Offline vs. Online Digital Forensics of Cloud-based Services

Nidaa F. Hassan[1] and Haider M. Jaber [2*]
[1] Department of Computer Science, University of Technology.
[2] Department of Computer Science, University of Alnahrain.
[*] Corresponding Author: haidermjaber@gmail.com.

Abstract

Digital forensics has become important due to the daily use of digital devices in our life, which may lead to evidence that can be useful to law enforcement in legal cases. When using cloud services, a lot of artifacts that might be useful for investigations are stored in the cloud which means that the extraction must be from cloud computers systems. So, the legacy forensics tools may not be usable or less effective due to the geographically distribution of data and legal issues among other issues. The challenges of the cloud digital forensics are discussed in this paper. Also, this paper shows the types of extracting data from cloud services that are offline and online extraction. The offline extraction is concerned with extracting data from a local device that used to access cloud services. While online data extraction is concerned with remote extraction of data from cloud services. A comparison between such two types are demonstrated from many ways according to the result of recent papers to show the situations that each type is useful in.

[DOI: 10.22401/JNUS.20.4.18]

Keywords: Digital Forensics; Cloud Forensics; Offline Data Extraction; Online Data Extraction.

Introduction

Cloud computing is a term that describes the services that can be accessed over an internal network such as a private network in corporations or over the internet such as public services. The National Institute of Standards and Technology (NIST) defines cloud computing as [1]: *“a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”.*

Cloud computing has three main service models [1]:

1. Software as a service (SaaS): The consumer can use the applications that run in the cloud infrastructure using a program interface or thin client such as web browser. The cloud infrastructure cannot be managed by the consumer. The consumer can change the settings of the programming tools and applications provided by the cloud but not the operating system nor the hardware resources.
2. Platform as a Service (PaaS): The consumer can upload, code, and run their programs on the cloud using the programming tools offered by the cloud infrastructure.
3. Infrastructure as a Service (IaaS): The consumer can control the underlying operating system, applications, network, and storage.

Digital forensics is a branch of forensic science that includes extraction and investigation of information from digital devices related to computer crime. Digital forensics has become very important due to the daily use of digital devices in our life which may lead to evidence(s) that can be useful to law enforcements [2].

Before the cloud computing, digital forensics was aiming to extract artifacts from storage devices and evolved over years to include all types of computer devices [3]. The main aim of forensic tools was to extract artifacts from standalone computer system. This means that artifacts are extracted from the same computer used or attacked by a criminal. For example, tools like EnCase, Oxygen, FTK, and others used to extract the digital artifacts in computer or mobile devices by examining the RAM and storage devices and generate a time line history on the activities. They all need the direct access to the device to start its process[4]. However, when using cloud services, a lot of artifacts that might be useful to investigations are stored in the cloud, so using the conventional digital forensics tools must be on cloud computer systems which are hard to achieve due to the properties of cloud computing [5].

Many challenges arose with cloud forensics were discussed in this paper. The main challenge among others is the usage of one person to a cloud service may not limit to one computer but it might be distributed to many computer systems which limit the conventional digital forensic tools and methodologies.

Digital Forensics Process

The core need of digital forensics is to find evidence that can be used in courts. Thus, the digital forensics process must be done according to what law accepts. McKemmish in [6] propose four phases to be done to make the digital evidence be legally accepted; which are identification, preservation, analyses and presentation of digital evidence. NIST in [7], proposed a much similar to McKemmish phases which are collection, examination, analysis, and reporting. Fig.(1) shows those phases. A description of each phase is given in the following:

i. *Collection*: in this phase, the data is identified, labeled, recorded, and acquired from different sources. However, this process must preserve the integrity of the data.
ii. *Examination*: It concerns with processing the large amount of data that collected from the previous phase without affecting its integrity.
iii. *Analysis*: It involves analyzing the data to derive useful data that can answer questions related to the case.
iv. *Reporting*: It includes the reporting of the results from the previous phase in addition to the procedures and tools used to conduct such results.

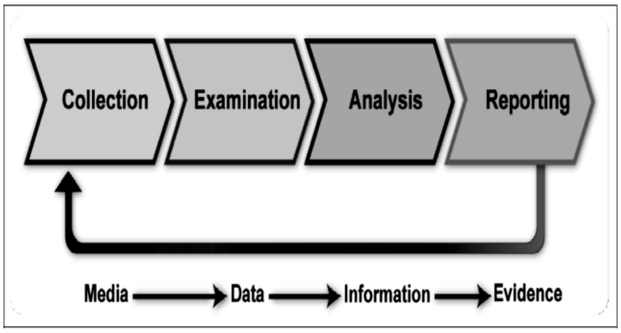

Fig.(1): Digital Forensics Process [7].

This paper is concerned with extracting data from cloud service which is done in the first phase.

Challenges of Digital Forensics of Cloud Services

Storing, processing, and transmitting methods of digital data is revolutionized by Cloud computing. The main difference of cloud computing services from other standalone computer services is the use of hypervisors and geographical distribution and independency. These characteristics make many challenges with digital forensics of cloud-based services to be accounted. The following are some of challenges given by NIST [5]:

i. Recovering deleted data.
ii. Unification of log formats.
iii. Distributed data collection.
iv. Evidence correlation across multiple cloud Providers.
v. Synchronization of timestamps between many sources.
vi. Cloud resources confiscation can affect the co-tenants' business continuity.
vii. Avoiding breaching the privacy of other tenants not related to the investigation.
viii. Data may quickly change or disappear and requestors lack knowledge of where and how data are stored.
ix. Identifying sources/traces of evidence.

Because of these challenges, it is difficult to use same standalone forensics tools with cloud servers and it is a must to develop new or adapt tools to be used with cloud forensics [5].

Cloud Forensics

In the recent years, cloud forensics papers suggest model, framework, or tools for extraction (collecting) and analyzing cloud data. As in [8], a framework for ID theft crimes digital forensics is suggested. Its approach leads the investigator according to specific procedures to an appropriate investigation process. In [9], a digital forensics model for online social networks is proposed; the model addresses two environments, physical and digital. It proposed a design of an application prototype to handle the digital environment. In [10], discusses the current frameworks that are not suitable for cloud computing because the data is usually stored at cross-border locations. It proposed a new iterated conceptual framework that inspired by the two frameworks, [6] and [7]. In which the various cloud computing types (IaaS, PaaS, and SaaS) are discussed in terms of different ways of data extraction. For example, the IaaS may offer Virtual Hard Disk export file that can be examined. But SaaS may only provide Application Programming Interface (API) to access the data.

Other papers are suggesting tools designed and developed by the authors or using a collection of free and commercial tools that used for standalone computer digital forensics. The data extraction in these papers may be automated or manual. Although, these papers are focusing on cloud services but they may categorize into two types: offline and online data extraction, some papers may use both types.

A. Offline data extraction is concerned with extracting data from a local device that used to access cloud services.

B. Online data extraction is concerned with extracting data from cloud services remotely.

The following subsections describe the concepts of offline and online data extractions:

A. Offline Data Extraction

The offline type of data extraction is concerned with the artifacts of using cloud services on a specific PC, mobile, or any other cloud-based devices. It may use the same forensics tools for standalone computer systems but directed to specific cloud services like Facebook, Google Drive, etc. This type of data collection does not concern with data that stored in the cloud computers but only in the computer or mobile device that used to access the cloud services.

The aim of this type is to extract security tokens, cashed messages, cashed files, cashed locations, cashed posts, cashed images, and any other cashed data from a cloud service.

For example, authors in [11], tested and analyzed the three main social network applications (Facebook, Twitter, and MySpace) on the three main smartphones operating systems (Android, IOS, and Blackberry). The paper showed that it could extract various information about the application account used from smartphone. However, it showed that they could not extract anything from Blackberry related to online social networks. It is noticed that there are a different amount of information being extracted from a different type of smartphones for the same service.

While in [12], the author tests and analyzes the three main social network applications (Facebook, Twitter, and LinkedIn) on the main smartphones operating systems (Android, IOS, Windows, and Blackberry). The paper showed that it could extract various information about the application account used from smartphone. Also, it could not extract anything from Blackberry as in [11]. The last applications version at the time of the paper was used in the test. It is worth noticing that both papers could not extract useful data from Blackberry devices which show it is a good for not retaining cashed data.

In [13], the authors show how to extract messages in Facebook chat using existing tools and how to identify the Arabic messages. It does not propose any automated system but depends on the human ability to detect and convert the chat messages to Arabic.

These papers show how it can extract cashed data of cloud services from PC or mobile. Although, it can recover cashed data but it is only part of the data in cloud computers and depends on the cloud application to what and when to store such data as cached. Also, it depends on the device used; as described before all researchers couldn't extract useful information from blackberry smartphones. Even though, if possible extraction process succeeded, usually, the most recent activities can be recovered.

If the forensics process is done as described in section II then the extracted data can be reproduced if the same source and procedures used as in the first time. The reproducibility is important in digital forensics to reproduced extracted evidence by third parity [14].

B. Online Data Extraction

The online data extraction concerned with how to extract the data from online services, i.e. the data that stored in the cloud servers. The techniques used are depending on the targeted services. There are services that offer a powerful API to access the data and others have a limited API, however, others do not have at all.

In [15], authors use Facebook API and web crawler to extract account information for further analyses. The information used to generate a timeline of activities of the user. Although, Facebook has API functions that could be used to access account information. The web crawling technique used may need to be updated when the Facebook changes its web interface.

The same authors in [15], published a new paper in which the same techniques are used but it extracts more information and graphs the relations between accounts according to the messages and picture tags in addition to generating the timeline of account activities [16].

In [17], a method has been developed for the extraction, analysis, visualization, and comparison of the user profile that online snapshotted from Twitter. It uses Twitter API to retrieve profile statuses and messages in addition to other information.

The data extracted from online extraction methods, mostly, are much more than the data extracted using offline extraction methods because the online one is connected to the service provider server which contains all the information. Although, there are cloud services that store most of the data offline in the user device like the chat service Viber [18], still most of the cloud services save most of the data at the cloud storage which makes the online extraction tools is the best option for collecting related data as much as possible. Figs. (2 and 3) show a graph that using the data extracted from online data extraction tools. Fig.(2) shows the interaction between the users that tagged in the same picture. The thicker the arc the more pictures in common. Figure 3 shows the interaction from another point of view. It shows the interaction between the users using the messages exchanged between them. The thicker is the arc the more messages in common. This example shows the huge amount of data that can be extracted using the online data extraction. It gets over the current account and gets the relation with other accounts that may at the end have relation with the incident [16].

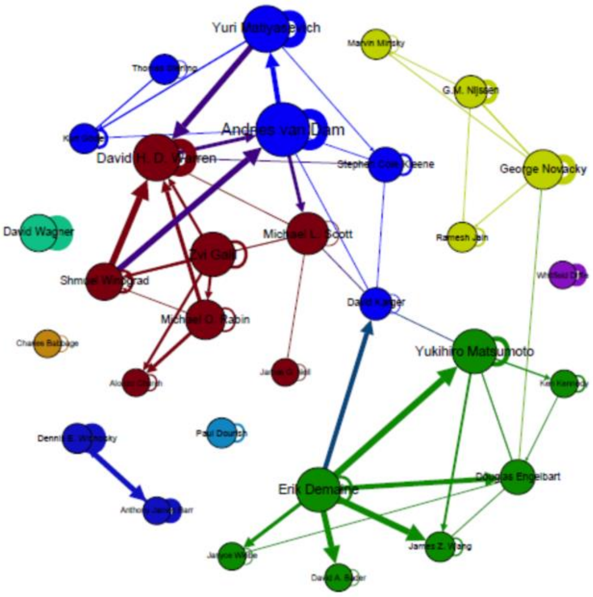

Fig.(2): Social intraction graph using extracted Facebook picture tags [16].

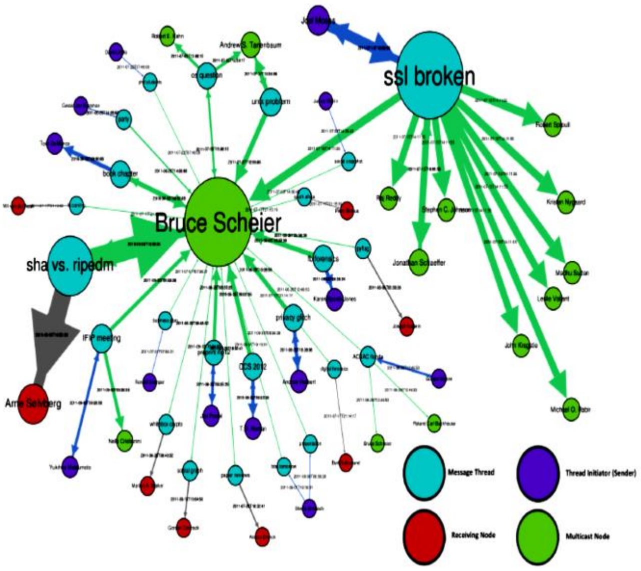

Fig.(3) :Social interaction graph using extracted Facebook messages [16].

Bjornson and Hunter [19] used Cellebrite's Universal Forensic Extraction Device (UFED) Cloud analyzer tool to extract data from cloud storage to examine its effectiveness. It showed that the data can be extracted under relatively strong assumptions, like the credentials must be preserved first to use the tool, in addition to the legal considerations and challenges.

All of the proposed online data collection methods can retrieve data from the different cloud services but there are many problems with them in respect to digital forensics. One of the main problems is the reproducibility. Not all data can be reproduced because there is no possible way to suspend the accounts for the most public cloud services (until the date of this paper), especially the social network services, from being updated and remain as read-only. Another problem is about the web structure and API functions of the cloud-based services. When the cloud service provider changes its way of showing the information in the web or upgrade or restrict the API definitions, the tools that depend on that API or web structure will fail partially or fully. For example, Facebook Graph API v1.0 allows programs to retrieve the friend list of the user including their IDs but it was restricted in Facebook Graph API v2.x to return only friends that gave permission to the application [20].

Also, there is the custody problem, it is not possible to block the criminal or one of his relates from changing the data in the cloud because it can be accessed from any device that is connected to the cloud service [8].

Another problem is how to get access the criminal's (suspect's) or victim's cloud service account. In the case of offline extraction tools, what needed is to access the device used to access the cloud service that may be acquired during investigator inspection. The investigator will take an image of the storage and RAM and run the tools to extract the data. But in the case of online extraction tools, there are two ways to get access to the cloud account, either by having the user name and password or using a sort of security token that the cloud service support. The user name and password could be taken from the victim or criminal directly if possible. Or either one the user name and password or the security token could be recovered using one of the offline extraction tools [21].

Such that, it is noticeable that the online extraction methods may need the offline extraction methods to start its job.

Comparison between Offline and Online Data Extraction

As shown previously, there are two types of data extraction from cloud services, offline and online. Each has its advantages and disadvantages. The offline is more popular because the investigator can use the conventional tools of digital forensics to extract information from the computer or mobile devices. Table (1) shows a comparison between the offline and online extraction for cloud-based services to be used in digital forensics.

It is shown, that the online data extraction is better from the point of information amount but it has problems with reproducibility and custody which it a concern in digital forensics. In addition, there is a problem with the durability of the extraction method where the upgrade or the change of the API may restrict or even fail the online extraction tool. However, the offline extraction tools can extract important information that recently used or may extract most of the information like in Viber.

Table (1)
Offline vs. Online Data extraction.

Issue	Offline	Online
Data Extracted	Depends on the device and only cashed data.	If possible, there are two ways: Using API: restricted by the service provider. Web Crawler: Can get all the information allowed to the regular used to see but it is more difficult to achieved.
Durability of extraction method	Depends on the updates on the application used to access the cloud service. Note that the application may not provide by the cloud service provider.	Depends on the upgrades done by the cloud service provider.
Amount of information	Small amount depending on the cash size.	Large depending on the amount stored in the cloud.
Reproducibility	Can be reproduced if standard digital forensics procedures are followed.	May or not, depending on the cloud service.
Custody	As soon the used device in custody, the criminal cannot tamper the data.	The cloud account can be accessed from any connected device to the cloud, so the criminal or one of his relates can log in and tamper with data. It is not possible to the make the account in custody without the help of the service provider.
Access	The device used to access the cloud service must be in custody	The user name and password or some sort of security token of the cloud service account must be available. The offline extraction tools may be used with devices in custody to extract the user name and password or the security tokens.

Conclusions

In cloud forensics, there are many challenges over the conventional digital forensics methods. These challenges include the geographical distribution of data, multi-tenant in the same computer, and the cloud service company agreement on investigating their computers. All these issues make the conventional forensic tools semi-obsolete. Although it can be used to extract cashed data from cloud applications in the PC or mobile, the data is not complete and the investigation may need more data that reside in cloud computers. Here come the online tools that specialized to collect data from online cloud services. The online extraction tools may use one of the two methods cloud services provided API or using web crawlers tools specifically designed for this targeted cloud service. But because of the changing natural of the cloud services, the online extracting tools may need an ongoing update to keep up to date with these changes.

Both offline and online cloud extraction tools are useful for extracting data. Each extracts useful information with respect to the type and amount of information. The online extraction tools may depend on the offline extraction tools to get access credentials, user name and password or security token, while the offline extraction tools can extract data from the digital devices as soon as the device in custody. Although a cloud service may provide API for accessing user's data but may provide restricted access to certain data such as log data. The offline extraction tools may be useful in these cases if the device is in custody and used frequently by the criminal or the victim.

As a conclusion, both offline and online extraction tools must be used in cloud-based digital forensics to get the best extraction results which may lead to more concrete digital evidence.

References

[1] Mell P., Grance T., "The NIST Definition of Cloud," National Institute of Standards and Technology, Gaithersburg, 2011.

[2]Carrier B., "Defining Digital Forensic Examination and Analysis Tools Using Abstraction Layers," International Journal of Digital Evid, 1(4), 2003.

[3]Casey E., "Handbook of Digital Forensics and Investigation", Elsevier Inc., 2010.

[4] Yates M., "Practical Investigations of Digital Forensics Tools for Mobile Devices", Info Sec CD, Kennesaw, GA, USA, October 2010.

[5]NIST Cloud Computing Forensic Science Working Group, "NIST Cloud Computing Forensic Science Challenges," National Institute of Standards and Technology, 2014.

[6]McKemmish R., "What is forensic computing?," Trends & issues in crime and criminal justice, 118, 1-6, 1999.

[7]Kent K., Chevalier S., Grance T. and Dang H., "Recommendations of the National Institute of Standards and Technology," National Institute of Standards and Technology (NIST), 2006.

[8]Angelopoulou O., "ID Theft computer forensics investigation framework," Australian Digital Forensics Conference, Edith Cowan University, Perth Western Australia, December 3rd 2007..

[9]Zainudin N. M., Merabti M. and Llewellyn-Jones D., "Online Social Networks As Supporting Evidence: A Digital Forensic Investigation Model and Its Application Design", Research and Innovation in Information Systems (ICRIIS), 2011.

[10] Martini B. and Choo K. R., "An integrated conceptual digital forensic framework for cloud computing," Digital Investigation, 9, 2012.

[11] Mutawa N. A., Baggili I. and Marrington A., "Forensic analysis of social networking applications on mobile devices," Digital Investigation, 9, 2012.

[12] F. A. Awan, "Forensic Examination of Social Networking Applications on Smartphones," Conference on Information Assurance and Cyber Security (CIACS), 2015.

[13] Mutawa N. A., Awadhi I. A., Baggili I. and Marrington A., "Forensic artifacts of Facebook's instant messaging service," in 6th International Conference on Internet Technology and Secured Transactions, Abu Dhabi, United, 2011.

[14] Garfinkel S., Farrell P. and Roussev V., "Bringing science to digital forensics with standardized forensic corpora," 6, S2-S11, September 2009.

[15] Huber M., Mulazzani M., Leithner M., Schrittwieser S., Wondracek G. and Weippl E., "Social Snapshots: Digital Forensics for Online Social Networks", Annual Computer Security Applications Conference, Orlando, 2011.

[16] Mulazzani M., Huber M. and Weippl E., "Social Network Forensics: Tapping the Data Pool of Social Networks", 2012.

[17] Chris H., Lu L., ZhiYuan L., JianXin L. and Nick A., "Virtual vignettes: the acquisition, analysis, and presentation of social network data," SCIENCE CHINA, Information Sciences, 57, 1-20, 2014.

[18] Viber, "Viber Support". [Online]. Available:https://support.viber.com/customer/portal/articles/1334452-create-a-backup-file-of-your-messages. [Accessed Jan 2017].

[19] Bjornson J. and Hunter A., "Mobile forensics for cloud data: Practical and legal considerations," in 14th Annual Conference on Privacy, Security and Trust (PST), Auckland, New Zealand, 2016.

[20] Facebook, "No way to get all friends of "me" with a permanent ID," Facebook, May 2014. [Online]. Available: https://developers.facebook.com/bugs/1502515636638396/. [Accessed Jan 2017].

[21] Martini B., Do Q. and Choo K. R., "Digital forensics in the cloud era: The decline of passwords and the need for legal

reform," Trends & issues in crime and criminal justice, 512, December 2016.

[22] Wong K., Anthony C. T. Lai, Jason C. K. Yeung, Lee W. L., Chan P. H., "Facebook Forensics", Valkyrie-X Security Research Group, July 5, 2011. [Online]. Available: https://sites.google.com/site/valkyriexsecurityresearch/announcements/facebookforensicspaperpublished. [Accessed Jan 2017].

[23] Garfinkel S., "Digital Forensics Research: The Next 10 Years," in The Digital Forensic Research Conference, Portland, OR (Aug 2nd-4th), 2010.