Cloud Storage Forensics Framework

Nay AungAung, MyatMyat Min
University of Computer Studies, Mandalay
nayaungaung.ucsm@gmail.com, myatiimin@gmail.com

Abstract

There are various types of cloud services with each type having a potentially different use in criminal activity. The storage as a service (StaaS) is showing significant growth as users adopt the capability to store data in the cloud environment across a range of devices. Cloud storage forensics has recently emerged as a salient area of inquiry. One area of difficulty is the identification and acquisition of potential data when disparate services can be utilized by criminals. There is a need for a sound digital forensic framework relating to the forensic analysis of client devices to identify potential data holdings. In this paper the cloud storage forensic framework is proposed which based on open source cloud StaaS application (OwnCloud). In this paper, 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 paper are to help forensic examiners and to solve criminal cases in the Cloud environment.

Keywords*: Cloud Storage, Digital forensic, Forensic, Provenance, Cryptography*

1. Introduction

Cloud computing is a relatively recent term to describe computer resources available as a service accessible over a network, such as internally to a corporation, or externally available over the Internet. The definition introduced by the National Institute of Standards and Technology is that: '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' [7]. There are a range of different cloud services, which are grouped as; infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS) [13]. Cloud storage, or file hosting, is the storage of electronic data on remote infrastructure, rather than local storage which is attached to a computer or electronic device.

Cloud storage is a new technology that makes it possible for users to upload data to the web, allowing for instant accessibility and the ability to share data with others at any time. Cloud technology is creating a challenge for forensic investigators, as data can be uploaded or shared from one computer and opened on another computer without leaving a large amount of traceable evidence. Google Drive, Dropbox, and SkyDrive are a few examples of these cloud storage services that need to be investigated further. The various cloud storage services can be undertaken in a variety of ways; a user can install client software on a personal computer (PC), use a web browser to access the cloud storage service, or use a browser on a mobile portable device, such as a web enabled mobile phone or tablet.

Cloud storage is used by criminals to store and distribute data, such as child exploitation material, terrorism-related material, and illicit drug material. Cloud storage offers criminals an ability to avoid the scrutiny of law enforcement and national security agencies, and to provide for difficulty in attributing ownership or an association with illicit data[2]. Cloud storage can also be subject to attacks by cyber criminals, who may be able to hijack and use resources for criminal purposes, thus adding to the challenge of growing volumes of digital evidence in cases under investigation.

2. Related Work

Academic publications in the area of cloud forensics remain somewhat elusive. Many of the published papers in the area have provided a sound grounding for the research required in cloud forensics by highlighting the issues for digital forensic practitioners. DominikBirk[4] et.al proposed the technical aspects of digital forensics in distributed Cloud environments. He contributed by assessing whether it is possible for the customer of Cloud Computing services to perform a traditional digital

investigation from a technical standpoint. Furthermore he discussed possible new methodologies helping customers to perform such investigations and discuss future issues.

Secure provenance that records ownership and process history of data objects is vital to the success of data forensics in cloud computing, yet it is stilla challenging issue today. The author proposed [8] a new secure provenance scheme based on the bilinear pairing techniques. This paper provided the information confidentiality on sensitive documents stored in cloud, anonymous authentication on user access, and provenance tracking on disputed documents.

Today's cloud computing architectures often lack support for computer forensic investigations. A key task of digital forensics is to prove the presence of a particular file in a given storage system. Unfortunately, it is very hard to do so in a cloud given the black-box nature of clouds and the multi-tenant cloud models. Shams Zawoad and RagibHasan et.al introduced [9] the idea of building proofs of past data possession in the context of a cloud storage service. They presented a scheme for creating such proofs and evaluated its performance in a real cloud provider. They also discussed how this proof of past data possession can be used effectively in cloud forensics.

3. Cloud Computing

The definition proposed by the National Institute of Standards and Technology (NIST) for cloud computing encompasses convenient, on-demand network access to shared configurable computing resources that can be rapidly provisioned and released with minimal management[7]. There are a range of different cloud services, which are grouped as; infrastructure as a service (IaaS), platform as a service (PaaS), and software as a service (SaaS)[13]. Infrastructure as a Service consists of the rental and use of hardware, with the user deploying an operating system and applications. Platform as a Service consists of the use of a supplied operating system, and the user deploying the software to run on virtually provisioned systems. Software as a service consists of the use of provided software over a network connection on a subscription basis, with limited control by the NIST also describe four deployment models for cloud services; private (within an organization), community (shared by organizations with common concerns), public (available to the general public), and a hybrid (composed of a mix of private, community, or public)[7].

3.1. Cloud Storage

Cloud storage provides users with virtual storage space to host documents, pictures, music, and other files[12]. Some services also offer the ability to work with the stored data, such as; editing documents, viewing pictures, or playing music files (i.e. Google Docs or Microsoft SkyDrive). According to Chung et al. [3] of the various cloud services, consumers mostly use storage as a service, which is available to client computers and portable devices.

Cloud storage can be used by criminals to store illicit data, and provide a distribution point which distances the owner or users from the illicit data. The latter can include a wide range of material, such as child exploitation material, terrorism-related material, or illicit drug material. Cloud storage offers criminals an ability to distance themselves from the service, and hence avoid the scrutiny of law enforcement and national security agencies. This also serves to provide a difficulty in attributing ownership or an association with illicit data[2]. Cloud storage can also be subject to attacks by cyber criminals, who may be able to gain access to a victim's account for the data contained therein, or hijack an account to use the resource for criminal purposes, such as distributing illicit data, thus increasing the challenge of investigation cyber crimes or traditional crimes carried out in the cyber domain.

4. Digital Forensics

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, 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 [10]. 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

4.1. Cloud Forensics

Cloud computing introduces a number of complications to traditional digital forensic practices, as cloud servers are generally physically located ina different jurisdiction from that of the investigating law enforcement agency (LEA) and/or suspect. In addition,

there are various methods of collection suited for the different cloud computing platforms and deployment models. For example IaaS may provide an export of the virtual hard disk and memory provided to the user while SaaS may only provide a binary export of the data stored on the hosted software environment. It is, therefore, important for the LEA collecting the evidence in one jurisdiction for use in a criminal prosecution taking place in another jurisdiction to work and cooperate closely with their foreign counterparts to ensure that the methods used in the collection are in full accordance with applicable laws, legal principles and rules of evidence of the jurisdiction in which the evidence is ultimately to be used.

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 [10, 11, 12] as shown in Figure.1.

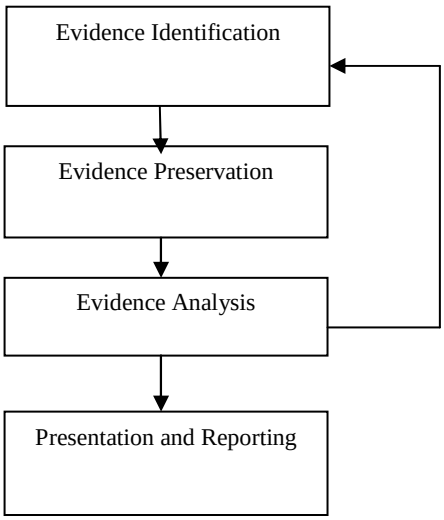

Figure 1. Cloud forensic processing phases

In this Figure.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.

5. Cryptography

The art of protecting information by transforming it (encrypting it) into an unreadable format, called cipher text. Only those who possess a secret key can decipher (or decrypt) the message into plain text. Encrypted messages can sometimes be broken by cryptanalysis, also called code breaking, although modern cryptography techniques are virtually unbreakable [14]. As the Internet and other forms of electronic communication become more prevalent, electronic security is becoming increasingly important. Cryptography is used to protect e-mail messages, credit card information, and corporate data. Cryptography systems can be broadly classified into symmetric-key systems that use a single key that both the sender and recipient have, and public-key systems that use two keys, a public key known to everyone and a private key that only the recipient of messages uses.

5.1. Hash Function

Hashing is the transformation of a string of characters into a usually shorter fixed-length value or key that represents the original string [14]. Hash functions are not reversible. Hash functions can be used to determine if two objects are equal (possibly with a fixed average number of mistakes). Other common uses of hash functions are checksums over a large amount of data (e.g., the cyclic redundancy check [CRC]) and finding an entry in a database by a key value.

5.1. 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] [5] [6].

6. Proposed SystemArchitecture

The proposed system architecture is as shown in Figure.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.

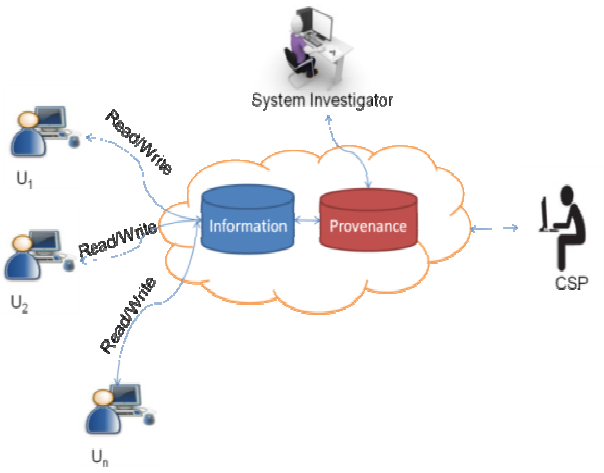

Figure 2. Proposed system architecture

The process of digital forensic for the proposed system is as shown in Figure.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.

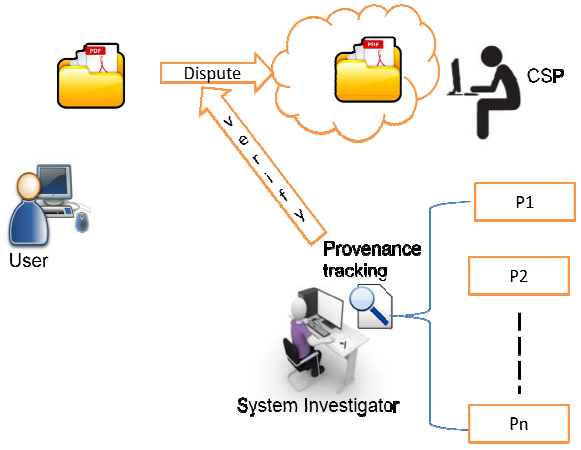

Figure 3. Forensic Investigation in proposed system

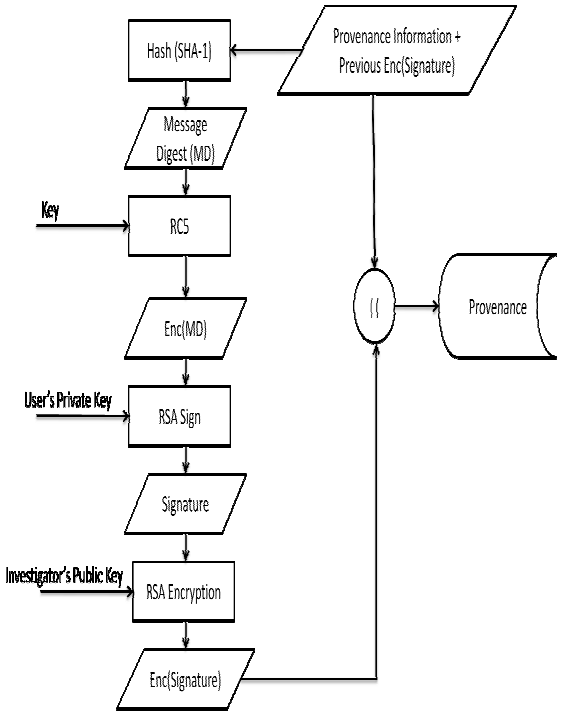

Figure 4. Provenance generation

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 Figure. 6. The digital provenance is generated by the following steps as shown in Figure.4 and the verification process is as shown in Figure.5:

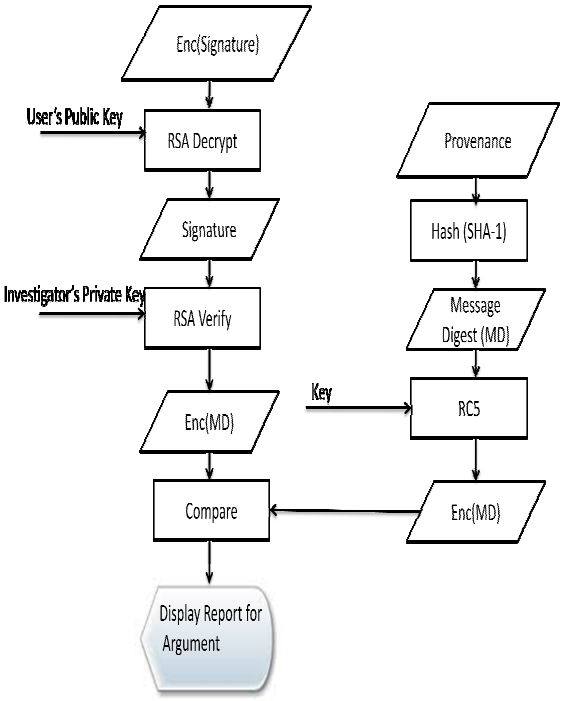

Figure 5. Provenance verification

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.

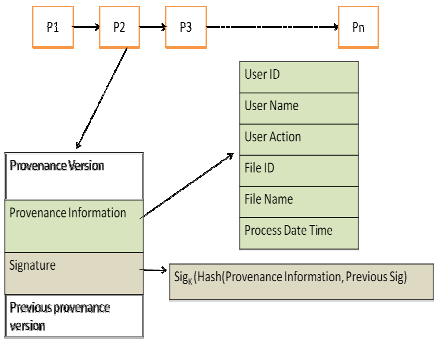

Figure 6. Provenance structure

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 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.

7. 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 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.

References

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

[2] Biggs, S &Vidalis , 'Cloud Computing: The Impact on Digital Forensic Investigations', paper presented at the IEEE International Conference for Internet Technology and Secured Transactions (ICITST 2009).

[3] Chung H, Park J, Lee S, Kang C., "Digital forensic investigation of cloud storage services. Digital Investigation", 2012;9(2):81–95

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

[5] 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.

[6] Mell, P &Grance, "The Nist Definition of Cloud Computing: Recommendations of the National Institute", NIST Special Publication 800-145, 2011.

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

[8] 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.

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

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

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

[12] Tadjer, "What Is Cloud Computing" PCMag.com, http://www.pcmag.com/article2/0,2817,2372163,00.asp

[13] Taylor, M, Haggerty, J, Gresty, D & Lamb, "Forensic Investigation of Cloud Computing Systems", Network Security, vol. 2011, no. 3, pp. 4-10.

[14] http://mathworld.wolfram.com/HashFunction.html

[15] http://www.webopedia.com/TERM/C/cryptography.html