DIGITAL FORENSIC TECHNIQUE BASED ON EXCEPTION LOGS FOR CLOUD

[1]SULABHA PATIL, [2]RAJEEV DHARASKAR, [3]VILAS THAKARE

[1]Prof., Tulsiramji Gaikwad-Patil College of Engg. & Tech., Nagpur, [2]Director, MGI, Nanded, [3]HOD, Department of Computer Science & Engineering, Amravati University
E-mail: [1]sulabhavpatil@gmail.com, [2]rajeev.dharaskar@gmail.com, [3]vilthakare@yahoo.co.in

Abstract - Cloud computing has introduced convenient ways of data storage and data sharing without having to invest in costly hardware and software. Ubiquitous computing made it even more simpler as users can use handheld devices such as mobile phones and smart phones to upload, download, share text, photos, videos easily on cloud application. But hackers and criminals committing digital crimes are equally benefitted as no standard guidelines, procedures, methods are in existence to tackle with the crimes committed in the cloud environment. Digital Forensic process applicable on static devices cannot be applied directly to cloud environment. Hence, a new approach is proposed in this paper. We are analyzing Exception Logs generated by the application uploaded in the cloud environment. Analysis of these logs will lead to the determination of malicious activities, malicious users, state of the machine at the time of the incident and many other factors.

Keywords - Exception Logs, Twitter, Virtual Machine Monitor (VMM), Forensic Integration Architecture (FIA).

I. INTRODUCTION

According to the Scientific Working Group on Digital Evidence (SWGDE), Digital Evidence is "information of probative value that is stored or transmitted in binary form"[1].

Though process of Digital Forensic is standardized to some extent but these processes are applicable to digital devices which can be seized and detained for forensics. The scenario of cloud is entirely different. Here no physical device is available for carrying out forensics. Moreover the data is scattered all over the world which makes it evitable to incorporate the laws and rules of all these nations while making policy of forensic investigation. Virtualization has further worsen the scenario. In order to tackle this issue forensic readiness approach is one of the approaches suggested by many researchers.

A. Digital Forensics

Digital forensics is a discipline which combines elements of law and computer science to collect and analyze data from variety of digital systems like computer systems, networks, wireless communications and storage devices in a way that is admissible as evidence in a court of law[3]. Emergence of cloud environment has added more challenges to the digital forensic techniques which was initially concentrating only on two categories namely stored, static or fixed data forensics and the changing or live data forensics.

B. Cloud Computing and its Security

As per NIST "Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable resources (e.g. networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [7].

Security and privacy issues have not remained confined to single device or user. In fact cloud environment has widened its scope to infinity. Online Economic Transactions are accepted world vide due to flexibility, less time requirement and ease of use, however, users confidence can be enhanced by adopting stringent security measures. In this paper we have proposed a technique for digital forensics in cloud environment. In Section II, we have carried out literature review of Digital frameworks for cloud environment. In Section III and IV, we have discussed sources of relevant data for cloud forensics and the established standard for the same respectively. In Section V we have proposed a technique for cloud forensic. Implementation of the proposed technique and the results thereof are discussed in Section VI.

II. LITERATURE REVIEW

Computer Science and Law are the two elements which constitutes Digital forensics The first one is utilized to collect and analyze data from the wired and wireless devices such as computer systems, networks and storage systems such as hard disks, memory in such a way that it is acceptable and admissible as a evidence in the court of law. According to [8], [20] a digital forensic process can be broken into four distinct phases:

1. Collection of artifacts (both digital evidence and supporting material) that are considered of potential value are collected.
2. Preservation of original artifacts in a way that is reliable, complete, accurate, and verifiable.
3. Filtering analysis of artifacts for the removal or inclusion of items that are considered of value.

4. Presentation phase in which evidence is presented to support investigation.

Traditionally, static forensics and live forensics are the two methods in which digital forensics is carried out. In [8] authors argue that these methods are result of evolution of forensic science for recreation and documentation of incidences. The process of forensics analysis of data stored in storage devices such as hard drives refers to static forensic. Traditional data acquisition techniques are adopted for getting the data stored in these devices. The process of analyzing the data when the system is in running state is termed as Live Forensic.

Three approaches of digital investigation are discussed by authors in [9]. The first one is the Digital Forensic Research Workshop [DFRWS] held in 2001 to provide a forum for a newly formed community of academics and practitioners for knowledge sharing on digital forensic science. Committee proposed a seven step process for digital forensics which includes Identification, Presentation, Collection, Examination, Analysis, Presentation and Decision.

Forensic Process Model proposed by U.S. National Institute of Justice in 2001 is a model for Electronic Crime Scene Investigation. They proposed a four step model i.e., Collection, Examination, Analysis and Reporting.

Abstract Digital Forensic Model is the concept proposed by authors at [9] which is enhancement of DFRWS. This model suggests nine steps beginning with Identification, Preparation, Approach, Strategy, Preservation, Collection, Examination Analysis, Presentation and Returning Evidence.

All these approaches presume that all the artifacts required for forensics be it static or dynamic i.e., all log files, network data etc is available for carrying out the process and hence these strategies can work on the available artifacts. But in the cloud computing scenarios, availability of these artifacts is the real crux. Considering digital forensic procedures and principles authors propose a Forensic Integration Architecture (FIA) for composing digital evidence in [10]. It consists of four layers (i) Evidence Storage and Access Layer (ii) Representation and Interpretation Layer (iii) Meta Information Layer (iv) Evidence Composition and visualization layer.

As discussed previously, here also the process of acquisition of data from memory devices and network traffic is not done instead it is assumed that forensically imaged copies are stored in persistent media.

Digital Forensic Readiness is defined as the ability of an organization to maximize its potential to use digital evidence while minimizing cost of investigation[11]. There is no digital forensic readiness framework for PKI system hence a new system is being proposed by authors in [11]. Proposed framework aims at [a] maximization of potential use of digital evidence, [b] minimization of investigation cost [c] minimization or prevention of interruption of PKI system business process [d] perseverance or improvement of current level of information system security of PKI system. Ten phases are proposed in this system.

Digital Forensic Readiness allows computer architectures to collect sensitive and critical information related to digital crimes before they happen, leading to save time and money during the investigations [12]. Cloud computing poses different challenges as compared to traditional setup. For example in case of computer forensics or mobile phone forensics the device is seized and is available for forensic but the scenario of cloud is entirely different as the infrastructure on which the data is stored is not known to the client. Moreover the data is scattered around the globe which gives rise to issues of time stamping, virtualization in cloud further complicates the issues. Thus main challenges in cloud forensics are [a] Reduced data access, Lack of physical control and Lack of standard because of elasticity, multiple locations and virtualization [b] Multiple log formats due to elasticity and virtualization. [c] No timestamps synchronization, No routing information, Legal measures, multiple jurisdiction as data is stored in multiple locations. An approach for dealing with these issues is the implementation of Digital Forensic Readiness (DFR) into the Cloud.[12].

A Cloud Forensic Reference Architecture has been discussed by author which takes care of all the issues discussed above in order to get a digital forensics readiness system for Cloud.For Cloud Forensics Readiness a model called Botnet As a Services [BaaS] is proposed. It is proposed that botnet can be implemented as a Service in Cloud environment. Botnet will be performing the similar task of infecting the instances of virtual computers in any cloud environment and harvesting the digital information but here it will be used in a non-malicious way. The harvested information should be preserved for the cloud forensic readiness purpose [13]. BaaS is embedded in SaaS module of the Cloud environment and is consisting of two processes namely Proactive Process where the virtual computers are being infected non-maliciously by the botnet and the Reactive process consisting of traditional Digital Forensic Investigation steps. In between these two processes exists the step of Harvesting of Digital Information as shown in Fig.1 [13].

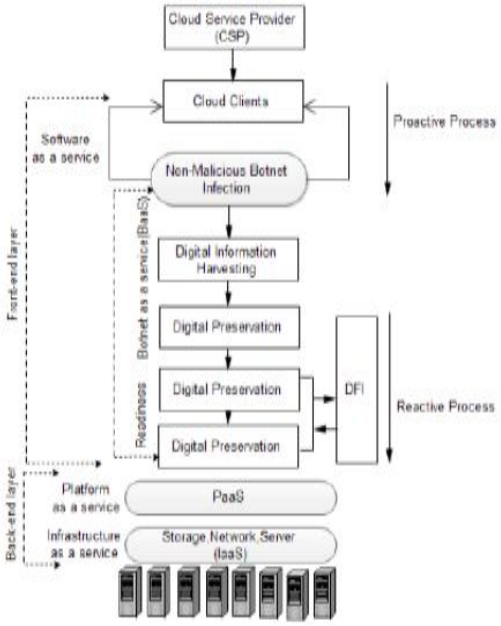

Fig.1. Cloud Forensic Readiness Model with Baas [13].

Although the concept proposed by author seems worthwhile from forensic point of view but in situations where there is no need of forensics, in that case such a system will cause violation of confidentiality, privacy. Though author proposes that the model will be implemented taking into account the legal acts and the data collected will be strictly for law enforcement purposes but implementing it will be difficult as the norms and laws pertaining to network security are very stringent. More over law differs from nation to nation, hence implementing such system globally would be intricate.

III. SOURCES OF RELEVANT DATA FOR CLOUD FORENSICS

Three Layers of Cloud Computing Architecture namely IaaS, Paas and Saas consists of forensically important data. Iaas contains snapshots and the system memory while System states and applications logs can be retrieved from PaaS. Saas stores Virtual Machines images and the Single SignOn logs. Cloud Consumers and Providers are connected via Cloud Carrier. Therefore, data such as network logs, activity logs, access record facility logs, virtual images and hypervisor events log are available with these carriers.

IV. ISO/IEC 27043

ISO/IEC 27043:2015 provides guidelines for common incident investigation processes applicable across various incident investigation scenarios involving digital evidence. Processes from pre-incident preparation through closure of investigation are covered under it. Processes and principles applicable to various kinds of investigations are included in the guidelines. It also includes guidelines to tackle unauthorized access, data corruption, system crashes, or corporate breaches of information security, as well as any other digital investigation [14].

V PROPOSED METHODOLOGY

In this paper we present exception log based technique for cloud forensics which will analyze exception logs generated by an application. Exception logs are generated when certain norms and rules of specific application are violated. Repeated violation of norms leads to abnormal behavior on the part of the user or machine. This behavior gives a clue about the malicious intention.To conduct our study we have utilized the dataset provided by cloud based application Twitter. the application designed by us is a twitter based clone web 2.2 application system. Twitter is a social networking site having millions of users sharing data, images etc. A special term Tweet is used in twitter. Tweet is a general expression with 140 characters limits. Registered users can read and post tweets, but unregistered users can only read them. It is of two types public and private. Public can be viewed by everybody whereas private by followers and users home page. As of May 2015, Twitter has more than 500 million users, out of which more than302 million are active users [15].

If we consider tweets as an entity then there will be some characteristics for identity. These characteristics are Length of the tweet. As said earlier the length of the expression should not exceed 140 character limits. The second characteristic is its Contents. There are certain restricted words like “bomb”, “kill” which are banned from its usage in our application [twitter clone]. The third characteristic is multiple occurrences of single tweet or # tags. This characteristic is included in order to protect the application from malicious spammers or spam-bots used in Distributed Denial of Service types of attacks. Hackers resort to Distributed Denial of Service attacks in order to disrupt the network traffic of a particular website. These types of attacks are carried out by them by employing bot nets which sends messages repeatedly so that the network is overloaded and thereby the functioning of the site is disrupted.We have considered these three characteristics of twitter to generate our exception logs. This application is executed in two parts. In the first part of the application dataset of about 3,50,000 lacs tweets is utilized for analysis. These tweets are received one by one and checked for violation of norms of length, contents and number of occurrences. During each check if it is observed that a particular tweet is violating norms then exception is generated accordingly and it is added to respective exception log. Model for generation of these logs is shown in Fig. 2. Snapshots of logs generated during our

experiment are in the form of graphs. They are shown in Fig 5.

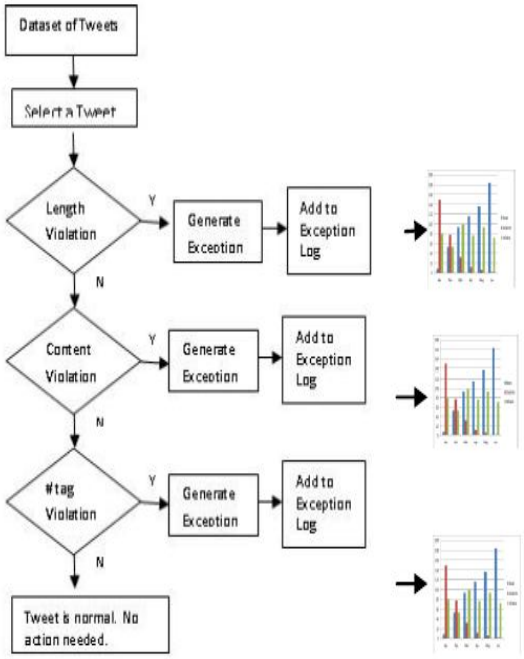

Figure 2. Exception Log Generation Model.

For carrying out digital forensics or cloud forensic investigators needs crucial information about the state of the machine, date and time log, information about applications running at the time of incident, users logged in etc. Cloud Service providers provide abstract information which is not sufficient for carrying out forensics. Eg. In situation when machine crashes cloud service providers does not provide the information about the cause of the crash. Whereas investigators are interested in finding out the reason hence additional information is needed to monitor the state of machine.In the second part of our application we are dealing with this issue. In this part a Master Server / Management server is created which receives exception logs generated by the application hosted in a separate cloud vendor environment. We will be monitoring multiple instances of this application at this Master server and will not be logging into application server every time. The machine state and the application error log is maintained by the application which pulls the machine specific data such as network bandwidth, RAM, Computer etc maintained at Platform As a Service [PasS] layer and also the application error logs [SaaS layer] as shown in fig.3. In this way Application Logs and State of Machine both are received by us. This data can be collected on demand or at regular interval. The data is also stored on remote cloud server for analysis. If at all situation arises when the server or application becomes unavailable, in that case event logs can be analyzed to obtain a greater insight on the chain of events leading to the failure of web application/server. This remote server is hosted in cloud

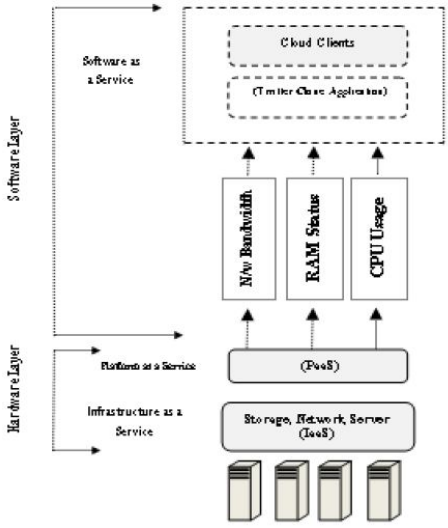

Figure 3. Model for recording machine state

VI. IMPLEMENTATION OF PROPOSED SYSTEM

Environmental set up is done by creating a cluster having two machines. Virtualization environment is used for managing hardware resources and providing cloud like environment. This is done using Type-1 Hypervisor / bare metal hypervisor to provide cloud like infrastructure and management console. A Type-1: native or bare-metal hypervisors run directly on the host's hardware to control the hardware and to manage guest operating systems. For this reason, they are sometimes called bare metal hypervisors. A guest operating system runs as a process on the host [16]. With the help of this environment multiple instances can be installed simulating different physical servers in cloud environment as shown in fig.4

Figure 4. Multiple Virtual Machines

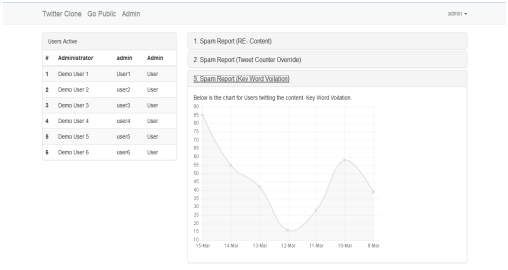

Figure 5. Generation of Exception Logs

CONCLUSION

Market trends of last five years reveal that organizations and people all over the world are shifting their work on the cloud because of its inherent characteristics. Cloud environment provides the flexibility of working without bothering about the Infrastructure, software or applications. It has made the working very simple as it is based on Pay Per Usage model. But the other side of the coin is cloud based crimes have also increased as no proper guidelines are available to tackle such crimes, Data in a cloud is scattered around the globe hence law of different nations needs to be considered while dealing with such crimes. Chain of Custody and Time standards are the issues which makes this problem more complicated.

Digital forensics is not in its infancy as far as forensics of digital devices is concerned such as computer, hard disk, cache, mobile phone forensics. These devices are physically available for carrying out forensics of the data available on it. But it is not true with the cloud. As the data in the cloud is scattered and the Cloud Service Provider himself is not aware many times as to which part of the data is stored in which machine. Virtualization has worsen the condition even more. Availability of physical artifacts for forensics is extremely difficult. Hence one has to rely on the network logs, transactions carried out by the cloud users etc for carrying out the forensics.

One of the approaches can be Forensic Readiness. This concept is dealt in depth in this paper. Forensic readiness helps to a considerable extent as the data required for forensics is available and the time required is minimized to a considerable extent.

The concept of digital forensics in cloud environment proposed by us in this paper is giving good results. Considering the popularity of social networking sites database of twitter was utilized for this study.

REFERENCES

[1] Robert Hegarty, Madjid Merabti, Qi Shi and Bob Askwith, "Forensic Analysis of Distributed Data in a Service Oriented Computing Platform" ISBN: 978-1-902560-22-9 ©, PGNet, 2009.

[2] Conan C. Albrecht, "Fraud and Forensic Accounting In a Digital Environment", White Paper for the Institute of Fraud Prevention, 2010.

[3] Robert Hegarty, Madjid Merabti, Qi Shi and Bob Askwith, "Pre-emptive File Signature Creation for Digital forenis, PGNet, " ISBN: 978-1-902560-24-3 ©, 2010.

[4] Keyun Ruan, Ibrahim Baggili , "Survey on cloud forensics and critical criteria for cloud forensic capability : A preliminary analysis", Journal of Network Forensics, vol. 3, issue 1, 2011.

[5] RongxingLu ,Xiaodong Lin and Xuemin (Sherman) Shen ,"Secure Provenance : The Essential of Bread and Butter of Data Forensics in Cloud Computing ",ACM,978-1-60558-936-7, 2010.

[6] George SIBIYA, Hein S. VENTER, "Digital Forensic Framework for a Cloud Environment ", IST Africa Conference, 978-1-905824-34-2, 2012.

[7] Xuefeng Liu, Yuqing Zhang, Boyang Wang, and Jingbo Yan, "Mona: Secure Multi-Owner Data Sharing for Dynamic Groups in the Cloud", 1182 IEEE Transactions on Parallel And Distributed Systems, Vol.24, No. 6., June, 2013.

[8] Inikpi O. Ademu, Dr Chris O. Imafidon, Dr David S. Preston, " A New Approach of Digital Forensic Model for Digital Forensic Investigation", International Journal of Advanced Computer Science and Applications, Vol. 2, No.12, 2011.

[9] Sriram Raghavan, Andrew Clark, George Mohay, FIA: "An Open Forensic Integration Architecture for Composing Digital Evidence", In: Sorell, Matthew, (ed) Forensics in Telecommunications, Information and Multimedia.Lecture Notes of the Institute for Computer Science, Social Sciences, Social Informatics and Telecommunications Engineering, 8. Springer Berlin Heidelberg, pp. 83-94.

[10] Aleksandar Valjarevic, HS Venter, "Towards a Digital Forensic Readiness Framework for Public Key Infrastructure Systems".

[11] Lucia De Marco, Filomena Ferrucci1, M-Tahar Kechadi, Reference Architecture for a Cloud Forensic Readiness System, EAI Endorsed Transactions on Security and Safety.(http://creativecommons.org /licenses/by/3.0/).

[12] Victor .R.Kebande, Hein.S.Venter, "A Cloud Forensic Readiness Model Using a Botnet as a Service", ISBN: 978-0-9891305-7-8 ©2014 SDIWC, pp:23-32.

[13] ISO/IEC 27043:2015, http://www.iso.org/iso / catalogue_detail.htm ?csnumber=44407.

[14] "Twitter MAU Were 302M For Q1, Up 18% YoY - Twitter (NYSE:TWTR) | Benzinga]". April 28, 2015. Retrieved May 2, 2015. http://en.wikipedia.org /wiki/Twitter.

[15] en.wikipedia.org/wiki/Hypervisor.

[16] Deoyani Shirkhedkar, Sulabha Patil, "Design of digital forensic technique for cloud computing", International Journal of Advance Research in Computer Science and Management Studies Volume 2, Issue 6, June 2014 pg. 192-194.

[17] Vaquero, L.M., Rodero-Merino, L., Mor_an, D.: Locking the sky: a survey on IaaS cloud security. Computing 91(1),93{118 (2011). DOI 10.1007 /s00607- 010-0140-x.

[18] Parag Shende, Sulabha Patil, "A Survey On Security Issues In Cloud Computing Environment", Proceedings of IRF International Conference, Bangalore 23rd March-2014, ISBN: 978-93-82702-68-9.

[19] Parag Shende, Sulabha Patil, "Enhancing Privacy in Intercloud Interaction", Proceedings of 7th IRF International Conference, 27th April-2014, Pune, India, ISBN: 978-93-84209-09-4

[20] R. Ahmed, R.V. Dharaskar, "Mobile forensics: an overview, tools, future trends and challenges from law enforcement perspective", 6th International Conference on E-Governance, ICEG, Emerging Technologies in E-Government, M-Government, 2008/12, pp. 312—23.

★★★