Review on Cloud Forensics: An Open Discussion on Challenges and Capabilities

Suchana Datta  
Department of Computer Science & Engineering  
Maulana Abul Kalam Azad University of Technology (formerly known as WBUT)  
Salt Lake City, Kolkata, India

Koushik Majumder  
Department of Computer Science & Engineering  
Maulana Abul Kalam Azad University of Technology (formerly known as WBUT)  
Salt Lake City, Kolkata, India

Debashis De  
Department of Computer Science & Engineering  
Maulana Abul Kalam Azad University of Technology (formerly known as WBUT)  
Salt Lake City, Kolkata, India

ABSTRACT

Amongst all recently emerging research paradigms, Cloud Computing is very much significant due to its utility services provisioning with shared and virtualized resources. Cloud is going to provide Everything-as-a-Service (EaaS) in very near future because all the services (Infrastructure, Platform, Software) will be made available as and when required and that too with high flexibility and low cost. Consumers can avail all the services without investing for infrastructures. There lies the spark of it which attracts the cloud attackers to get indulged in malicious activities and this creates a threat for this technology. Cloud Forensics is a new outlook introduced to identify, analyze and investigate these security threats. This paper insights a better awareness about cloud forensics illustrating all its related technical aspects, few of the suggested architectures and thus it identifies the major research scopes and challenges as well so that Cloud technology can be made secure from various threats and attacks.

General Terms
Cloud Forensics.

Keywords
Digital forensics; cloud computing; cloud forensics; SaaS; PaaS; IaaS; virtualization;

1. INTRODUCTION

Cloud Computing [1-6] is the most challenging and interesting technology that has emerged in recent years. All the provisions provided by this technology have made this not only a popular one but also a platform that has made our life very easy and convenient. But all its beauty becomes a menace whenever any cloud crime is performed by some of the malicious users or hosts. At early age, whenever any cyber-crime occurred, investigators were devoted to identify the malicious user or host by probing each and every IP addresses one by one[7-8]. Though it has too much overhead from the investigators’ aspect, but still investigation can be done anyway. But with the advent of cloud technology, this type of one to one investigation will make this process a tedious one. It is extremely difficult to identify malicious users or hosts from such a huge amount of connections and distributive architectures. There lies the reason where researchers came up with a new research domain called Cloud Forensics. This paper contributes towards the introduction of this challenging technology, frameworks, its challenges and some of the proposed solutions. Finally, little research directions have been identified to make readers acquainted and attracted towards this research field.

2. PREREQUISITES

2.1 Cloud Computing

Cloud computing, shortly expressed as ‘on-Demand Computing’ is mainly based on Internet. In this computing technology, data, information and various required shared resources are provisioned to computers or various other electronics devices on-demand. Different users and enterprises are provided cloud computing and storage solutions so that they are able to store and process data in several third party data centers. Shared services and converged infrastructure is the main concept of this new technology. So, basically information resources and underlying infrastructure and mechanism of delivering those data to the client are separated in this technology. In 2009, the US National Institute for Standards and Technology (NIST) proposed the definition of cloud computing as “…a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.”[1-4]

2.2 Digital Forensics

A step by step mechanism aimed to analyze digital data and preserve the integrity and confidentiality of the chain of custody is expounded as Digital Forensics. This include several steps, like- proper identification of evidences, collection, preservation of sized media extracted from the
crime scene, validate those evidences, analyze and interpret accordingly, document those interpreted results and presenting the documented results to the court room. Information based on probative value stored or often transmitted digitally based on SWGDME (Scientific Working Group on Digital Evidence) definition generally, is referred to as Digital Evidence. It is quite obvious that the evidences those are collected and examined maintaining all the steps is under the control of specific law enforcement which is quite infeasible when the cloud environment is concerned due to the distributed and black-box architecture of Cloud.

2.3 Cloud Forensics

Utility services on shared and virtualized resources are provisioned by a completely new operational paradigm named as Cloud Computing. It can be strongly assured that Cloud will achieve the capability to provide Everything-as-a-Service (EaaS) [7-14] to all the cloud consumers. CSPs are making all the software, platforms and infrastructures available as and when required over the internet very flexibly and at a very low cost. All the consumers are provisioned in such a way that they are able to access various resources for computing at a remote location. It is not necessary for users to have the infrastructure with them. From the financial benefits aspect, it is becoming a significant field that tempts the attackers more and more CSPs and users are participating in this field. Several security approaches are taken into account to mitigate these kinds of threats and ensure protection to the resources of cloud so that its potential can be exploited to the fullest. Cloud forensics [9-12] [15-18] is such an approach that attempts to investigate as well as analyze cloud security threats. It plays the role of a deterrent, improving security and reducing network crime rate.

2.4 Cloud Crime

Cloud crime [4][7][19] can be defined as the extension of the computer crime. The name itself says that cloud computing environment is involved in this type of crime, i.e. Cloud may act as the subject or object or tool, those are related to the performed crime. In many cases, it may be observed that Cloud Service Provider himself is the target of the crime currently being investigated. Cloud can play the role of an object when CSPs are directly affected by the Distributed Denial of Services (DDoS) attacks. Certain sections of the cloud are targeted by these types of attacks or the entire cloud can be a target as a whole. On the other hand, if the criminal activity is committed within the cloud, in that case we can consider the cloud as the subject of the crime. Theft of identity of the accounts of cloud users is such type of crime. And finally there may be some cases where cloud is used to conduct any crime, in such cases cloud plays the role of a tool [10][16].

2.5 Migrating Forensics to the Cloud

In these circumstances, it is quite obvious that there must be some ways which will bridge the gap between traditional Digital Forensics procedure and Cloud Computing technology. There lies the need where researchers came up with the completely new domain, Cloud Forensics. All the IT services [19-20] are now-a-days being dominated by this Cloud Computing’s highly dynamic architecture. With the rapid flourish of cloud computing, the traditional digital forensics methodology is becoming insufficient. This is because unlike digital forensics, both the machine and the process of investigation are out of the reach of investigators. A complete dependence upon Cloud Service Providers is faced by the investigator at each and every step of the investigation process from accessing the machines to the collection of evidences. The main difficulty arises when the virtual evidences, those are in the form of VM snapshots, are to be digitized and investigated. This is due to the lack of proper knowledge of investigators. Whenever investigators seize any virtual instances for the investigation purpose, CSPs are supposed to shut all other virtual machines forcefully which is completely against their terms and policies [5].

3. FORENSICS METHODOLOGY

The traditional digital forensic process [3][6-7][9][11] undergoes the following steps which can be incorporated in cloud forensics considering its different service and deployment models:

- **Identification**: Reporting against malicious activities is considered as identification which arises when any individual or CSP authority places complaints against undesirable issues. This phase comprises with two types of identification, i.e. Incident Identification and Evidence Identification.

- **Collection & Preservation**: Due to the distributed architecture of cloud, the traditional digital forensics process faces lots of challenges. Since data collection is nothing but the physical acquisition of investigation related data, in most of the cases
investigators are supposed to be dependent upon CSPs. This dependence never guarantees 100% availability of the resources, neither its preservation after the collection of data. The storage capacity of the collecting device is another important issue since no data is put in a single location in cloud architecture.

- **Examination**: After collecting the desired large amount of data with the help of CSPs, these are to be processed through a combination of manual and automated processes too. The main motto of examining is to extract and assess data of the particular interest of the classified incident scene. The integrity must be preserved through this entire process.

- **Analysis**: All the relevant data are analyzed using suitable and legally justified techniques so that the proper suspected hosts or data can be identified through this investigation procedure. Investigators must be able to meet up with all queries those are raised during the presentation of the analyzed report to the court.

- **Reporting & Presentation**: These are the final stages of any investigation process. Report must be comprised with all the details of this investigation process (explanation against what, why and how). The detail report is to be presented to the jurisdiction section with authenticity and accuracy without tampering the evidences which is the most crucial part of the investigation.

4. RELATED FRAMEWORKS
The most challenging part is to incorporate traditional digital forensics into black box architecture of the cloud. Numerous researches are being done to make cloud forensics system an efficient and robust one. As a consequence, few frameworks have also been proposed by several researchers in order to mitigate the gaps between digital forensics and cloud computing. Some of those proposed models are being discussed below.

Keeping the numerous challenges and variety of attacks faced by the cloud environment in mind, a framework is proposed by the authors in [17]. Some of the modules of this architecture have already been implemented whereas some are still in progress. Meghdoot presented BOSS cloud stack which is used to deploy the private cloud that helps CDAC, Chennai to carry out the experiment. The entire architecture has been divided into four different layers, like- abstract layer, front end, middle end and back end. All the prerequisites of this architecture form the abstract layer. Front end is mainly the interacting layer of the model, whose main component is the API interface. Middle end is concerned with the database maintaining all the relevant data for the forensic process. Several dynamic data mining techniques are the main components of this layer of the proposed model. These mining techniques mainly segregate the relevant evidences which are the proofs for a specific crime scene and deliver those to the presentation layer.

Another framework proposed by [18] where the authors emphasized on the admissibility of the evidences. They also mentioned that the efficacy of the evidences must also be guaranteed for ensuring their acceptance in litigation. They presented a new process model for forensics which differs significantly from that of the traditional forensic process. First identifying the purpose of the investigation and then choosing the device, software and platform accordingly are the main building block of this proposed model. The technology behind the concerned cloud is also to be verified so that the specific investigation process can be executed smoothly and perfectly. Users’ role at the specific local terminals, negotiation with the Cloud Service Providers and collection of potential evidences etc. have been discussed thoroughly by the authors.
presentation, each of these phases are further subdivided into several modules as depicted in fig: 6. Each and every module bridge the gap using these hypervisors. This model also goes through several phases but the main motto is to identify the proper host, and then collect data from that particular host. All the relevant evidences are then presented to the respective jurisdiction by preserving all its integrity and authenticity without being hampered by any contamination and temperament. The preparative phase comprises with identification, preparation and approach strategy. Whereas, the investigation module has been subdivided into several phases like, preservation, collection, examination and analysis. And finally the presentation phase mainly deals with the provenance of the evidences from the identification to the court of law.

According to [19], users are provisioned numerous cloud services by the Cloud Service Providers. Various confidential and sensitive information is thus stolen by the malicious users and as a consequence, the trust of the CSPs is affected badly. In this respect, for the purpose of monitoring VMs of the consumers and detecting malicious activities, CSPs must be provisioned with either introspection mechanism [20] or Intrusion Detection System [21].

[22] represents such a mechanism in order to monitor and track different malicious activities. An intrusion detection system (IDS) has been incorporated within all VMs (Virtual Machine) and VMM (Virtual Machine Manager). CSPs are ought to manage and monitor the system besides deploying it at clients VMs. Suspected VMs are monitored for a large period of time than the identification phase since the more the investigator monitors the suspected VM, the more they will be sure about the presence of the required evidences and the possibility of its malicious behavior. Investigators then move suspicious VMs to other nodes once they are identified so that the rest of the VMs’ confidentiality, integrity and authenticity can be preserved. Moving or isolating from the original position is the more effective way to protect evidences from temperament and contaminations.

A new technique has been introduced by Delport et al. [23] through which VM instances, those to be investigated can be isolated in the Cloud so that evidences can be collected accurately. After authentic evidences are collected, they pass through various forensic tools for analysis and then placed to the jurisdiction section. This investigation methodology on the basis of VM snapshots are depicted in the Fig-7. Eucalyptus private cloud has been used as the experimental setup. The running snapshots of all instances are collected first where any sort of change is found in the source data. The proposed model is dedicated to reduce time and space of the investigation exponentially. If investigators can identify malicious VMs, they take snapshots properly and store those evidences persistently.

5. CHALLENGES & SOLUTIONS

Cloud forensics has immense challenges due to its abstract and black box architecture. Since it is a technology in its budding ground, researchers are quite interested to expose this new technology in a full-fledged mode. Here some of those challenges rather research issues have been addressed which can be considered as future research scopes for the betterment of this grooming technology in order to live in a safe and sound cloud environment. Besides addressing the challenges and their proposed solutions, the following table helps researchers to find at which level of cloud services these models have been proposed.
## Table 1. Challenges and Solutions of Cloud Forensics

<table>
<thead>
<tr>
<th>Phase</th>
<th>Challenges</th>
<th>Solutions</th>
<th>IaaS</th>
<th>SaaS</th>
<th>PaaS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence Identification</td>
<td>Finding out relationships amongst several evidence related files.</td>
<td>[24] proposes a metadata based association to identify the relations on the syntactic and semantic level.</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Identify and rank remote malicious hosts authentically.</td>
<td>[25] presents a model that selects hosts based on probability and thus minimizes the cost of investigation.</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Identifying cloud attacks and applying suitable forensics.</td>
<td>[26] proposes a process to identify Syslogs relevant to that attack and apply forensics using Eucalyptus, open source cloud computing platform.</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>Evidence Collection</td>
<td>Robust Service Level Agreement (SLA)s [27].</td>
<td>SLA has been described as robust and a CIA (Confidentiality-Integrity-Availability) system for information security [28].</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Collecting consistent evidences from VMs.</td>
<td>In [29] an efficient Data Mining and log managing technique has been proposed to come out of the addressed problem.</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Automated system with the ability of reconstruction of user activities.</td>
<td>[30] proposes SigDiff, an automated reconstruction model of user activities which is to be incorporated in cloud system.</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Identifying as well as localizing SSL/TSL attacks</td>
<td>[31] reports the development and implication of Cross Bear, a tool that identifies MitM attacks on SSL/TLS, in cloud computing platform.</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>A suitable monitoring system based on virtualization.</td>
<td>VAIL [32] has been proposed which mainly monitors mini intrusive live system based on virtualization.</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Accessing outsourced data, collected for forensic, very easily.</td>
<td>A Leaked Access Credential Tracing (LACT) has been proposed [33] which ensures the security of the outsourced data which helps in evidence collection.</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Security and reliability issues over Cloud storage and outsourced data.</td>
<td>Authors propose [34] a probabilistic scheme of challenge-response to prove the availability of all the clients’ files, kept in a specific cloud server.</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>An automated forensic data collection system [35].</td>
<td>Investigators will get adequate artifacts that help them to reconstruct malicious activities and attacks by the proposed model [36].</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Forensic Analysis of textual information.</td>
<td>[37] presents a cloud-based framework that retrieves textual evidences either from a picture or a PDF document by Optical Character Recognition (OCR) algorithms.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evidence Integration</td>
<td>Designing suitable model for verifying data integrity and fault tolerance [38-39].</td>
<td>[40] proposes a TP based fault-tolerant and data integrity verification scheme for secure and effective performance.</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Data integration with the help of public auditability.</td>
<td>A dynamic Merkle Hash Tree [41] has been constructed to propose an advanced dynamic evidence integrity model over this tree structure.</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Verifying integrity for secure forensic analysis.</td>
<td>A hassle-free and fixed rate cloud model has been proposed [42] including the mechanisms for integrity verification.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ensuring protection of evidences from being tampered and contaminated [43].</td>
<td>A cloud separation idea [44] has been introduced so that the movements of cloud instances can be accomplished and the cloud architecture can be divided too.</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>Evidence Preservation</td>
<td>Ensuring secure logging and management in cloud.</td>
<td>Homomorphic Encryption has been proposed [45] so that the privacy and confidentiality of the log data can be ensured.</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>Evidence Analysis</td>
<td>A suitable and sound forensics tool for Open Stack.</td>
<td>Authors [46-47] proposes a platform where FROST has been incorporated that works upon management of cloud instead of guest virtual machines.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Botnet suppression at the time of investigation.</td>
<td>GARLIC has been proposed by the author [48] with the ability of automation and distributed Botnets suppression as and when required.</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td></td>
<td>Implementation of anti-phishing technique while investigation.</td>
<td>LARX (Large-scale Anti-phishing by Retrospective data-exploration) has been introduced [49] towards an offline phishing identification system.</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Forensic computation in a very large scale.</td>
<td>MPI MapReduce (MMR) has been proposed [50] based on the suitable implementation of the MapReduce model.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
6. FUTURE RESEARCH DIRECTIONS

In the cloud architecture, consumers are served by several Cloud Service Providers according to their demands. Since cloud is a distributed platform, it is quite evident that consumers are supposed to get cloud services from several CSPs rather than any singleton provider. All the CSPs basically reside within a network or any sub-networks of those networks. In this respect when a malicious activity is reported, all those inter and intra related hosts must be taken under the consideration of the investigation process. In this scenario if an investigator is supposed to examine all the IP addresses, it becomes a menace. To come out of this particular overhead there must be some automated system which will identify all the probable malicious hosts with the help of the previous history. By training the system based on the log of the malicious hosts, an investigator can identify all the malicious hosts whenever a set of hosts, with the same incident and attributes are to be examined by calculating the probability of being malicious. This reduces the investigation time and costs in a major rate.

It is quite obvious that if the CSPs cannot ensure the probity of the information, consumers will not leave the complete computation control over cloud. At the same time it is particularly true from the cloud forensics investigators’ aspect too. Keeping this in mind, authors [42] proposed a job-based SaaS cloud model including integrity verification mechanism. Since investigators have to collect all the evidences from a distributed architecture of cloud, therefore there is always a question of evidence integration verification properly and accurately due to much dependence on CSP. Therefore future researches can be concentrated upon making a trustworthy and effective framework for the investigator so that they won’t be misguided at the time of evidence collection and integration.

In cloud environment, investigators are supposed to analyze a large volume of collected evidences using all cloud service models. OPENSTACK is such a cloud computing software platform which is free and open source. Users primarily deploy it as an IaaS. In [46] authors proposed FROST which is a new forensic tool for this open stack cloud computing platform that supports mainly an IaaS cloud and provides quite trustworthy forensic acquisition. Unlike traditional acquisition tools, its dependency upon CSPs is lesser, since it overcomes nontrivial challenges in the accumulation of remote evidences when log data are stored in the hash table. But in case of SaaS and PaaS investigators are supposedly more dependent upon CSPs as they give less control to cloud users where the proposed model faces challenges. As in these two service models, investigators need to be more dependent upon CSPs, so a model can be incorporated that build trust on the CSP as well as enhance this trust.

7. CONCLUSION

With the advent and massive uses of Cloud Computing technology, the users are being benefitted in one hand; on the other hand it creates a threat to all the cloud technology users. The interesting and multi-tenancy behavior of cloud makes it prone to various malicious activities, where lies the necessity of migrating digital forensics techniques to the world of cloud. In this paper, the main motivation of the Cloud Forensics technology, the building blocks and various proposed frameworks have been explained in brief in order to find out the basic challenges and issues of this emerging technology. Finally the paper came up with certain future research areas, such as- identifying and ranking different remote hosts using classification algorithm, incorporating evidence integration verification model to ensure the trust on CSP from the client side, embedding a trustworthy CSP within the SaaS and PaaS platform, so that this new technology which is very sensitive in nature can gain more accuracy and authenticity.

8. REFERENCES


