A Trusted Third Party System for Collection of Verified Evidences from Big Data

N. V. More[1] Onkar Bhanarkar[2] Omkar Bhalerao[3] Ashwin Gophan[4]
[1]Professor [2,3,4]Student
[1,2,3,4]Department of Computer Engineering
[1,2,3,4]PVPIT, Pune, Maharashtra, India

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.

Key words: Digital Forensics, Network Forensics, Live Network Evidence (LNE), Big Data Forensics, Digital Investigations

I. INTRODUCTION

In current state-of-art, if an investigator needs an 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.

II. BACKGROUND & BASIC

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, Velocity, and 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.

III. LITERATURE SURVEY

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.

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].

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.

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].

In this paper, Digital droplets: Microsoft SkyDrive forensic data remnants, to find that a 6. 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].

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

The system will work in three operating modes:

A. Server Mode

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.

In this, that we called LNE-Proxy, the TTP acts as a HTTP/HTTPS proxy. The investigator can configure its local browser in order to use the TTP to acquire digital evidence. All the network traffic flowing through the TTP proxy is captured, signed and finally sent to the investigator. The network traffic captured by the TTP is saved in the standard pcap format, which is supported by most of network analysis tools. It is worth noting that network packets contain lot of low-level information which can be useful for the subsequent analysis phase. Log4j API is use to track the information in the form of logs. Here log file is maintained to track all the operation that are performed on client and server machine. Based on these logs system can fetch the records that are related to the suspect. The output of this mode will be collection on logs that are already present in the log file.

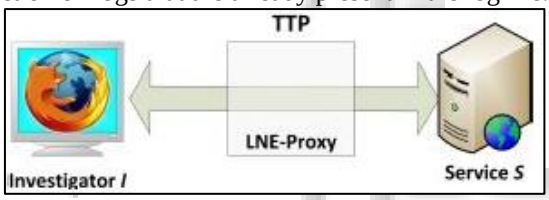

Fig. 1: LNE-Proxy Operating Mode

B. Intermediate Mode

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. the LNE-Agent plays an active role in the communication with the target server, since it is in charge of generating requests on behalf of the investigator. The general architecture of LNE-Agent is depicted in Figure. Also in this case the acquisition is driven by the investigator. The service uses a Remote Desktop Protocol (RDP) to provide the investigator with a remote control interface. In fact, in order to gather the target information, the investigator I remotely accesses the TTP server console. The functionality of the agents are important in our system. Agents must be active when investigator fires a request to collect evidences. Then agents will upload evidences in the form of text files. These files will also be uploaded to cloud for safety purpose. Any number of agents can upload evidences for single request raised by investigator. After uploading phase, all these files for particular request will be process by admin to get the relevant information. This relevant information will be written to other text file and this file will be considered as output of this mode.

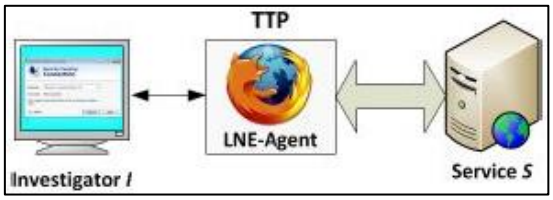

Fig. 2: LNE-Agent Operating Mode

C. Investigation Mode

In above two modes of operation, investigator does not participate in investigation process unlike third operation mode. Savvy users, who can leverage all their experience to perform the acquisition. In this case, the TTP provides a fully-fledged virtual environment where the investigator can go through all the data which is related to request. In this mode, investigator is provided with collection of evidences and from this collection investigator can select the evidences that matches the request. This selected information will be written to a file so that it will be output of this mode.

After all the modes will be executed successfully, investigator will get signatures. These signatures will be used while downloading the evidences. If investigator will enter signature values correctly the system will unite all the output files of three modes into zip file. And then user can download this zip file which contains evidences.

V. ALGORITHM

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

A. 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.

B. Product Functions

Propose 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.

Our application will be client-server-based web application. As it is a web application user can access it from anywhere, anytime. This gives portability to the system.

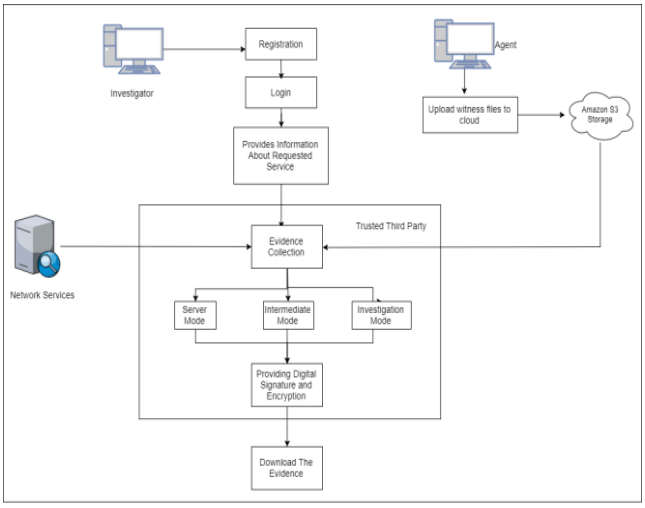

Fig. 3: System Flow Diagram

VI. 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.

ACKNOWLEDGEMENTS

We would like to take this opportunity to thank our guide Prof. N. V. More for giving us all the help and guidance we needed. We are really grateful to him for his kind support. His valuable suggestions were very helpful.

REFERENCES

[1] 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 from: http://nij.gov/topics/forensics/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 SemanticallyRelevantWeb ContentonBigData Framework," IEEE Transactionson 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

