International Journal of Current Trends in Engineering & Technology
ISSN: 2395-3152
Volume: 03, Issue: 02 (MARCH– APRIL, 2017)

Cloud Forensic: An Approach to Crowd Source Forensic Evidences in Cloud Environment

Amerendra Narayan Mishra[1], Prof. Ratnesh Kumar Dubey[2], Dr. Vineet Richharia[3]
Department of computer science & engineering, LNCT, Bhopal
[1]amrendra.narayan.mishra@gmail.com, [2]ratneshcse@lnctgroup.co.in, [3]vineetrich100@gmail.com

Abstract- *the rapid growth of cloud services in internet environment, computer attacks are intensifying the exploitation and may easily cause millions of dollar damage to an organization. Cloud is a very easy way to reach any system. If confidential data is not properly protected, then it becomes opens to vulnerable access and misuse. Cloud forensics relates to cyber-crime on the Internet. Some criminal activities like Leak of personnel images, child pornography, hacking, and identity theft can be traced and the criminals can be punished if proper evidence is found against them. Cloud forensic analysis brings out some details like when and in what sequence did somebody access a cloud services. Cyber-crime can cause varying degrees of damage by hackers. Hence, detailed forensic analysis of cloud computing is required to come to a conclusion about an incident and to prove or disprove someone's guilt.*

I. INTRODUCTION

Cloud computing is emerging as a new paradigm for next generation computing in the field of computer science and information technology because of their attractive services such as adaptive, online, value added and pay as use scheme. Cloud can be defined in a number of ways. It is a business model, which provides the on demand hardware and software as services to the client through the internet. With respect to each of the individual areas touched by the subject matter, this work looks to provide the researchers with sufficient understanding of the methodology chosen for investigation of digital data for attacks in cloud environment. The areas discussed are cloud-forensics, Forensic models, challenges in cloud forensics, cloud environment and their components. The complexity of cloud forensics increases with the number and variety of underlying services end user depends on.

II. CLOUD COMPUTING MODELS

According to NIST cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources. Theses computing resources include networks, servers, storage, applications, and services. This cloud model is basically composed of five essential characteristics, three types of service models, and four deployment models.

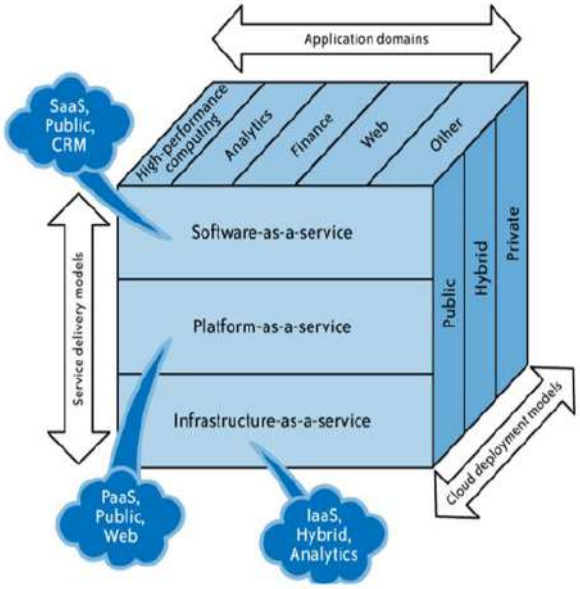

Figure1: Cloud Computing Models in 3D

Cloud computing provide three type of services i.e. software as a service (SAAS), platform as a service (PAAS), and infrastructure as a service (also known as hardware as a service).

2.1 Digital Forensic Models

It is the process of uncovering and interpreting electronic data. The goal of the process is to preserve any evidence in its most original form while performing a structured investigation by collecting, identifying and validating the digital information for the purpose of reconstructing past events.

"Forensic computing is the process of identifying, preserving, analyzing and presenting digital evidence in a manner that is legally acceptable." [3]

"Gathering and analyzing data in a manner as freedom distortion or bias as possible to reconstruct data or what has happened in the past on a system."

Forensic Computing, also known as Evidential Computing and even sometimes Data Recovery, is the specialist process of imaging and processing computer data which is reliable enough to be used as evidence in court. Common digital forensic model is shown in figure 2. The major requirement of a successful forensic analysis is its model. A forensic analyst need to incorporate an appropriate model based on the scenario. Therefore, applying feasible forensic model in a scenario has always been a concern of researchers and analysts. The following section includes the

revival of existing forensic models.

The first forensic model has been proposed by Politt in 1995 [4]. The work is based on the use of computer forensics to exploit stored digital information. It includes four steps namely- Acquisition, Identification, Evaluation and Admission, of evidences. The work is based on the use of computer forensics to exploit stored digital information. McKemmish [3] proposed a forensic model in 1999. The steps involved in it are shown in Figure 2.

In order to make forensic process more effective by providing clear definitions and complete terminology, researchers proposed different forensic models. These models comprise seven phases- Identification, Preservation, Collection, Examination, Analysis, Presentation and Decision. Modern digital forensic model is shown in Figure 3.

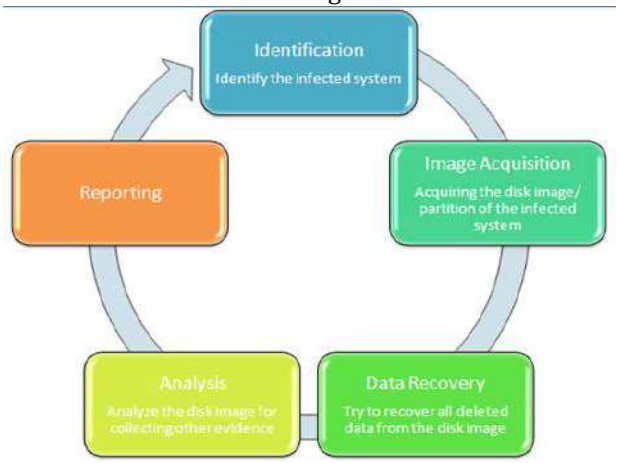

Figure 2: Common Digital Forensic Model

2.2 Cloud Forensics

It is a part of digital forensics and refers to the investigation and acquisition of artefacts in cloud environment. New threats cloud computing made forensic science a challenging endeavour in the last couple of years. The data stored at cloud servers such as personnel images, documents, e-mails, videos and short messages (SMS) may accessed remotely if the device is connected to the Internet. It poses a major challenge for forensics investigators.

a. Evidences in Cloud Computing

Cloud computing contains evidences mostly at two locations client-side and server-side. Which may be user created, internet related or third party information. Client side evidences are the evidences which are stored at; +Resources include networks, servers, storage, processing, applications, services and Cloud Service Providers (CSP).

The network layer in the cloud instance can also contain potential evidence.

Volatile and non-volatile data: On the Cloud customer device, the client used to access the cloud (web browser) is usually the only application that contain evidence

User-created information: It includes Photographs (including EXIF data); video/audio; maps, MMS; GPS waypoints; stored voicemail; files stored on system and connected computers.

Internet-related information: It covers online accounts; purchased media (often discoverable in embedded metadata); e-mail; Internet usage and social networking information.

Installed third-party applications: It includes alternate messaging and communication systems; additional capabilities; malware applications; IM applications—which contains traces and evidences stored in logs and databases.

b. Forensic Issues in Cloud Computing

Forensics investigation in the brings

1. Decentralization of Data Centers:- Data in cloud computing is stored at different locations of the different servers. It is due to providing the security and backup the data. To create the backup cloud servers creates the replica of the data hence it is available at different sources. Also for securing the data CSP creates several layers of the data and stored at different cloud servers existing at various locations. Storage depends on the scheme of the cloud service providers.
2. Physical Locations Unknown :- Due to the remote locations of cloud servers and the end users, searching exact location of the data is tedious task. And due to the security reasons data storage location is not provided to the end users.
3. Specific Logging:- Logging mechanism of the cloud computing services also depends on the CSP. Schema and the no of entries are also fixed as per the rules of the CSP. CSP also may delete these entries after specific time slot.
4. Inaccessibility to virtual instance:- Cloud services are dependent on the virtual machine which creates the inaccessibility of all data. In IaaS, the customer's virtual machine (VM) is controlled by the cloud service provider (CSP). The CSP is responsible for the hypervisors, network infrastructure right down to the physical hardware of the data centre. In cloud computing the virtualization of data storage complicates the identification and isolation of physical storage devices where the cloud customer's data may be stored and processed.
5. CSP Dependency on Next CSP:- CSPs have dependencies on other CSPs as data is pushed further back into the cloud computing environment. These dependencies are very dynamic and complex, thus increasing the

dependency of the investigation on the CSP. Any break in the chain of dependencies between the CSPs will have a serious impact on the collection of evidential data.

6. Dependency on Cloud Service Provider:- In a cloud computing environment the Cloud Service Provider (CSP) has all the power over the environment and therefore controls the source of the evidential data. The process of preserving digital evidence in the cloud highly depends on the support that the investigator receives from the CSP. Employees of CSP are not forensic expert hence they do not know the importance and technique of the data extraction and its integrity. Hence, there are several chances where data may be altered when imaging. There might also be a requirement that the CSP should put a litigation hold on the data to prevent the data from being destroyed or changed prior to the time that the data can be collected.
7. Metadata protection:- Metadata is the data about data. It provides the information of the data such as access time, modification time and creation time. Metadata describes ownership and the process history of the data objects. Metadata may the significant evidence in digital forensics to recognize the ownership and access information about the data objects. Cloud servers are established at different location zones that also creates a difference in time stamp metadata of data objects. Uniformity time stamp or conversion into same time stamp value in the process of investigation requires that the correct time and time-zones to be established. Hence, it may be challenging to digital forensic expert in a cloud environment due to the decentralization of data servers.
8. Cache data:- Cloud servers store frequently used data into cache as usual as in computer system. Also cloud may store it at client's machines during the interaction with the cloud server. Cache would be remarkable evidence in digital forensics. Cache memory is volatile memory hence, evidence exist in cache memory may lost in cloud computing if the virtual machine is powered down or rebooted or if the incorrect preservation process was followed by the CSP in the process of evidence collection on behalf of the investigator.

III. Background and Literature Survey

The focus of the presented work is to exploring digital evidences present in cloud computing technologies by analyzing and extracting its features. The background and literature survey deals with the areas which are related to concepts and techniques applied in the work. Section 1.3.1, deals with Cloud forensics and its approaches used for forensic analysis. Section 1.3.2, examines the machine learning techniques used in forensic process. Cloud computing is an evolution of technology in a model of multiple stakeholders, location independent, elastic, on demand metered supply of computing resources. Cloud forensics is a technique of gathering digital evidences where the data is extracted from the technologies used in cloud computing. It relies on evidence extraction from the internal memory of cloud servers, log files, request packets etc. The challenges present in area of cloud forensics attract analysts and researchers towards this field. The noteworthy works done in this field is discussed in this section.

Author survey and uncovered the unaddressed issues related to the forensic analysis of the cloud computing services that influence the digital investigations process such as Decentralization of data centers, Physical locations unknown or not accessible, Specific logging volatile, Dependency on CSP, Metadata/Provenance protection, Volatile data.

The decentralization of data centers in other countries may create jurisdictional challenges during the search and seizure process to locate digital evidence.

Physical extraction of the data from the cloud servers is in most of the cases technically not possible due to the remote location of the data or the fact that the location of the data cannot be determined.

There might also be a requirement that the CSP should put a litigation hold on the data to prevent the data from being destroyed or changed prior to the time that the data can be collected.

In this paper author analyses traditional digital forensic procedures for each phase in context of the cloud forensics. Deduced that traditional digital forensic procedures addressing several issues which influence the forensic investigation in cloud environment. Finally, author identified and discussed the challenge occurred in cloud forensics such as Distributed, virtualized and volatile storage, Cross-jurisdictional standards, Imaging all physical media in a cloud is impractical, Evidence from multiple time zones. In tradition digital forensics investigator creates a byte-for-byte copy (Image) of the entire memory device from the existing tool such as 'dd' an open-source tool as shown in Figure. The virtualization of data storage in a cloud makes it complex to identify and isolate the portions of the one or more physical storage devices owned by a cloud provider as shown in Figure. Author proposed an approach to ensure the authenticity and integrity of cloud evidences through encrypt-then-sign and sign-then-encrypt method. The encryption and decryption

operations are the inverse of each other. Hence-

[{Msg} Kpri] Kpub = {[Msg] Kpub} Kpri = Msg

{ Where-

Kpri = Private Key

Kpub = Public Key

Msg = Plaintext Massage }

Encrypt-then-sign is suffers from the plaintext-subsection attack and cipher text stealing attacks. Sign-then-encrypt approach suffers from a forwarding attack.

Alice will encrypt the message M using Victor's public key and then sign the result using her secret key. Then, Alice sends [{M} Victor] Alice to Victor. However, Bob can set himself as a man-in-the middle and intercept messages from Alice to Victor. Bob can then use Alice's public key to compute {M} Victor. Then, Bob signs it and sends [{M} Victor] Bob to Victor. When Victor receives [{M} Victor] Bob and verifies Bob's signature on it, Victor will assume that Bob has made this astonishing discovery and Alice cannot disprove Bob's claim. Author introduces an approach for digital evidence acquisition in cloud computing environment to maintain the authenticity and integrity of the extracted data. Authenticity and integrity of evidence are two most important fundamental requirements in digital forensic investigation for admissibility of evidence in court of law. Author uses encryption and digital signature algorithms in different patterns to ensure both fundamental requirements. Author proposed a sign-encrypt-sign and Encrypt-sign-encrypt approach to maintain the authenticity, confidentiality and integrity of the evidences in cloud forensics which is shown in Figure. Author also focuses on digital forensic issues in cloud computing such as shared hosting and disk cloning of big data.

- **Shared hosting**: It is common in the cloud. The shared host contains both suspicious material data to the cybercrime and sensitive immaterial information.
- **Cloning of large database:** Traditional disk cloning in digital forensics might be unbearable in cloud due to the lager database and distributed storage. It is a major issue in cloud to gather evidence from the larger storage.

In this paper author reviews the digital forensics analysis process and investigation steps in cloud computing in context of existing digital forensics tool kit. Author also discusses the challenge occurred during cloud forensic investigation such as Service Level Agreements, Transparent Behaviour, Internal Staffing, External Chain of Dependency. Author suggested the need of developing the cloud forensic tools for the investigator that may extract and image data as the evidence on the cloud. Existing common forensics tool kits are-

IV. Machine Learning Techniques in Forensics

Security-related applications require careful consideration of their adversarial nature and novel learning methods with improved robustness against potential attacks. As it becomes increasingly difficult to reach the desired properties solely using statically designed mechanisms, learning methods are being used more and more to obtain a better understanding of various data collected from these complex systems. The section deals with the works carried out in this field. Author analyzed the role that machine learning might play in computer forensics, by considering the understanding and modelling simplification provided by machine learning for complex problems of computer security. A brief literature review has been quoted illustrating the areas of computer forensics where machine learning techniques have been used such as e-mail, network forensics etc. Application of machine learning in security provides enhanced simplification and effective problem solving. Author focused on the idea of implementing Machine learning for computer security in different fields such as Malware detection. Simple classification of malware cannot go beyond hashes, simple rules, or heuristic fingerprints. Anomaly based classification of malwares has been identified as the best suitable technique for full-fledged classification.

V. PROPOSED WORK

Proposed work contains six phases as Log File Repository, Preprocessing of Log File, Fine-grained Log Feature Selection, Feature Modeling, Classification, and Abnormal Behavior Analysis. The framework is as follows.

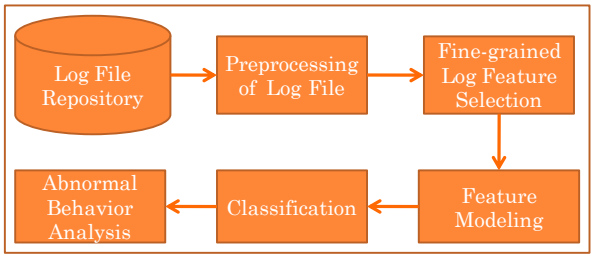

Figure3. Proposed Framework for Malware Detection through Log File

VI. CONCLUSION

Extracted data from the cloud computing will be treat as the evidence through mine the information from the data. Hence, data mining technique has been applied. This work proposed a model to forensically investigate the cloud services and mitigate the issues occurs during the evidence extraction, correlation and investigation process.

REFERENCES

[1.] Gertruida Meyer and Adrie Stander, "Cloud Computing: The Digital Forensics Challenge", in

Proc. of Informing Science & IT Education Conference (InSITE), Tampa, Florida, USA, pp. 285-299, 2015.

[2.] George Grispos, Tim Storer and William Bradley Glissons, "Calm Before the Storm: The Challenges of Cloud Computing in Digital Forensics", International Journal of Digital Crime and Forensics, Volume: 4, Issue: 2, pp. 28-48, 2012.

[3.] R. McKemmish, "What is Forensic Computing," Canberra Australian Institute of Criminology, 1999.

[4.] M. M. Pollitt, "Computer Forensics: An Approach to Evidence in Cyberspace," in National Information Systems Security Conference, Baltimore, 1995.

[5.] Mahmoud M. Nasreldin et Al., "Digital Forensics Evidence Acquisition and Chain of Custody in Cloud Computing", International Journal of Computer Science Issues (IJCSI), Volume: 12, Issue: 1, No: 1, pp. 153-160, 2015.

[6.] S. Hou1, R. Sasaki, T. Uehara, and S. Yiu, "Verifying Data Authenticity and Integrity in Server-Aided Confidential Forensic Investigation," Lecture Notes in Computer Science 7804, Springer, 2013, pp. 312-317.

[7.] G. G. a. F. R. D. Ariu, "Machine learning in computer forensics (and the lessons learned from machine learning in computer security)," in ACM workshop on Security and artificial intelligence, New York, 2011.

[8.] A. Joseph, "Machine Learning Methods for Computer Security", Dagstuhl Manifestos Perspectives Workshop, vol. 3, no. 1, 2013.

[9.] M. P. Mohite and S. B. Ardhapurkar, "Overcast: Developing Digital Forensic Tool in Cloud Computing Environment", In IEEE sponsored 2nd Int. Conf. on Innovations in Information Embedded and Communication Systems ICIIECS'15.