International Journal of Computer Architecture and Mobility (ISSN 2319-9229)
Volume 7 -Issue 4, March 2019

Digital Forensics and Analysis of Intrusion Detection Techniques in Cloud Environment

Prachi Vaishnav
Shri Vaishnav Vidyapeeth
Vishwavidyalaya Indore-Ujjain
Road, Indore – 453111, India
pvaishnav.98@gmail.com

Romil Jain
Shri Vaishnav Vidyapeeth
Vishwavidyalaya Indore-Ujjain
Road,Indore – 453111, India
romilj01@gmail.com

Somya Lalwani
Shri Vaishnav Vidyapeeth
Vishwavidyalaya Indore-Ujjain
Road,Indore – 453111, India
somyalalwani9@gmail.com

***Abstract*-Today, Cloud Computing provides elastic and pay-per-use based services to its users and therefore is the preferred choice of every IT organization. However, security and privacy is a major barrier in its attainment because of its open and distributed architecture that is vulnerable to intruders, cyber criminals to perform fraudulent activities without leaving any traces behind. It has become a tough job for the cyber forensic investigators to find out who the real culprit or attacker is, due to various challenges faced during the forensic investigation like dispersion of data. The innovative nature of cloud computing has created unique challenges in the field of digital forensics, faced by the investigators. In this paper, we have also reviewed the approaches to investigate the forensic issues in cloud computing using intrusion detectionsystems**.

Keywords: Cloud computing, Digital Forensics, Intrusion Detection System

I. INTRODUCTION

A. Cloud Computing

Cloud computing refers to applications and services that run on a distributed network using virtualized resources and accessed by common Internet protocols and networking standards [1]. Cloud computing is a model for enabling global, convenient, on-demand network access to a shared pool of configurable computing resources (like applications, networks, storage, servers, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction [2]. It offers computing resources to the clients, who pay for these services per use.

There are specific deployment and delivery models through which the services are delivered to the clients. The deployment models provided are public, private, community and hybrid. In public model, the cloud infrastructure is available for publicuse, alternatively for a large industry group and is owned by an organization selling cloud services. On contrary, in private model, cloud infrastructure is operated for the exclusive use of an organization. The cloud is managed by that organization or a third party. Whereas a community cloud is the one where the cloud has been organized to serve a common function or purpose. It may be for one organization or for several organizations, but they share common concerns such as their mission, policies, security, regulatory compliance needs [1]. In the last model, a hybrid cloud combines multiple clouds (private, community of public) where those clouds retain their unique identities, but are bound together as aunit.

The delivery models provided in cloud computing are Software as a Service (SaaS), Platform as a Service (PaaS) and Infrastructure as a Service (IaaS).In SaaS, the customer's responsibility begins and ends with entering and managing its data and user interaction. Everything from the application down to the infrastructure is the vendor's responsibility [3]. Whereas in PaaS, the service provider manages the cloud infrastructure, the operating systems, and the enabling software. The client is responsible for installing and managing the application that he is deploying. In IaaS, service provider manages the entire infrastructure, while the client is responsible for all other aspects of the deployment that includes operating system, applications and interaction with thesystem.

Cloud computing architecture is divided into two sections, Front end and Back end. These two are connected through a network, usually Internet. Client resides on the front end, which includes the client's computer and the application required for accessing the cloud computing system. On the back end, there are various servers, computers and data storages that are responsible for creation of cloud.

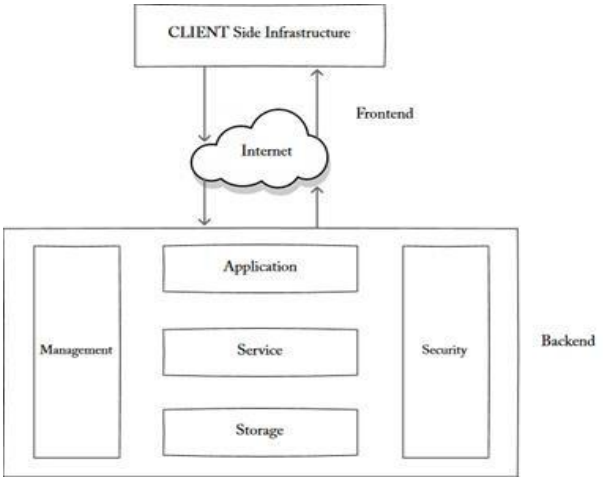

Figure 1: Cloud Architecture [14]

B. DIGITAL FORENSICS INCLOUD COMPUTING

Digital forensics is based on Locard's exchange principle. It states "perpetrator of a crime will bring something into the crime scene and leave with something from it, and that both can be used as forensic evidence. [4]" Digital forensics is a forensic science branch that consists of identification, investigation, presentation, recovery and management of facts regarding the digital evidences found on computer systems.

Digital forensic investigation is an approach of collecting, examining, analyzing digital evidences and managing the case. The aim of forensic investigation is to extract evidences from this information and use these evidences in the court of law.

Figure 2: Digital Forensic Process Model

II. LITERATURESURVEY

A. *"INTRUSION DETECTION IN CLOUD COMPUTING ENVIRONMENT USINGNEURAL NETWORK"*

This research work proposed a hybrid algorithm PCANNA (principal component analysis neural network algorithm) for reducing the number of computer resources, both memory and CPU time required to detect attack and the PCA (principal component analysis) transform is used for reducing the features. The trained neural network is used to identify the any kinds of new attacks [10].

B. "INTRUSION DETECTION AND PREVENTION IN CLOUD COMPUTINGUSING GENETICALGORITHM"

They proposed an intrusion detection system which is based on the cloud computing to reduce the risk of intrusion on the cloud networks and cover up the deficiency of already existing intrusion detection systems. This design is based on Software-as-a service (Saas) model for detection and prevention of intrusion for cloud based users. [11]

C. "INTRUSION DETECTION ONCLOUD USING HYBRID MACHINE LEARNING TECHNIQUES"

This paper work proposed hybrid machine learning based Intrusion Detection System (IDS) which is a combination of supervised and unsupervised machine learning algorithm. This system uses Artificial Neural Network (ANN) and K-means algorithm which is supervised machine learning algorithm and unsupervised machine learning algorithm respectively for detecting known and unknown attacks respectively.[12]

D. *"COMPARING DIFFERENT SOFT COMPUTING TECHNIQUES FOR CLASSIFICATION OF SATELLITEIMAGES"*

This paper work provides a comparison between the techniques for satellite image classification based on fuzzy set theory, Support vector machines, genetic algorithms and ANN (artificial neural network) [13].

III. CHALLENGESFACEDDURINGFORENSIC INVESTIGATION IN CLOUDCOMPUTING

A. IDENTIFICATION STAGE: The machine has to be first identified where the illegalactivities

Happened. The dynamic nature of cloud leads to several obstacles that oppose the investigators to conduct this stage. They are:

1. *Access to the evidence in logs*: The identification of evidence via various sources could be challenging within the cloud environment [6]. Due to distributed nature of cloud (i.e. data is distributed among many hosts in multiple data centers), it may happen that the investigators do not even know the data's location.
   On the basis of the cloud service, system statutes and logs files are available. From the log information, it is possible to identify crucial information such as the IP address of the attacking machine, browser type etc [6].

2. *Volatile Data*: Volatile data doesn't sustain as soon as the power is turned off. Likewise, when a VM is turned off or restarted, all the data will be lost unless was stored somewhere [6].

3. *Lack Control of The System:* A number of obstacles occur when digital investigators carry out the evidence acquisition. Consumers have limited access and control at all levels within the cloud environment and have no knowledge where its data is physically located [7]. This, the opportunity to perform a physical acquisition of the disk isremoved.

4. *Lack of Customer Awareness:* Lack of CSP transparency leads to loss of important terms regarding forensics investigations in the Service level Agreement (SLA) [6]. This issue is possible to all three services models [6].

B. DATA COLLECTION & PRESERVATION

1. *Dependence on Cloud Forensics Providers*: Both customers and investigators have limited control on the system and thus heavily depend upon the CSP in collecting the digital evidence from cloud computing environment. This dependence introduces serious issues of the CSP's trust and evidence integrity [6]. These include but are not limited to:
i. Most CSPs keep only a limited amount of backups due to large amount of data causing problems if data getsdeleted.
ii. In case of an incident, the cloud provider will focus upon restoring the service rather than preserving the evidence and handling it in a forensically sound manner[6].

2. Isolating a CloudInstance
For any forensics process, it is vital to isolate the incident environment and that particular instance connected with the incident in the cloud environment. However, due to the fact that data instances share storage with multiple instances, so achieving such a task in the cloud environment is not a trivial task.

3. *Data Integrity:* The original evidence shouldn't be changed at all. A piece of incident related information has to be listed in the chain of custody registers [6]. This maintains the integrity of the digital evidence, including how, where and by whom the evidence was collected, how the evidence was stored. If the evidence is preserved improperly it might become valueless in the court oflaw.

4. Time Synchronization

The time stamps play very significant role in evidence. Though the date and time stamps of the data are questionable when they are from multiple systems [6]. Furthermore, the difference in time zones between cloud servers and cloud clients can affect the integrity, reliability and admissibility of evidence [6].

5. *Cloud Literacy of Investigators*: The training provided to investigators is not up to mark and little training materials and forensic procedures are made available.

6. *Chain of Custody*: The most critical problems in digital forensics arena as it illustrates how the evidence was collected, analyzed and preserved to further present it in admissible way at the court of law.

C. ANALYSIS & EXAMINATION

1. *Lack of Forensics Tools*: The available forensic tools have various limitations and cannot cope up with the distributed and elastic characteristic of the cloud computing [6].

2. *Crime Scene Reconstruction*: In order to understand how illegal activities were committed, reconstruction of the crime scene is done, which is a problem in the cloud environment. However, regeneration event can be done where a snapshot is done due to occurrence of every attack [6].

D. PRESENTATION

How the physical location must be specified is still not clear due to distributed and shared resources.

IV. INTRUSION DETECTION SYSTEM(IDS)

An Intrusion Detection System (IDS) can be a hardware device or software application which monitors system or host or network activities for fraudulent threats or policy violations, creates and sends reports to a Management Station or System Administrator who decides whether to take an action on the intrusion or not[8] . IDS can be classified into two - Host based and Network based

A. Network Intrusion Detection System(NIDS)

It gathers the traffic of entire network and analyze it to detect possible intrusions or check for any malicious activities like port scanning, DoS attacks etc. NIDS usually performs intrusion detection by processing the IP and transport layer headers of captured network packets [8].

B. Host based Intrusion Detection System (HIDS)

It collects information from a particular host and analyzesittodetectintrusiveevents.Thiscollected information can be related to system logs or audit trails of operatingsystem.[8]

V. IDS IN CLOUDCOMPUTING

For an IDS to be deployed on a system it is mostly at the border of the networking infrastructure that is to be protected from the external attacks [9]. In a cloud computing environment, where computing and communication resources are shared among several users on an on-demand, pay-per-use basis, such strategy will not be successful and effective. Thus proper strategies in the cloud need to be implemented to prevent the attacks that originate within the cloud itself and also from the users using the cloud technology from different locations all over the globe through the Internet [9].

A. Types of Intrusions incloud

1) Attack on or threat to Virtualmachines

The attacker controls the virtual machines by compromising the hypervisor. Common attacks include SubVir, BLUEPILL, and DKSM which enable hackers to supervise host through hypervisor [8].

2) User to root (U2R)attacks

Remote to local attack (r2l) has been widely known to be launched by an attacker to gain unauthorized access to a user's system in the entire network. In a similar way user to root attack (u2r) is usually launched for illegally obtaining the roots privileges when legally accessing a local machine[8].

3) Insiderattack

This attack is done by the authorized users who try to exploit the files that are assigned or not assigned to them officially. This attack impacts the confidentiality of cloud user[8].

4) PortScanning

Port scanning is used by attackers to obtain list of closed ports, open ports and filtered ports and then launch attacks against the services running on open ports.

Port scanning involves techniques like SYN scanning, ACK scanning, TCP scanning etc[8].

5) Backdoor channelattacks

Attackers use backdoor channels to get control of users resources and utilize it as a host to launch DDoS attack. It targets the availability and confidentiality of the user[8].

6) Denial of Service (DoS)attack

The attacker sends a large number of network packets to overwhelm the available resources. The attacker may dispatch huge number of requests through zombies to access VMs thus disabling their availability to legitimate users which is called DoS attack. It targets the availability of cloud resources [8].

B. Detection Techniques used by cloud computing inIDS

1) Signature Based Detection

Signature based detection is carried out by comparing the information collected from a network or system against a database of signatures [8]. A signature is a already defined set of rules that corresponds to a known attack. This method helps network managers with average security expertise to identify intrusions accurately. It is a pliable approach since new signatures can be added to database without modifying existing ones.

2) Anomaly Based Detection

Anomaly based detection compares current user activities against already loaded profiles of users or

networks to detect abnormal activity that may be intrusion.

Types of anomaly based detection:

a. Hybrid Machine Learning Techniques: Machine Learning Algorithms are broadly classified into 3 categories:
1. Supervised Learning
2. Unsupervised Learning
3. ReinforcementLearning
Hybrid machine learning algorithm is based on supervised and unsupervised learning.
Supervised learning: for detecting known attacks.
Example: Artificial Neural Network
Unsupervised learning: for detecting new attacks.
Example: K-means algorithm

b. Fuzzy Logic: Fuzzy logic system is an automatic system that is able to imitate human actions for a certain task. These systems involve 3 major operations:
The first operation being fuzzification, which is the mapping from a crisp input to a fuzzy value.
The second is inference, which involves the making of inference in accordance to fuzzy rules in the form of IF-THEN.
The third step is defuzzification, which again transforms the fuzzy output of the expert system into a crisp output. Fuzzy logic removes uncertainty or it converts the uncertain data into a certainone.

c. Artificial Neural Network: Artificial Neural Network (ANN) is simulation of the capability of human brain mathematically. The category of supervised method requires inputs and target outputs. The Back-Propagation algorithm, which is an ANN, processes training tuples, compares the predicted value with actual values and learns iteratively. The weights are modified such that the mean squared error is minimum. This modification is done in backward direction. Hence, this technique can be used only for the knownattacks.

d. Genetic Algorithm: Genetic algorithm is an evolutionary algorithm. The genetic unit compares the pattern and is also a decision making unit. The data is analyzed in detail and then the algorithm matches the data using Fitness Function with known behaviors that are stored in the existing knowledge base. If existence check is passed it isfurther analyzed and matched with the voice records for the proper recognition according to the fitness function. If the fitness is proved, then different authentication checks are applied to reach the result. If the desired record is not matched it reproduce the new record by mutation, performs fitness test and again compares with the data present in knowledge base. The above is continued for three generations as per termination criteria. Based on the results, it identifies the malicious behavior, provides access or exit and generates the alerts usingnotifier.

3) Hybrid Detection

To improve the efficiency of IDS the combination of signature based and anomaly based detection technique is used which is called Hybrid detection technique [8]. This hybrid technique allows the investigators to detect both known and unknown attacks.

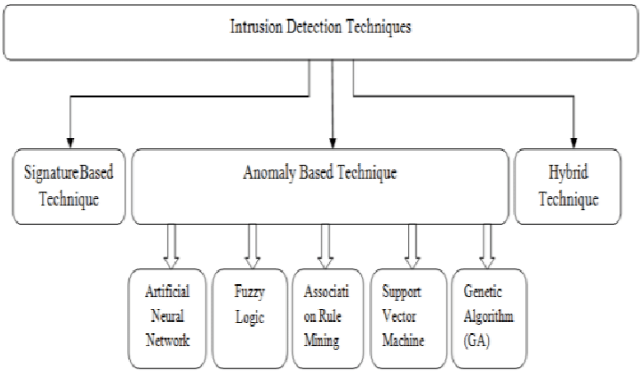

Figure 3: Types of Intrusion Detection Techniques [15]

VI. COMPARISON CHART

TABLE I. **Analysis of IDS Techniques**

Author	ZeenatMahmood, ChetanAgrawal,Syed ShadabHasan, SyedaZenab[10]	Umar Hameed, ShahidNaseem, FahadAhamd, TahirAlyas, Wasim-Ahmad Khan[11]	LaxmiMuttappanavar, Praveen S. Challagidad[12]	Sonika Jindal And Richa Jindal[13]
IDS Technique	Neural Network	Genetic Algorithm	Hybrid Machine Learning Technique	Fuzzy logic
Characteristics	Algorithm used to reduce the number of computer resources, both memory and CPU time required to detect attack.	Proposal based on SaaS model for detection and prevention of intrusion cloud based users by analyzing request patterns against malicious requests	Detection of known and unknown attacks using supervised and unsupervised algorithm	Provides a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy or missing input information.
Algorithms used	PCANNA (principal component analysis neural network algorithm), Back Propagation	Genetic Algorithm	Artificial Neural Network, K-means algorithm	Firefly algorithm
Advantages	• Enhanced security and accuracy • Provides high degree ofsecurity • It increases the detection speedwhich meets the requirements of cloud communication.	• Does not provide access to the data unless security checks arecompleted. • Focuses on voice data validity • Good at refining irrelevant and noisy features selectedfor classification.	• Easy to control complexity and provides better performance results	• Used for risk assessmentand management process • Different stochastic relationships can be identified to describe properties.
Limitations	• Structure is difficult to understand. • Requires a lot of computation	High definition visual or biometric data and simulation test is prospective goals.	• False alarm alerts are more thanexpected	• Prior knowledge is very important to get goodresults. • Precise solutions are not obtained if direction of decision is notclear

VII. OUTCOME

In cloud network, there are different intrusion detection techniques present that are used to detect anomalous activities. With these different types of techniques, unwanted activities can be traced, detected and stopped.

The main advantage of anomaly based detection is its ability to abstract information about the normal behavior of a system and hence, regardless of whether or not system has seen the abnormal activity before, attacks aredetected.

The main flaw of anomaly based detection systems is that they are vulnerable to malicious activities that are launched by intruders during the learning procedure.

The proposed genetic based IDS was handling voice intrusion only. It doesn't provide access to the data unless security checks are completed. There should be an upgradation in model to take care of all other data aspects forIDS.

Due to the self-learning nature, neural networks can detect malicious user behaviors based on information fed into them. They are also tolerant to noisy input as well as provide high degree of security but require a lot of computation.

In fuzzy logic IDS technique, precise solutions are not obtained if the direction of decision is notclear.

A hybrid machine learning based IDS is a good strategy as it allows the detection of both known and unknown attacks using both supervised and unsupervised machine learning algorithm.
The proposed system which uses combination of ANN and K-means is expected to achieve better performance than the existing intrusion detection systems.

Thus, all the mentioned techniques have their own advantages and limitations and further modifications are required in these techniques to prevent and detect intrusions in cloud.

VIII. CONCLUSION

In the past few years, the use of cloud services has increased extensively and moreover, its consequences have also increased. It has become a challenge for the forensic investigators to extract evidences related to any malicious or fraudulent activities. In this paper we have discussed the basics of cloud computing, described digital forensics and listed various challenges faced by cyber forensic investigators. The integration of IDS in cloud is a new research field which is gaining interest as the number of cloud users and the number of attackers targeting the cloud increase. We listed some of the few intrusion detection techniques and prepared a comparison chart which could be of great help to cyber forensic investigators as it would be a lot easier for them to categorize the intrusion and use the appropriate detectiontechnique.

REFERENCES

[1]https://www.oreilly.com/library/view/cloud-computing-bible/9780470903568/ch01.html
[2]https://www.nist.gov/programs-projects/nist-loud computing-program-nccp?lectureFacile=true
[3]https://www.sanfoundry.com/cloud-computing-questions-answers-entrance-exams/
[4]https://en.wikipedia.org/wiki/Locard%27s_exchange_principle.
[5]https://core.ac.uk/download/pdf/74389545.pdf
[6]Alqahtany, Saad& Clarke, Nathan &Furnell, Steven & Reich, Christoph. (2015). Cloud Forensics: A Review of Challenges, Solutions and Open Problems. 2015 International Conference on Cloud Computing, ICCC 2015. 10.1109/CLOUDCOMP.2015.7149635.
[7]J. Dykstra and A. T. Sherman, "Design and implementation of FROST: Digital forensic tools for the OpenStack cloud computing platform," Digit. Investig.,vol. 10, pp. S87–S95, Aug. 2013.
[8]YasirMehmood, UmmeHabiba, Muhammad AwaisShibli, RahatMasood, "Intrusion Detection System in Cloud Computing: Challenges and Opportunities", 2013 2nd National Conference on Information Assurance(NCIA).
[9]S N Dhage, B BMeshram, R Rawat, S Padawe, M Paingaokar, AMishra.
[10]Mahmood, Zeenat&Agrawal, Chetan& Syed, Shadab& ,Hasan&Zenab, Syeda. (2019). Intrusion Detection in Cloud Computing Environment using NeuralNetwork.
[11]Hameed, Umar &Naseem, &Ahamd, Fahad&Alyas, Tahir& Khan, Wasim-Ahmad. (2014). Intrusion Detection and Prevention in Cloud Computing using Genetic Algorithm. International Journal of Scientific and Engineering Research.
[12]Challagidad, Praveen. (2018). "INTRUSION DETECTION ON CLOUD USING HYBRID MACHINE LEARNING TECHNIQUES."
[13]Sonika Jindal &RichaJindal()2012. "COMPARING DIFFERENT SOFT COMPUTING TECHNIQUES FOR CLASSIFICATION OF SATELLITEIMAGES".
[14]http://teoriasdadenny.com/stylish-cloud-computing-architecture/perfect-cloud-computing-architecture-on-regarding- mobile-download-scientific-diagram/
[15]https://www.researchgate.net/figure/Types-of-Intrusion-Detection-Techniques_fig2_279916728.