Cloud Forensic Frameworks based on Machine Learning Techniques

Nandita Goyal, Kanika Gupta, Munesh Chandra Trivedi

Abstract: ***Cloud computing is considered to be one of the most significant and influential topics in the field of computing sciences. With time, cloud computing has paved its way in almost every aspect of human life. With the significant hike in number of users and service provider in the cloud environment, attackers are also increasing malicious activities in this area. Due to criticality in the area of the cloud computing, it is necessary that cloud environment should be safe . The concept of Cloud forensics has been introduced to establish a well-defined forensic capability in cloud environment. Although a lot of work has been carried out in the area of cloud forensic challenges and solutions, but the research on its frameworks and methodologies is still to be explored. The major challenge lies in providing a framework for analysis of massive amounts of forensic data in limited period. As proposed by many researchers, one of the best solutions for such analysis is the use of machine learning methods. This paper provides the study on the methodological aspect of cloud forensic analysis using various machine learning approaches. It gives a critical review of existing cloud forensic methodologies making use of machine learning for investigation of security related incidents in cloud. Furthermore, it provides a comprehensive study and comparison of existing frameworks using machine learning for digital and cloud forensic analysis, their drawbacks and scope for novel future research directions in thisarea.***

Index Terms: ***Cloud forensics, machine learning, cloud forensic methodologies, review, machine learning methodologies, cloud forensic solutions.***

I. INTRODUCTION

Cloud computing is considered to be thenextgenerationof information technology framework.

Cloudcomputing gives a extensive repository ofreliableITcomponents that can be used on shared basissuchascomputingresources,networks,softwareanddatabases.Customers are able to use virtualized resources inas callable and elastic fashion through cloudcomputingenvironment.Cloud computing is said to be defined as the group technologies and a means for providing the use of large scale Internet services for the remote applications with high quality of service[1]. Major benefits of cloud computing includes scalabiliy, elasticity, self serviceprovisioning,andperusepayetc.Cloudcomputingservice scanbepublic,community, hybrid and private.Service modelsofcloudcomputing offer application platform to be aservice(PaaS),infrastructuretobeaservice(IaaS)andsoftwareto beaservice (SaaS).

Revised Manuscript Received on December 22, 2018.

NanditaGoyal, ABES Engineering College, Ghaziabad

Kanika Gupta, ABES Engineering College, Ghaziabad.

Munesh Chandra,Chandra Trivedi Rajkiya Engineering College, Azamgarh.

Computing includes various security issues. The main reason for these issues is the incorporation of different technologies like operating systems, resource scheduling, networks, virtualization etc. This results in applicability of the threats and challenges in these technologies to cloud computing also[3]. The network which interrelates the systems in a cloud has to be confidential. Furthermore, virtualization paradigm shift in cloud computing results in several security issues, like, mapping the virtual machines to the physical machines has to be carried out timely. Data security involves encryption of the data as well as ensuring that appropriate rules are enforced for data sharing. In addition, memory management and resource allocation algorithms have to be safe.

As a number of service providers and users are getting added to the cloud environment, it has become a important field for performing malicious activities for financial gain by attackers[2]. It is a significant need to securely store, analyze and share complex (e.g., semi-structured and unstructured) data which is large in volume. Because of the critical nature of the applications, it is important that clouds be secure. The major security challenge with clouds is that the owner of the data may not have control of where the data is placed. This is because if one wants to exploit the benefits of using cloud computing, one must also utilize the resource allocation and scheduling provided by clouds [5]. Therefore, we need to safeguard the data in the midst of entrusted processes.

Several frameworks of security are being considered in order to solve the danger and threats to provide security for the resources so that the potential of cloud computing can be considered to the maximum level. In spite of various security measures if an attack is encountered then the role of cloud forensics comes into existence. Cloud forensics is a strategy that provides a platform to investigate and analyze cloud security threats. Cloud forensics is the deployment of digital forensics in cloud computing environment as a subset of cyber and network forensics. Cloud forensics involves investigation after the attack in order to know the source and detect malicious cloud criminals associated with the attack so that they could be punished in a justified manner [4] [6].

Basically, it is a inter-discipline between cloud computing and digital forensics. The major problem lies in analysis of each and every local host to detect malicious actors. Here comes the role of machine learning in analyzing complex data in

limitedtime.

The main objective of this paper is to present a detailed literature review on the existing frameworks and methodologies of cloud forensics environment. Itcritically reviews cloud forensics' existing frameworks that make use of machine learning approaches and it explores based on a detailed review of the area, major work that has been carried out in the area of cloud forensics using machine learning.

The paper is organized as follows: In Section II, related work on current frameworks and methodologies in cloud forensics environment. Section III presents a review of existing presented methodologies along with various challenges. Section IV, discusses the proposed system for identification of malicious hosts in cloud environment using machine learning approach. Finally, SectionV concludes the paper and raises future researchdirection.

II. RELATEDWORK

Cloud Forensics

Cloud computing offers benefits to companies and organizations as it leads to improve profit margins in business while controlling the costs incurred. With the use of cloud computing, companies are able to outsource required services while paying more attention to their core development areas. The Cloud service Providers (CSPs) are responsible for providing services and maintaining infrastructure after signing contracts with these companies. The various advantages of cloud computing include on demand self-service, elasticity, resource pooling, efficiency etc.

There are three service models available in cloud computing namely: PaaS (Platform as a Service), IaaS (Infrastructure as a Service) and SaaS (Software as a Service).

PaaS: As per the NIST's definition of cloud computing defines Platform to be a Service as:[15]
It provides ability to the customer to create applications using different tools and programming languages and make use of libraries without controlling the deployed infrastructure.

IaaS: The NIST's definition of cloud computing [16] describes IaaS as capability to customers to make use of any software or infrastructure even an operating system. While using the deployed infrastructure the customer is free from the responsibility of managing or controlling the cloud infrastructure.

SaaS: The NIST's definition of cloud computing defines Software as a Service [15] as:It provides ability to run applications on cloud infrastructure. The customers can access these applications through their devices without any requirement to manage involved infrastructure including operating systems, storage network, servers, or even individual application capabilities, with the possible exception of limited user-specific application configurable settings.Cloud computing can be deployed by taking one of the four installation models: Private, community, public and Hybrid.
Public clouds are operated and taken by third-party a cloud service provider, who gives their computing resources like servers and storage over the net community. Microsoft Azure is an example of a public cloud. With a public cloud, all software computing resources and other supporting infrastructure is owned and managed by the cloud subscriber. Users can take the access of these services and manage the account using a web browser[7].

In a private cloud computing environment and framework, resources are used only for a one business or firm. The physical site of a private cloud can be company's live datacenter. Some companies also deploy intermediate providers to install their own cloud on network. In a privately owned cloud, the infrastructure and services are settled on a private network[8].

Hybrid clouds are mixed combination of privately and publicly owned clouds. They allow the users to take the advantage of both the categories while allowing sharing of data and applications between them[9]. They generally provide greater flexibility with a number of options for deployment in business. It also gives optimization to the present infrastructure along with security.

A community cloud is a solution to the needs of organizations and individuals which are confined in number These are managed and governed by an intermediary organization or service provider. [10].
Intoday'sscenarioasthenumberofCSPsandusersisgrowing,then umberofcloudcrimesisalsoincreasingonthe same pace. Here comes the role of cloud forensics.

Cloud forensics is the applicationofforensicscience in cloud environment. It is the postinvestigationthatis carried out after a security attackhas taken place incloudenvironment.Itisasubdisciplineofdigitalforensicsbuthe re the investigation is more complex duetovirtualization.There are many more issues like volume ofdata,multi-jurisdiction and lack of forensic tools etc. that add ontothiscomplexity. The major phases of cloudforensicsinclude: evidence identification, collection, preservation, examination,interpretation and reporting [11].

Various digital and cloud forensics methodologies have been developed and studied for fulfillment of above objectives but rapid advances in cloud computing require development of more sophisticated methods and tools for carrying out forensic analysis in cloud environment..

Existing Methodologies and Frameworks for Cloud Forensics

In this section, we present a comprehensive review on latest work in cloud forensics methodologies that are making use of machine learning approaches. The goal of comparison is to address the challenges of existing frameworks by assigning them in proposedframework.

Shravana Kumar Chinnikatti [13] in his review discussed the use of Artificial Intelligence in supporting digital forensics. AI algorithms can be used provide fast solutions to support comprehensive and written basedcommunicationwith evidences proved with statistics and analysis without judicial errors. With AI it is also possible to do meta- analysis

from different data sources. In this paper, author also emphasizes that artificial intelligence has a well-established importance in forensic science.

Prerak Bhatt [14] et al. in his paper discusses the role of machine learning in digital crime analysis. The author has also presented an elaborate discussion on steps for setting up an environment to train artificial neural networks and use the ANNs to investigate and analyze artifacts for forensic investigation.

Suchana Dutta [17] et al. has proposed a model Cloud Malicious Actor Identifier which focuses on legitimate demand of cloud forensic investigators. This model ranks the malicious actor in a particular crime scene on basis of probability of being malicious which is computed using a machine learning technique called boosting. This model attempts to reduce the overhead of probing each and every IP address during investigation. This model also optimize the investigation process, cost andtime.

J. K. Alhassan [18] et al. proposed a method called Fuzzy Classifier-based Vulnerability and Assessment Testing (FCVAPT) to protect confidential information available on web applications. In this paper, the researchers used XSS and SQL to evaluate their technique. They estimate the classification performance of their technique in terms of MSE, MAPE, and RMSE and they achieved about 33%, 14%, and 5% respectively. The authors of this paper claimed this technique (FCVAPT) to be best for identifying the threads on webapplications.

DanialJavaheri [19] et al. suggested a method to escape from malware threads, by taking into consideration the behavior of malware. The technique proposed in this paper identifies the incident period of Trojan at user level and kernel level of OS, by clearing the storage space of effected malware on right time and particular hook installing. The authors aim to provide an effective technique for the identification of malware behavior by using the metamorphic engine, packer and protector tools. These tools make decisions on obfuscation and metamorphosis situation caused bymalwares.

KamalakantaSethi [20] et al designed a structure for the identification and classification of various malicious files with the extension like exe, php, pdf, etc. In this article, the researchers used a two-level classifier called Macro and Micro. The Macro helps to identify the malware and Micro helps to classify the malwares like spyware, Trojan, adware, and so on. This method makes a static as well as dynamic enquiry and creates report by executing the data present in the virtual environment with the help of Cuckoo Sandbox technique. They also include a module for the extraction of feature, which works based on the report created by the Cuckoo Sandbox. Here, they used Weka framework for the development of machine learning models by including the dataset used for the training process.

Imansharafaldin [21] et al. proposed a system calledBotViz for the classification of malware with the support of data collected on analysis of forensics and the domain generation algorithm detector.

Their system works based on the techniques of machine learning and the performance of their system is calculated by using a live Zeusbotnet.S Saibharath [22] et al. has proposed the implementation of a web software tool for cloud forensics. It helps in data collection and rendering mechanism for cloud environment. This mechanism is implemented through Hadoop using struts 2.0 MVC framework. It makes use of clustering for preprocessing of evidence files. Cross drive analysis is also performed between drives using correlationfunction.

III. FRAMEWORK COMPARISON

In Table 1, a comparison of different frameworks for digital and cloud forensics making use of machine learning approaches is discussed. From the comparative analysis, illustrated in Table 1, most of the frameworks and methodologies used for cloud forensics make use of different machine learning approaches for forensics analysis purpose. The table also shows the challenges and solutions provided by various frameworkdiscussed.

Table. 1 Digital and cloud forensics methodologies

Name of Author	Objective	Technique used	Solutions	Challenges
SuchanaDatta, PalashSantra, Koushikand Debashish De(2018)	To mitigate the overhead of probing each and every IP address while forensic investigati on.	Boosting	Proposed a model 'Cloud MaliciousActor Identifier' to rank malicious actors for a particular crime scene on the basis of probabilityof beingmalicious usingboosting	To increase the accuracy of proposed model using other classificatio n methods
J.K. Alhassan, Sanjay Misra, A. Umar, RytisMaskeliūnas , RobertasDamaševičius, and AdewoleAdewumi.(2 018)	To provide security to sensitive data/information in web applicati on s.	Fuzzy Classification	Proposed a Fuzzy Classifier-based Vulnerability and Assessment Testing (FCVAPT) model. It provides penetration testing for web applications.	This model detects vulnerability for web applications and penetration levels for recognized cases. So, it could be further extended for other types of applications.
DanialJavaheri and Mehdi Hosseinzadeh(2018)	Establishment of an efficient platform for detection and analysis of novel	Memory dumping and hook installing	Proposed an efficient platform for detection and analysis ofnovel malwares. It identifies malware behavior by using	To improve success rate for identificatio n of kernel level malwares.

			and protector tools.	
Kamalakantha Sethi, Shankar Kumar Chaudhary, Bata Krishan Tripathy, and Padmalochan Bera (2018)	To provide a framework for detection of malware using machine learning approach.	Weka framework and Cuckoo sandbox	Proposed an intelligent framework for identification and classification of malware. It makes use of micro and macro models developed using Weka for identification of malwares and Cuckoo sandbox for analysis.	To increase the number datasets to improve accuracy of results as only 220 datasets were used in this framework.
S. Saibharath and Geethakumari G. (2015)	To provide a web software tool for cloud forensics.	Clustering and cross drive analysis using correlation function.	Implemented a framework for data collection and rendering in cloud environment that makes use of clustering for pre-processing evidence files.	To improve accuracy and time for analysis.

IV. PROPOSED METHOLOGY

In this section we present our proposed method for classifying and detecting malicious hosts in cloud environment. The main challenge is that we try to analyze the identification of host which is providing malicious data to the users in the cloud. So, we present a framework for identifying these malicious hosts by first extracting features from the host list by applying a feature extraction algorithm. The extracted features help to classify the different malicious host. Then finally, we will classify these hosts as malicious or non-malicious by using Artificial Neural Networks(ANNs).

Data collection phase: In this phase, the data that are needed for the process will be collected. Analysing and collecting data in the cloud is one of the challenging tasks. These data include both normal and malicious host list. The collected data will be used for extracting the features.

Analysis Phase: In this phase, the extracted features of the host list will be used training the machine and then testing. In this phase will Artificial Neural Networks for the purpose of classification of a host as malicious or non- malicious.

V. CONCLUSION

Although cloud computing have many advantages but there are lot of security issues accompanied with it due to incorporation of various technologies [12]. In this paper, a detailed present view of different cloud forensic frameworks has been presented. We have presented various methodologies for dealing with cloud forensics. An extended discussion is also given regarding the solutions and challenges provided by the variousdiscussedMethodologies. A greater emphasizes is laid on frameworks making use of different machine learning approaches for forensic investigation. This review provides a direction to researchers by identifying the various challenges in the cloud forensic analysis and presenting efforts which have already been conducted in this field. Our main goal is bring forward a new solution framework for identifying malicious host cloud environment using approach of machine learning, in order to improve precision and minimize the amount of time required for analysispurpose.

REFERENCES

1. Christos Sgaras, M-TaharKechadi, and Nhien-An Le-Khac. "Forensics acquisition and analysis of instant messaging and VoIP applications." In Computational forensics, pp. 188-199. Springer, Cham,2015.
2. Suleman Khan, Abdullah Gani, Ainuddin Wahid Abdul Wahab, and Mustapha AminuBagiwa. "SIDNFF: Source identification network forensics framework for cloud computing." In Consumer Electronics- Taiwan (ICCE-TW), 2015 IEEE International Conference on, pp. 418-419. IEEE,2015.
3. Li, Jin, Xiaofeng Chen, Qiong Huang, and Duncan S. Wong. "Digital provenance: Enabling secure data forensics in cloud computing." Future Generation Computer Systems 37 (2014): pp.259-266.
4. Alluri, BKSP Kumar Raju, and G. Geethakumari. "A digital forensic model for introspection of virtual machines in cloud computing." In SignalProcessing,Informatics,CommunicationandEnergySystems (SPICES), 2015 IEEE International Conference on, pp. 1-5. IEEE, 2015.
5. Zawoad, Shams, RagibHasan, and Anthony Skjellum. "OCF: an open cloud forensics model for reliable digital forensics." In Cloud Computing (CLOUD), 2015 IEEE 8th International Conference on, pp. 437-444. IEEE,2015
6. Rani, DeeviRadha, and G. Geethakumari. "An efficient approach to forensic investigation in cloud using VM snapshots." In Pervasive Computing (ICPC), 2015 International Conference on, pp. 1-5. IEEE, 2015.
7. Perumal, Sundresan, NoritaMdNorwawi, and Valliappan Raman. "Internet of Things (IoT) digital forensic investigation model: Top- down forensic approach methodology." In Digital Information Processing and Communications (ICDIPC), 2015 Fifth International Conference on, pp. 19-23. IEEE,2015.
8. Samy, GanthanNarayana, BharanidharanShanmugam, NurazeanMaarop, PritheegaMagalingam, SundresanPerumal, and Sameer HasanAlbakri. "Digital Forensic Challenges in the Cloud Computing Environment." In International Conference of Reliable Information andCommunicationTechnology,pp.669-676.Springer,Cham,2017.
9. Meera, G., BKSP Kumar RajuAlluri, DigambarPowar, and G. Geethakumari. "A strategy for enabling forensic investigation in cloud IaaS." In Electrical, Computer and Communication Technologies (ICECCT), 2015 IEEE International Conference on, pp. 1-5. IEEE,2015.
10. Alqahtany, Saad, Nathan Clarke, Steven Furnell, and Christoph Reich. "A forensic acquisition based upon a cluster analysis of non- volatile memory in IaaS." In Anti-Cyber Crimes (ICACC), 2017 2nd International Conference on, pp. 123-128. IEEE,2017.
11. Kebande, Victor R., and Indrakshi Ray. "A generic digital forensic investigation framework for internet of things (iot)." In Future Internet of Things and Cloud (FiCloud), 2016 IEEE 4th International Conference on, pp. 356-362. IEEE,2016.
12. Spiekermann, Daniel, Tobias Eggendorfer, and Jörg Keller. "Using network data to improve digital investigation in cloud computing environments." In High Performance Computing & Simulation (HPCS), 2015 International Conference on, pp.98-105. IEEE, 2015.
13. ShravanaKC.ArtificialIntelligenceinForensicScience.ForensicSci Add Res. 2(5).FSAR.000554.2018

14. P. Bhatt and P.H. Rughani. “Machine learning forensics: a new branch of digital forensics”. In International Journal of Advanced Research in Computer Science, ISSN No. 0976-5697
15. N.C.C. F. S. W. Group, NIST cloud computing forensic science challenges. Draft NISTIR 8006,2014.
16. MellP,GranceT.TheNISTdefinitionofcloudcomputing,NIST special publication . Gaithersburg2011.
17. Datta,Suchana,PalashSantra,KoushikMajumder,andDebashisDe. "An Automated Malicious Host Recognition Model in Cloud Forensics." In Networking Communication and Data Knowledge Engineering, pp. 61-71. Springer, Singapore,2018.
18. Alhassan, J. K., Sanjay Misra, A. Umar, RytisMaskeliūnas, RobertasDamaševičius, and AdewoleAdewumi. "A Fuzzy Classifier-Based Penetration Testing for Web Applications." In International Conference on Information Theoretic Security, pp. 95-104. Springer, Cham,2018.
19. Javaheri, Danial, and Mehdi Hosseinzadeh. "A Framework for RecognitionandConfrontingofObfuscatedMalwaresBasedon Memory Dumping and Filter Drivers." Wireless Personal Communications 98, no. 1 (2018):119-137.
20. Sethi, Kamalakanta, Shankar Kumar Chaudhary, Bata KrishanTripathy, and PadmalochanBera. "A Novel Malware Analysis Framework for Malware Detection and Classification using Machine Learning Approach." In Proceedings of the 19th International Conference on Distributed Computing and Networking, p. 49. ACM, 2018
21. Sharafaldin, Iman, AmirhosseinGharib, ArashHabibiLashkari, and Ali A. Ghorbani. "BotViz: A memory forensic-based botnet detection and visualization approach." In Security Technology (ICCST), 2017 International Carnahan Conference on, pp. 1-8. IEEE,2017.
22. Saibharath S, Geethakumari G “Cloud Forensics: Evidence Collection and Preliminary Analysis” IEEE,2015