 1, no. 1, pp. 26–34, March 2011.

[4] D. Birk and C. Wegener, “Technical issues of forensic investigations in cloud computing environments,” Systematic Approaches to Digital Forensic Engineering, 2011.

[5] J. Dykstra and A. Sherman, “Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques,” DoD Cyber Crime Conference, January 2012

[6] Zawoad, Shams, and Ragib Hasan. “Digital Forensics in the Cloud ”.available from : http://secret.cis.uab.edu/media/secretlab-journal-2013-1.pdf

[7] Thethi, Neha, and Anthony Keane. "Digital forensics investigations in the Cloud." Advance Computing Conference (IACC), 2014 IEEE International. IEEE, 2014.

[8] Aljaedi, Amer, et al. "Comparative analysis of volatile memory forensics: live response vs. memory imaging." Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. IEEE, 2011.

[9] Carvajal, Leonardo, Cihan Varol, and Lei Chen. "Tools for collecting volatile data: A survey study." Technological Advances in Electrical, Electronics and Computer Engineering (TAEECE), 2013 International Conference on. IEEE, 2013.