Chapter 3: A Secure Cloud Storage System Combining Checking Data Integrity and Digital Forensic

Cloud storage has become very important for providing storage services for individuals and companies. There are two types of services, free or paid up to use. Several storage providers are providing these services such as Box, Google Drive, Microsoft OneDrive, Dropbox, and ADrive. Criminals can use these storage services for performing illegal activities like storing their criminal data. One of the attractive domains in security is digital forensics that concern with find digital evidence about crimes. Digital investigators and examiners are suffering from lack available information about digital investigation process of cloud storage crimes due to dynamic nature of cloud infrastructure, virtualization, multi-tenancy, legal and trust issues, physical inaccessibility of digital evidence and volatile data.

This chapter provides an approach for assisting cloud storage forensics. This approach is divided into two stages. The first stage of this approach is to develop a forensic analysis approach for digital objects such as digital photos and documents in the cloud where these objects contain vital metadata that can be used to help cloud investigators to investigate cloud storage based crimes. The metadata can be used also by attackers for performing illegal activities so that there is a serious need to protect the metadata because it provides the investigators with reliable information for performing a 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 the integrity of uploaded data to cloud storage services like cloud storage services. 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.

In the second stage which based on the first stage, a forensic approach for assisting cloud storage forensics is proposed. The proposed approach depends on check integrity of
uploaded data to cloud storage then if there are any modifications (i.e. illegal activities), the investigators will start the investigation process to find any evidence related to the committed crime as well as provide a forensic analysis of data artifacts and remnants that remain in user machine when using Box cloud storage as a case study. The Box is one of the most popular cloud storage providers that enable users to store, manage and share files and folders in the cloud. There is a free service which allows 10 GB of data storage, and extra storage space can be obtained by referring or signing up new users or subscribe to the paid service of up to use.

The rest of this chapter is organized as follows: Section 3.1 provides an introduction to cloud storage forensic and related issues while the related work in this area is presented in section 3.2. Section 3.3 provides a forensic analysis approach for digital objects such as digital photos and documents in the cloud. Section 3.4 introduces an approach for assisting cloud storage forensics to examine data artifacts and remnants that remain in user machine when using Box cloud storage as a case study. Finally, the chapter summary is presented in section 3.5.

3.1 Introduction
Cloud computing is a revolutionary technology that provides individuals, organizations and countries with enormous computing, memory, network bandwidth and storage resources as services accessible via network internally or externally over the Internet. It is an innovative technology that changes people life through providing massive services on request from users with the method of pay-per-use. Nowadays, Traditional digital storage systems are shifting to cloud storage providers where providing an immense amount of storage capabilities from anywhere and at any time. Cloud storage has become very important for providing storage services for individuals, companies, and governments. And its availability is becoming a popular option for consumers and users to store data that is accessible via arranging of devices such as personal computers, tablets, and mobile phones.

According to Hooper cited in Quick et al [27], Cloud computing is subject to attacks by cybercriminals, who may be able to steal and use resources for illegal activities, thus adding to the challenge of growing volumes of digital evidence in cases under investigation. Cloud services can also be used as a launching pad for new attacks, or to store and distribute
criminal data by them, organized crime groups and politically-motivated actors to avoid the scrutiny of law enforcement and national security agencies. Use of cloud computing by attackers’ means that data of interest may be virtualized, geographically distributed and ephemeral, presenting technical and jurisdictional challenges for identification and seizure by law enforcement and national security agencies. These issues can hamper digital forensic investigators and potentially prevent law enforcement from acquiring digital evidence and forensically analyzing digital content in a timely fashion.

Also, the cloud storage becomes subject to criminals and attackers to keep their illegal data in cloud storage and can destroy all evidence from their local storage to remain clean. Cloud storage services are part of cloud computing services that are subject to exploit by attackers to change or modify data in the cloud storage. Store data in the cloud storages that are remotely distributed on cloud servers in overseas jurisdictions rather than in local machines make a new challenging for forensic practitioners and law enforcement agencies to acquire and collect digital evidence for analysis and examination them in a forensically manner to be admissible in the court of law.

Cloud storage can be used for organizations and individuals to store and access their data remotely anywhere anytime on nearly any device without any extra drain. The advantages brought by cloud storage such as scalability and accessibility are driving rapid adoption at business organizations around the world. However, the key problem of cloud data storage is security. Moreover, cloud users must be able to use the cloud storage just like the local storage, without worrying about their data.

There is a serious need for secure cloud storage because of conducting business in the cloud means that confidential files and sensitive data are exposed to new risks, as cloud-stored data resides outside of the limits of many safeguards used to protect sensitive data held on-premise. As such, enterprises must take additional measures to secure cloud storage beyond its basic protections offered by providers.

Recently, organizations recognize that it’s perilous to protect sensitive data while enabling employees to enjoy the performance and flexibility of cloud services so that Cloud storage security is becoming a top priority in organization’s IT architecture and information security plans. Cloud storage providers, individuals, and organizations share responsibility
for cloud storage security. Cloud storage providers implement various protection methods for their platforms and stored data, such as authentication, access control, and encryption. Most of the organizations enhance these methods with added security measures of their own to support cloud data protection and straiten access to critical data on the cloud storage site.

Distributed and dynamic nature of cloud computing makes the forensic investigation process more difficult work for the digital investigator to investigate a cloud-based crime, extract a digital evidence and reconstruct a cybercrime events then send a final report to a court of law as admissible proof about the committed crime that occurred in the cloud environment. To solve the sophisticated nature of it, the hashing and metadata are two important topics in digital forensics that can help in performing the investigation process. Hashing is the process of generating a number from a string of text which called hash value. This hash value uses for security purpose to ensure the integrity of data [55][56]. On the other side, Metadata is data that describe data. The metadata is used to facilitate the discovery of relevant information and help to classify and organize electronic resources by providing digital identification, and supporting archiving and preservation of this resource [56].

This chapter introduces an approach for assisting cloud storage forensics. The proposed approach extracts metadata and generates hash values. The metadata provides investigators with reliable information for performing a forensic investigation. 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 storage. Then if there are any modifications (i.e. illegal activities) the investigators will start the investigation process to find any evidence related to the committed crime. Therefore, this chapter provides highlights current related work in cloud storage forensics, proposed an approach for assisting digital forensics of cloud storage services such as Box cloud storage as a case study. And to determine data artifacts when conducting box cloud storage service through the use of client software called Box Sync and browsers such as Google Chrome.
The motivation for conducting work into cloud storage forensic can be summarized in the following three points:

- Cloud storage is increasingly being used by users, companies, and governments to store rising amounts of data, which can be accessed remotely through using various devices such as mobile phones, computers, and tablets.
- Malicious users are taking up the opportunity to store criminal data in cloud services, which contributes to difficulties in tracing and finding criminals.

The aim of this work is to determine whether there are any cloud storage data artifacts on popular client devices. When commencing this work it was determined that there was a need to have a procedure to guide the investigation process, and hence a procedure was developed and assessed using the practical research to examine the benefits of using a procedure when undertaking a forensic analysis of cloud artifacts. To achieve the aims of the research, the following research objectives are defined:

- To study and identify present related work relating to cloud storage and identified cloud storage analysis procedures.
- To develop a digital forensic analysis procedure that will digital investigators and examiners, follow a typical method when undertaking a forensic analysis of cloud storage services.
- To conduct research using one of popular cloud storage services; Box, and determine whether there are any data artifacts which assist digital forensic analysis and investigations.

The primary research question is thus defined as: “What data artifacts result from the use of Box cloud storage to identify its use”. This question leads to the following hypotheses:

**H0:** There are no data artifacts from Box cloud storage use to identify the service provider, username, or file details.

**H1:** There are artifacts from Box cloud storage use which enable the identification of the service, a username, or file details.
These lead to the following sub-questions;

- What data artifacts on machine hard drive after Box cloud storage client software and web browser are installed and used to upload and store data with the cloud storage provider.
- What data is observed in network traffic and system memory when client software or browser access is undertaken.

### 3.2 State-of-the-Art in Cloud storage Forensics

The digital forensic investigation is the process of collecting digital evidence related to cybercrime events to present in the court of law as admissible proof. Cloud storage forensics become one of the important topics in the digital forensic domain for law enforcement and forensic practitioners. Forensic tools and techniques to collect digital evidence from cloud storage accounts are little available to cover this area. There is some research done in this area to explain how to extract and collect the digital evidence related to crimes in the cloud especially in cloud storage services such as Dropbox, Microsoft OneDrive, and Google Drive. Thus, several researchers worked in studying on artifacts and data remnants on devices and accounts of clients in cloud storages.

Quick and Choo [27-30] introduced a study about 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. 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 related to the modification made in his/her account’s contents. Quick and Choo also explored methods to preserve cloud-stored information because there is critical information in the cloud account which may not available in user machine. This information may help investigators to collect vital evidential data to reconstruct crime event related to cloud storage.

Da-Yu Kao [57], introduced a novel cybercrime investigation countermeasure using a Created-Accessed-Modified (CAM) model to improve the effectiveness of forensic analysis. CorradoFederici et al. [58], described the concepts and internals of the Cloud Data Imager
Library, a mediation layer that offers a read-only access to files and metadata of selected remote folders and currently supports access to Dropbox, Google Drive and Microsoft SkyDrive (i.e. OneDrive) storage facilities. They built an application, called Cloud Data Imager, which leverages the library to securely browse a remote account and make a logical copy of all retrievable objects and their metadata in a raw NTFS volume transferred to an expert witness container. Mohammad Shariati et al [59] prepared a study on SugarSync as a case study, where their research was undertaken to determine the types and nature of volatile and non-volatile data that can be recovered from Windows 8, Mac OS X 10.9, Android 4 and iOS 7 devices when a user has carried out different activities such as upload and download of files and folders.

The existing literature has a paramount contribution in providing basic knowledge, new insight, vision, and ideas how to study cloud storage services and related forensic study. Still, there are gaps in their works or literature, for instance, lack of exploration and check-up of the integrity of stored data in the cloud storage side, before and after data uploaded, whether the existing stored data are changed, modified, or hacked. The process of checking for data integrity in ahead is like an alarm for users and cloud storage providers. Such kind of preventive check-up helps the user or providers to identify the problem and in turn helps them to take automatic corrective action by applying the forensic investigation process.

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

In this part, a forensic analysis approach for digital objects such as digital photos and documents in the cloud where these objects contain vital metadata that can be used to help cloud investigators to investigate the cloud-based crime.

3.3.1 Metadata and Hashing for Digital Objects

3.3.1.1 Metadata and Forensic Investigation

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

There are two types of metadata which are descriptive and structural as follows [56]:

- **Descriptive Metadata**: Descriptive metadata contains vital information about a digital object such as author, title, organization, and keywords. The descriptive metadata is used for producing and managing a group of digital objects, such as searching published chapters.
- **Structural Metadata**: This metadata describes how compound digital objects are put together and the relation between the parts. The structural metadata is used for the digital object presentation and navigation through its several parts.

### 3.3.1.2 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 [56]:

- **EXIF Metadata**: The 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 model, camera settings, time and other vital information. These metadata can give the digital investigator the ability to collect vital evidence 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 such as title, author, keywords associated with the document, subject, creation date and time and last modification date and time.

### 3.3.1.3 Hashing and Forensic Investigation

Hashing is the process of generating a number from a string of text which called hash value. This hash value uses for security purpose to ensure the integrity of data. There are many cryptographic hashing algorithms that can be used for forensic investigation purpose as shown in Table 3.1.
Table 3.1: Hashing Algorithms for Forensic Investigation Purpose.

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Length in bits</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD5</td>
<td>128</td>
</tr>
<tr>
<td>SHA-1</td>
<td>160</td>
</tr>
<tr>
<td>SHA-2</td>
<td>224-256-384-512</td>
</tr>
<tr>
<td>SHA-3</td>
<td>224-256-384-512</td>
</tr>
<tr>
<td>RIPEMD-160</td>
<td>160</td>
</tr>
</tbody>
</table>

Hashing algorithms can be used in digital forensics for many purposes as follows [55]:

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

- **Modification and Change Detection:** Hash values can be used to defend against malicious activities to configuration and installation files in systems by generating hash values for original files which called “White Lists” that periodically can rescanning 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 the forensic investigation process, hash values can be used to perform searches of known file objects. Hash values generated for a 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.

### 3.3.2 First Proposed Approach

This is an analytical 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 to malicious users.
3.3.2.1 Proposed Approach Flowchart

The flowchart of the proposed approach is shown in Figure 3.1.

![Flowchart of the proposed approach]

**Figure 3.1:** First Proposed Approach.

3.3.2.2 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.

3.3.2.3 Analysis 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 depend on each type which will be useful for the investigation purpose. Table 3.1 shows some of the extracted metadata and hash values for each file before uploading to the cloud and Table 3.2 shows some of the 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 mean no modifications occurred to the uploaded data.
The proposed approach provides a method to check the 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 helps 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 a searching process to help the digital investigators.
- **Detecting Modifications in Files:** Compare generated hash values and metadata for each file helps to detect any changes in files in the 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 the digital camera.

<table>
<thead>
<tr>
<th>File Name</th>
<th>Metadata</th>
<th>Hash Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>File_1.pdf</td>
<td>Size: 1532927, File Type: PDF Document</td>
<td>6C5806A9DB0F9268BCFA35AEAA2DAEB9E</td>
</tr>
<tr>
<td>File_2.pdf</td>
<td>Size: 238664, File Type: PDF Document</td>
<td>271BB7384C7537756A1DAEAA09B45A93</td>
</tr>
<tr>
<td>File_3.pdf</td>
<td>Size: 969643, File Type: PDF Document</td>
<td>FE4F3FD6D105E539E08A937F2B211D03</td>
</tr>
<tr>
<td>File_4.pdf</td>
<td>Size: 3055285, File Type: PDF Document</td>
<td>DBA2BF042BE5AC0AB75E50DE86494C76</td>
</tr>
<tr>
<td>File_5.pdf</td>
<td>Size: 287431, File Type: PDF Document</td>
<td>5FE2FDDEE8754DF5C218B440D5BE5F67</td>
</tr>
<tr>
<td>Pic_1.jpg</td>
<td>Size: 624744, File Type: JPEG Image</td>
<td>87166EE99B90E51C69AF02F77F021AA</td>
</tr>
<tr>
<td>Pic_2.jpg</td>
<td>Size: 1224201, File Type: JPEG Image</td>
<td>0A23F62DC9ED694CA80E3CA97F2D8996</td>
</tr>
<tr>
<td>Pic_3.jpg</td>
<td>Size: 559224, File Type: JPEG Image</td>
<td>E59026E9440751A93C9A5144B363083</td>
</tr>
<tr>
<td>Pic_4.jpg</td>
<td>Size: 446759, File Type: JPEG Image</td>
<td>1894912F5030242D93E45E370F5D3BD5</td>
</tr>
<tr>
<td>Pic_5.jpg</td>
<td>Size: 96831, File Type: JPEG Image</td>
<td>B0E335DE41D1CF5ADF6DFE0A8F7E3B88</td>
</tr>
</tbody>
</table>
Table 3.3: Some of the extracted metadata and hash values for files after downloading from the cloud.

<table>
<thead>
<tr>
<th>File Name</th>
<th>Size</th>
<th>File Type</th>
<th>Hash Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>File_1.pdf</td>
<td>1532927</td>
<td>PDF Document</td>
<td>6C5806A9DB0F9268BCFA35AE2DAEB9E</td>
</tr>
<tr>
<td>File_2.pdf</td>
<td>238664</td>
<td>PDF Document</td>
<td>271BB7384C7537756A1DAEEAA09B45A93</td>
</tr>
<tr>
<td>File_3.pdf</td>
<td>969643</td>
<td>PDF Document</td>
<td>FE4F3FD6D105E539E08A937F2B211D03</td>
</tr>
<tr>
<td>File_4.pdf</td>
<td>3055285</td>
<td>PDF Document</td>
<td>DBA2BF042BE5AC0AB57E50DE86494C76</td>
</tr>
<tr>
<td>File_5.pdf</td>
<td>287431</td>
<td>PDF Document</td>
<td>5FE2FDDEE8754DF5C218B440D5BE5F67</td>
</tr>
<tr>
<td>Pic_1.jpg</td>
<td>624744</td>
<td>JPEG Image</td>
<td>871666E99B90E51C69AF02F77F021AA</td>
</tr>
<tr>
<td>Pic_2.jpg</td>
<td>1224201</td>
<td>JPEG Image</td>
<td>0A23F62DC9ED694CA80E3CA97F2D8996</td>
</tr>
<tr>
<td>Pic_3.jpg</td>
<td>559224</td>
<td>JPEG Image</td>
<td>E59026E9440751A93C9A5A144B363083</td>
</tr>
<tr>
<td>Pic_4.jpg</td>
<td>446759</td>
<td>JPEG Image</td>
<td>1894912F5030242D93E45E370F5D3BD5</td>
</tr>
<tr>
<td>Pic_5.jpg</td>
<td>96831</td>
<td>JPEG Image</td>
<td>B0E335DE41D1CF5ADF6DFE0A8F7E3B88</td>
</tr>
</tbody>
</table>

3.4 An Approach for Assisting Cloud Storage Forensics: Box as a Case-Study

In this part, a forensic approach for assisting cloud storage forensics is proposed. The proposed approach depends on check integrity of uploaded data to cloud storage then if there are any modifications (i.e. illegal activities), the investigators will start the investigation process to find any evidence related to the committed crime as well as provide a forensic analysis of data artifacts and remnants that remain in user machine when using Box cloud storage as a case study.

3.4.1 Second Proposed Approach

This section provides the idea and architecture of the proposed system. As shown in Figure 3.2, the proposed architecture consists of three entities as follows:

- **Cloud Storage Provider (CSP):** Cloud service provider (CSP) manage cloud server to provide important storage space and computation resources to anyone desiring to store data in the cloud.
- **Cloud User (CU):** The customer who has data to be stored in the cloud server.
- **Cloud Investigator (CI):** The person who has expertise and capabilities that users may not have and is trusted to assess and expose risk and handle criminal cases against cloud infrastructure on behalf of the users upon request.
Figure 3.2: Proposed System Architecture.

Briefly, the proposed system is constructed in two phases: security auditing and forensic investigation as shown in Figure 3.3. The details of these two phases will illustrate in Figure 3.4. The proposed approach is capable of providing a secure cloud storage system with support digital forensic to protect user’s data from the cloud provider and provide automatic forensic investigation process when detecting any malicious behavior on data inside cloud storage servers.

The proposed system increases the level of automatic to accomplish cloud storage forensics by using data integrity mechanism for fully protect the system from unauthorized users and conduct digital investigation process for trace malicious users in given frame time. The experimental results demonstrate the effectiveness and efficiency of the proposed system when performing forensic investigation process and determine data artifacts and remnants that remain on user machine during conducting cloud storage services such as our case study; box cloud storage.
3.4.1.1 Flowchart of Proposed Approach

The proposed approach works as a first-hand preventive check-up in the integrity of user data in cloud storage such as Box cloud storage, then, if there is any changes or modifications and tampering with the user data that stored in cloud storage provider, then the digital investigators and practitioners will start performing the digital investigation process. This approach will help to secure user’s data with saving time and cost of the whole investigation process. The proposed system is able to provide notification while some kinds of change or altering of data occurred, because of the notification, it is easy to identify the criminal activity against cloud user’s data.

The proposed approach is for testing user’s data by extracting metadata and generating hash values of them before uploading each data, and then downloading them, extracting the metadata, generating the hash values to check the integrity of uploaded data and monitoring any modifications that may be occurred by malicious users. In conclusion, the purpose of this research is to fill gaps of previous work in cloud storage forensics area that did not undertake in consideration automatic check-up of data integrity that can speed up of taking action against criminals as well as provide an efficient method for execution forensic analysis of data artifacts and remnants that remain in user machine when using cloud storage as a case study; box. The proposed approach flowchart is shown in Figure 3.4.
Figure 3.4: Flowchart of Proposed Approach.
3.4.1.2 Procedure of Forensic Investigation

The flowchart for performing digital investigation for cloud storage data is illustrated in Figure 3.5.

![Flowchart](image)

**Figure 3.5:** Proposed Digital Forensic Procedure for Cloud Storage.

3.4.2 Box as a Case Study

3.4.2.1 Box Cloud Storage Provider

Box [60] is a cloud storage service that allows users to store, manages and shares their data in the cloud with efficient and secure manner. There is a free service which permits 10 GB of data storage, and extra storage space can be obtained by referring or signing up new users or subscribe to the paid service of up to use. Users can define how their data can be shared with other users. Box offers three account types: Personal, Business, and Enterprise. Depending on the kind of account, Box has features such as custom branding, unlimited storage, and administrative controls. The box can be accessed using a using both
options the web browser (e.g. Google Chrome, Mozilla Firefox, and Internet Explorer) and client software application.

3.4.2.2 Box Forensic Investigation

3.4.2.2.1 Criminal Scenario

To understand the significance of the proposed approach, a criminal scenario for hacking box cloud storage user shown in Figure 3.6 to simulating how can attackers stole sensitive information such as username and password from the cloud storage user and then used this information to do illegal activities on the cloud user. If the investigators apply the proposed approach on suspect user machine can find valuable information related to the box cloud storage which can be as admissible proof of the cloud storage based crime.

![Figure 3.6: Criminal Scenario for Attacking Box Storage User.](image)

3.4.2.2.2 Experiment Preparation

To collect the data (i.e. digital evidence) to satisfy and achieve research objectives in relation to the use of Box cloud storage. The experiment preparation setup on The proposed approach prepared on Intel(R) Core(TM) i3 CPU with 4 GB RAM and 500G Hard Disk running Windows 7. In the experiment, we used a virtual machine with windows 7 which enable the quick setup and analyze different configurations without having to re-configure and prepare physical hard drives. This help to perform the experiment to conduct a variety of tests of forensic investigation of Box cloud storage. The experiment preparation can summarize as follows:

- Create one account to access Box cloud storage service for testing purpose.
- Install VMware® Workstation on the host machine.
- Install Virtual Machine (VM) using VMware® Workstation with Windows 7 Ultimate on 15 GB virtual hard drive with 1 GB RAM to reduce the time during the forensic analysis phase.
- Setup Web Browsers for in the VM for accessing user account such as Google Chrome.
- Download and install Box Sync software for Box cloud storage in the VM for testing purpose. This software to synchronous with the Box cloud storage account.
- Create test folder called TestData which contains 15 pdf and 9 picture files.
- Calculate MD5 hash values for TestData folder files before uploading to the Box storage services as tabulated in Table 3.4.
- Install Wireshark in host machine to monitor the network traffic by capture packets for analysis purpose.
- Install AccessData Forensic Imager (FTK) on the host machine for forensic analysis purpose.
- Install Autopsy on the host machine for forensic analysis purpose.
- The benefit of using a VM is easy of capturing memory by imaging the VMEM file as a VM was running. It was also easy to capture network traffic by running network monitoring programs like Wireshark that running on the Host machine.

Table 3.4: MD5 hash value for the uploaded data.

<table>
<thead>
<tr>
<th>File Name</th>
<th>MD5 Hash Value</th>
<th>File Name</th>
<th>MD5 Hash Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>File_1.pdf</td>
<td>6C5806A9DB0F9268BCFA35AEA2DAEB9E</td>
<td>File_13.pdf</td>
<td>B7F0E153C4C64D33764AD35759B0A35B</td>
</tr>
<tr>
<td>File_3.pdf</td>
<td>FE4F3FD6D105E539E08A937F2B211D03</td>
<td>File_15.pdf</td>
<td>86DBDF3751CA95C19ED6141E0ECEB837</td>
</tr>
<tr>
<td>File_4.pdf</td>
<td>D8A2BF042BE5AC0AB57E50DE86494C76</td>
<td>Pic_1.jpg</td>
<td>871666EE99B90E51C69AF02F77F021AA</td>
</tr>
<tr>
<td>File_5.pdf</td>
<td>5FE2FD6DE8754DF5C21B440D5BE5F67</td>
<td>Pic_2.jpg</td>
<td>0A23F62DC9ED694CA80E3CA97F2D8996</td>
</tr>
<tr>
<td>File_6.pdf</td>
<td>AE47F7CB2D279492BD8E3EEF08189A006</td>
<td>Pic_3.jpg</td>
<td>E59026E940751A93C9A514B363083</td>
</tr>
<tr>
<td>File_7.pdf</td>
<td>A33AEB323471E4FC04FD1C240748FA</td>
<td>Pic_4.jpg</td>
<td>1894912F5030242D93E45E370F5D3BD5</td>
</tr>
<tr>
<td>File_8.pdf</td>
<td>8C0676BB121334FB9D4C32561026B33</td>
<td>Pic_5.jpg</td>
<td>B0E335DE41DF5AD06D2A8F7E3B88</td>
</tr>
<tr>
<td>File_9.pdf</td>
<td>4B9091B48771D7F2242AB2027697824A</td>
<td>Pic_6.jpg</td>
<td>732A289B7DD8C7DD28D4D73ED2480BCF</td>
</tr>
<tr>
<td>File_10.pdf</td>
<td>F74C4A5D119423C99A66A461F8A2843D</td>
<td>Pic_7.jpg</td>
<td>C3681F1A5A50BDECF958B0018CA71F66</td>
</tr>
<tr>
<td>File_11.pdf</td>
<td>DAA15D753583D7737AE58CAEB7C3BBA7</td>
<td>Pic_8.jpg</td>
<td>E408440D3078D399ADD50BE999BAE99</td>
</tr>
<tr>
<td>File_12.pdf</td>
<td>A48C66508637F86E9A28CCA0B6D07186</td>
<td>Pic_9.jpg</td>
<td>40374D33463DFE213D31CCB0E1DED22</td>
</tr>
</tbody>
</table>
3.4.2.2.3 Experiment Steps

McKemmish et al [54], determined four stage of forensic investigation of digital evidence which is identification, preservation, analysis, and presentation of the digital evidence. McKemmish mentioned that there are many requirements when dealing with the digital evidence such as minimal handling of original evidence comply with the rules of the evidence and not exceed the knowledge. The forensic investigation process is performed as follows:

1. **Identification**: The process of identification of data that will contain information needed for analysis phase. There are many valuable data such as the VMDK hard drive file, the VMEM memory file, and the network capture (PCAP) file.

2. **Preservation**: The preservation of digital evidence is done by taking a copy from the original source. In this research the copy taking from identified data which are VMDK hard drive file, the VMEM memory file, and the PCAP file. A forensic copy from the VMDK file is taken by FTK imager in E01 format. MD5 cryptographic hash value is calculated for the original forensic data and verified for each copy from them. Figure 3.7 image verification for VMDK file.

![Image Verification for VMDK File](image)

**Figure 3.7**: Image Verification for VMDK File.

3. **Analysis**: For this research, each of the forensic copies of the VM hard drive (VMDK file), Memory (VMEM file) and Network Capture (PCAP file) examined using a range of forensic analysis tools including AcessData FTK Image, Autopsy, and Wireshark. These tools have been used widely for digital investigation. Figure 3.8 shows the forensic analysis process in this experiment.
4. **Presentation:** Presentation is the phase of presenting analysis findings that are collected during the analysis phase to present them in the court of law as admissible evidence about the committed crime that occurred in the crime scene. After that investigator makes an organized report to state his findings of the case. This report should be appropriate enough to present to the jury. In this research, various types of data remnants and artifacts were collected when the user uploaded, accessed, stored and downloaded his/her data from box cloud storage service.

### 3.4.2.2.4 Experimental Results and Findings

This part introduces the findings of an experiment on Box cloud storage.

- **Web browser box data:** We observed when the Box account accessed via ‘https://account.box.com/login/’ displays the login page to enter username and password. After login in the page ‘https://app.box.com/files’ appears where the user can access his/her files. You find the username in the top browser page when selecting files in a Box account using browser there an option to display the metadata of files. By clicking Right-Mouse-Button, and click properties – general info. When downloading files through the browser if it single file, it will be downloaded as ‘filename. Extension’. But if it folder will be download as ‘foldername.zip’ and finally if select all uploaded data. They will be downloaded as ‘mybox-selected.zip’. Also, whilst accessing the box account using web browsers, the account retains account information (account type, storage, bandwidth used, max. file size and cancellation) as shown in Figure 3.9. Access a box account via web browser also has an option to show deleted files from trash and the user has two options for a single file; restore or delete and for multiples files; restore all or delete all as shown in Figure 3.10.
Box Sync Software: In the VM located the “Box Sync.exe” that download from box website to the user machine in folder ‘Download’. When the executable file ‘Box Sync.exe’ runs copied to ‘c:\Program Files\Box’ folder. Created in the user's folder is a directory as follows: ‘c:\Users\[username]\AppData\Local\Box Sync’. Box synchronized files and folders observed at default box Sync folder located at ‘c:\Users\[username]\Box Sync’ as tabulated in Table 3.5.
Table 3.5: Box Sync Client Software Locations in Client Machine.

<table>
<thead>
<tr>
<th>Location 1</th>
<th>Location 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>“C:\Users[username]\AppData\Local\Box Sync”</td>
<td>“C:\Users[username]\Box Sync”</td>
</tr>
</tbody>
</table>

- **Prefetch Files:** Prefetch files are used by Windows to store information in relation to program activity including the number of time that software has run and associated files used by the software. Information located with prefetch files included the file and folder path the number of times runs and last time and date. A program called ‘WinPrefetchView’ [61] can be used that reads the prefetch files stored on the user machine and display information related to them. This information helps digital investigators while the investigation process. Figure 3.11 shows some prefetch files related to Box cloud storage.

- **Link Files:** Analysis of link files undertaken using autopsy. The link files observed related to the filenames and folder names for box executable file, uploaded data such as TestData.link. These located in the users ‘AppData’ directory in the ‘Windows/Recent’ and ‘Windows/Start’ folder. The link files observed related to the filenames and folder names for the box executable in \Users\[username]\Links\Box Sync. Link as shown in Figure 3.12.

Figure 3.11: Prefetch Files related to Box.
<table>
<thead>
<tr>
<th>Name</th>
<th>Modified Time</th>
<th>Change Time</th>
<th>Access Time</th>
<th>Created Time</th>
<th>Size</th>
</tr>
</thead>
</table>

**Figure 3.12:** Box Sync Link File.

- **Recycle Bin:** Files that deleted from the user machine easily located in $Recycle Bin folder. Figure 3.13 shows files found in a $Recycle.Bin folder during the analysis of the VM using Autopsy.

<table>
<thead>
<tr>
<th>Name</th>
<th>Modified Time</th>
<th>Change Time</th>
<th>Access Time</th>
<th>Created Time</th>
<th>Size</th>
</tr>
</thead>
</table>

**Figure 3.13:** Files found in $Recycle.Bin related to Box Storage.

- **Metadata:** Using autopsy it is possible to extract metadata from files such as extract EXIF metadata from images like JPEG images. This step can help the digital investigators to find any information related to any crime.

- **Network Traffic Capture:** One of the important and valuable information during the analysis phase is the information extracted from network traffic capture files which contain potential information about the communication between client system and the box storage service provider on the cloud side. For the research purpose, Wireshark program was installed on the host machine to monitor the traffic between the client system and the provider. The Wireshark program is operated before accessing, uploaded downloaded data. The output file (i.e. PCAP) which is generated for the analysis purpose and there was information about accessing Box website. The network traffic observed on ports 80 for HTTP and 443 for https.
When accessing Box account using web browsers, it observed that there a connection established with Box website where some IP addresses related to Box website observed in the network capture like ‘107.152.24.197’ and ’74.122.184.85’. Also, these IP addresses appeared in network connection when analysis the memory dump as shown in Figures 3.16 and 3.17.

- **Memory/Volatile Data:** Volatile data gives us events during the system was live such as usernames, passwords, encrypted data, sockets, process and much other valuable information. The memory of VM is stored in a file with extension VMEM. In this research, the dumpIt program was used to dump the memory of the VM, then we use Volatility framework to analysis the memory dump file produced by the dumpit tool to extract sensitive information that will help investigators during doing their mission. From this analysis, there was evidence about using Box software on the client machine as shown in Figure 3.14 from memory analysis. Figure 3.15 shows hex display for the username. Also, from memory dump, we observed there a connection with IP addresses ‘107.152.24.197’ and ’74.122.184.85’ where the owner was BoxSync.exe as shown in Figures 3.16 and 3.16.

![Figure 3.14: Running processes related to Box Storage extracted from the memory dump.](image)

![Figure 3.15: Hex Display of Username.](image)
From the extracted IP addresses (i.e. 74.112.85.85 and 107.152.24.197) from memory analysis, we can identify they are related or not for Box cloud storage through using IP Address Lookup tool as shown in Figure 18 and Figure 19. From that, we can prove the IP addresses 74.112.85.85 and 107.152.24.197 are related to Box cloud storage provider.

Figure 3.16: Network connection established with IP Address 74.112.184.85.

Figure 3.17: Network connection established with IP Address 107.152.24.197.

Figure 3.18: IP Address Lookup for IP Address 74.112.184.85.
Figure 3.19: IP Address Lookup for IP Address 107.152.24.197.

A comparison study between the proposed approach and other related work in the area of cloud storage forensics is tabulated in Table 3.6. The Table 3.6 shows that proposed approach adds security plus forensics together feature in comparison with related work. This feature helps in secure cloud storage with performing forensics analysis in case modification on cloud data which uploaded to cloud storage provider.

Table 3.6: Comparison Study between Proposed Approach and Related Work.

<table>
<thead>
<tr>
<th></th>
<th>Darren Quick et al. [27]</th>
<th>Darren Quick et al. [28]</th>
<th>Darren Quick et al. [29]</th>
<th>Shariati et al [59]</th>
<th>Proposed Work</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cloud Storage (Case Study)</td>
<td>Dropbox</td>
<td>Sky Drive</td>
<td>Google Drive</td>
<td>SugarSync</td>
<td>Box</td>
</tr>
<tr>
<td>Security Plus Forensic</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Forensic Analysis</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>
3.5 Chapter Summary

This chapter provides an approach for secure cloud storage which is based on checking data integrity and digital forensics. The proposed approach extracts metadata and generates hash values of user data. The metadata provides the investigators with valuable information for performing a forensic investigation. The metadata and hash values are then stored in the local storage for forensic investigation purpose because if there is any illegal activity was 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. After they discovered there are a modification and tampering in the uploaded data, the digital investigators will start performing the investigation process to find any digital evidence about the committed crime related to the cloud storage in where the crime occurred. The proposed approach provides a method to check the integrity of the uploaded data to the cloud storage by using metadata and hash values that help cloud investigators to investigate the uploaded data to the cloud side.

As a case study, we used Box cloud storage to examine data artifacts and remnants on the client machine when the user uploading, storing and downloading data to and from box cloud storage service. This data includes network traffic capture and memory volatile data in addition to some other related data. The determination of service provider and user account information is a necessary step that will enable examiners to identify potential and valuable information about box cloud storage when any crime occurred and legal entry points to collect and acquire evidential data in timely fashion and sound forensically manner. The experiment results showed that there are data artifacts that remain the user machine that uses windows 7 about using box cloud storage such as IP address, and user account information like the username. The proposed approach can use the potentially valuable procedure for performing cybercrimes investigation related to cloud storage services.