Journal of Basic and Applied Engineering Research
Print ISSN: 2350-0077; Online ISSN: 2350-0255; Volume 1, Number 8; October, 2014 pp. 21-26
© Krishi Sanskriti Publications
http://www.krishisanskriti.org/jbaer.html

# A Study on Tools and Techniques used for Network Forensic in a Cloud Environment: an Investigation Perspective

**J. Rajeshwar Rao[1], Shweta Jain[2], Siby Samuel[3]**

[1]*St. Aloysius College (Auto.) Jabalpur, M.P.*
[2]*Student, M.Sc. CS [Final] St. Aloysius College (Auto.) Jabalpur, M.P.*
[3]*St. Aloysius College (Auto.) Jabalpur, M.P.*

***Abstract:*** **The modern computer environment has moved past the local data center with a single entry and exit point to a global network comprising many data centers and hundreds of entry and exit points, commonly referred as Cloud Computing, used by all possible devices with numerous entry and exit point for transactions, online processing, request and responses traveling across the network, making the ever complex networks even more complex, making traversing, monitoring and detecting threats over such an environment a big challenge for Network forensic and investigation for cybercrimes. It has demanded in depth analysis using network tools and techniques to determine how best information can be extracted pertinent to an investigation. Data mining technique providing great aid in finding relevant clusters for predicting unusual activities, pattern matching and fraud detection in an environment, capable to deal with huge amount of data. The concept of network forensics in cloud computing requires a new mindset where some data will not be available, some data will be suspect, and some data will be court ready and can fit into the traditional network forensics model. From a network security viewpoint, all data traversing the cloud network backplane is visible and accessible by the cloud service provider. It is not possible to think now that one physical device will only have one operating system that needs to be taken down for investigation. Without the network forensics investigator, understanding the architecture of the cloud environment systems and possible compromises will be overlooked or missed. In this paper we focus on the role of Network Forensic in a cloud environment, its mapping few of the available tools and contribution of Data Mining in making analysis, and also to bring out the challenges in this field.**

***Keywords: Network Forensic Investigation, Cloud Environment, Virtualization, and Cybercrime.***

## 1. INTRODUCTION

Cloud computing is an emerging research infrastructure that builds on the achievements of different research areas, such as service-oriented architecture (SOA), grid computing, and virtualization technology. It offers infrastructure as a service that is based on pay-as-you-use and on-demand computing models to the end users (exactly the same as a public utility service like electricity, water, gas, etc.). With this flexibility comes few of the aspects becoming bottle neck for the Network forensic and investigation when it comes to fraud detection, tracing of unusual activity and other network operation that are part of a monitoring system.**[1]**

Virtualization has revolutionized data center's technology through a set of techniques and tools that facilitate the providing and management of the dynamic data center's infrastructure. It has become an essential and enabling technology of cloud computing environments. Virtualization can be defined as the abstraction of the four computing resources (storage, processing power, memory, and network or I/O). It is conceptually similar to emulation, where a system pretends to be another system, whereas virtualization is a system pretending to be two or more of the same system. The figure **1** below shows a layered Virtualization architecture. **[2]**

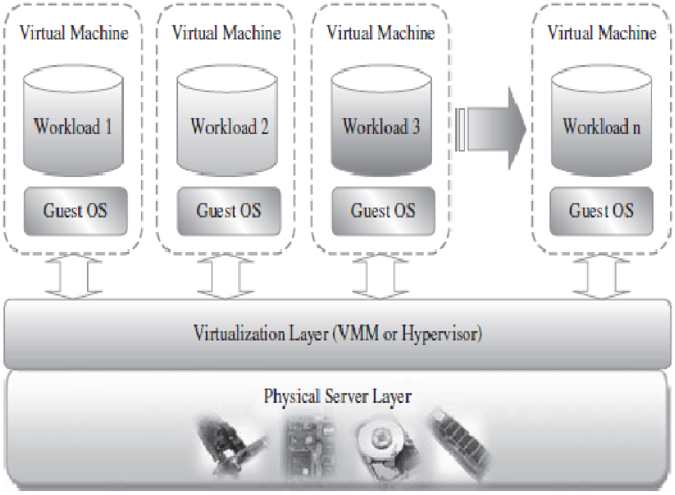

**Fig. 1: Layered Virtualization architecture.**

## 2. THE ROLE AND LIMITATION OF NETWORK FORENSIC IN CLOUD COMPUTING

Having an effective management's suite for managing virtual machines' infrastructure is critical for any cloud computing infrastructure as a service (IaaS) vendor. Now imagine some billions of users accessing their data from those rented interfaces with each users, probably equipped with two to three devices like PC, Laptops, Smartphones, PDA and many more gadgets that allow accessing those data over the network using Apps, which allow them to perform almost all transactions, banking operations, and many more. And the job of Network monitoring, the data which flow like a stream with threats and fault becomes everything equivalent to impossible, and not only this, the architecture, so called "Cloud Computing" has its own **limitation** to provide the basis for Network forensic and investigation process. As said before, with numerous entry and exit point for transactions, online processing, request and responses traveling across the network, making the ever complex networks even more complex, making traversing, monitoring and detecting threats over such an environment a big challenge for Network forensic and investigation for cybercrimes. It has demanded in depth analysis using network tools and techniques to determine how best information can be extracted pertinent to an investigation.

Network forensic **helps** tries to analyze traffic data logged through firewalls or intrusion detection system or at network devices like routers and switches. A forensics investigation requires the use of disciplined investigative techniques to discover and analyses traces of evidence left behind after a committed crime.

Network forensics in the cloud computing environment could be focused only on data that go to and from the systems that the company has access to, but that would miss the rest of the picture. Network forensics needs to be part of and work with all the other components that comprise the entire system within the cloud environment. Without the network forensics investigator, understanding the architecture of the cloud environment systems and possible compromises will be overlooked or missed. The network forensics investigator also needs to understand that the cloud environment is the space that the company rents on another company's computer systems to perform the work. **[3]** The rented space in the cloud can be in a globally connected data center with many other companies where the user network entry point can be at any point on the Internet. Data in the cloud environment can be replicated to any data center in the world that is owned and operated by the cloud provider. The cloud providers have their own series of policies, security systems, hardware, and software packages that are independent of what a company is doing in the cloud space. Cloud computing customers **may or may not have access** to the data that relates to them specifically if a computer is suspected to have been compromised by a hacker or if data is stolen by an insider or outsider.

Cloud computing can also frustrate network forensics because of the **lack of direct access** to the physical machines that are suspect. This can also frustrate network-based forensics because the way that the cloud environment is set up at the hosting facility. Accessible evidence may only be limited to data on the virtual machine or system and not across the entire network path from end-to-end. Networks are only a logical hierarchy in the investigation rather than being able to directly monitor the data from a span port off the network device. Complicating the process is that the current set of skills and tools are **still being developed**. This means that for now, the tools and skills are still relatively immature in relationship to the current tool sets that are available when the systems and networks are fully owned by a company.

**Memory in role:** The study of forensics to branch out in new directions and encourages the development of more technology-specific methodologies and techniques. As technology continues to evolve and become more varied and complex, so must the toolkit available to a forensic examiner. Many researchers believe that volatile memory is becoming the location of choice for attackers and other malicious users to stores information they do not want found, or execute harmful code that they do not want reverse-engineered. It is convincing simply because the majority of incident response teams do not collect volatile memory, and the majority of analysts are not equipped to decipher relevant information from it. In addition, several recent pieces of malware have been written to exist solely in memory and never touch the hard drive, effectively evading any analysis that only looks at the physical disk. **[5]** It is logical to assume that malicious users will continue to move toward using physical memory instead of the hard disks, and analysts will need to adapt to that movement**.**

There are a lot of ways that their cloud and your data can be compromised.

**Hackers** attempt to break secure network or try to fit in the header of a secure connection to get involved in the activity, which sometimes means, they can sell your proprietary information to your competition to surreptitiously encrypting your storage until you pay them off or they may just erase everything to damage your business and justify the action based on their ideological beliefs. Either way, hackers are a real concern for your data managed on a cloud. Because your data is held on someone else's equipment, you may be at the mercy of whatever security measures they support. **[6]**

**Bot Attackers:** In a commonly recognized worst-case scenario, attackers use botnets to perform distributed denial of service (DDOS) attacks. In order to get the hackers to stop attacking your network, you face blackmail**.** In fact, in Japan, blackmail involving DDOS is on the rise. One major Tokyo firm had to pay 3 million yen (about U.S. $31, 000) after the network was brought to a screeching halt by a botnet attack. Because the attack was so dispersed, police have been unable to track down the attackers. In the world of cloud computing, this is clearly a huge concern. Network forensics deals with the capture, recording and analysis of network events in order to discover evidential information about the source of security attacks in a court of law.**[4]** If your data is on the cloud being attacked, who or will the ransomers come to for money? Will it be you? Will it be the vendor? And will the ransom even be paid? All these questions are unanswered in most of the cases.

## 3. MAPPING WITH AVAILABLE TOOLS

The good part of the cloud computing environment is that not everyone is aware of its potential and not everyone is aware of the risks involved in using the cloud-based systems. As more and more data is accessed and shared between systems, usually a user interface has cached credentials involved with the process somewhere in the software. Cached credentials are beneficial for an investigation, allowing the investigation to proceed unhindered, but it also means that anyone who had access to that machine in the context of the user also might have had access to any system that had its credentials cached. This provides a fertile source of information and associated systems that could be a part of the investigation being conducted.

**Tools:** There are **tools** like **Hyperic HQ,** which enables the modern enterprises to monitor their Amazon Web Services securely alongside internal infrastructure. It is also the first enterprise-class monitoring and management software offered for deployment and payment directly through Amazon Web Services. Hyperic also offers a free cloud monitoring tool, **CloudStatus**. Their most recent addition to the tool is continuous monitoring of Google. Google App Engine is the second significant cloud service to be monitored by CloudStatus, which launched in June 2008 with support for Amazon Web Services. Applying the cloud computing model to distributed development, **CollabNet CUBiT 2.0** enables teams to access on-demand servers from private corporate datacenters or public clouds. **Cassatt** Corporation offers several products to help internal cloud computing—an IT approach that delivers the benefits of cloud computing using the resources that organizations already have inside their datacenters. Cassatt Active Response enables datacenter managers to use policies to control and optimize the multiple diverse components of their IT infrastructure. Cassatt Active Response can monitor and automatically provision or decommission physical and virtual server, software, and network resources as appropriate to meet the application demand.

The cloud still has physical security constraints. After all, there are actual servers running somewhere. The above mentioned tools and many more are not centralized to a few of the basic concerns like data security in terms of lawsuits, physical security protocol, access control, network and host security, encryption technique used for file system, regulatory and standards compliance where standards written for Internet applications predate the acceptance of virtualization technologies. Currently, the law enforcement agencies such as polices are conducting the computer forensics work. The Association of Chief Police Officers (ACPO) Guidelines **[7]** for Computer Investigations and Electronic Evidence is a thorough complete document, which specifies the procedures and steps that officers should take in dealing with a variety of situations associated with computers and digital evidence. This guideline provides the necessary information to ensure that each investigation is performed to the highest level of standards. **[8]**

All the six Layers (Network, Physical Hardware, Host OS, Virtualization, Guest OS and Guest Application/data) of the IaaS Cloud Environment carries the network as a trust required, making the Network as a crucial part of the Cloud Environment. Thus requires a special attention by all the network and Cloud management tools. This paper focus on the fact that there is a need of greater mapping of network in the tools that deals with Cloud environment, especially for the process of investigation. Whether it's a user-end, third party or Service Provider, the investigators need to have compiled information of the network activity such as logs of user terminals (it may be any device used to access Cloud storage), network units like Routers, Switches, Firewalls, etc., System logs (which in many cases becomes unproductive because of virtualization) and operating system concern.

**Network Intrusion Detection:** Perimeter security often involves network intrusion detection systems (NIDS), such as **Snort**, which monitor local traffic for anything that looks irregular. Examples of irregular traffic includes Port scans, Denial-of-service attacks, and Known vulnerability exploit attempts. **[9]**

Network forensics can also have an influence on the outcome of an investigation into an event as long as data was collected at the box and at the entry and exit points of the company network. The use of in-built firewall logs, system logs, and other logs will generally point to an entry time, place, and IP address that can be used to help determine how the event was propagated through the network and what steps can be taken to help minimize any future event by providing solid data on the event. A large part of network forensics is being able to monitor the network traffic in order to isolate the number of servers that need to be taken down for the traditional forensics process. This is where the process gets problematic – there are **porous boundaries** with any system that can access those systems that have been compromised. Network forensics can be likened to deep packet inspection of the packet header and no encrypted payload, as well as stateful packet inspection.

This is much like any good market intrusion detection system (IDS) or intrusion prevention system (IPS). The problem with this is that in the cloud data center, the network investigator is limited to the data that can be recovered at the server. The traffic that is on the backplane of the network is **not** going to be available because of the manner in which the virtualized systems work. It is **impossible** to isolate a series of compromised computers, and it is impossible to "sniff the local network" in the cloud because of the way that the hypervisor and virtualization systems work.

**Wireshark** for both Windows and Linux, **WinPcap** for Windows, and **Snort** for both Windows and Linux, as well as the in-built firewalls for those two operating systems are few among the available tools that plays a greater role in the Network forensic.

***Need for Security Policies, Standards, and***

**Procedures for the IT Infrastructure.** The need for security policies, standards, and procedures becomes readily apparent when one recognizes that an IT infrastructure is only as secure as the weakest link in the chain of security responsibility. For example, one could secure data resources and maintain the confidentiality by storing it in an encrypted database stored on a server positioned on a LAN behind a firewall. However, if that same server is not installed in a data center or hosting center with proper physical security measures where access is limited to authorized personnel, the same becomes insecure.

**Implementing Standard and Enhanced Security Solutions.** There are many security countermeasures that can be implemented and deployed throughout a LAN and WAN network environment. Some of the tools, requirements, and items to incorporate into a security counter measure strategy as contributed by **Clint P. Garrison** in "Digital Forensics for Network, Internet and Cloud Computing: A forensic evidence guide for moving targets and data" are:

- Internet-facing systems should run the vendor's most stable and secure operating systems and have any necessary patches applied.
- Network-based intrusion detection systems (IDS) should monitor the outside perimeter segment to provide details on outside system attacks taking place before their entry into the firewall. These IDS should report activity back to the security information management system.
- As budgets allow, key network devices should be deployed in pairs with automatic fail over capability to increase overall network availability.
- VPN clients with built-in software firewalls should be used to prevent tunneling attacks. Incoming remote access data streams should be decrypted outside of or at the corporate firewall, optionally analyzed by an IDS system, run through the firewall rule base, and analyzed by an IDS system once inside the system.
- Key network devices should employ the use of out-of-band management channels that only allow a minimum of management communication from a dedicated management segment to the devices. Devices that are remotely located and cannot be connected to the management segment should employ VPN tunnels for their management. Other out-of-band management techniques (e.g., asynchronous terminal servers) should be considered.
- The organization's servers and workstations should employ the use of centrally managed and updated antivirus systems.
- Network-based IDS should monitor key segments for post-firewall data traffic intrusions and internal intrusions. These IDS should report activity back to the security information management system.
- Host-based IDS should be utilized on key hosts, if not all hosts, and be configured to report back to the security information management system.
- Internal WAN data traffic should be encrypted over private links if the data's value is sufficiently high and risk of exposure exists.
- VPN connections and internal network users are authenticated using one or more types of authentication servers.

As more data-visualization tools and flow-control tools can be made to work in the cloud computing environment, independent of the operating system limitations of Azure or hypervisor limitations of any virtualized system, the more effective and capable network forensics will be.

## 4. NETWORK ANALYSIS: DATA MINING APPLIED TO CONTRIBUTE NETWORK FORENSIC

When it comes to investigation, especially in case Network Forensic, analysis is the most important part of the process as all other stages like Case design, are dependent on the analysis report only.

**Analysis** is very easy when data is available in concrete format. But not the case here, analysis is most of the critical thing as far as Network forensic is concerned, since data here is Numeric, Alphanumeric, Character based and many a times mixed with such a noise that classification itself becomes a tedious approach. Good news is we have a scientific stream which is capable enough to deal with such a noisy, volatile and non-sequenced data, known as Data Mining.

Data mining is about solving problems by analyzing data already present in databases or any storage. Data mining is defined as the process of discovering patterns in data. The process must be automatic or (more usually) semiautomatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic one. The data is invariably present in substantial quantities.

We can apply or integrate few of the algorithms provided for the use of Data Mining, so as to make possible predictions based on Clusters or Association Analysis, captured as outcome using the Data Mining tool **Weka**. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from user designed program in Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization and is also well-suited for developing new machine learning schemes.

The top ten data mining algorithms as per 2006 poll conducted by the International Data Mining Conference are **C4.5, k-means, SVM, Apriori, EM, PageRank, Adaboost, kNN, Naïve Bayes** and **CART**.

But data is a massive dataset, and require development of algorithms that treat input as a continuous data stream. **Naïve Bayes** is a rare example of an algorithm that needs no adaptation to deal with data streams. Training is incremental: It merely involves updating a fixed set of numeric parameters. Memory usage is small because no structure is added to the model. Other classifiers with the same properties include 1R and the basic perceptron. Multilayer neural nets usually have a fixed structure as well, stochastic backpropagation updates weights incrementally after each training instance has been processed, rather than in a batch operation, and thus is suitable for online learning. Rules with exceptions make modifications incrementally by expressing exceptions to existing rules rather than reengineering the entire set, and thus could be rendered suitable for data stream learning—although care would need to be taken to ensure that memory usage did not increase inexorably as the number of exceptions increased. **[10]**

We propose, keeping the Network forensic point of view, the development of a fusion version of Aprori **[11]**, k-Means and Naïve Bayes algorithm, in a sequence considering analysis based on the association rule applied on the Captured data and then to determine whether to apply k-means (if belongs to Clustering) or Naïve Bayes (if belongs to Classification) based on the type of data retrieved during the Capturing process of investigating the Network resources using tools like Wireshark or NIKSUN's Puma. Other combination of tools (as listed in http://www.forensicswiki.org/wiki/Tools:Network_Forensics) can be applied depending on the type of investigation and issue respectively.

**Challenges:** In the boundaries of the content discussed in this paper, the challenge(s) in Network Forensic is the development of algorithms and tools for analysis of the Captured record, which prepares foundation for the investigation. The core challenge, as mentioned, is the development of Algorithms that incorporate the changing needs of the Network forensic process to throughput an efficient analysis based on which further investigation can be processed. Secondly the development of Forensic tools which are scalable to retrieve information, relevant information from the wide range of devices and terminals used or engaged in a Cloud Environment. The future of network forensics depends on resolving the challenges faced by the criminal justice system regarding the admission of evidence and use of forensically sound procedures. It also depends on the organizations implementing devices to reduce and mitigate the challenges with existing devices to perform network forensics.

## 5. CONCLUSION

In this paper we tried to bring out the logical and conceptual part of investigation, challenges, tools and techniques used in Network forensic in a Cloud Environment. As the world changes, the data, type of network and network units, its architecture and the environment, especially the Cloud changing, there arises the need of Compatible tools and techniques for this changing scenario in Computing world. We also propose finding a solution algorithm using Data Mining technique (an optimal and scalable version of Aprori, Naïve Bayes and k-means) to deal with massive dataset Captured from network so as to give a relevant Clusters or Classification out of those Captured Noisy and incomplete data. In an investigation perspective, the Network forensic as a process and the investigator with the required tools and techniques plays a vital and urgent role in any sorts of Cybercrime attempted, predicted or committed over the Network and Cloud. To address these challenges, a network forensics community must be created. The goal should be to provide the basis for the network forensics examiner to address the criminal justice system's mandate, the driving demands of organizations seeking effective and efficient network forensics tools, and the on-site and off-site investigative tools for the examiner based on crimes committed with the use of networking technology.

## REFERENCE:

[1] Rajkumar Buyya, James Broberg and Andrzej Goscinski, (2011), "*CLOUD COMPUTING Principles and Paradigms*" John Wiley & Sons, pp. 123

[2] Rajkumar B, pp. 126

[3] Clint P. Garrison, (2010) "*Digital Forensics for Network, Internet and Cloud Computing: A forensic evidence guide for moving targets and data*." Elsevier Inc. pp. 7-12

[4] Natarajan M, Sumanth R and Loretta A, (2009) "*Tools And Techniques For Network Forensics*" (IJNSA), Vol .1, No.1, April 2009 Issue.

[5] Kristine A, "*Techniques and Tools for Recovering and Analyzing Data from Volatile Memory,* " (2009), www.sans.org Reading Room section.

[6] Anthony T. Velte, Toby J. V, Robert Elsenpeter, (2010) "*Cloud Computing: A Practical Approach*" pp.36

[7] ACPO Good Practice Guide for Digital Evidance, (2012), www.acpo.police.uk

[8] S. Biggs and S. Vidalis, "*Cloud Computing: The Impact on Digital Forensic Investigations*, " International Conference for Internet Technology and Secured Transactions (ICITST), pp. 1-6, November 2009.

[9] George Reese, *(2009) "Cloud Application Architectures Building Applications and Infrastructure in the Cloud, " First Edition,* O'Reilly, pp.110-112.

[10] Ian H. (2011) "*Data mining: practical machine learning tools and techniques*, " Third Edition, pp.380-383.

[11] Akshay Zadgaonkar, Vijaya Balpande, *"Optimizing Live Digital Evidence Mining Using Structural Subroutines of Apriori Algorithm, "* International Journal on Computer Science and Engineering (IJCSE), Vol. 3 No. 4 Apr 2011, pp.2

### *Web Resources:*

[12] *www.cyberforensics.in*

[13] *www.edecision4u.com/FIT.html*

[14] *www.digi-forensics.com/home.html*

[15] *www.mantaro.com/products/MNIS/collector.htm*

[16] *www.cisco.com/warp/public/732/Tech/nmp/netflow/index.shtml*

[17] *www.sandstorm.net/products/netintercept*

[18] *www.niksun.com/*

[19] *www.netgrab.co.uk/*

[20] *www.pcap2wav.xplico.org/*

[21] *www.snort.org/*

[22] *www.nature-soft.com/forensic.html*

[23] *www.apnic.net/*

[24] *www.wireshark.org/*

[25] *www.xplico.org/*

[26] *www.expert-team.net/*