Challenges of Forensic Investigation in Cloud Computing

Singh Jyoti

Model College

Abstract:- Cloud computing is a collection of services which enable resource sharing of different entities such as processing units, storage devices and software which are connected via the internet. It is the most discussed information technology sector of particular use and interest for business owners, cyber thefts and forensic investigators. The information technology used in cloud computing derives live resources of stored data such as from different data sources connected online as per individual requirements to arrive at the possible conclusions computed live from the cloud computing environment. The services provided may be economical as well expandable. But many customers are reluctant to move their business infrastructure to cloud because of concerns regarding cloud security and threat of the unknown. The cloud service providers increase such a twisted perception by not letting customers see behind the virtual curtain. Now in normal computing used by organizations, it is required to procure, deploy and manage physical IT infrastructure to host the software application in limited environment of the individual organization. But with cloud computing, firms deploy their IT infrastructure in remote, virtualized environments often hosted and managed by third parties. Also the decentralized nature of data processing in cloud computing is very different from traditional approach to evidence collection and recovery are no longer applicable. Digital forensics steps into play to find evidences against criminals and needs to assume careful control and management of IT assets like data storage while conducting the investigation process. Now with regards in particular to forensic investigations in the digital domain with the help of cloud computing, due to usage of new technology and methods, the main issues are difficulties to deal with the different rulings obliged on a variety of data saved in different locations, limited access to obtain evidences from cloud and seizing the physical evidence for the sake of integrity validation of evidence presentation. Cloud Forensics constitute new and disruptive challenges for investigators.

Keyword:- Cloud Computing, Forensics investigation, Security, Forensic challenges

1. INTRODUCTION

Computing is a diverse collection of information technologies designed to support different services that are delivered to an individual based on their requirements. In modern computing, users need quick access to these services regardless of where the services are hosted. cloud computing can be used to provide high availability to the servers of organization. It then became the most popular concept of computer networks: The Internet. Hence it is also referred as an Internet-centric way of computing. cloud computing refers to both the applications delivered as services over the Internet and the hardware and system software in the datacenters that provide those services.

Cloud computing is a novel method for computing being used by small as well as big organizations and corporates. The Cloud Service Provider(CSP) provides services like databases, hardware, networking etc. in live environment on internet, which is a reason for few unique questions and concerns on the part of the consumers. The physical locations of the above cloud infrastructure provided by the CSPs are unknown, creating concerns regarding it's security. Herein come the issues regarding acquiring, storing and processing large amounts of data for forensic purposes which is the main issue for about a decade. Now since the sources of stored data, the physical hardware etc. being used by the consumer organizations is in a widely distributed pattern, if and when any security incidents occur, it becomes very hard for the corporate security team of the consumer organizations/firms to perform an independent investigation without depending on third parties. A lot of data is under usage in the cloud making it that much complicated for digital forensic investigators to retrieve evidence of security breaches or cyber crimes in terms of scope of three aspects viz. Technological, legal and organizational processes.

Cloud forensics needs to use a hybrid approach that taps into devices used to access cloud services viz. remote, virtual, network, live, large-scale, thin-client, thick-client or end-point to discover the digital artifacts. Cloud computing has become a household term in recent year. Cloud computing is not just for businesses capable of purchasing their own server. Online storage utilities are thought of single block of storage. It is used by end user download files and upload a files. An forensic investigation will be aware of requirement of cloud based forensic practices. The cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources for eg. Networks, storages, services, servers and applications that can be rapidly provisioned and released with minimal management or service provider interaction, cloud forensics is a subset of digital forensics based on investigation cloud environments. Cloud computing services enable vendors to provide a on-demand services to the users.

Cloud poses various security threats and attacks in the cloud. It's moves application software and databases to large data centers. Data collection technique play a major role to identify the source of attacks by acquiring evidence from various sources like cloud log analysis, web browser, physical evidence acquisition process. A criminal can also keep secret files in the cloud storage. Cloud computing is a technology that involved from technologies of the field of distributed computing, grid computing. Cloud computing is from technical point of view and combination of existing technologies for people these is difficult task to capture a big pictures. The cloud forensic major three dimensions: technical, organizational and legal. The legal dimension refers to multi-jurisdiction and multi-tenancy challenges.

1.) Technical Background:-

According to the NIST, Cloud Computing is 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.

This new raw definition of Cloud Computing brings in several new features such as multi-tenancy, elasticity, pay-as-you-go and reliability. Within this work, the following three models are used in the context of Cloud Computing:

a.) Infrastructure as a Service (IaaS):- This is a model, in which the customer is uses the virtual machine provided by the CSP to install his own system on it. The system can then be used like any other physical computer with a certain limitations. However, such additive power over the original system comes along with additional security obligations.

b.) Platform as a Service (PaaS):- Offers the capability to deploy application packages created using the virtual development environment supported by the CSP. For the efficiency of Software Development Process this service model can be of particular advantage.

c.) Software as a Service (SaaS):- This is a model, in which the customer borrows a service run by the CSP on a Cloud infrastructure to use it for own individual/organizational purpose. In a lot of cases this service can be accessed through an API for a thin client interface such as a web browser. Usage of closed-source public SaaS offers such as Amazon S3 and Google Mail is limited to the public deployment model, which leads to further issues concerning security, privacy and the gathering of suitable evidences. Furthermore, the two main deployment models viz. private and public Cloud, need to be distinguished. Common public Clouds are made available to the general public or a large industry group. The corresponding Cloud infrastructure is owned by an organization acting as a CSP and offering services to its customers.

In contrast, the private cloud is solely operated for an organization but may not provide the scalability and agility of public Cloud offers. The additional notions of Community Cloud and Hybrid Cloud are not exclusively covered within this work. However, independently from the specific model used, the movement of applications to the Cloud offered as services by a CSP, comes along with limited control for the customer about the application itself, the data pushed into the applications and also about the underlying technical infrastructure.

2.) Technical Issues:-

Digital investigations are about control of forensic evidence data. From the technical point of view, this evidence data can be available in three different states: at rest, in motion or in execution. Data at rest is represented by allocated disk space. Whether the data is stored in a database or in a specific file format, it allocates disk space. Furthermore, if a file is deleted, the disk space is de-allocated for the operating system but the data is still accessible since the disk space has not been re-allocated and overwritten. This fact is often exploited by investigators which explore these de-allocated disk space on hard-disks. In case the data is in motion, data is transferred from one entity to another e.g. a typical file transfer over a network can be seen as a data in motion scenario. Several encapsulated protocols contain the data each leaving specific traces on systems and network devices which can in return be used by investigators. Data can be loaded into memory and executed as a process. In this case, the data is neither at rest or in motion but in execution. On the executing system, process information, machine instruction and allocated/de-allocated data can be analyzed by creating a snapshot of the current system state.

2.1 Sources and Nature of Evidence:-

Concerning the technical aspects of forensic investigations, the amount of potential evidence available to the investigator strongly diverges between the different Cloud service and deployment models. Independently from the used model, the following three components could act as sources for potential evidential data.

2.1.1 Network Layer:-

The different ISO/OSI network layers provide several information on protocols and communication between instances within the

Cloud as well as with instances outside the Cloud [12, 9, 8]. Unfortunately, ordinary CSP currently do not provide any log data from the network components. This means, that in case of malware infection of an IaaS VM, it will be difficult to get any form of routing information. This situation gets even more complicated in case of PaaS or SaaS. Hence, the situation of forensic evidence is again strongly affected by the support the investigator receives from the customer and the CSP.

2.1.2 Client System:-

On the system layer of the client, it completely depends on the used model (IaaS, PaaS, SaaS) if and where potential evidence could be extracted. In most of the Cloud scenarios, the browser on the client system is the only application that communicates with the service in the Cloud. This especially holds for SaaS applications which are used and controlled by the web browser. Hence, in an exhaustive forensic investigation, the evidence data gathered from the browser environment [15] should not be omitted.

2.1.3 Virtual Cloud Instance:-

The virtual instance within the Cloud, where i.e. data is stored or processes are handled, provides potential evidence [3, 2, 6]. In most of the cases, it is the place where an incident happened and hence provides a potential starting point for a forensic investigation. The instance can be accessed by both, the CSP and the customer who is running the instance. Snapshots provide a powerful technique for the customer to freeze specific states of the virtual machine. Therefore, virtual instances can be still running which leads to the case of live investigations or can be turned off leading to static image analysis. In SaaS and PaaS scenarios, the ability to access the virtual instance for gathering evidential information is higly limited or simply not possible.

Digital forensic challenges are categorized into three major heads as per Fahdi, Clark, and Furnell (2013)[2] these are:-

1.)Technical challenges,

2.)Legal challenges,

3.)Resource Challenges

1.)TECHNICAL CHALLENGES:-

With new developments in technology, crimes and criminals also parallely develop and update their methods with it.

Digital forensic experts adopt forensic tools for collecting pieces of evidence against criminals, while the criminals use such tools for hiding, altering or removing the traces of their crime, in digital forensics this process is called Anti-forensics technique which is considered as a major challenge in digital forensics world.

Other Technical challenges are:

1.Operating in the cloud.

2.Time to archive data.

3.Skill gap

4.Steganography

Anti-forensics techniques are categorized into the following types:-

1 Encryption:- It is legitimately used for ensuring the privacy of information by keeping it hidden from an unauthorized user/person. Unfortunately, it can also be used by criminals to hide their crimes.

2 Data hiding in storage space:- Criminals usually hide chunks of data inside the storage medium in invisible form by using system commands, and programs.

3 Covert Channel:- A covert channel is a communication protocol which allows an attacker to bypass intrusion detection technique and hide data over the network. The attacker used it for hiding the connection between him and the compromised system.

2.)LEGAL CHALLENGES:-

An area presenting new opportunities for both legitimate business, as well as criminal organizations, is Cloud computing. This work gives a strong background in current digital forensic science, as well as a basic understanding of the goal of Law Enforcement when conducting digital forensic investigations. These concepts are then applied to digital forensic investigation of cloud environments in both theory and practice, and supplemented with current literature on the subject. Finally, legal challenges with digital forensic investigations in cloud environments are discussed.

Legal issues have risen with the changing landscape of computing, especially when the service, data and infrastructure is not owned by the user. With the Cloud, the question arises as to who is in the possession of the data. The Cloud provider can be considered as a legal custodian, owner or possessor of the data thereby causing complexities in legal matters around trademark infringement, privacy of users and their data, abuse and security. By introducing Cloud design focusing on privacy, legal as a service on a Cloud and service provider accountability, users can expect the service providers to be accountable for privacy and data in addition to their regular SLAs.

3.)RESOURCE CHALLENGES :-

Resource management is the primary issue as the demand grows for provisioning resources and computation in cloud systems. This article presents various research issues pertaining to the management of cloud resources while a comparison is made between existing resource allocation systems. The issues and challenges discussed in this paper are resource provisioning, job scheduling, load balancing, scalability, pricing, energy management and availability.

2. MOTIVATION

With the increasing demand for using the power of the Cloud for processing lots of sensible information and data, enterprises face the issue of Data and Process Etymology in the Cloud . Digital etymology, means, the meta-data that describes the ancestry or history of a digital object, is a pivotal feature for forensic investigations. In combination with a suitable authentication scheme, it provides information about who created and who modified what kind of data in the Cloud. These are essential aspects for digital investigations in distributed environments such as the Cloud. Unfortunately, the aspect of forensic investigations in such distributed environment has so far been mostly forsaken by the research community. Current discussion centers mostly around Cloud Security and Privacy/Data Protection [16]. The impact of forensic investigations on Cloud environments was little noticed. In 2009, the authors of [4] stated that ”[...] to our knowledge, no research has been published on how cloud computing environments affect digital artifacts, and on acquisition logistics and legal issues related to cloud computing environments.”. At the same time, massive investments are being made in Cloud technology. Combined with the fact that information technology increasingly exceeds peoples’ private and professional life, thus mirroring more and more of peoples’ actions, it becomes apparent that evidence gathered from Cloud environments will be of high significance to litigation or criminal proceedings in the future.

3.2 Investigations in XaaS Environments:-

Within this section specific issues of investigations in SaaS, PaaS and IaaS environments will be discussed.

3.2.1 PaaS Environments:-

One of the main advantages of this model is that the core application is under the control of the customer. Given these circumstances, the customer obtains theoretically the power to dictate how the application interacts with other dependencies (databases, storage entities etc.). Moreover, depending on the runtime environment, logging mechanisms can be implemented which automatically sign the information and transfer it to a third party storage. Additional encryption could prevent potential eavesdroppers from being able to view log data information on the way to storage server. CSP normally claim that this transfer is encrypted but this statement can hardly be verified. Since the customer has the ability to interact with the platform over a prepared API, system states and specific application logs can be extracted. However potential adversaries, which can compromise the application during runtime, should not be able to alter these log files afterwards which could be realized by push-only mechanisms. Unfortunately, the customer has no direct control of the underlying runtime environment. As in the SaaS scenario, this strongly depends on the configuration done by the CSP. Concerning the Microsoft Azure platform, the environment is made of an virtualized OS (Microsoft Windows), a webserver (Internet Information Server) and the runtime environment (.NET). Windows Azure Diagnostics, a new feature released in November 2009, gives developers the ability to collect and store a variety of diagnostics data in a highly configurable way.

3.2.2 SaaS Environments:-

Especially in the SaaS model, the customer does not obtain any control of the underlying operating infrastructure such as network, servers, operating systems etc. or even the application that is used. This means that no deeper view into the system and its underlying infrastructure is provided to the customer. Only limited user-specific application configuration settings can be controlled. In a lot of cases this urges the investigator to rely on high-level logs which are eventually provided by the CSP. Given the case that the CSP does not run any logging application, the customer has no opportunity to create any useful evidence by himself. The installation or configuration of any toolkit or logging tool is impossible. These circumstances do not allow a valid forensic investigation and lead to the assumption that customers of SaaS models do not have any chance to analyze potential incidences. Moreover, evidence data has to be interpreted by an investigator in a proper manner which is hardly be possible due to the lack of circumstantial information. For auditors, this situation does not change: Questions who accessed specific data and information cannot be answered by the customers, if no corresponding logs are available. Moreover, a lot of SaaS CSP like Google offer Single sign-on (SSO) access control to the complete set of their services. Unfortunately in case of an account compromise, most of the CSP do not offer any possibility for the customer to figure out which data and information has been accessed by the adversary. In private SaaS scenarios this situation is tremendously improved by the fact that the customer and the CSP are probably under the same authority. Hence, logging mechanisms could be implemented which contribute to potential investigations. Additionally, the exact location of the servers and the data is known at any time. Due to the limited ability of receiving forensic information from the server in SaaS scenarios, the client has to contribute to this process. This can be achieved by implementing Proofs of Retrievability (POR) in which a verifier (client) is enabled to determine that a prover (server) possesses a file or data object without actually downloading it [11]. In [17], the authors introduced a new data integrity verification mechanism for SaaS scenarios which could also be used for forensic purposes.

3.2.3 IaaS Environments:-

From the forensic point of view, IaaS instances provide much more information that could be used as forensic evidence in case of an incident than the PaaS and SaaS models do [6]. This fact is caused through the ability of the customer to install and set up the image for forensic purposes. Hence, log data and other evidence information could be transferred to other hosts in a frequent manner for providing the ability to perform an investigation if needed.

- Snapshots:-

Traditional forensics expect target machines to be powered down to collect an image. This situation completely changed with the advent of the snapshot method which is supported by all popular hypervisors such as Xen, VMware ESX and Hyper-V1. Snapshots, also referred to as forensic image, of virtual machines provide a powerful tool with which a virtual machine can be cloned by one click including also the running system's memory. This leads to the main benefit that systems hosting crucial business processes do not have to be shutdown for performing a forensic analysis. This could also affect scenarios in which a downtime of a system is not feasible or practical due to existing SLAs. Due to the fact that the customer is responsible for the security of the virtual instance, the system itself can be prepared for forensic investigation purposes. RFC 3227 [7] contains several best practices for responding to a security incident especially in the case of live investigating systems. According tob[3], log data information concerning currently logged users, open ports, running processes, system and registry information etc. should be gathered. These log data should be transferred to an external system mitigating the chance that a maliciously motivated shutdown process destroys the data. Encrypting and signing these log files can be helpful for providing security and integrity of the created log files. Unfortunately, it has to be emphasized that each process such as an encryption process run on the virtual instance, can be controlled by the hypervisor or the CSP respectively. Although this risk can be disregarded in most of the cases, the impact on the security of high security environments is tremendous. Generally, for an investigator it is important to know if the virtual machine was properly shutdown or is still running [2]. Hence, virtual instances have to be divided into two different categories concerning the forensic analysis of the system: shutdown (dead) and running (live) virtual systems. In general, live investigations on running virtual machines become more common providing evidence data that is not available on shutdown systems. The technique of live investigation is mostly influenced by the huge amount of evidence data that has to be stored and processed in case of shutdown instances. Nevertheless, it cannot be denied that live investigations change the state of the investigated system and the results of the investigation may not be repeatable. However, this does no prevent a lot of SMCs from mostly performing live investigations due to the bond of legislation and government-contracting agreements.

- Volatile Data:-

Depending on the Cloud offer used, virtual IaaS instances do not have any persistent storage. In the specific case of an AWS EC2 cloud instance, all volatile data is lost if the instance is rebooted or shutdown. Persistent data has to be stored in long time storage environments like Amazon Simple Storage Service (S3) or Amazon Elastic Block Store (EBS). This situation leads to several issues: In case an adversary compromises a virtual IaaS instance with no persistent storage synchronization, the adversary could shutdown the system leading to a complete loss of volatile data. Additionally, the instance could be abused for sending spam, attack further external and internal targets, join botnets and steal volatile data of the running system. After the attack, the attacker can cancel the contract with the corresponding CSP forcing the virtual machine to shutdown and destroy most of the evidence data which is inevitable for further forensic investigations. This problem mainly results from the unclear situation how CSP handle the termination of customer contracts. In real world scenarios, this process is not transparent for the customer bringing up further questions e.g. does data on virtual systems in the Cloud get exhaustively deleted and how is this done? File deletion is all

about control and this used to not be an issue till the advent of Cloud Computing. In current Cloud environments CSP do not offer any verification process providing the ability for the customer to verify that the sensitive data stored on a virtual machine has been deleted exhaustively. Moreover, an interesting perspective is the case in which the real owner of the image decides to engage in malicious activities through his EC2 machine from a veiled IP address and afterwards claims, someone compromised the password or key pair to his EC2 machine. In a subsequent forensic investigation, it will be difficult to prove the opposite due to the lack of evidences.

- Virtual Introspection:-

As expected, even virtual instances in the Cloud get compromised by adver saries as happened to Amazon EC2 instances in the end of 20092. Hence, the ability to determine how defenses in the virtual environment failed and to what extent the affected systems have been compro mised is crucial not only for recovering from an incident. Also forensic investigations gain leverage from such information and contribute to resilience against future attacks on the systems. Virtual Introspection (VI) is the process by which the state of a virtual machine is observed from either the VMM or from some virtual machines other than the one being examined [10]. However, the fact that the VMM has full access to the resources of all VMs represents a significant risk to customers' data. The issue whether VMs can ever be managed by a VMM, while simultaneously being protected from a compromised VMM remains an open research problem.

3. CONCLUSION

Cloud computing comes with numerous economical security benefits to firms /organisations which have limited security resources due to budget constraints in normal circumstances. But, there is an absence of standards for processes used within the Cloud, causing various problems such as security, proper deployment, compliance and the very basic issue of how can am investigation within such an environment be processed. Further there is loss of control over the data and computing resources being used by the consumers and provided by Cloud environments and vendors, presenting great challenges to digital forensic investigators. Investigators need to be clever enough to reconstruct the corresponding Cloud environment in order to recreate scenarios and test hypothesises. The computer forensics communities' preliminary findings need to be revised and adapted to the new environment in the field of digital forensics. Also the investigators need to identify and collect collateral data to prove or disprove hypothesis created by them for various possible scenarios in cloud environment with reference to the analysis of any particular case. The investigators need to know whether there is any knowledge of CSP logs and how long they keep the information. Digital forensics is defined as the usage of scientifically derived and proven methods towards the identification, collection, preservation, validation, analysis, interpretation, and presentation of digital evidence derivative from digital sources to facilitate the reconstruction of events found to be criminal. But such digital forensics investigation methods face some major challenges at the time of practical implementation.

4. REFERENCE

[1] Cloud computing: Business benefits with security, governance and assurance perspectives. Technical report, ISACA, 2009.

[2] R. A. Bares. Hiding in a virtual world: using unconventionally installed operating systems. In ISI'09: Proceedings of the 2009 IEEE international conference on Intelligence and

security informatics, pages 276–284, Piscataway, NJ, USA, 2009. IEEE Press.

[3] D. Barrett and G. Kipper. Virtualization and Forensics: A Digital Forensic Investigator’s Guide to Virtual Environments. Syngress, 6 2010.

[4] N. Beebe. Digital forensic research: The good, the bad and the unaddressed. Advances in Digital Forensics V, pages 17–36, 2009.

[5] Ruan, K., et al. Cloud forensics. in IFIP International Conference on Digital Forensics. 2011. Springer.

[6] D. Bem and E. Huebner. Computer forensic analysis in a virtual environment. International Journal of Digital Evidence, 6(2), 2007.

[7] D. Brezinski and T. Killalea. Guidelines for evidence collection and archiving, 2002.

[8] V. Corey, C. Peterman, S. Shearin, M. Greenberg, and J. Van Bokkelen. Network forensics analysis. IEEE Internet Computing, 6(6):60–66, 2002.

[9]Taylor, M., et al., Digital evidence in cloud computing systems. Computer Law & Security Review, 2010. 26(3): p. 304-308.[10] B. Hay and K. Nance. Forensics examination of volatile system data using virtual introspection. SIGOPS Oper. Syst. Rev., 42:74–82, April 2008.

[11] A. Juels and B. S. Kaliski. Pors: proofs of retrievability for large files. In In CCS ’07: Proceedings of the 14th ACM conference on Computer and communications security, pages 584–597. ACM, 2007.

[12] Martini, B. and K.-K.R. Choo, An integrated conceptual digital forensic framework for cloud computing. Digital Investigation, 2012. 9(2): p. 71-80.

[13] Nasreldin, M.M., et al., Digital Forensics Evidence Acquisition and Chain of Custody in Cloud Computing. International Journal of Computer Science Issues (IJCSI), 2015. 12(1): p. 153.[15] M. T. Pereira. Forensic analysis of the firefox 3 internet history and recovery of deleted sqlite records. Digital Investigation, 5(3-4):93–103, 2009.

[14] T. Ristenpart, E. Tromer, H. Shacham, and S. Savage. Hey, you, get off of my cloud! Exploring information leakage in third-party compute clouds. In S. Jha and A. Keromytis, editors, Proceedings of CCS 2009, pages 199–212. ACM Press, Nov. 2009.

[15] Y. Shi, K. Zhang, and Q. Li. A new data integrity verification mechanism for saas. In F. Wang, Z. Gong, X. Luo, and J. Lei, editors, Web Information Systems and Mining, volume 6318 of Lecture Notes in Computer Science, pages 236–243. Springer Berlin / Heidelberg, 2010.