ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

Forensic Analysis Approach Based on Metadata and Hash Values for Digital Objects in the Cloud

Ezz El-Din Hemdan[1], Manjaiah D.H[2]

Research Scholar, Department of Computer Science, Mangalore University, Mangalore, India[1]

Professor, Department of Computer Science, Mangalore University, Mangalore, India[2]

ABSTRACT: In the recent time, cloud security has become an important research area for many researchers especially in the field of digital forensic investigation in the cloud which called Cloud Forensic where little research has done to cover this critical area. In this paper, a forensic analysis approach for digital objects such as digital photos and documents is proposed. These objects contain vital metadata that can be used to help cloud investigators to investigate crimes related to the cloud. Metadata can be used also by attackers to perform illegal activities so that there is a serious need to protect the metadata because it provides investigators with reliable information for performing forensic investigation. In this approach, the metadata is generated from these objects and also a hash algorithm is applied to generate hash values to guarantee integrity of uploaded data to cloud services like cloud storage services such as ADrive, Box, Microsoft OneDrive, Google Drive, Copy and Dropbox. The metadata and hash values are then stored in local storage for forensic investigation purpose because if there is any illegal activity done against the uploaded data from malicious users then the digital investigators will investigate this case by using these values that are stored in the local storage to check the integrity of the uploaded data to the cloud.

KEYWORDS: Cloud Computing, Digital Forensic, Forensic Analysis, Cloud Storage, Digital Objects, Hash Values and Metadata.

I. INTRODUCTION

With the appearance of cloud computing that depends on theory of distrusting systems which are disturbed around the world to provide services for companies and individuals with a method of cost per-use. This new technology changes the thinking of criminals because traditional cybercrime techniques will change to adapt with a dynamic nature of the cloud computing. In the other hand, digital investigators should change and expend their tools and techniques to deal with this new type of cybercrimes in the cloud to reconstruct the crime event that occurred. One of the important services that are providing by the cloud providers is storage services. There are many cloud storage service providers such as ADrive, Box, Microsoft OneDrive, Google Drive, Copy and Dropbox that provide storage services for companies and individual users in a cost-effective way, and in some times free. In the cloud storage, criminals can keep their secret files in cloud storage and can destroy all evidence from their local storage to remain clean. The cloud storage services are part of cloud computing services that are subject to exploit by attackers to stole or modify data in the cloud storage. Storing data in the cloud storages that are remotely distributed on cloud servers in overseas jurisdictions rather than in local machines make a new challenge for forensic practitioners and law enforcement agencies to acquire digital evidence for analysis and examination in forensically manner to be admissible in the court of law.

Many researchers worked in studying of data remnants on devices and accounts of clients in cloud storages. Quick and Choo [1]-[3] make a study on data remnants on client devices and found that there is information in cloud storage accounts (i.e. Dropbox, Microsoft SkyDrive and Google Drive) which is not available on user machine which may either accessed an account through web browser or is synchronized to an account using the client software [4]. This information includes previous and historical versions of files and information that identify the cloud storage user such as computer name, IP address, times and dates associate to modification made in his account's contents. Quick and

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

Choo also explored methods to preserve cloud stored information because there is critical information in the cloud account which may not available in user machines. This information may help investigators to collect vital evidential data to reconstruct crime event related to the cloud storage. In the cloud, malicious users can extract significant sensitive information from the metadata embedded in digital objects that are uploaded to the cloud storages then use this information for illegal purpose so that there is a serious need to study data upload to the cloud and related metadata as an important step for cloud investigators because this study will help them to understand and undertake steps to secure digital evidence in the cloud.

The rest of this paper is organized as follows: Section II provides cloud computing definition, characteristics and models. Section III introduces digital forensic definition and digital forensic investigation process while metadata and hashing for digital objects presented in Section IV. Section V describes the proposed approach while the paper conclusion is presented in Section VI.

II. CLOUD COMPUTING

A. Cloud Computing Definition

The National Institute of Standards and Technology (NIST) defined cloud computing as follows: "Cloud computing is a model for enabling ubiquitous, convenient, on demand network access to a shared pool of configurable resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [5].

B. Cloud Computing Characteristics

Cloud computing has five main characteristics that are identified from NIST's definition for the cloud computing as follows [5]:

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

C. Cloud Computing Models

Cloud computing has two types of models as follows [5]:

1. **Deployment Models**
   In deployment models there four types as follows:
   a. Private cloud. b. Public cloud. c. Community cloud. d. Hybrid cloud.

2. **Service Models.**
   In service models there are three types as follows:
   a. Software as a Service (SaaS). b. Platform as a Service (PaaS). c. Infrastructure as a Service (IaaS).

III. DIGITAL FORENSIC

A. Digital forensic definition

Digital forensic defined as: "The use of scientifically derived and proven methods toward the preservation, collection, validation, identification, analysis, interpretation, documentation, and preservation of digital evidence derived from digital sources for the purpose of facilitating or furthering the reconstruction of events found to be criminal, or helping to anticipate unauthorized actions shown to be disruptive to planned operations" [6].

B. Digital Forensic Investigation Process

Digital forensic investigation process involves several steps follows:

- **Identification**: Identification process is comprised of two main steps: identification of an incident and identification of the evidence, which will be required to prove the incident.
- **Collection**: In the collection process, an investigator collect the digital evidence from crime scene these digital evidence such as hard disks, cell phones and any related evidences.
- **Extraction**: In the extraction phase, an investigator extracts the digital evidence from different types of media e.g., hard disk, cell phone, e-mail, and many more. Additionally, he needs to preserve the integrity of the evidence.
- **Analysis**: In the analysis phase, an investigator interprets and correlates the available data to come to a conclusion, which can prove or disprove civil, administrative, or criminal allegations.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

- **Examination**: In the examination phase, an investigator extracts and inspects the data and their characteristics.
- **Report**: In this process, an investigator makes an organized report to state his findings about the case. This report should be appropriate enough to present in the jury.

IV. METADATA AND HASHING FOR DIGITAL OBJECTS

A. Metadata and Forensic Investigation

Metadata is data that describe data about any digital object such as digital photos. The metadata is use to facilitate in the discovery of relevant information and help to classify and organize electronic resources by provide digital identification, and supporting archiving and preservation of these resource. Metadata can be used in forensic investigation process where an investigator needs to reconstruct a cybercrime event to draw a conclusion about what happen in the incident. There are four important questions the forensic investigators need to answer which are **4W** questions as follows: (1).**W**ho/ **W**hat: Who /what did cause an incident to occur. (2).**W**hen: When did an incident occur. (3).**W**here: Where did an incident occur and (4).**W**hy: Why did an incident occur.

There are two types of metadata as follows [7]:

1. **Descriptive Metadata:** Descriptive metadata contains vital information about a digital object such as title, author, organization, creation date and keywords. The descriptive metadata is used for creating and managing a collection of digital objects, such as searching published papers.
2. **Structural Metadata:** this metadata describes how compound digital objects are put together and the relation between the parts. The structural metadata is helpful for presentation of the digital object and navigation through its various parts.

B. Metadata Associated with Digital Objects

There are many types of metadata that associated with digital objects such as digital images and PDF documents as follows [7]:

- **EXIF Metadata:** a standard Exchangeable Image File Format (EXIF), allowing camera manufacturers to embed and store camera and image metadata into JPEG and TIFF files such as camera make and model, camera settings, time, author, copyright and other vital information. These metadata can give the forensic investigator the ability to extract vital evidences such as when the picture was taken, who took the image and where the image was captured.
- **PDF Metadata:** PDF document metadata can be information about the PDF document and can be stored as entries in the information dictionary associated with the document. These entries include title, author, subject, keywords associated with the document, creator, creation date and time and last modification data and time.

C. Hashing and Forensic Investigation

1. Cryptographic Hashing Algorithms

Hashing is the process of generating a number from string of text which called hash value. This hash value uses for security purpose to ensure integrity of data. There are many cryptographic hashing algorithms that can be used for forensic investigation purpose as shown in table1.

Table 1: Hashing Algorithms for Forensic Investigation Purpose.

Algorithms	Length in bits
MD5	128
SHA-1	160
SHA-2	224-256-384-512
SHA-3	224-256-384-512
RIPEMD-160	160

2. Hashing Algorithms for Digital Forensic

Hashing algorithms can be used in digital forensic for many purposes as follows [8]:

- **Preservation of Evidence:** hashing algorithms used in the digital forensic for preserving digital evidence from tampering and modification by generating hash values which are unique values. If the new values

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

match the original, this can prove that the evidence has not been modified. These hash values has been safeguarded against modifications and tampering.

- **Modification and Change Detection:** hash values can be used to defense against malicious activities to configuration and installation files in systems by generate hash values for original files which called " White Lists" that periodically can rescan them to ensure no files have changed from attackers by using White Lists to make sure no files have been modified or deleted.
- **Searching:** in forensic investigation process, hash values can be used to perform searches of known file objects. Hash values generated for collection of confirmed child-pornography files. Then any suspect system could be scanned for the presence of these malicious files by calculating the hash values of each file and comparing the resulting values to the known list. If matches are found, then the files on the suspect system matching the hash values would be examined further.

V. PROPOSED APPROACH

This is an analysis approach for testing digital objects by extracting metadata and generating hash values for them before uploading to the cloud then downloading them and extracting the metadata and generating the hash values to check the integrity of uploaded data and monitor any modifications that may occur by malicious users.

A. Proposed Approach Flow Chart

The flow chart of proposed approach is shown in figure1:

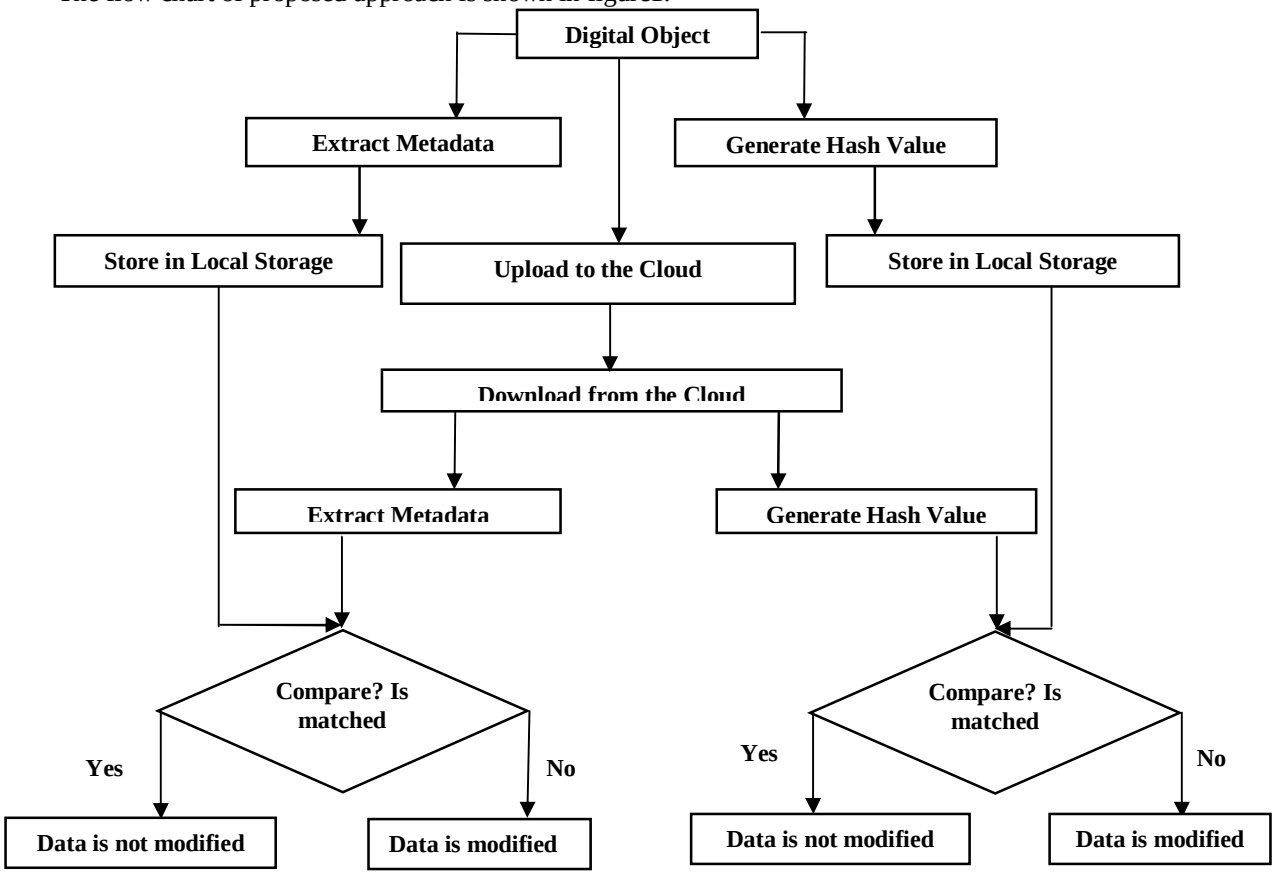

Fig.1. Proposed Approach.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

B. Proposed Approach Description

The proposed approach steps as follows:

1. Start.
2. Enter the input "digital object".
3. Extract metadata from digital object.
4. Store the extracted metadata in local storage.
5. Generate hash value for the digital object.
6. Store the hash value in local storage.
7. Upload the digital object to the cloud.
8. Download the uploaded digital object from the cloud.
9. Extract metadata from the digital object.
10. Compare the extracted metadata before uploading with the extracted metadata after downloading as follows:

    If (**metadata before uploading == metadata after downloading**) then
    (**There is no modification in the uploaded digital object**)
    Else
    (**There is modification in the uploaded digital object**)
11. Generate hash value for the digital object.
12. Compare the hash value before uploading with the hash value after downloading as follows:

    If (**Hash value before uploading == Hash value after downloading**) then
    (**There is no modification in the uploaded digital object**)
    Else
    (**There is modification in the uploaded digital object**)
13. End.

C. Results and Discussion

An experiment is carried out to evaluate the proposed approach using different digital objects such as images and PDF file (i.e. 5 Images and 5 PDF Files). These files are uploaded to the cloud and then downloaded for the experimental purpose. Some information related to metadata such as file name, file size and file type are extracted and also hash value is generated before uploading the files to the cloud. There various metadata information that can be extracted from each digital objects that depends on each type which will be useful for the investigation purpose. Table 2 shows some of extracted metadata and hash values for each file before uploading to the cloud and table 3 shows some of extracted metadata and hash values for each file after downloading from the cloud. **From the experimental test**: noticed that during the uploading and downloading of digital objects to the cloud, there is no change has occurred in the extracted metadata and hash values that means no modifications occurred to the uploaded data.

The proposed approach provide a method to check integrity of the uploaded data to the cloud by using metadata and hash values that help cloud investigators to investigate the uploaded data to the cloud side. From this study, the metadata and hash values are unique information for each digital object that help digital investigators in many purposes such as follows:

- **Improving Search Process:** hash values can be calculated for each file and compared the resulting hash values to the known list of known hash values. If matches are found, then the files on the suspect system matching the hash values would be examined further. Also, the metadata can be used for searching process to help the digital investigators.
- **Detecting Modifications in Files:** compare generated hash values and metadata for each files helps to detect any changes in files in local system or in the cloud.
- **Insurance of Investigation:** for example, metadata of digital photos can be used to determine and identify a location of the suspect who has stolen digital camera.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Special Issue 7, October 2015

Table 2: Some of extracted metadata and hash values for files before uploading to the cloud.

Metadata			Hash Value
File Name	Size	File Type	
File_1.pdf	1532927	PDF Document	6C5806A9DB0F9268BCFA35AEA2DAEB9E
File_2.pdf	238664	PDF Document	271BB7384C7537756A1DAEAA09B45A93
File_3.pdf	969643	PDF Document	FE4F3FD6D105E539E08A937F2B211D03
File_4.pdf	3055285	PDF Document	DBA2BF042BE5AC0AB57E50DE86494C76
File_5.pdf	287431	PDF Document	5FE2FDDEE8754DF5C218B440D5BE5F67
Pic_1.jpg	624744	JPEG Image	871666EE99B90E51C69AF02F77F021AA
Pic_2.jpg	1224201	JPEG Image	0A23F62DC9ED694CA80E3CA97F2D8996
Pic_3.jpg	559224	JPEG Image	E59026E9440751A93C9A5A144B363083
Pic_4.jpg	446759	JPEG Image	1894912F5030242D93E45E370F5D3BD5
Pic_5.jpg	96831	JPEG Image	B0E335DE41D1CF5ADF6DFE0A8F7E3B88

Table 3: Some of extracted metadata and hash values for files after downloading from the cloud.

Metadata			Hash Value
File Name	Size	File Type	
File_1.pdf	1532927	PDF Document	6C5806A9DB0F9268BCFA35AEA2DAEB9E
File_2.pdf	238664	PDF Document	271BB7384C7537756A1DAEAA09B45A93
File_3.pdf	969643	PDF Document	FE4F3FD6D105E539E08A937F2B211D03
File_4.pdf	3055285	PDF Document	DBA2BF042BE5AC0AB57E50DE86494C76
File_5.pdf	287431	PDF Document	5FE2FDDEE8754DF5C218B440D5BE5F67
Pic_1.jpg	624744	JPEG Image	871666EE99B90E51C69AF02F77F021AA
Pic_2.jpg	1224201	JPEG Image	0A23F62DC9ED694CA80E3CA97F2D8996
Pic_3.jpg	559224	JPEG Image	E59026E9440751A93C9A5A144B363083
Pic_4.jpg	446759	JPEG Image	1894912F5030242D93E45E370F5D3BD5
Pic_5.jpg	96831	JPEG Image	B0E335DE41D1CF5ADF6DFE0A8F7E3B88

VI. CONCLUSION

In this paper, an approach for forensic analysis of digital objects such as digital photos and PDF documents is proposed. This approach can be used to investigate the digital objects that are uploaded to the cloud side. These objects contain vital metadata that can be used to help cloud investigators to investigate crimes related to the cloud by checking integrity of them after downloading from the cloud. The proposed approach provides two way of checking and insurance of integrity of the uploaded data to the cloud from tampering, modification and illegal activities through extracting metadata and generates hash value for each digital object.

REFERENCES

1. Darren Quick and Kim-Kwang Raymond Choo,"Dropbox analysis: Data remnants on user machines", Elsevier, Digital Investigation, vol.10, pages: 3–18, 2013.
2. Darren Quick and Kim-Kwang Raymond Choo," Digital droplets: Microsoft OneDrive forensic data remnants", Elsevier, Future Generation Computer Systems, vol. 29, pages: 1378–1394, 2013.
3. Darren Quick and Kim-Kwang Raymond Choo,"Google Drive: Forensic analysis of data remnants", Elsevier, Journal of Network and Computer Applications, vol.40, pages: 179–193, 2014.
4. Darren Quick and Kim-Kwang Raymond Choo," Forensic collection of cloud storage data: Does the act of collection result in changes to the data or its metadata? ", Elsevier, Digital Investigation, vol.10, pages: 266–277, 2013.
5. P. Mell and T. Grance, "The NIST definition of cloud computing" (NIST SP 800-145), National Institute of Standards and Technology, U.S. Department of Commerce, 2011.
6. DFRWS technical report, "A road map for digital forensic research", Digital Forensic Research Workshop. G. Palmer. Utica, New York, 2001.
7. Usama Salama, Vijay Varadharajan, and Michael Hitchens, "Metadata Based Forensic Analysis of Digital Information in the Web", Annual Symposium on Information Assurance & Secure Knowledge Management, June 5-6, 2012.
8. Chet Hosmer,"Python-Forensics-A-Workbench-for-Inventing-and-Sharing-Digital Forensic-Technology", Elsevier Inc, Syngress, 2014.