A TRUSTED THIRD PARTY SYSTEM FOR COLLECTION OF VERIFIED EVIDENCES FROM BIG DATA.

- **Omkar Bhalerao, Onkar Bhanarkar, Ashwin Gophan, Prof. N.V. More**

Abstract:

Digital forensics experienced many issues while collecting the network evidences. Here it is difficult to collect the network evidences because it changes as the time passes and these evidences are far from the crime scene. And it is also difficult to gather the evidences from remote storage so here we are mainly focusing on novel methodology to gather these evidences. And this information changes with time so it's important to store this information so that we can gather this as network evidence. For this we are using cloud platform so that we can collect these evidences from remote server. In this process it gives detail information regarding user's access such as network packets and whatever information is generated during the scene. And in the last third party verifies all this network evidences digitally.

Keyword- *Digital Forensics, Network Forensics, Live Network Evidence (LNE), Big Data Forensics, Digital Investigations.*

Introduction

In current state-of-art, if an investigator needs information, he/she will have to collect by his/her own. Even after collecting the information, there's no guarantee that collected information or evidences are fully verified or authenticated. For that an online notary system came into existence. Even this system has its limitations as it needs a person with deep knowledge regarding these collected evidences in order to verify them. To avoid this, a trusted third-party data acquisition system is proposed. This system will collect the evidences and provide digital signature, Encryption to preserve integrity of data. Thus, once the data is acquired by system, it can be considered that the collected data is not only authenticated but also the digital signature ensures that it is robust against attacks like spoofing or man-in-the-middle attack. Big data is a term for data sets that are so large or complex that traditional systems are not capable to deal with them. Big data has also been defined by the three Vs: Volume, Variety. Cloud computing allows users, and enterprises, with various computing capabilities to store and process data in either a privately-owned cloud, or on a third-party server located in a data center in order to make data accessing mechanisms more efficient and reliable. Data mining is an interdisciplinary subfield of computer science and the analysis step of the "Knowledge Discovery in Databases" process and it is the approach of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, statistics, machine learning, and database systems. Data mining is an important part of knowledge discovery, is defined as the automated method of finding previously unknown, nontrivial, and useful information from databases. Database mining is the process of generating high-level patterns that have acceptable certainty and are also interesting from a database of facts Knowledge Discovery of patterns has been applied effectively to solve many diverse problems.

Problem Definition:

Previous information gathering tools proposed in the last years suffer from various limitations. Firstly, they have drawback of non-repudiation and data-integrity solutions to prevent the collected information, which means that the result of the acquisition could be disturb by a hacker. Also, we cannot provide guarantee about reliability of source of information. Lastly, this information gathering processes was vulnerable to 'man-in-the-middle' attack. So, we proposed a method through which we can acquire the reliable and valid data.

Objectives

The major objective of the project is to acquire information with verified methods and get evidences such that no one can question about its integrity. Also, system concentrates on providing security features to gather evidences.

System design:

System Architecture

Proposed system should be able to gather information regarding given request from data stored on logs. 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**.**

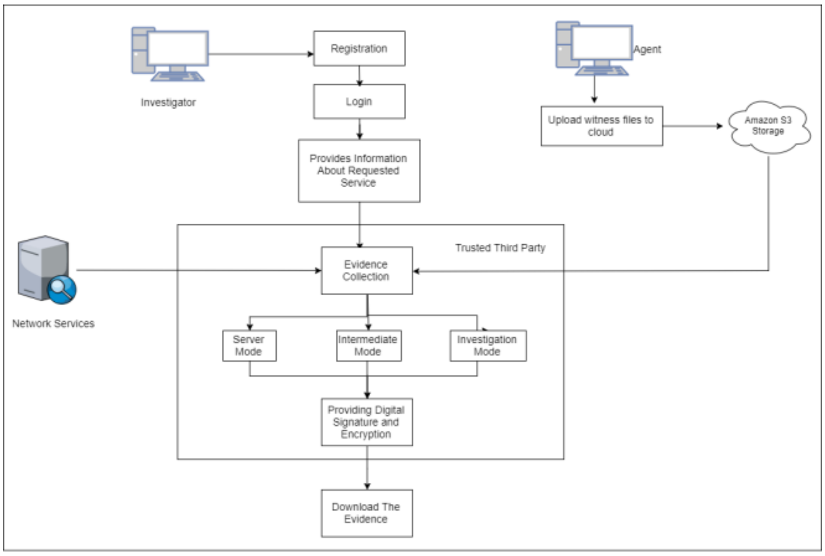

Fig 1: System Overview

Related Works

1.The paper, Cloud incident handling and forensic-by-design: cloud storage as a case study is an integrated cloud incident handling and forensic by design model, and that incorporates digital forensics practices have been presented. This system failed in 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 [10], first analyzed the relations among the cost, performance, and availability of one cloud based big data application, and built three models [1]. Limitations of this system is, first one is to add more constraints, including the security and data processing preference [3]. The second one is to test our approach on advanced networking environments, such as Software-Defined Networking (SDN) [4].

3.Cloud Storage Forensic: hubiC as a Case-Study 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? 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: ownCloud as a case study aims to use 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 [6] 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 [7].

5.In this paper, Digital droplets: Microsoft SkyDrive forensic data remnants, to find that a 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 [8]. 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 [9]. Future research opportunities include conducting research in to the remnants of other cloud storage services such as Google Drive [2].

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	:	Intel i3/i5/i7
RAM	:	4 GB (min)
Hard Disk	:	20 G/B(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:

```
• 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 agents will upload proofs 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 acquire evidences by three different methods. These evidences will be packaged into zip file.

Future scope:

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

Acknowledgment: (optional)

It gives us great pleasure in presenting the preliminary project report on ‘**A** Trusted Third Party System for Collection of Verified Evidences from Big Data’. I would like to take this opportunity to thank my internal guide Prof. Nitin More for giving me all the help and guidance I needed I am really grateful to them for their kind support. Their valuable suggestions were very helpful.

I am also grateful to Prof. B.K Sarkar, Head of Computer Engineering Department, *PVPIT, Pune*, for his indispensable support and suggestions.

Omkar Bhalerao

Onkar Bhanarkar

Ashwin Gophan

Conclusion:

The idea of collection of Live Network Evidence (LNE) from online services is based on Trusted-Third-Party (TTP). TTP collects the evidences on behalf of investigator. Evidences 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:

- ACPO Computer Crime Group, “Good practice guide for computer-based evidence,” Association of Chief Police Officers, Tech. Rep., 1999.

- NIST, “Disk imaging tool specification,” Computer Forensics Tool Testing (CFTT) Project, Tech. Rep., 2001.

- 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.

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

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

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

- 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

- “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

- 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.

- 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.