International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-5, January 2020

Role of Data Hiding Techniques for Detection of Virtual Disks in Cloud Environment

Yunus Khan, Sunita Varma

Abstracct- In today's era, cloud computing is very popular and the most widly used technique to store the data. As we know more than 75% of the data that is used in internet services and applications is being stored on the maximum cloud only. Where our data is stored in the cloud, it is called data center, there are two important roles in cloud computing technology, one is cloud customer and the other is cloud service provider. The complete control and monitoring at the public data center is of the service provider itself, the user is kept away from the information of the location of the data center and its access credentials. This means that the user has absolutely no information about the virtual machine hard disk, and their access locations. Whenever any forensic inquiry comes in the cloud environment, the Investigator and forensic expert first have to find out about the virtual machine disk and its location in the cloud, which is a very challenging and difficult task in the cloud environment. In this paper we have developed a new process that detects virtual machines using data hiding techniques. To prove this new algorithm, we have performed an experiment using Oracle VirtualBox 6.0 on OpenSUSE virtual machine.

Keywords: Cloud Forensics, Virtual Machine, Evidence Collection, Magic Numbers and SADS.

I. INTRODUCTION

The digital forensic is a process to apply the forensic operations in a cloud computing environment. The digital forensic process has basically four steps: a collection of digital evidence, the examination of evidence, analysis of evidence and finally reporting of evidence as proof of law [1].

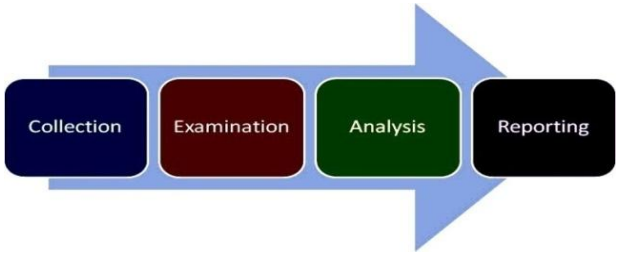

Figure 1. Process of Forensic Evidence Collection

Manuscript published on January 30, 2020.
Corresponding Author
Yunus Khan*, Ph.D. Scholar, Shri G.S. Institute of Technology and Science Indore India callyunuskhan@gmail.com
Sunita Varma, HOD and Professor, Shri G.S. Institute of Technology and Science Indore India

Identification, Collection and Acquisition phase: The identification of digital evidence consists of identifying the sources of evidence [2]. The sources of evidence could be from the client-side or service provider side. Desktop computer, laptop, mobile device or any other computing device could be a client device.

When a cloud crime is reported client device can be identified by the network [18]. The collection phase includes the collection of data from data repositories like logs files, different applications, registry files, activities of virtual machines and installed system and application software's. The data collection in a cloud environment is very difficult because the evidence is physically out of reach of the investigator [3]. If any error or fault may be detected during the detection and collection the whole process may be affected. Multitenancy and jurisdiction are also creating hurdles in the detection and collection of digital evidence in the cloud environment. The forensic technique is how the device is connected to the cloud can be known using the connection mechanism like a web browser. It is also important to know the time, format and location of the evidence belong to the CSP's or client-side [2]. In this paper, we also address the problem of how multitenancy and jurisdiction affect the detection and collection of evidence in the cloud environment. After the collection of digital evidence, preservation is the most important phase to protect the integrity of the evidence in a proper way. Hashing is used to authenticate the integrity and authenticity of the evidence. The preservation of the digital evidence is widely done by the MD4, MD5, SHA-1 and other integrity check mechanism available in security. The collection of data in a cloud environment is fully dependent on the cloud platform and deployment model is used because the evidence is collected either from cloud service providers (CSP's) data center or client-side artifacts [4].

So the collection and preservation are done in two ways:

a. User or customer's side data collection and preservation.
b. Cloud service providers (CSP'S) side data collection and preservation

User or customer's side data collection and preservation: In the forensic process when the user of customer device is identified, data of the physical memory should be collected before shut down the client device using the tools like dd-data duplication, FTK imager OSforensics, Lime, etc. We can also collect the data from the power off devices with the help of tools FTK Imager, Encase of hardware tools like tableau forensic duplicator, hard copy, etc [4].

Client-Side Data Analysis: Once the digital evidence is collected at the client-side the analysis process may begin. The investigator performs the examination and analysis of all cookies, log files, database files, registry files, prefetch files, browser history, operation system and virtual machines file, memory files, page files, library link files, and network traffic files [19]. Investigator can perform the analysis task to any operating system like Macintosh, macOS, Ubuntu, Windows, Kali Linux, and android client device [5].

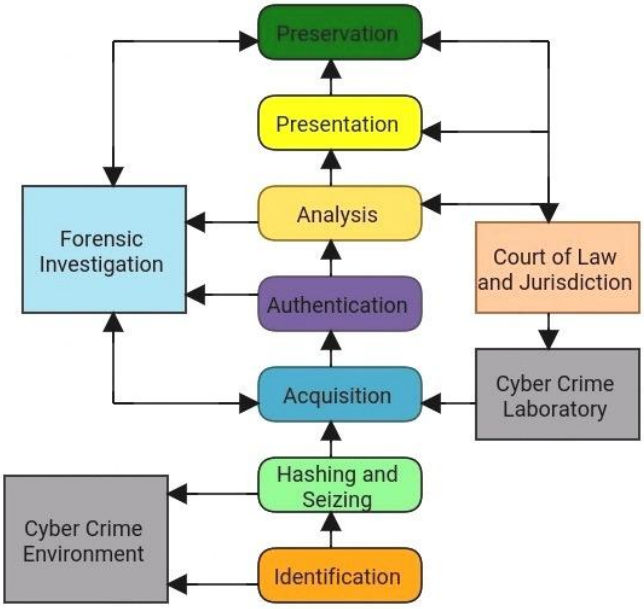

Figure 2: Digital Forensic Investigation Process and Scenario

- **Cloud service providers (CSP'S) side data collection and preservation:** It is a very difficult and challenging task for the investigator to collect, preserve the digital evidence in the cloud or virtual environment because of multiple geo-locations of cloud or virtual machine data. Cloud crime is dependent on the analysis of virtual disk and virtual machine memory, also the logs generated by the virtual machine like API logs and host logs [6]. The number of virtual machines runs under a data center may depend on the requirements and capacity of the data center. In a private cloud platform log files and data generated by virtual machine and cloud services cannot be shared with the investigator due to the privacy and security issues of multiple tenets of private cloud.

The data collection in the public deployment model is more difficult as compared to the private cloud deployment models because of multiple geo-locations of the data. If any service used by the client machine about the usage of the cloud, the forensic investigator has to understand which service was used [13]. If it is used for the storage purpose, the investigator can obtain the use of credentials from the user's machine and collect the data stored in the cloud [7]. In the other case it can be used by any virtual machine of the cloud, so in this case investigator have two choices, either download the virtual disk file or make a service request to the cloud service provider for accessing the virtual disk.

Analysis of the Digital Evidence: After successful completion of the identification, data collection and preservation of digital evidence from the cloud computing platform, evidence examination, and analysis phase will commence. Theprocess of digital evidence extracted actionable evidence from the collected data on the basis of nature of crime and inputs of law enforcement agencies [7]. In cloud side data analysis there is no specific data analysis facility available in cloud computing architecture. When the data is collected from the cloud environment, the data analysis method is based on the type of data collected. If an investigator wants to analyze the virtual machine disk data traditional digital forensic process and tools will be used [8].

- **Reporting of Digital Evidence:** The reporting of digital evidence is the process of preservation and documentation found after the analysis before sharing as evidence with the court of law to enforce the action against the cybercriminal based on country rules and regulations of the act [20].

II. THE COMPARATIVE ANALYSIS OF EXISTING WORK:

Mark Tangiwai Mathew Piwari develop a model for the evidence forensic named development and implementation of evidence collection strategy in cloud environment within cloud client and service provider using Microsoft server 2012 database and hyper-V [6]. An implementation of cloud forensic done by Katz and Ryan at LCDI in November 2013 which is based on skydrive, google drive and dropbox on some CSV file artifacts.

Ahamad Jasim and Almarzooqi develop a model which provide roadmap and proper guideline to the cloud forensic area researchers that how to build digital forensic framework and apply in cyber crime actual scene [2] [1] . The author aimed and providing a solution to the isolation and segregation of evidence in cloud environment [4] [10][2].

Magnesia Oralese and Helwing designed a novel approach for the computer based cyber crime malicious activities which manage computer security using digital forensics [3]. Gaurav Chaurasia address the issue of evidence collection in cloud environment like identify the target, jurisdiction problem, collection of evidence, chain of custody, third party issues and privacy and security of data[9]. The researcher implemented AUDIT tool which is designed to configure and integrate open source tools for investigation of disc and ranking of forensic tools (Linux based) [16]. Victor R. Kebane implement a novel cloud forensic readiness service model and researcher focus the forensic application in cloud environment and develop a model cloud forensic readiness as a service software prototype [17]. Zareefa and Mustafa discovered in investigated the evidence recovered from the (Zen Cloud Platform) using available tools [21]. Basically the research focus on the three areas like apply existing tools in cloud environment, collect artifacts and evidences from cloud and analyze the value in collected evidence. As a part future direction we can implement in near future existing tools with platform as a services(PaaS) and software as a service (SaaS) (or all service type models in one framework). Finally this research focused and recovery XCP with file system based and LVM based storage repositories (SRS) [7]. Philip and Clark contributed in their work mainly on exif metadata contained in JPEG image files. In near future all testing perform in various other file formats like pdf, word, excel, ppt and others .

Ramakrishnan Krishnan discuss about the major current trends issues of security and privacy in the cloud computing, also categories security and privacy issues in to security only issues, privacy only issues and intertwined security privacy and security issues. Mhlupheki George and Sibiya in their research describe the requirements for a cloud forensics system and what standard procedures followed during the cloud forensic process and how a cloud forensics system can be designed and also cloud forensic as a service CFAAS architecture [13] .Lucia De Marco implement in their research digital forensic readiness capability in the cloud using natural language based on service level agreement (SLA's) clauses, cloud logs and utilize tuple, set theory and functions this research is totally based on service level agreements [11]. Sheikh Kadar, Ahmed Manoj and D. Lalitha bhaskari presents secure cloud framework for the cybercrime with the help of team support of cloud user, cloud service providers, trusted third party and forensic investigators [16]. Elington Alex and R. Kishore develop a system in case of denial of service(DDoS) attack whether the forensic management plane (FMP) collect the data about the fraudulent activities for forensic analysis. In near future we can implement the whole attack scenario will be implemented in cloud platform [6][12]. Ameer Pichan, Mihai Lazarescu and Sie Teng Soh in their research provide a systematic solution of analysis of cloud forensic challenges, possible solution of each phase and summary of forensic as a service model [9]. Vassil, Irfan, Andres and Shane in their investigation applied analysis and acquisition on SaaS and test the result in their case studies: Kumodd: tools used for cloud drive acquisition, Kumoocs: a tool for Google Docs acquisition and analysis and Kumofs: a tool for remote preview and screening of cloud drive data [18].

III. EXPERIMENTAL RESULT OF OUR APPROACH:

In this experiment result, we use windows 10 operating system for the host machine and oracle VM virtual box software for creating cloud virtual environments. We have used TotalUninstall 6.23 software tool for analysis purpose. It is freeware software which we used to uninstall a computer program and analysis of data.

For detection of the virtual machine environment we have used the following process:

Step 1: Install TotalUninstal 6.23 on windows 10 host machine successfully.

Step 2: Capture the current system snapshot of the current host machine running using TotalUninstall 6.23 software.

Step 3: Install Oracle VM VirtualBox 6.0 on windows 10 host system and create virtual machines with the help of it.

Step 4: Perform the daily life activities on created virtual machines like internet access, email, office work, or playing an audio or video.

Step 5: After all these again capture the snapshot of the system using total TotalUnInstall 6.23 software.

After the whole experiment, we notice that so many changes and manipulations are done in the host operating system due to the creation of virtual environment, for the future work this experiment also can be simulated using other available virtual machine environment line Virtual PC, Quick Emulator, Microsoft Hyper-V, parallel Desktop, Citrix and VMlite workstation. With the help of these mechanisms the administrator of the virtualized data center can keep a record of the different levels of virtualizations and monitor the activities of virtual machine users.

IV. HIDDEN VIRTUAL MACHINE DETECTION:

Data hiding using stealth alternative data streams: So many sets of reserved names are used by the windows in their operating systems for its connected peripherals hardware devices output operations. These reserved names are also used by Microsoft to enable software components to communicate. These names are CON, PRN, AUX, NUL, COM1, COM2, COM3, COM4, COM5, COM6, COM7, COM8, COM9, LPT1 to LPL9.

We can create more stealth ADS files undetectable using software tools like lns.exe, stream.exe, adsy and sfind.

Stealth alternative data streams files are created by standard DOS commands by naming of alternative data streams file with one of the reserved names and they will service and remain undetectable even after using dir and dir/p command-line switches. At last, after applying all the ADS detection tools, lads.exe is the only way to detect the stealth alternative data files.

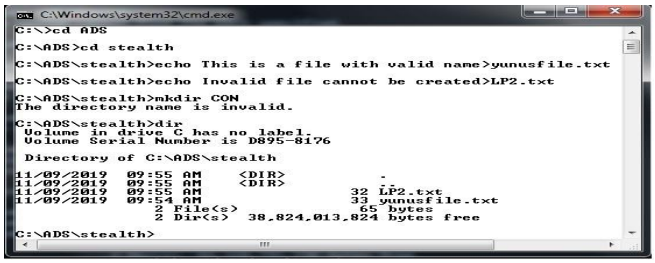

Figure 3: Error generated with reserve window names.

As clear to perform in the command prompt if we create a file or directory with any of the reserved name generated an error. When we save a file or directory in windows operating system, the operating system checks the names of the file or directory with the help of NT-style name rules which analyze the path length and names.

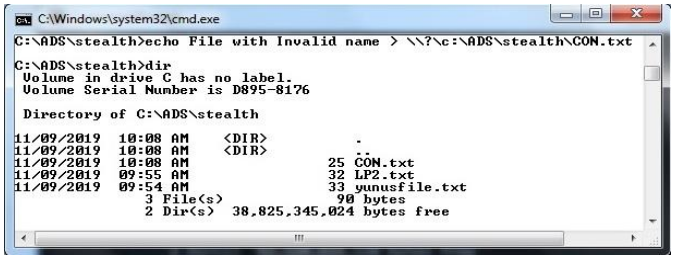

Figure 4: Reserved window name file creation

CON.txt file is created with a reserved window name and when we open this file, a message is displayed "Cannot find Con.txt file" to read this file we use the prefix (?) again.

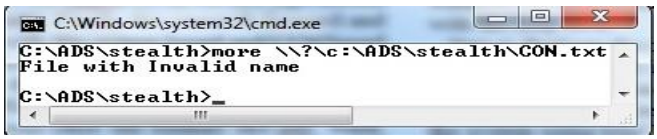

Figure 5: Reading file with reserved window names

In case of stealth alternative data streams (SADS) we combine both methods of file with reserved window device name and attach an ADS file stream on it.

This approach allows creating fully hidden ADS file which is not easily detected. Alternatives data streams(ADS) detection tools, DOS commands do not work to detect these files, even antivirus software not scan the hidden stealth ADS. Stealth alternative data streams are also used to hide malicious attacks, software, malware, and executable files. For example, in below command execution, we hide the Microsoft paint executable file (Mspaint.exe) using stealth ADS and access by the WMIC command.

Figure 6: Hiding executable file in SADS

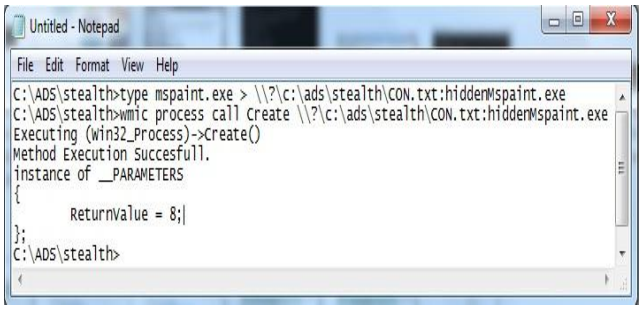

Figure 7: Access of executable file in SADS by WMIC command.

Experimental Result of Hiding and Detect a Virtual Machine Hidden using Stealth ADS: The detection of a hidden virtual machine using stealth ADS applies to only windows operating system environment. For this experiment, we create a private cloud server of the set of different operation systems virtual machines. In this experiment, we use Oracle VM VirtualBox 6.0 as virtualization software installed on Windows 10 host machine. We created four virtual machines Kali Linux, OpenSuse 13.1, Ubuntu 18.04.3 and Windows 7.

We experimented with OpenSUSE 13.1 virtual machine named Opensuse Yunus and when we start this Opensuse Yunus virtual machine it will create lots of supporting files for a specific use. The file which is necessary for performing the digital investigation is .vdi(virtual disk image) file in case of Oracle VM VirtualBox 6.0 and may be different in other virtualization software tools. The virtual disk image (.vdi) file is a system file generated by the Oracle VM VirtualBox 6.0 software during the creation of virtual machines and easily hide and see with the help of host operation system Windows 10.

Hiding and Accessing a Virtual Machine using Stealth ADS:

When we create a virtual machine of Opensuse operation system, it creates a virtual dist image file Opensuse Yunus.vdi in the following folder:

C:\Users\JIT\VirtualBox VMs\Opensuse Yunus

We are hiding Opensuse Yunus.vdi file in a temporary file vmhide.txt in following way:

C:\Users\JIT\VirtualBox VMs\Opensuse Yunus> Type Opensuse Yunus.vdi>vmhide.txt: yunusvmfile.vdi

After execution of all these commands, we use the file name vmhide.txt: yunusvmhide.vdi for the testing motive we deleted the file Opensuse Yunus.vdi from its original location and try to access the OpenSUSE virtual machine, it will display an error.

Figure 8: Error Message

The meaning of the error message is that one of the supporting files of the virtual disk is missing. Then we do the changes in the configuration file(.VBOI-PREV) to locate the path of Opensuse Yunus.vdi as shown in Figure 8 and figure 9.

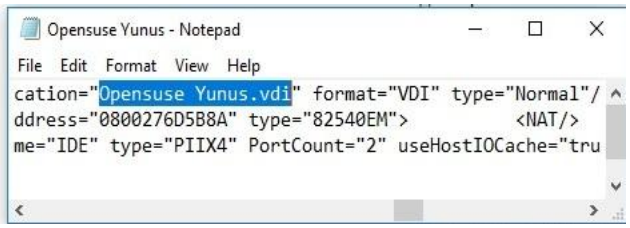

Figure 9: Virtual Machine Configuration file

It's an impossible task to retrieve data from the hidden virtual machine because the hacker will use the permanent data deletion software for the Opensue Yunus.vdi. Now we calculate the hash value for assuring the integrity of the file vmhide.txt before and after the adding of stealth alternative data streams (SADS).

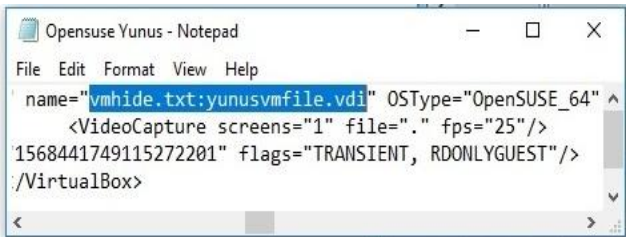

Figure 10: Modified Virtual Machine Configuration file

For this purpose, we use Hash Tool 1.2 software to calculate the hash value using the SHA-1 algorithm. The hash values before the stealth alternative data streams attachment and after are the same, so the investigator does not understand that vmhide.txt is the altered file of OpenSUSE virtual machine. Now the challenge is how to detect the presence of SADS during the investigation.

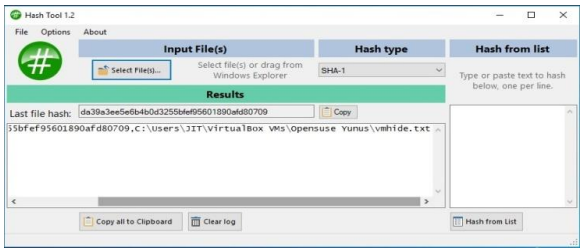

Figure 11: Hash value before the SADS attachment

Figure 12: Hash value after the SADS attachment

V. HIDDEN VIRTUAL MACHINE DETECTION METHODS:

In new technology file system (NTFS) master file table plays the role of main data structure which keeps record about the retrieving files. The first record in MFT specifies the layout of MFT, the total size of MFT and information about the current record is in use or not. The bitmap attribute in the first record shows the status of an MFT record, each bit in a sequence of the attribute represents the allocation status of MFT record. If the bit value is 1 means that this MFT record is in use. It represents that this is a normal undeleted file, and when the bit value is 0 means record is not currently used and conclude that it contains the details about the deleted file. In our research area interest is to detect the hidden virtual machines with the help of data hiding techniques like stealth alternative data streams.

If we want to detect hidden virtual machine within the NTFS file system, we have to scan all the records in MFT. If present following filters are applied to check the metadata information of the file. The four filters grantees that, the hidden file is, in fact, a virtual machine file as show in figure 13.

1. the Checking of file extensions like .vdi, .vmdk, .qcow2 and .vhd.
2. If the size limit is greater than 1 gigabyte.
3. With the help of header Signature mechanism.
4. Apply mirroring, Parity check, CRC and Magic Number algorithm.

The first and second conditions are not sufficient to check that the given file is a virtual machine file. Every file has a unique header value, so will match the header of stealth alternative data streams and header of the virtual machine hard disk file. If we don't detect the virtual machine using the first three steps we applied in our research work to apply mirroring algorithm, parity check algorithm, CRC and magic number algorithm.

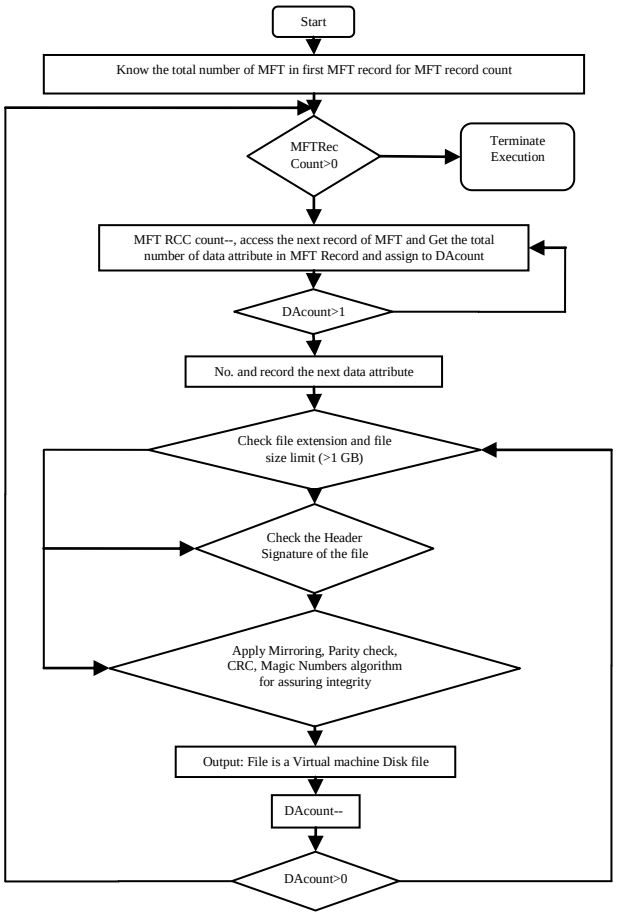

Figure 13: Hidden virtual machine detection algorithm

File Extensions	Header Signature
.vdi	5644492E(VDI)
.vhd	636F6E6563746978
.vmdk	4B444D56(KDMV)
.qcow2	514649FB(.QFI)

Table 2: File extensions and header signatures

VI. CONCLUSION

In this paper, we have discussed and experimented the ways how forensic analysis should be done in the virtual machine cloud environment. We also underline the difficulties and problems of applying digital forensic in the virtual cloud environment. We design and explain a new solution to detect and analyze the virtual cloud environment using Oracle VM VirtualBox 6.0. We also discuss and perform an experiment on how to detect the hidden virtual machine using data hiding techniques like stealth alternative data streams (SADS). Also, we presented a four-step algorithm solution to detect such hidden virtual machines. In next step we will perform a partial analysis on collected data and design and implementation of a novel digital evidence collection framework for the cloud environment. For the future work, this experiment is also performed in other virtual machine environments like virtual pc, quick emulator, Microsoft hyper-v, parallel desktop with different data detection techniques and algorithms.

REFERENCES:

1. Ahmed N. et al., "CFaaS: bilaterally agreed evidence collection", Journal of Cloud Computing: Advances, Systems and Applications, (2018) 7:1 DOI 10.1186/s13677-017-0102-3.
2. Alex and Kishor, "Forensic Model For Cloud Computing", IEEE Wispnet Conference 2016.
3. Alecsandru and Victor, "Implementation Of Cloud Computing Framework For Cloud Forensics", 978-4799-4601-3/14 IEEE 2014.
4. Ameer P. et al., "Cloud forensics: Technical challenges, solutions and comparative analysis", Digital Investigation 13 (2015) 38e57, Digital Investigation, Science Direct Elsevier.
5. Daniel S. et al. "Network forensic investigation in OpenFlow networks with ForCon", DFRWS 2017 Europe d Proceedings of the Fourth Annual DFRWS Europe, 20 (2017) S66eS74.
6. Edington M. et al., "Forensics framework for cloud computing", Computers and Electrical Engineering 60 (2017) 193–205, ScienceDirect.
7. Emi and Mehrdad, "Digital Forensic Research On Cloud Computing:An Investigation Of Cloud Forensics Solutions", 978-1-5090-0770-7/2016 IEEE.
8. Filipio Sharevski "Digital Forensic Investigation In Cloud Computing Environment: Impact On Privacy" International Conference IEEE Louissville Chapter 2013 Pp1-6.
9. Gaurav C. ,"Issues In Acquiring Digital Evidence From Cloud", Journal Of Forensic Research Chaurasia, J Forensic Res 2015.
10. Kim K. et al., "Evedence And Forensic In Cloud: Challenges And Future Research Dierctions", IEEE Cloud Computing Published By Ieee Computer Socity 2325-6095/2017.
11. Lucia De Marco,"Forensic Readiness Capability For Cloud Computing", 2015 University Degli Studi Di Salerno.
12. Manabu H. et al., "LogDrive: a proactive data collection and analysis framework for time-traveling forensic investigation in IaaS cloud environments', Journal of Cloud Computing: Advances, Systems and Applications (2018) 7:18 https://doi.org/s13677-018-0119-2, Springer Open.
13. Mhlupheki George and Sibiya,"Digital Forensic Model For A Cloud Environment", University Of Pretoria February 2015.
14. Muhammad I. et al., "A framework for cloud forensics evidence collection and analysis using security information and event management", Security and Communication Networks Security Comm. Networks 2016; 9:3790–3807 Published online 14 July 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/sec.1538.
15. Saad Said Alkahtny, Thesis On "A Forensically Enabled IaaS Cloud Computing Architecture", University Of Plymouth, Http://Hdl.Handle.Net/10026.1/9508. Jan 2017.
16. Sheik K. et al., "Cloud Forensics-A Framework For Investigating Cyber Attacks In Cloud Environment", International Conference On Computational Modeling And Security (CMS 2016), Procedia Computer Science 85 (2016) 149 – 154.
17. Sameera A. et al., "Cloud forensics:A research perspective",DOI:10.1109/Innovations.2013.6544395
18. Vassil R. et al., "Cloud forensicseTool development studies & future outlook", Digital Investigation 18 (2016) 79e95, Digital Investigation, Science Direct Elsevier.
19. Yanwei XU et al.," The Key Technology of Electronic Evidence Collection Research Based on Cloud Computing", 4th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering (ICMMCCE 2015).
20. Zachary R. et al., "Automatic Forensic Data Acquisition In The Cloud", IEEE International Conference Computer Society 2014.
21. Zareefa and Mustafa,"Assessing The Evidential Value Of Artifacts Recovered From The Cloud", Cranfield University 2016.

AUTHOR'S PROFILE

Yunus Khan, is an Ph.D scholar, in Computer Science and Engineering Department of Shri G.S. Institute of Technology and Science Indore India MP, India. He has done his Bachelor of Engineering degree in Computer Science and Engineering and Master of Engineering in Computer Science and Engineering with specialization in software engineering. He has published one book on cloud computing and many national and international research papers in various journals and conferences. His areas of interest are Cloud Computing, Computer Forensic, Big Data, Machine Learning, Soft Computing Techniques, Computer Architecture, Software Engineering and Storage Management.

Dr. Sunita Varma, is Professor and Head, in Computer Engineering Department of Shri G.S. Institute of Technology and Sciences, Indore, MP, India. She received his Bachelor of Engineering and Master of Engineering Degrees in Electrical Engineering and Computer Science and Engineering from RGPV University and Ph.D. in computer science and engineering, DAVV University Madhya Pradesh India in 2011.She is the member of institute of engineers and current research interests include ad hoc wireless networks. She has published several research papers in various journals and conferences. The areas of interest of are- Adhoc Wireless Network, Enterprise Resource Planning, Knowledge Management, E-Commerce, Cloud Computing, Digital Forensics, GreenIT and Energy Management. He has delivered various technical and motivational lectures in academic institutions and conferences.