Forensics Investigation System for Cloud Storage based on Provenance Scheme

Nay Aung Aung, Myat Myat Min
University of Computer Studies, Mandalay
nayaungaung.ucsm@gmail.com, myatiimin@gmail.com

Abstract

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. This paper proposes a forensic investigation framework for cloud storage environment. In the proposed system, the evidences are created and collected according to the cloud user's actions using digital provenance scheme. However, provenance information cannot be trusted unless its integrity is assured. So, the cryptographic algorithms are used to be the reliable and trusted provenances (evidences) at preservation phase and used to verify the evidence at examination phase. By using this scheme, the proposed system is faster than traditional investigation system for the forensics purpose. It also helps forensic examiners to examine sensitive file from the cloud storages.

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

1. 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 [10 and 12]. 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 [4]. 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 [9]. 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 [2]. Provenance is still an unexplored area in cloud computing, in which we

need to deal with many challenging security issues [7].

In this system, we propose a secure digital provenance scheme based on cryptographic and digital signature technique to provide trusted evidences for data forensics in cloud storage environment. 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.

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. Dominik Birk [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 still a 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 Ragib Hasan et.al introduced [12] 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

Cloud computing global definition announced by NIST as follows:

"*Cloud computing is a model for enabling 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*" [13].

Cloud computing [6] is a collection of hardware, networks, interfaces, services and storage providing feasibility to deliver everything such as social networks (Facebook) or collaboration tools as a service over internet whenever and wherever you need on-demand.

3.1. Cloud Storage

Cloud storage provides users with virtual storage space to host documents, pictures,

music, and other files[14]. 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.

4. Digital Forensics

Digital forensics [5] is the science of obtaining, preserving, analyzing, and documenting digital evidence from electronic devices, such as tablet PC, server, digital camera, PDA, fax machine, iPod, smart phone, and various memory storage devices. Digital forensics can be performed in four distinct phases of collection, preservation, analysis, and presentation [1], illustrated in Figure 1.

Figure.1 Digital forensic processing phases

4.1. Cloud Forensics

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 and 14] as shown in Figure.2.

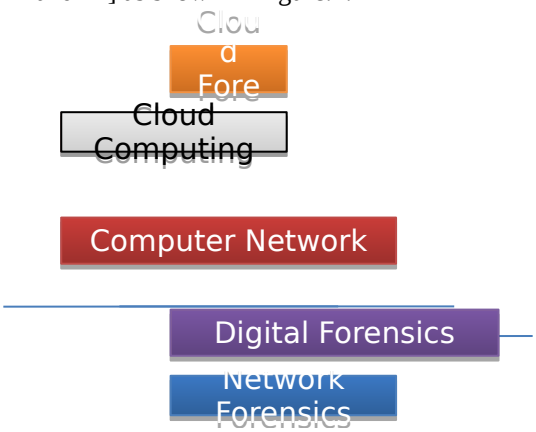

Figure.2 Cloud forensic

5. Provenance

Provenance is the chronology of the ownership, custody or location of a historical object [15]. The term was originally mostly used in relation to works of art, but is now used in similar senses in a wide range of fields, including archaeology, paleontology, archives, manuscripts, printed books, and science and computing. The primary purpose of tracing the provenance of an object or entity is normally to provide contextual and circumstantial evidence for its original production or discovery, by establishing, as far as practicable, its later history, especially the sequences of its formal ownership, custody, and places of storage. The practice has a particular value in helping authenticate objects. Provenance is particularly crucial in the cloud, because data in the cloud can be shared widely and anonymously; without provenance, data consumers have no means to verify its authenticity or identity.

6. The Proposed System Architecture

The proposed system architecture is as shown in Figure 3. 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.

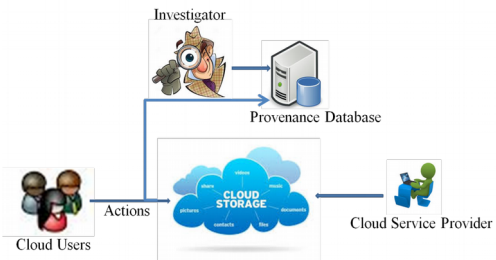

Figure.3 Proposed system architecture

The process of digital forensic for the proposed system is as shown in Figure 4. 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 verification algorithm..

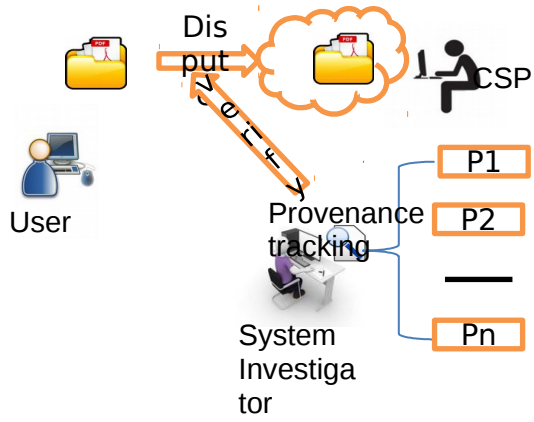

Figure.4 Forensic Investigation in proposed system

6.1. Provenance Structure

The structure of provenance involves four portions; Provenance Version ($P_{1,P2,..}$), 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 5.

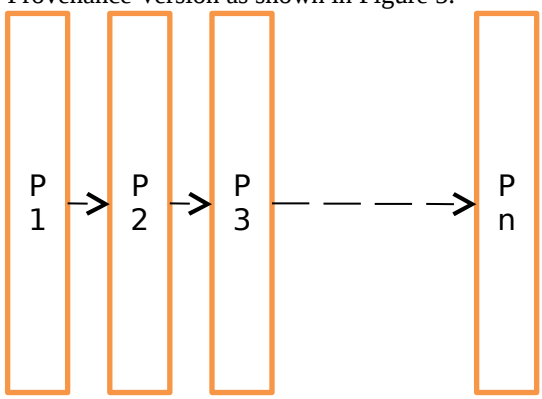

Figure.5 Provenance structure

6.2. Provenance Creation

In the provenance creation, firstly the provenance information (P_{info}) and the previous

provenance signature (Sig_{n-1}) are hashed by using SHA1 hash function which produces message digest (MD).

$$MD=H(P_{info} \| Sig_{n-1})$$

The digital signature (Sig_n) is produce by singing the MD with RSA digital signature algorithm and user's private key (PK_{user}).

$$Sig_n=RSA_{sign}(MD, PK_{user})$$

Finally the provenance is generated by concatenation of Sign, P_{info} and P_{n-1}.

$$P_n=(Sign \| P_{info} \| P_{n-1})$$

The overall flow diagram of provenance creation is as show in Figure 6.

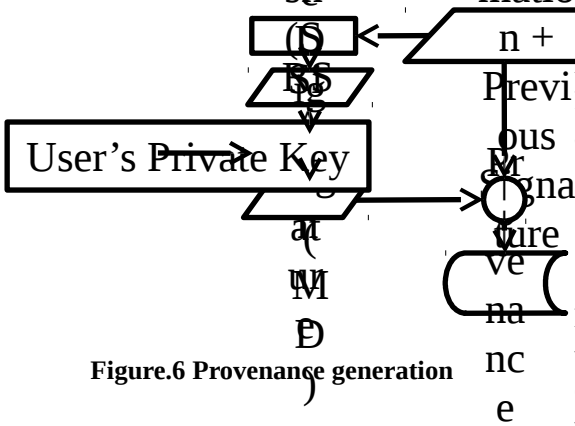

Figure.6 Provenance generation

6.3. Provenance Verification

The provenance verification process is as shown in Figure 7. In this process, firstly the provenance information (P_{info}) and the previous provenance signature (Sig_{n-1}) are hashed by using SHA1 hash function which produces message digest (MD).

$$MD=H(P_{info} \| Sig_{n-1})$$

The provenance is verified by using RSA digital signature algorithm and user public key PU_{user}.

$$Verify=RSA_{sig}(Sig_n, MD, PU_{user})$$

The report for argument between cloud user and cloud service provide is produced according to verification result.

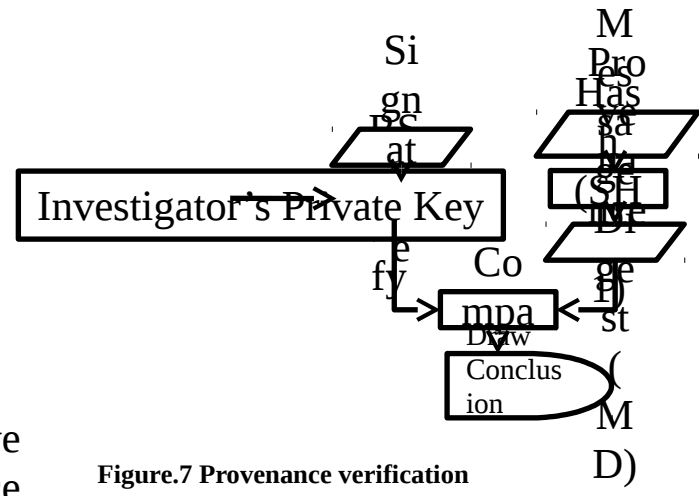

Figure.7 Provenance verification

7. The Role of the Proposed System

There are three main roles in the forensics investigation framework: Cloud Service Provider, Forensics Investigator and Cloud Users as shown in Figure 3.

7.1 Cloud Service Provider

The Cloud Service Provider provides storage location for cloud users to store data according their desired spaces. The Cloud Service Provider creates user account for Cloud Users who want to use cloud storage and also generates the RSA key pairs which are used for provenance creation and verification for forensics.

7.2 Forensics Investigator

The investigator consults the argument between Cloud Service Provider and Cloud User using digital provenance scheme. It provides logs for the each file stored in cloud storage by user actions such user id, process date time and process action. The evidences are created by using digital provenance according each user actions on each stored file. When argument occurs the investigator tracks the logs of user action and verifies the argument of Cloud Service Provider and cloud user by using

provenances of suspected file on storage. The investigator also draws the conclusion about the suspected file which it is user fault or CSP fault.

7.3 Cloud User

To access the cloud storage, the user must have user account that is registered by the Cloud Service Provider. The users can upload, edit and delete their files in cloud storage. The provenance is created using user's private key while the user uploaded or edited or deleted the file. The action of user about stored file can be verified by using provenance of that file.

7.4 Case Study of the Proposed System

Sometime the file may be edited or deleted from storage manually by Cloud Service Provider who is dishonest person. It can be occurred argument between Cloud Service Provider and Cloud User when the user suspects the file. At that time the suspected file is verified by the investigator in this system. But editing process of Cloud Service Provider cannot be concluded by using file's action logs. It can be verified by using digital provenance scheme. If the verifying process is fail, it can be concluded the Cloud Service Provider is edited the file by manually.

The Cloud User can delete or modify the file from storage via web application. Some time they may be forgotten their action of accessing file. It can be also argument between Cloud Service Provider and Cloud User when the user suspects the file. At that time the investigator verifies the suspected file by using digital provenance. If the verifying process is success, it can be concluded the user fault.

7. Experimental Results

The processing time is very important in digital environments because it should be faster than traditional system. At the logs based forensics investigation system, the processing time is so long to verify the forensics cases. In this section, we analyze the processing time of provenance creation and verification.

The total space of the proposed system is about 400GB. It allows storage space of 100MB for each user. So there are about 4000 Cloud Users can be accessed in this system. In this analysis, it includes two parts: provenance creation and provenance verification. The provenance creation contains three parts: file uploading, modification and deleting.

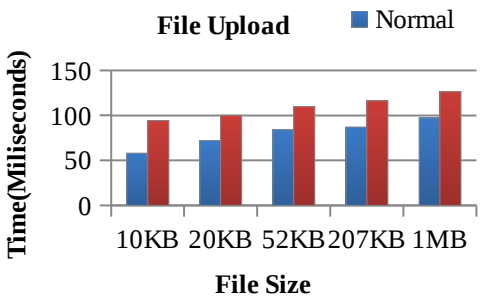

Figure.8 the Provenance Creation Time for File Uploading

The processing time of provenance creation for file uploading is as shown Figure 8. In this figure, we show the tested provenance creation time of different file sizes for file uploading. The processing time of provenance creation is about 95 milliseconds for 10KB and the 1MB file is about 126 milliseconds. According to the tested results, the processing time is suitable for file uploading like traditional file uploading system.

Figure 9 shows the provenance creation time for file modification. The provenance creation

time is about 35 milliseconds for 10KB and about 99 milliseconds for 1MB.

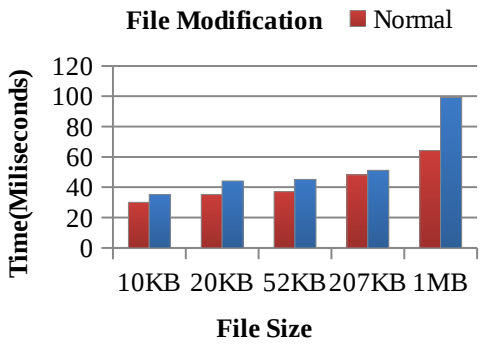

Figure.9 the Provenance Creation Time for File Modification

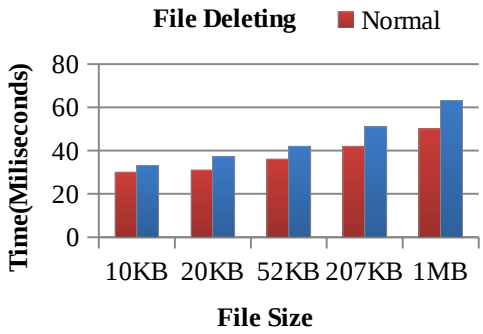

Figure.10 the Provenance Creation Time for File Deleting

The processing time of provenance creation for file deleting is as shown Figure 10. The processing time is 10KB for 33 milliseconds and 1MB for 63 milliseconds.

The Figure 11 shows the provenance creation time for all action on different file size. According to the results, the provenance time of file uploading is twice as much as the time of file modification or file deleting.

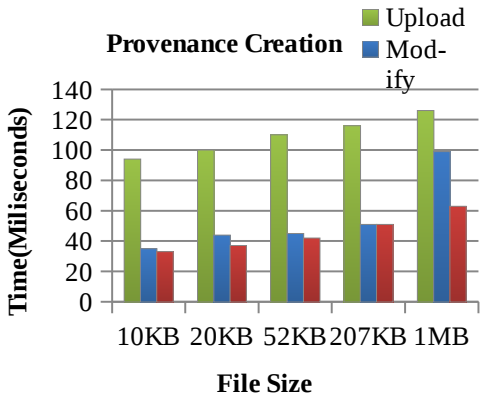

Figure.11 the Provenance Creation Time for all User Action

In the provenance verification, we tasted the verification time on different number of provenance on same file size as shown in Figure 12. According to the experiment, the verification time for 1P that means it has one action is about 103 milliseconds and the time for 112P that means it has 112 action is about 35500 milliseconds. So, it is very faster than manual forensics investigation system. Moreover the provenances are collected in accordance with user id and file id. Therefore, we can find quickly the provenance of suspected file in this system.

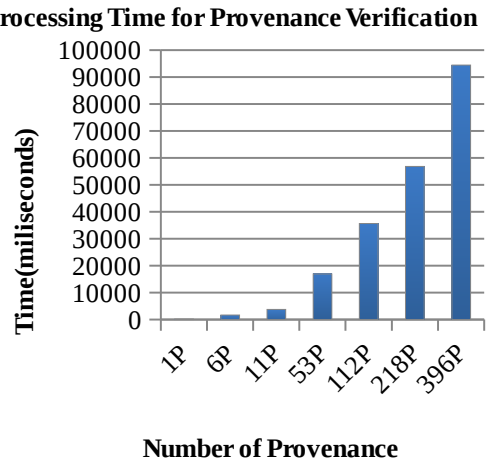

Figure.12 the Provenance Verification Time

9. Conclusion

In conclusion, there are two main part of proposed cloud forensics investigation framework such as provenance creation and provenance verification. There are three states such as file creation (uploading), file modification and file deleting in the provenance creation. The provenances are created based on clod user actions using RSA signing and SHA1 hashing algorithm. At the creation, we analyze the time for all user actions based on different file size. We compare with the time of normal storage system. According to the results, it is sufficient time for file creation, modification and deleting as normal system.

The investigator verifies the suspected file of cloud user in the provenance verification using provenances of that file. The investigator draws the conclusion about file that is the user fault if the verification process is success or that is the Cloud Service Provider fault if the verification process is fail. The verification time is faster than normal verification system for forensics investigation framework.

References

[1] A. Jones and C. Valli C, "Building a Digital Forensic Laboratory", Elsevier, Inc., 2009.

[2] 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.

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

[4] 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.

[5] E. Casey, "Handbook of Digital Forensics and Investigation", Academic Press, 2010. Mell, P & Grance, "The Nist Definition of Cloud Computing: Recommendations of the National Institute", NIST Special Publication 800-145, 2011.

[6] Frank Gens, "Cloud Services and Cloud Computing", 22 October, 2010.

[7] 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.

[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, "Cloud Forensics: A Meta-Study of Challenges, Approaches, and Open Problems,"26, Feb, 2013.

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

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

[12] T. Grance, P. Mell, “The NIST Definition of Cloud Computing,” National Institute of Standards and Technology, Information Technology Laboratory, Technical Report, 2009.

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

[14] Taylor, M., H Aggerty, J., G resty , D., and Lamb, D., “Forensic investigation of cloud computing systems.”, Network Security, 3 (2011), 4–10, 2011.