Grenze International Journal of Engineering and Technology, June Issue

Forensic Data Storage on Cloud for Secure Access: An Overview

Preeti Dudhe[1] and S.R.Gupta[2]
[1]Assistant Professor, Department of Information technology
Prof Ram Meghe Institute of technology & Research, badnera
[2]Assistant Professor, Department of Computer science and engineering
Prof Ram Meghe Institute of technology & Research, badnera
Email: preetidudhe21@gmail.com, sunilguptacse@gmail.com

***Abstract*— The Internet's popularity has skyrocketed in recent years, and it is critical for providing a wide range of services to people worldwide. Cloud computing, which has risen dramatically in popularity to effectively provide services at a low cost, is one of the most well-known technologies. It provides consumers with on-demand services while minimizing their operating expenses by allocating virtual instances and software services. Due to the availability of incredibly powerful processing and storage at a cheap cost, an adversary or attacker may choose to launch attacks from devices inside or outside the cloud. As a result, resources are being used more frequently, and cloud services are becoming unavailable for long periods of time. This chapter delves into the forensic procedure, the challenges of cloud forensics, and the methods utilized to obtain information in a cloud scenario. The collection of evidence from a variety of sources, including web browsers, cloud storage (such as Google Drive, Dropbox, and Microsoft SkyDrive), cloud log analysis, and corporal indication recording, is critical for discovering the perpetrators of attacks.**

***Keywords*: Web browser, cloud forensics, security, and forensics.**

I. INTRODUCTION

Users today rely extensively on the internet for all of their daily activities. Because it provides services to clients, cloud computing is critical to the mainstream use of Internet technology. Cloud computing infrastructure allows providers (Amazon EC2, Google, and others) to give on-demand services to users. Cloud computing jeopardize the security of critical data and services by transferring sensitive software and records to massive server farms. As a result, the cloud is vulnerable to various attacks and security challenges. Attackers, for example, could use employee login details to get remote access to the account using cloud computing [4]. In addition to focusing on cloud infrastructure, attackers may leverage the cloud to target other configurations. For example, an opponent might conduct a distributed denial-of-service (DDoS) attack employing hundreds of virtual machine (VM) simulations. A criminal may also use cloud storage to conceal materials such as child pornography, extreme literature, etc. Investigators must undertake cloud-based forensic investigations in order to conduct forensic investigations into such cloud-based offenses. This necessitates the application of cloud forensics, a subset of system forensics. Cloud forensics uses scientific principles, methods, and procedures to reconstruct events through the identification, gathering, preservation, investigation, and reporting of crucial evidence [5]. A cloud server

Grenze ID: 01.GIJET.9.2.338_1

stores identified evidence, and an indicator can be saved anywhere in the cloud.
Because of the emergence of new knowledge, frameworks, and technologies, the detective can now obtain information from respectable third parties, particularly cloud service providers (CSP). Cloud forensics refers to a set of approaches generated from cloud service and deployment procedures. Clients who use the Infrastructure as a Service (IaaS) concept can access a virtual machine (VM) image and logs.
To isolate the issue configuration in the virtualized environment, the forensic investigator secures and isolates data from many sources, including log records, RAM pictures, and hard disc data. This signal is evaluated using the attacker's attack artefacts [6, 7]. A variety of facts is used by the forensic investigator to identify the location, cause, timing, persons involved, type of the occurrence, and other factors. This chapter examines the limitations of cloud forensics as well as the ways for acquiring cloud data.

II. Related Work

To combat cybercrime, or crime in cyberspace [1,] law enforcement officials must immerse themselves in a digital ocean of large amounts of information, and attempt to gather and objectify evidence of illicit conduct. Every important piece of electronic data is treated as evidence in criminal prosecutions and must be processed using exact scientific methodologies in order to retain its probative value. If an item is acceptable, authentic, trustworthy, and complete [2], a judge can use it to form a judgement and create a ruling without risk. The core of criminal procedures, according to article 177 of the Greek Code of Criminal Procedure (Greek Law 4620/2019), is the possession of legally admissible, incontrovertible evidence that results in fact-based verdicts. As a result, all electronic evidence must adhere to the "Rules of Evidence," a set of procedural principles and legal doctrines guiding the use of evidence in court proceedings. These guidelines govern how evidence should be presented and which sorts of evidence a judge or jury must or may not examine when making a decision [3,4]. When data is moved to a cloud-storage environment, the forensic techniques described above, which need interaction with a tangible object within reach, become obsolete, and new legal difficulties emerge. Law enforcement officials require innovative tactics for investigating online criminal conduct while taking the suspect's right to due process and privacy into account [5].

A. The Nature and Challenges of Digital Evidence

The National Standard ISO/IEC 27037:2012 [6] defines digital evidence as "information or data saved or transmitted in binary form that can be relied on as evidence and serves as an incredibly significant tool in the resolution of cybercrimes." It also provides guidance for specific activities in dealing with digital evidence (identification, gathering, acquisition, and preservation in a way that improves their evidential value). Digital evidence is both extremely weak and extremely robust. Despite the fact that their position and content can be easily changed, they can nonetheless give essential evidence in a conclusive and unarguable manner while remaining in the same condition in which they were seized [8]. Furthermore, because data-containing information systems employ integrity assurance mechanisms such as redundancy and fault tolerance, removing digital evidence often necessitates a hands-on approach to the physical media on which it is stored [9].
According to the SANS Institute for Information Security Training and Security Certification [10], there are five guidelines that must be followed when seizing digital or electronic evidence, and each guideline relates to a counterpart quality that evidence must possess in order to be recognised as valid:

- Admissibility: In order to be admissible in court, digital evidence must be collected in a legally acceptable and allowable manner.
- Authenticity: Digital evidence must be directly connected to and relevant to the incident under investigation.
- Completion: Digital evidence must offer all relevant facts about the incident under investigation and serve as both exculpatory and incriminating evidence [11].
- Reliability: Digital evidence must be collected and evaluated in such a way that its authenticity and truthfulness are validated. The appropriate technique must produce a singularity and uniqueness that distinguishes that specific piece of evidence from other digital entities that are morphologically and technologically identical [12].
- Credibility: Digital evidence must be presented in court in a comprehensible, compelling, and straightforward manner.

B. Types of forensics

- The forensic investigation starts as a post-incident activity soon after the corruption. The method used to establish the grounds for indication is preset. The five sub-categories include mobile forensics, cloud

forensics, online forensics, and cardinal forensics [13].

- In system forensics, network data is gathered and examined. It compiles data on the system ports that are utilized for data admission.
- Cloud forensics is the primary forensics claim in the cloud and is a subset of system forensics. Finding indications is more difficult because cloud organizations have internationally dispersed information.
- Indicators from mobile devices are identified using a subset of classical forensics known as mobile forensics. The call history, SMS, and memory of the mobile device are used to collect the data. [14].

C. Cloud forensic development movement

- The progress of cloud forensics is shown in Figure 1 as follows:
- Documentation: The detective assesses whether corruption took place.
- Data collection: SaaS, IaaS, and PaaS providers for three different cloud facility models are consulted by the researcher [8]. The SaaS platform shows information about each user's virtual machine by gaining access to log records.
- Inspection and examination: The forecaster assesses the information they have been given and contrasts, incorporates, and combines data to draw a valid inference. The analyst evaluates the evidence found in logical and physical documents alike [15].
- Data preservation: Unauthorized access is prevented to the data. The protection chain has been kept since log records are regionally stored.
- Reportage and presentation: A detective puts together a well-organized report of his findings in a case.

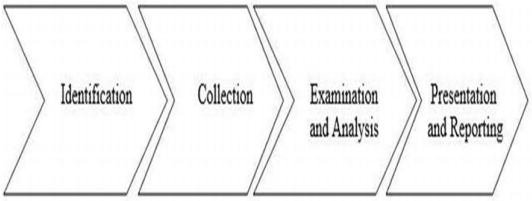

Figure 1. Cloud forensic process flow.

III. Evidence Collection

Evidence collecting is critical for forensic inquiry since it allows for identifying and accessing data from several bases in a cloud environment [15]. The data from the indicator is now diffused across a new geographic region and is no longer restricted to a single corporeal multiplicity. As a result, it might be quite challenging to locate the evidence after a crime [16]. The information is collected from a variety of bases, routers, switches, servers, virtual machines, browser articles, and internal storage devices like hard discs, RAM imageries. Additionally, data on cloud loading, Web browser artefacts, log files, and physical memory inspections are used to gather proof [17].

A. Cloud log examination

A safety measure known as logging can identify operational problems, event violations, and fraudulent activity [9, 10]. Logging's main objectives are to show the setup and look into different hostile attacks. The source of information generated at different time intervals by machines like routers, switches, servers, virtual machine representations, and other inner mechanisms like hard discs, RAM images, corporeal memory, and log records is identified using cloud log analysis [18].

- Developers hypothesize claim logs by including events in their applications. Application logs inform configuration administrators of how an application currently functions on a server.
- The system log includes details like the time and dates the log was created, the type of communication (debug, fault, etc.), emails sent by the system referencing the incident, and the procedures that were impacted by the event.
- Information on packages routed from their origin, and unsuccessful login attempts are all recorded in the firewall log.
- The network log file includes comprehensive details on a variety of network occurrences. Possible events include malicious traffic, package losses, and bandwidth interruptions. The network administrator monitors and resolves everyday operations by examining system logs for various attack attempts.

- Information about the Web pages that the server provides is kept in the Web server log file. The admissions include details about a page application's past, including the client's IP address, the application's date and time, the HTTP code used to make the request, and the total number of bytes assisted.
- The VM log stores details about the VM illustrations, including initial setup, processes, and completion. In order to assist CSP in identifying malicious activity throughout the attack, it additionally keeps track of the number of illustrations executing on the VM, the implementation time of each request, and claim relocation.

Numerous log records show that the number of cloud-based vulnerabilities or attacks has increased due to increased network traffic or new software releases in the cloud. The access log, the system log, the authentication log, and the numerous log folder trails of the Apache server are only a few of the log files that store information about application layer events. To determine whether application layer assaults have occurred, these logs are forensically examined. The various attack types and the methods applied to analyze their logs are listed in the table below. A nice example is the access log trace.

TABLE I. DIVERSE TYPES OF LOGS, OCCURRENCES, AND THE LOG EXAMINATION DEVICE.

Types of log	Occurrences	Apparatuses for log examination
DMesg log	This is not a log folder; rather, it is intended to track unusual bot behavior.	—
Debugging log	Stack tracing is used to ascertain application- and service-based occurrences' nature.	—
Firewall log	Unambiguously auditing the firewall.	Event Log Analyzer, event logging, and intensive care facilities
Arrangement log	Regulates if somebody is attempting to perform a buffer overflow or has already done so.	Syslog-ng, Log & Incident Executive
System log	Identifying web-based and distributed denial-of-service (DDoS) occurrences.	Splunk, Log4j2
VM log	Identifying occurrences against hypervisors.	Virtual Machine Log Assessor, JVM manager
Validation log	Audits credential occurrences and identifies unlawful admittance.	
Audit log	Defining who has admission to the arrangement and system without authorization. Comprises terminus discourses, usernames and passwords, and a date.	WP Safety Audit Log, auditpol.exe
Database log	Defining occurrences on databases.	Splunk, Nihuo Web Log Analyzer

B. Indication assortment from cloud storing

Through the use of a web browser and current software solutions. When folder contents are uploaded or transferred to storage media, this tool identifies unlawful cloud storage modification or access and decides whether the intruder changed the timestamp data in user financial records. By interacting with the preset value of the VM copy using a Web browser that operates in a cloud environment, the evidence is accessible for each account. System packet recorders like Wireshark, snappy, and other tools are used to capture the packets from each VM instance operating on the hosts. Through the customer access software that is installed on each device used to identify the basis of indication, the account data is synced and replicated. The indicator is separated from the virtual machine's data by using the Dropbox directory "C: Users[username] Dropbox," which is made public. The path to a file that may be obtained through a browser to evaluate the effect of a date on a disc is provided in the zip folder. The rigid drive of the virtual machine, the history of cloud-stored files, and a cache are all examined to determine whether an intruder updated the contents of a folder. Additionally, it is possible to inspect it using the VM copy's hash cost.

C. Indication assortment via a Web browser

Customers interact with the cloud server through a Web browser to complete a number of tasks, such as data retrieval, online shopping, and the inspection of emails and newsflashes [14, 15]. The history of the web browser is an essential piece of proof. Evidence is found and recovered from deleted files by examining URLs in the Web browser's history, doing a timeline analysis, monitoring user browsing behavior, and encrypting the URLs. This image shows how a Web browser's URL structure looks.

Similar to this, evidence found in the root directory of a Web application's Web browser cache is utilized to identify the attack's origin. explains how to get and retrieve information for different Web browsers.

IV. Cloud Forensics Contests

The forensic issues that arise when working in private and public cloud environments are covered in this module. The literature claims that there are more issues with public clouds than private clouds.

A. Availability of logs

Cloud computing platforms produce many levels of logs. Forensic investigators need vital logs to look into the issue, designers need them to remedy problems, and arrangement superintendents need them to troubleshoot the arrangement. Logs from all revelries, including the user, the CSP, and the forensic investigator, can be found using an admission control system.

B. Physical unapproachability

The data is spread among the hardware maneuver's geographical regions. Due to the fact that the information is spread across several CSPs and that the configured device cannot provide evidence, it is challenging to acknowledge these physical characteristics. Because the business maintains complete control over the resources, in the event of an emergency, all devices are quickly built in a separate cloud environment. The same methods won't work in a public cloud environment to make information available.

C. Instability of information

The data saved in a cloud-based virtual machine is lost when it is turned off. As a result, crucial signs like Syslog, system logs, registry admissions, and temporary Internet data are destroyed. To recover logs from finished virtual machines, it is essential to maintain the virtual machine illustration. The attacker starts a simulation of the virtual machine, triggers an event, and then ends it, making these traces useless for forensic analysis.

D. Documentation of indication at the customer side

Evidence is discovered on both the employee and customer sides. The user can communicate with another client by using a web browser. An attacker used a web browser to distribute malicious software that talks with third-party providers in order to access cloud-based services. As a result, the cloud evidence is erased. Cookies, user agents, and other data can be utilized to gather evidence; however, obtaining all of this data is difficult due to the location of the client-side VM instance.

E. Necessity of CSP conviction

Customers rely on CSPs to access documents anonymously for investigations. When CSPs fail to provide accurate information to customers that use their services, a problem occurs. Sensitive data is lost due to agreements CSPs make with other CSPs to use their facilities.

F. Multitenancy

Because numerous VMs in cloud infrastructures share the same physical infrastructure, logs are dispersed throughout the VMs. The detective must produce the logs to the court in order to prove that several service providers were acting hostilely. It also protects the privacy of other tenants.

G. Decentralization

Log data is stored on many servers throughout cloud topologies. Grouping log data from many users together or distributing it among different cloud storage tiers and levels is possible. The operating system log, record log, application log, and system log all include relevant data for forensic investigations. The cloud's scattered distribution makes it more challenging to synchronize the cloud.

H. Nonappearance of the typical format of logs

CSP provides access to logs from multiple cloud tiers via various methods. The documents describe specific events' what, when, where, and why. A general explanation that covers all CSPs and log formats has this important restriction. The procedure of acquiring evidence in cloud forensics is the focus of the poll's questions.

V. Forensic Apparatuses

There are numerous technologies available to assist investigators with the identification, gathering, and processing of forensic data. Juel et al. created the PORs device to discover safe and reliable internet archives. Dykstra et al. [6] introduced a cloud-based data-gathering tool for the supervision plane. More trust in cloud-

based infrastructures will result from this. The FTK suite of tools, including those with "Access," is used to find and gather trustworthy information. Similar to this, F-response and indication finder can be used to detect social network indicators. FROST, an open-source OpenStack cloud device for analyzing signals from a variety of sources, including computer-generated discs, API logs, and firewall logs, was introduced by Dystra et al. in their work.

VI. Open Investigate Difficulties In Cloud Forensics

Numerous scholars have suggested various strategies for addressing cloud forensics problems. Innovative tactics for replicating actual attacks in the field have been developed by a number of researchers. CSPs have not yet incorporated the suggested adjustments. Customers or investigators can access data kept in the registry, rigid drive, RAM, and log records using CSP. The CSP dependence remains a mystery despite the adoption of many forensic gathering tools. The crucial issue is how bandwidth resources are being used. When there is an excess of cloud storage, bandwidth usage increases. Provenance has not been adequately protected by keeping the chain of custody up to date. There is no one method that can be used to reconstruct cybercrime scenarios and safeguard evidence. Another serious concern is the loss of evidence brought on by improvements in forensics technology. Logging has been endorsed by several researchers as a strategy for preserving confidentiality, authenticity, and verifiability [3]. Sadly, this is not a workable cloud infrastructure as a service (IaaS) option.

VII. Research Methodology

Cloud services must now address additional issues like how to securely store and distribute digital records as more data in the cloud is converted to digital form. In order to access digital sharing records, users must expend a lot of effort querying the relevant data, but the results are not always accurate, access is occasionally limited, and the records themselves might not be secure. Managing cloud-based digital record storage and making sure that data can be safely shared among all parties involved are two of the most crucial aspects of digital forensics. In this system, a loss of confidentiality has no appreciable impact on the security of the digital record that is being saved, but a violation of integrity might have serious repercussions, including the hacking of secure data by an unidentified third party.

Therefore, it is crucial to saving digital data securely in the cloud. Modern cloud systems are infamously complex to set up and manage, prohibitively expensive, and fraught with problems relating to data privacy, security, and integrity. However, by stepping up monitoring and control of retail records, these complexity and security issues will be lessened.

Due to the decentralized and reliable nature of cryptographic cloud security, the cloud sector has a lot of potentials. The management of retail records is plagued by problems with data management and the issue of how to make data reliable, unchangeable, and accessible. One significant benefit of using Cryptographic Cloud security in the cloud database is the ability to update record interoperability systems, which enhances access to records, tracking, secure systems, user assets, etc. Access to virtual store records is crucial because cryptographic cloud security has the ability to greatly enhance the architecture of cloud services. A system must be created using cryptographic cloud security to give authentication and offer integrity to cloud records and resource management. The main goal of the proposal is to present a hybrid Cryptographic Ciphertext Policy-Attribute-based encryption method based on permissioned secure blocks, with the aim of safeguarding cloud data's security and privacy while simultaneously enabling fine-grained control over who has access to what as shown in following fig 01.

A hybrid heuristic algorithm, a permission cloud block system, and information collecting on virtual clouds and resource management are among the phases of this technique. Here, a better meta-heuristic will be developed during the encryption stages, leading to records in the cloud that are optimized all around. As a result, by applying the optimization principle and reducing the length of ciphertext, both the encryption and communication costs can be reduced. The suggested method's dependability and robustness are confirmed by comparing it to industry norms.

VII. Conclusion

With virtual examples and software tools, cloud computing offers consumers on-demand services. Data security is a big worry in the cloud due to the growth of cloud-based attacks and corruption, yet it is difficult to investigate security breaches and crimes. Because the client (end-user) is oblivious of the actions of the VM illustrations and because cloud logs are dispersed over numerous pieces of virtual and physical machinery (VM

illustrations), switches, and routers, hackers utilize these bases to deplete all cloud properties. The indication assembly is therefore quite interested in identifying the responsible people. Gathering logs from the cloud is an immensely difficult procedure because the investigator/safety analyst relies on CSPs with limited control over the infrastructure. This section examines the many forensic approaches, indication-gathering strategies for cloud forensics, and the multiple obstacles associated with forensic research in the cloud environment in order to uncover suspicious activity in the cloud.

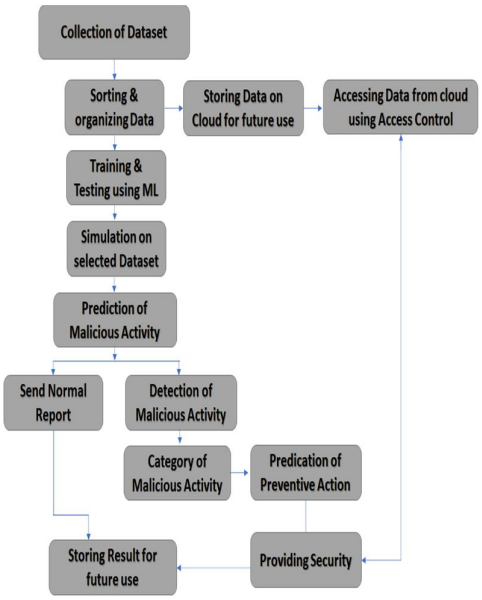

Figure 01: Proposed framework

REFERENCES

[1] Zhao, Z., Zhou, H., Li, C., Tang, J., Zeng, Q.: Deepemlan: deep embedding learning for attributed networks. Inf. Sci. 543, 382–397 (2021).

[2] Pooranian Z, Shojafar M, Garg S, Taheri R, Tafazolli R (2021) LEVER: secure Deduplicated cloud storage with EncryptedTwo-party interactions in cyber-physical systems. IEEE Transact Industrial Informatics. https://doi.org/10.1109/TII.2020.3021013

[3] Saad M, et al. (2020) Exploring the Attack Surface of Blockchain: A Comprehensive Survey, IEEE Communications Surveys & Tutorials, 22(3):1977–2008

[4] Chen, C.H., Lu, C.Y., Lin, C.B.: An intelligence approach for group stock portfolio optimization with a trading mechanism. Knowl. Inf. Syst. 62(1), 287–316 (2020)

[5] Liu Y, Yu F, Li X, Ji H, Leung VM (2020) Blockchain and machine learning for Communications and networking systems. IEEE Commun Survey Tutorials 22(2):1392–1431. https://doi.org/10.1109/COMST.2020.2975911

[6] Zhao, Z., Zhang, X., Zhou, H., Li, C., Gong, M., Wang, Y.: Hetnerec: Heterogeneous network embedding based recommendation. Knowl. Based Syst. 204, 106218 (2020)

[7] Fu X, Yu FR, Wang J, Qi Q, Liao J (2019) Resource Allocation for Blockchain-Enabled Distributed Network Function Virtualization (NFV) with Mobile Edge Cloud (MEC), IEEE INFOCOM 2019. In: IEEE conference on computer communications workshops (INFOCOM WKSHPS), Paris, France, pp 1–6

[8] Belotti M, Bozic N, Pujolle G et al (2019) A Vademecum on Blockchain Technologies: When, Which and How. IEEE Commun Surveys Tutorials 21(4):3796–3838. https://doi.org/10.1109/COMST.2019.2928178

[9] Ali M, Vecchio M, Pincheira M, Dolui K, Antonelli F, Rehman M (2019) Applications of Blockchains in the Internet of Things: A Comprehensive Survey. IEEE Commun Survey Tutorials 21(2):1676–1717

[10] Mell P, Grance T. The NIST Definition of Cloud Computing. Computer Security Division, Information Technology Laboratory, National Institute of Standards and Technology, United States Department of Commerce. Gaithersburg. 2011:1-7 DOI: Special Publication 800-145

[11] Liu F, Tong J, Mao J, Bohn R, Messina J, Badger L, et al. NIST Cloud Computing Reference Architecture. Gaithersburg: NIST Special Publication. 2011. pp. 1-28. DOI: NIST SP 500-292

[12] Pichan A, Lazarescu M, Soh ST. Cloud forensics: Technical challenges, solutions and comparative analysis. Digital Investigation. 2015;13:38-57. DOI: 10.1016/j.diin.2015.03.002
[13] Guo H, Jin B, Shang T. Forensic investigations in cloud environments. In: 2012 International Conference on Computer Science and Information Processing (CSIP); 2012 Aug 24; IEEE. 2012. pp. 248-251. DOI: 978-1-4673-1411-4/12/
[14] Zawoad S, Hasan R. Cloud forensics: A meta-study of challenges, approaches, and open problems. 2013. arXiv preprint: 1302.6312
[15] Dykstra J, Sherman AT. Acquiring forensic evidence from infrastructure-as-a-service cloud computing: Exploring and evaluating tools, trust, and techniques. Digital Investigation. 2012;9:S90-S98. DOI: 10.1016/j.diin.2012.05.001
[16] Marty R. Cloud application logging for forensics. In: Proceedings of the 2011 ACM Symposium on Applied Computing; 2011 Mar 21; ACM. 2011. pp. 178-184. DOI: 10.1145/1982185.1982226
[17] Anwar F, Anwar Z. Digital forensics for eucalyptus. In: 2011 Frontiers of Information Technology; 2011 Dec 19; IEEE. pp. 110-116. DOI: 10.1109/FIT.2011.28
[18] Khan S, Gani A, Wahab AW, Bagiwa MA, Shiraz M, Khan SU, et al. Cloud log forensics: Foundations, state of the art, and future directions. ACM Computing Surveys (CSUR). 2016;49(1):7. DOI: 10.1145/2906149