A Tri-layered Data Acquisition System for Evidence Gathering from Cloud.

ShitalDudhane, KajalShinde, Gauri Bahirat, Roma Kudale
Department of Computer Engineering,
STES Smt. KashibaiNavale College of Engineering,Vadgaonbk, Pune-41

Abstract:
Digital forensics have plenty of applications. Digital evidence in the field of forensic investigation has become very important. There are many issues in dealing with network evidence. As network is volatile in nature it becomes difficult to gather network evidence. Sometimes, such information may change with the time, may be located on server which needs authority to get access or far away from the crime scene. In this paper, a novel methodology is presented to collect network evidence. Precisely, the online services like web pages, chats, photos or videos would be source for collecting information. This method is suitable for both experts and non-experts as it takes user through whole process of obtaining evidences. During this process, the information received from remote source is automatically gathered. This information consists of network packets and any information generated by user. A trusted-third party, works as a digital notary to verify both, obtained evidence and the acquisition process.

Keywords:
Data Encryption: Code breaking, Data encryption standard (DES), Public key cryptosystems Standards (e.g., DES, PGP, RSA)

Introduction

Digital forensics have plenty of applications. Digital evidence in the field of forensic investigations has become very important. There are many issues in dealing with network evidence. As network is volatile in nature it becomes difficult to gather network evidence. Sometimes, such information may change with the time, may be located on server which needs authority to get access or far away from the crime scene. In this paper, A novel methodology is presented to collect network evidence. Precisely, the online services like web pages, chats, photos or videos would be source for collecting information. This method is suitable for both experts and non-experts as it takes user through whole process of obtaining evidences. During this process, the information received from remote source is automatically gathered. This information consists of network packets and any information generated by user. A trusted-third party, works as a digital notary to verify both, obtained evidence and the acquisition process.

Problem Definition:

Previous acquisition tools proposed in the last years suffer from various limitations. Firstly, they typically lack of non-repudiation and data-integrity solutions to protect the collected information, which means that the result of the investigation could be interfere by an attacker. Also, we cannot be assured about reliability of source of information. Lastly, this

type of acquisition processes was vulnerable to 'man-in-the-middle' attack. So, we proposed a method through which we can get the reliable and valid data.

Objectives

The main objective of the project is to collect information with verified techniques and create evidences such that no one can question about its integrity. Also, system focuses on providing security features to collected evidences

System Architecture

Proposed system should be able to collect information regarding given topic from data stored on cloud. This information is acquired using three operation modes namely, LNE-Proxy, LNE-Agent, Savvy users. These three OMs ensures verified data is collected. After collection of this data evidence packing process is carried out in order to preserve the integrity of data. Packing process consist of encryption and providing digital signature to gathered evidences. The system can be used for collection of evidences of a kind whose integrity cannot be questioned.

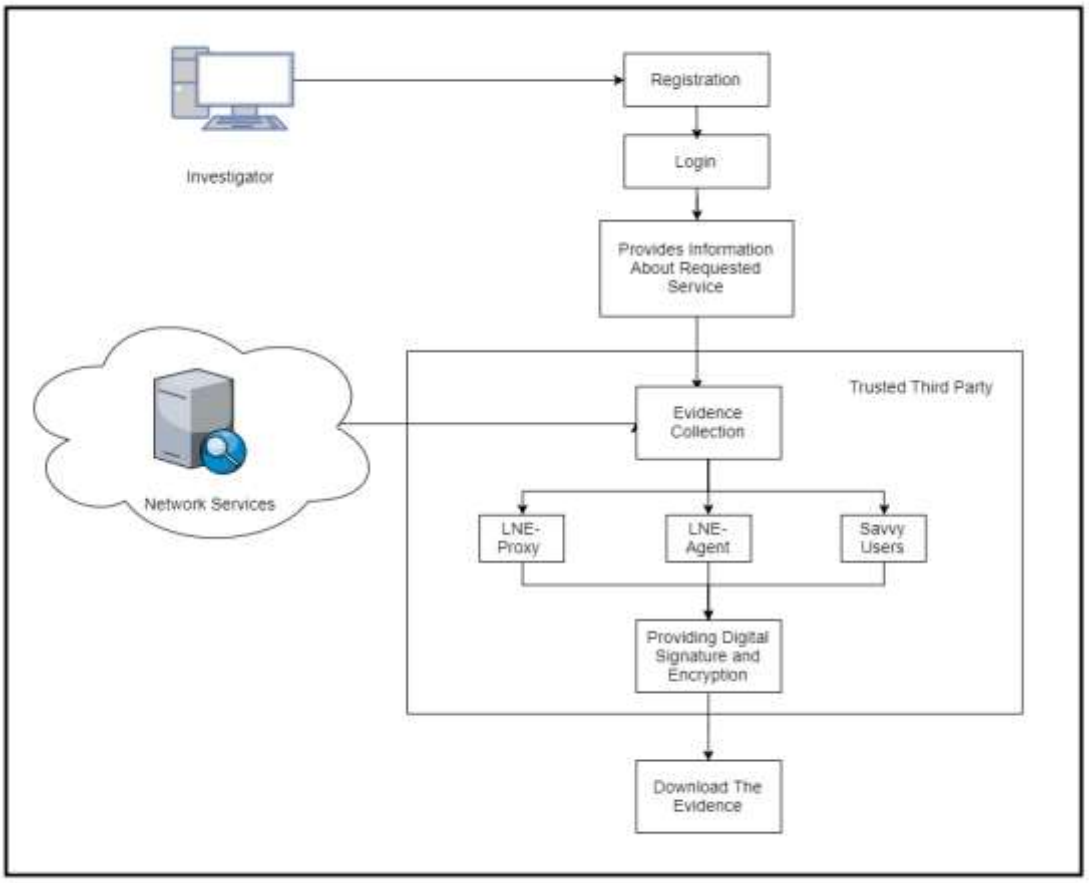

Fig 1: System Overview

Related Works

1) Cloud incident handling and forensic-by-design: cloud storage as a case study:

Year: 2016

Author Name:

- NurulHidayah Ab Rahman
- NikenDwiWahyuCahyani
- Kim Kwang Raymond Choo

Description: An integrated cloud incident handling and forensic-by-design model, and that incorporates digital forensics practices have been presented.
Limitations: Deploying the proposed model in a real-world setting, with the aims of validating and refining the model.

2) Cloud Infrastructure Resource Allocation for Big Data Applications:
Year: 2016
Author Name:

- Wenyun Dai
- LongfeiQiu
- Ana Wu,4.Meikang Qiu

Description: First analyzed the relations among the cost, performance, and availability of one cloud-based big data application, and built three models.
Limitations: First one is to add more constraints, including the security and data processing preference. The second one is to test our approach on advanced networking environments, such as Software-Defined Networking (SDN).

3) Cloud Storage Forensic: hubiC as a Case-Study:
Year: 2015
Author Name:

- Ben Blakeley
- Chris Cooney
- Ali Dehghantanha
- Rob Aspin

Description: This research paper aims to answer following questions:
• What data can be recovered on the hard drive of a Windows 8.1 machine after the use of the hubiC cloud storage service?
• What data can be recovered from the physical memory (RAM) of a Windows 8.1 machine after the use of the hubiC cloud storage service?
Limitations: Direct future applications of this research could apply similar methodology to the investigation of other cloud platforms, (e.g. ADrive, eCloud, etc) on Windows 8.1. This would increase the scope of the investigation and provide greater insight to an investigator, potentially revealing common flaws across several cloud platforms.

4) Cloud storage forensics: own Cloud as a case study:
Year: 2013
Author Name:

- Ben Martini
- Kim-Kwang Raymond Choo

Description: Using ownCloud as a case study, we successfully undertook a forensic examination of the client and server components of an ownCloud installation and discussed the relevance of a number of artefacts to a forensic investigation.

Limitations: Further work on the potential for network interception as a method of forensic collection should be pursued especially as a method of identification of potential evidence sources.

5) Digital droplets: Microsoft SkyDrive forensic data remnants.

Year: 2013

Author Name:

- Darren Quick
- Kim-Kwang Raymond Choo

Description: To find that an examiner can identify SkyDrive account use by undertaking keyword searches, hash comparison, and examine common file locations in Windows 7 systems to locate relevant information.

Limitations: A future research opportunity would be to undertake the experiments with the Windows 8 operating system to determine if the same data remnants are present. Future research opportunities include conducting research in to the remnants of other cloud storage services such as Google Drive.

6) Distributed filesystem forensics: XtreemFS as a case study:

Year: 2014

Author Name:

- Ben Martini
- Kim-Kwang Raymond Choo

Description: Aim to address this gap in knowledge. Using our previously published cloud forensic framework as the underlying basis, we conduct an in-depth forensic experiment on XtreemFS, a Contrail EU-funded project, as a case study for distributed filesystem forensics.

Limitations: Future work includes validating our framework and the proposed process with other similar distributed filesystem products such as GlusterFS, FhGFS and Ceph. Another aspect of future work would be to develop forensic processes for cloud/distributed filesystems where APIs can be used for object storage and retrieval.

7) Forensic data acquisition from cloud-of-things devices: windows Smartphones as a case study:

Year: 2016

Author Name:

- NikenDwiWahyuCahyani
- Ben Martini
- Kim-Kwang
- Raymond Choo1
- AKBP Muhammad Nuh Al-Azhar

Description: In this paper, we examined forensic data acquisition results from three Windows phone devices, which are commonly used to connect to (interconnected) cloud services. We found that the capacity of Windows phone data acquisition remains limited. Only one of the three forensic tools supported physical acquisition.

Limitations: Future work will include examining other WP8 OS devices and new releases of mobile forensic tools. A forensic data acquisition model for Windows phones would also assist to guide practitioner's decision making as part of the acquisition process.

8) Geographical information system parallelization for spatial big data processing: a review.

Year: 2015

Author Name:

- Lingjun Zhao
- Lajiao Chen
- Rajiv Ranjan
- Kim-Kwang Raymond Choo
- Jijun He

Description: This paper surveys the current state of parallel GIS with respect to parallel GIS architectures, parallel processing strategies, and relevant topics. We present the general evolution of the GIS architecture which includes main two parallel GIS architectures based on high performance computing cluster and Hadoop cluster.

Limitations: The key problems and future potential research directions are addressed, such as a parallel GIS architecture incorporated with efficient storage and computing, dynamical business work flow combination, a multiplicity balanced data partition strategy and a relatively easy parallel way in the system level.

9) Google Drive: Forensic analysis of data remnants

Year: 2013

Author Name:

- Darren Quickn
- Kim-Kwang Raymond Choo

Description: Using Google Drive as a case study, the following questions are examined:

1.What data remains on a computer hard drive after a Google Drive user has used client software or accessed cloud storage

via a browser, and the location within the Windows 7 operating system of data remnants?

2.What data can be seen in network traffic, and what data Remains in memory?

3.What data remains on an Apple iPhone running iOS version 4.2.1 after a user has used the inbuilt browser to access Google Drive cloud storage?

Limitations: A future research opportunity is to undertake similar research using an iOS5 later device, and also Expand the type of devices examined to include other popular mobile device operating systems, such as Google Android or Microsoft Windows. Future research could also include the comparison of a physical extract of an iPhone to a logical extract to Determine the information available, and also a comparison with other iPhone forensic software and hardware.

10) Kvasir: Scalable Provision of Semantically, Relevant Web Content on Big Data Framework

Year: 2015

Author Name:

- Liang Wang
- Sotiris Tasouli
- TeemuRoos
- JussiKangasharju

Description: This paper presents the architecture of Kvasir, which is able to seamlessly integrate LSA-based content provision in web browsing by using state-of-art technologies. We propose a parallel version of the randomized partition tree algorithm which provides fast indexing in high dimensional vector spaces using Apache Spark.
Limitations: Currently, Kvasir does not provide full-fledged security and privacy support. For security, malicious users may launch DDoS attacks by submitting a huge amount of random
requests quickly. Though limiting the request rate can mitigate such attacks to some extent, DDoS attacks are difficult to defend against in general. For privacy, Kvasir needs tracking a user's browsing history to provide personalized results.

Limitation of Study:
The only limitation of the system is working of agent. Agent will have to be active to upload the proofs for the requested acquisition case. If agents will not upload the proofs or evidences then mode 2 operation will be failed.

Design of the Study

- Input: Request details
- Output: Evidences
- Functions :
  1. Identify the posts posted by suspect on social media.
  2. Finding correlated information from files uploaded by agents.
  3. Allowing investigator to get the evidences.
- Success Conditions: Evidences Zip file that contains evidences from three modes
- Failure Conditions: Agents will have to be active.

Tools Used

- **Software Requirement:**
  - Operating System : windows 8 and above.
  - Application Server : Tomcat5.0/6.X
  - Language : Java
  - Front End : HTML, JSP
  - Database : MySQL
- **Hardware Requirement:**
  - Processor - Pentium –III
  - RAM - 1 GB (min)
  - Hard Disk - 20 GB(min)

Statistical Technique Used
We have developed Login and Registration which manages the user profiles(Investigator, Agent, and Admin), so that the users can post request and get its output according to his role. Database stores the information of all users, requests, files uploaded by agents, also these files will be uploaded to cloud.

Algorithm

- **Log4j**: This API allows us to keep track of all the activities performed by various users.

- **Advance Encryption Standard**: In our system, we have used AES to provide encryption to gathered evidences. Along with that we will be using digital signature to conserve the integrity of data.
- **Secure Hash Algorithm** 1: In cryptography, **SHA-1** is a cryptographic hash function which takes an input and produces a 160-bit (20-byte) signature value known as digital signature.

Advance Encryption Standard:

```
Cipher(byte in[16], byte out[16], key_arrayround_key[Nr+1])
begin
byte state[16];
state = in;
AddRoundKey(state, round_key[0]);
fori = 1 to Nr-1 stepsize 1 do
SubBytes(state);
ShiftRows(state);
MixColumns(state);
AddRoundKey(state, round_key[i]);
end for
SubBytes(state);
ShiftRows(state);
AddRoundKey(state, round_key[Nr]);
end
```

Secure Hash Algorithm:

1. add some extra data to the end of the input
   set the initial sha-1 values
   for each 64-byte chunk do
   extend the chunk to 320 bytes of data
   perform first set of operations on chunk[i] (x20)
   perform second set of operations on chunk[i] (x20)
   perform third set of operations on chunk[i] (x20)
   perform fourth set of operations on chunk[i] (x20)
   end
   return sha-1 values as a hash

Our Approach:

The system will work in three operating modes:

1. LNE-Proxy:

In this mode of operation, information related networks, servers, etc. is collected. This information can be used in order to identify attacks like man-in-the-middle and spoofing etc.

2. LNE-Agent:

Number of different sources collect information regarding requested topic/incident. This information then put under process of finding co-relations to extract relevant content. This operation mode thus, verifies source of information.

3. Savvy Users:

In above two modes of operation, investigator does not participate in investigation process unlike third operation mode.

Experiment Result:

This system will collect evidences by three different modes. These evidences will be collected into zip file.

Future scope:

In future, we will develop a flexible system which can work with real time social media applications.

Acknowledgment: (optional)

Conclusion:

The idea of acquisition of Live Network Evidence (LNE) from online services is based on Trusted-Third-Party (TTP). TTP collects the information on behalf of investigator. Data will be obtained by using three different modes. The evidence collected by TTP are strong and its validity can be checked at any time after acquisition process. This data is more accurate and its integrity and authenticity can be guaranteed.

Reference:

1. IRJET, "DAS: A Novel Approach to Gather Evidence from Cloud", (Available:https://irjet.net/archives/V5/i3/IRJET-V5I3682.pdf)
2. ACPO Computer Crime Group, "Good practice guide for computer-based evidence," Association of Chief Police Officers, Tech. Rep., 1999.

2. NIST, "Disk imaging tool specification," Computer Forensics Tool Testing (CFTT) Project, Tech. Rep., 2001.

3. Computer Crime and Intellectual Property Section, Criminal Division, "Searching and seizing computers and obtaining electronic evidence in criminal investigations," U.S. Department of Justice, Tech. Rep., 2002.

4. National Institute of Justice (USA), "Digital Forensics Standards and Capacity Building," (Available on evidence/digital/standards/welcome.htm), [Accessed on 17 October 2012].

5. W. Kruse and J. Heiser, Computer Forensics: Incident Response Essentials. Pearson Education, 2001. [Online]. Available: http://books.google.it/books?id=-qWa5Svv7BIC

6. M. Sheetz, Computer Forensics: An Essential Guide for Accountants, Lawyers, and Managers. Wiley, 2007.

7. D. Quick and K.-K. R. Choo, "Digital droplets: Microsoft SkyDrive forensic data remnants," Future Generation Computer Systems, vol. 29, no. 6, pp. 1378 – 1394, 2013. [Online].
Available:http://www.sciencedirect.com/science/article/pii/S0167739X13000265

8. "Google Drive: Forensic analysis of data remnants," Journal of Network and Computer Applications, vol. 40, pp. 179 – 193, 2014. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1084804513002051

9. L.Wang, S. Tasoulis, T. Roos, and J. Kangasharju, "Kvasir: Scalable Provision of Semantically Relevant Web Content on Big Data Framework," IEEE Transactions on Big Data, vol. PP, no. 99, pp. 1–1, 2016.

10.W. Dai, L. Qiu, A. Wu, and M. Qiu, "Cloud Infrastructure Resource Allocation for Big Data Applications," IEEE Transactions on Big Data, vol. PP, no. 99, pp. 1–1, 2016.