Cloud Forensic Investigation using Digital Provenance Scheme

Nay Aung Aung, and Myat Myat Min

***Abstract*—**In recent year, Cloud computing has become popular as a cost-effective and efficient computing paradigm. Many customers remain reluctant to move their business IT infrastructure completely to the Cloud. Cloud storage is increasingly being used by consumers, businesses, and government users to store growing amounts of data. One of the main concerns of customers is Cloud security and the threat of the unknown. However, criminals are embracing the opportunity to store illicit data in Cloud. There is a need for a sound digital forensic framework relating to the forensic analysis of client devices to identify potential data holdings. A key task of digital forensics is to prove the presence of a particular file in a given storage system and to track who is the past data possession. We propose the digital provenance based Cloud forensic framework in this paper. In this system, the evidences are created and collected according to the Cloud users and Service Provider actions using digital provenance scheme. The cryptographic and hash algorithms are used to be the reliable and trusted provenances at the evidence generation and verification. The aims of this are to help forensic examiners and to solve criminal cases in the Cloud environment.

***Keywords*—**Cloud computing, Cloud Storage, Digital forensic, Digital provenance, Cryptography.

I. Introduction

ALTHOUGH the cloud might appear attractive to small as well as to large companies, it does not come along without its own unique problems. Outsourcing sensitive corporate data into the cloud raises concerns regarding the privacy and security of data. Security policies, company's main pillar concerning security, cannot be easily deployed into distributed, virtualized cloud environments [6] [7]. This situation is further complicated by the unknown physical location of the company's assets. In the cloud, this is not possible anymore: The CSP obtains all the power over the environment and thus controls the sources of evidence [2]. In the best case, a trusted third party acts as a trustee and guarantees for the trustworthiness of the CSP.

The rise of Cloud computing not only exacerbates the problem of scale for digital forensic activities, but also creates a brand new front for cyber crime investigations with the associated challenges. Digital forensic practitioners must extend their expertise and tools to Cloud computing. Cloud-based entities, Cloud Service Providers (CSPs) and Cloud customers must establish forensic capabilities that can help reduce Cloud security risks. To tackle this dilemma, Cloud computing should also provide provenance [5]. The concept of provenance has been extensively studied for a long time, and widely used in the archival theory to denote the documented history of some data objects. Given its provenance, a data object can report who created and who modified its contents. Once a dispute rises in a document stored in a cloud, provenance is important for data forensics to provide digital evidences for post investigation [1]. Provenance is still an unexplored area in cloud computing, in which we need to deal with many challenging security issues [3].

II. Cloud computing

A. Cloud Computing

Cloud computing is a model for enabling ubiquitous, 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 cloud model is composed of five essential characteristics, three service models, and four deployment models [4].

B. Cloud Storage

Cloud storage is a model of networked enterprise storage where data is stored not only in the user's computer, but in virtualized pools of storage which are generally hosted by third parties. Its services may be accessed through a web service application programming interface (API), a cloud storage gateway or through a Web-based user interface [1] [6].

C. Five Essential Characteristics

1. On-demand self-service
2. Broad network access
3. Resource pooling
4. Repaid elasticity
5. Measure service

D. Service Models

In the Infrastructure as a Service (IaaS) model, the customer is using the virtual machine provided by the CSP for installing his own system on it. The system can be used like any other physical computer with a few limitations. However, the

Nay Aung Aung is with the University of Computer Studies, Mandalay, Myanmar (corresponding author's phone: +959402632243 ; e-mail: nayaungaung@gmail.com).

Myat Myat Min was with the University of Computer Studies, Mandalay; Myanmar. She is now with the Head of Department of Software Engineering, University of Computer Studies, Mandalay (e-mail: myatiimin@gmail.com).

http://dx.doi.org/10.15242/IIE.E0314136

additive customer power over the system comes along with additional security obligations. Platform as a Service (PaaS) offerings provide 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 propellent. In the Software as a Service (SaaS) model, the customer makes use of a service run by the CSP on a cloud infrastructure. In most of the cases this service can be accessed through an API for a thin client interface such as a web browser. Closed-source public SaaS offers such as Amazon S3 and Google Mail can only be used in the public deployment model leading to further issues concerning security, privacy and the gathering of suitable evidences [2].

E. Deployment Models

In the Private Cloud model, the cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them, and it may exist on or off premises.

The cloud infrastructure of Community model is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them, and it may exist on or off premises [4].

In the Public Cloud model, the cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization, or some combination of them. It exists on the premises of the cloud provider.

The Hybrid cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities, but are bound together by standardized or proprietary technology that enables data and application portability (e.g., cloud bursting for load balancing between clouds).

III. Digital Forensics

A. Digital Forensic

Digital Forensics is the science about how to obtain, preserve, analyze and document digital evidences from electronic devices such as: Tablet PCs, Servers, PDAs, fax machines, digital cameras, iPods, Smart phones (Mobile Forensics) and all of those storage devices. The purpose of this digital forensics is to improve and to acquire legal evidence found in digital media. Digital investigations are about control of forensic evidence data [9]. From the technical standpoint, this evidence data can be available in three different states:

i. at rest - represented by allocated disk space
ii. in motion - data is transferred from one entity to another
iii. in execution - loaded into memory and executed as a process

B. Cloud Forensic

Cloud forensics likes the application of computer forensic principles and procedures in a cloud computing environment. Since cloud computing is based on extensive network access, and as network forensics handles forensic investigation in private and public network, it can be defined cloud forensics as a subset of network forensics. So, Cloud forensic process can be defined as Network forensic phases [9] as shown in Fig.1

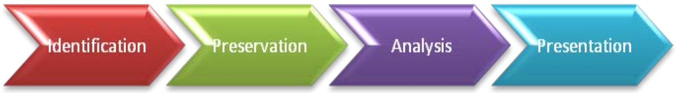

Fig. 1 Cloud forensic processing phases

In this Fig.1:

1. Identification of evidence
   - Evidence must be able to distinguish between evidence and junk data
   - We should know what the data is, where it is located, and how it stored
2. Preservation of evidence:
   - It must be preserved as close as possible to its original state
   - Any changes made during this phase must be documented and justified
3. Analysis of evidence:
   - The stored evidence must be analyzed to extract the relevant information and recreate the chain of events
4. Presentation of evidence:
   - The manner of presentation is important, and it must be understandable by a layman to be effective.

IV. Provenance

Provenance is one kind of metadata which tracks the steps by which the data was derived and can provide significant value addition in such data intensive scenarios. In other words, who owned it, what was done to it, how it transferred. It was widely used in arts, archives, and archeology. Provenance provides a record of the ownership and operations on an object throughout its existence. It can be used to verify authenticity of the object [1] [3] [5].

V. Usages of cloud forensics

Cloud Forensics has numerous uses [9], such as:

1. Investigation
   - On cloud crime and policy violations in multi-tenant and multi-jurisdictional environments
   - On suspect transactions, operations, and systems in the cloud for incident response

http://dx.doi.org/10.15242/IIE.E0314136

- Event reconstructions in the cloud
- On the acquisition and provision of admissible evidence to the court
- On collaborating with law enforcement in resource confiscation.

2. Troubleshooting

- Finding data and hosts physically and virtually in cloud environments
- Determining the root cause for both trends and isolated incidents, as well as developing new strategies that will help prevent similar events from happening in the future
- Tracing and monitoring an event, as well as assessing the current state of said event
- Resolving functional and operational issues in cloud systems
- Handling security incidents in the cloud

3. Log Monitoring

- Collection, analysis, and correlation of log entries across multiple systems hosted in the cloud, including but not limited to: audit assists, due diligence, and regulatory compliance

4. Data and System Recovery

- Recovery of data in the cloud, whether it's been accidentally or intentionally modified or deleted
- Decrypting encrypted data in the cloud if the encryption key is already lost
- Recovery and repair of systems damaged accidentally or intentionally
- Acquisition of data from cloud systems that are being redeployed, retired or in need of sanitation

5. Due Diligence/Regulatory Compliance

- Assist organizations in exercising due diligence as well as in complying with requirements related to the protection of sensitive information, maintenance of certain records needed for audit, and notification of parties concerned when confidential information is exposed or compromised.

VI. Proposed System Architecture

The proposed system architecture is as shown in Fig. 2. In this architecture involves three portions; Cloud Users (Ui), System Investigator and Cloud Service Provider (CSP). The Cloud Users access the data in the Cloud Storage via the Internet. The CSP provides the Storage services for Cloud Users to store their data, information and so on. The System investigator creates the digital provenance according to Cloud User actions such as (creating, deleting, modifying, etc.) to prove the criminal activities in Cloud.

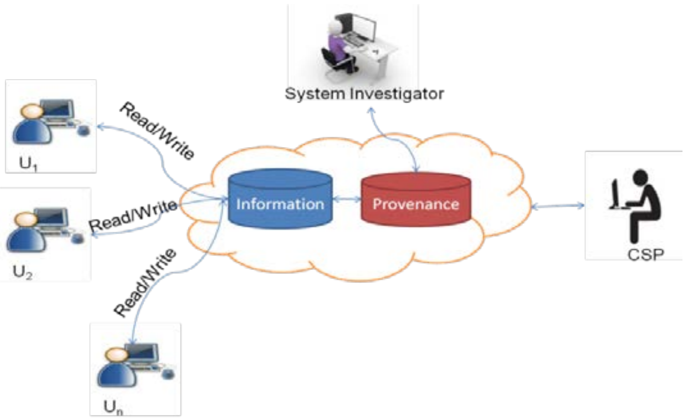

Fig. 2 Proposed system architecture

The process of digital forensic for the proposed system is as shown in Fig. 3. In this system, data and documents of Cloud Users are normally stored in Cloud storage. Sometimes, it can be dispute between Users and CSP about stored data. At that time, the System investigator can draw a conclusion about dispute by using the provenance information relating to the document and User and provenance tracking algorithm.

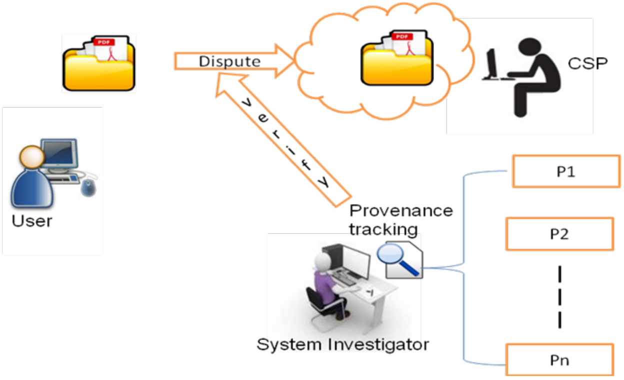

Fig. 3 Digital forensic in proposed system

The structure of provenance involves four portions; Provenance Version (P1,P2,..), Provenance Information which contains user information (User ID, User Name,..) and file information (File Name, File Type , Process Date Time and so on), Signature and Previous Provenance Version as shown in Fig. 6. The digital provenance is generated by the following steps as shown in Fig. 4:

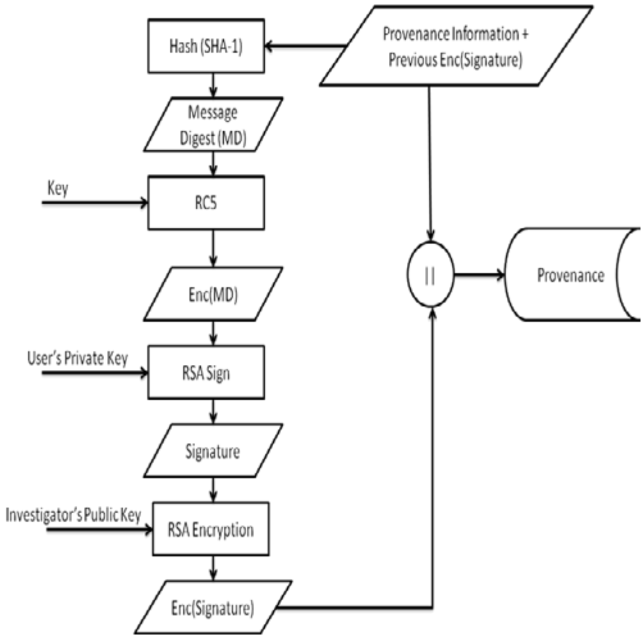

Fig. 4 Provenance generation

The verification process is as shown in Fig. 5:

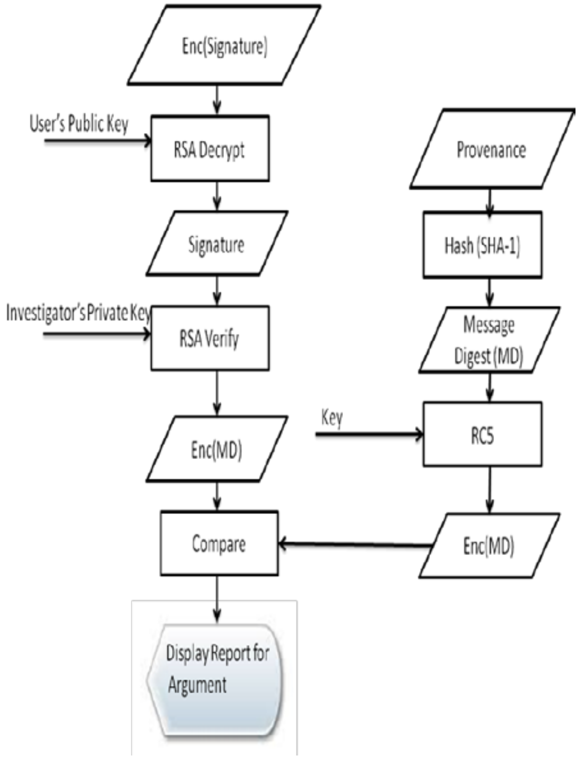

Fig. 5 Provenance verification

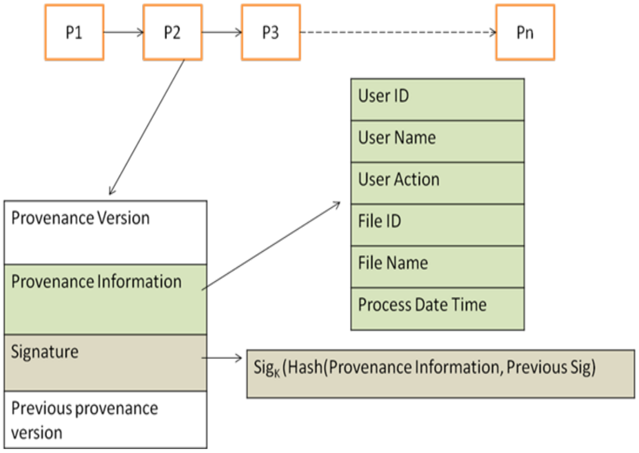

Fig. 6 Provenance structure

Cloud forensics procedures will vary according to the service and deployment model of cloud computing. For Software-as-aService (SaaS) and Platform-as-a-Service (PaaS), we have very limited control over process or network monitoring. Whereas, we can gain more control in Infrastructure-as-a-Service (IaaS) and can deploy some forensic friendly logging mechanism. In cloud infrastructure, log information is not located at any single centralized log server; rather logs are decentralized among several servers. Multiple users' log information may be co-located or spread across multiple servers. To acquire the logs, we extensively depend on the CSPs. The availability of the logs varies depending on the service model. In SaaS, customers do not get any log of their system, unless the CSP provides the logs. In PaaS, it is only possible to get the application log from the customers. To get the network log, database log, or operating system log we need to depend on the CSP. There is no standard format of logs. Logs are available in heterogeneous formats from different layers and from different service providers. Moreover, not all the logs provide crucial information for forensic purpose, e.g., who, when, where, and why some incident was executed. By using digital provenance as shown in Table 1 as forensic evidence we will solve some issues of Cloud forensic such as evidence acquisition and collection. And also this model will state the crucial information and actions of users and incidents such as who, when, why and so on.

VII. Conclusion

With the increasing use of cloud computing, there is an increasing emphasis on providing trustworthy cloud forensics schemes. Researchers have explored the challenges and proposed some solutions to mitigate the challenges. In this system, the secure digital providence based cloud forensic investigation is proposed for cloud environments. This system creates secure digital evidences according to the actions of cloud users or cloud service provider such as writing, reading, modifying or deleting data in the cloud storage using cryptographic algorithms and digital provenience scheme. The system manager (investigator) can tracks and verifies their

http://dx.doi.org/10.15242/IIE.E0314136

action using this provenience for cloud forensic. It provides trusted evidences for data forensics in cloud computing environments and also it overcomes some issues of cloud forensic investigation.

Acknowledgment

We gratefully acknowledge our rector, Dr. Mie Mie Thet Twin because of the feedback from the discussions of our research topic. So special thanks to my supervisor, Dr. Myat Myat Min, Head of Software Department, UCSM for her kind support.

References

[1] Adam Bates, Ben Mood, Masoud Valafar, and Kevin Butler, "Towards Secure Provenance-Based Access Control in Cloud Environments," Department of Computer and Information Science University of Oregon, Eugene.

[2] Doinik Birk, Ruhr-University Bochum, "Technical Issues of Forensic Investigations in Cloud Computing Environment," Ruhr-University Bochum, Horst Goertz Institute for IT Security, Bochum, Germany.

[3] Kiran-Kumar Muniswamy-Reddy, Peter Macko, and Margo Seltzer, Harvard School of Engineering and Applied Sciences, "Provenance for the Cloud," Harvard School of Engineering and Applied Sciences.

[4] P. Mell and T. Grance, "The NIST Definition of Cloud Computing,"Version15, 2009.

[5] Rongxing Lu, Xiaodong Lin, Xiaohui Liang, "Secure Provenance: The Essential of Bread and Butter of Data Forensics in Cloud Computing," ASIACCS'10 April 13–16, 2010, Beijing, China.

[6] Shams Zawoad, University of Alabama at Birmingham, "Providing Proofs of Past Data Possession in Cloud Forensics," 19, Nov, 2012.

[7] Shams Zawoad, University of Alabama at Birmingham, "Cloud Forensics: A Meta-Study of Challenges, Approaches, and Open Problems,"26, Feb, 2013.

[8] Shams Zawoad, University of Alabama at Birmingham, "Digital Forensic in the Cloud".

[9] Xath Cruz, "The Basic of Cloud Forensic".

http://dx.doi.org/10.15242/IIE.E0314136