A New Approach for Evidence Gathering Using Data Mining Techniques

Anjali Kadam, Shalvi Jain, Ruchita Mane, Sampada Jain, Sakshi Bhagat

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

***Keywords*: Digital Forensics, Network Forensics, Live Network Evidence (LNE), Big Data Forensics, Digital Investigations.**

Introduction:

Live Network Evidence (LNE) consists of any kind of information that can only be accessed through a network. Nowadays, the common way to produce a trustworthy LNE involves the participation of a "human" Trusted-Third-Party (TTP) in the acquisition process, such as a notary. In that case, the notary should certify with reports, photos, and videos any information of interest obtained by the investigator from the inquired online service. The collected evidence is digitally signed, timestamped and provided to the investigator. The collector produces a report containing high-level information, which can be accessed by non-technical parties (such as judges, juries, lawyers, etc.) without the assistance of technical consultants and/or advanced analysis tools. Finally, the implementation and evaluation of a fully-fledged prototype to collect LNE, called Live Network Evidence Acquisition (LINEA for short), is presented. It has been proven to be forensically-sound, i.e., the collected evidence is robust, and its reliability can be verified at any time after the acquisition. A typical example of LNE may be an information contained in a web page. Digital evidence is supposed to reside on a digital device. Traditional storage media forensics assume that such device is available to the investigator for the analysis. On the contrary, a LNE has been defined as any kind of information flowing through a network. Communications on computer networks are typically based on the client-server model. As a consequence, most of the relevant information on a network is maintained by servers, which can be requested by the clients through a particular protocol. This work focuses on the acquisition of LNE from online services. It is physically stored in a file on the web server, then it is transmitted over the network as a sequence of packets upon a client's request. On the client side, the network flow is reassembled, processed and rendered by the web browser. During this process the initial information (the one stored on the web server) may be transformed by the various elements along the communication path. Sometimes the information presented to the user may be totally different and unrelated to the original. This may happen, for example, in case of a man-in-the-middle attack, where the attacker replaces the server response before redirecting it to the client. This case may be detected if the analyst is able to capture the transient information from the different elements along the communication channel. It allows to correlate this information at different levels in order to increase both the accuracy and the reliability of the collected evidence.

Related work

A. LIVE NETWORK EVIDENCE:

The definition of LNE given so far is very general. The purpose of this section is to clarify the objective of this work. In this sense, programmer need to give a more precise definition of LNE and to evaluate all the issues related to its acquisition and management. In this context, an LNE can be defined as a digital information that holds the following properties: the system containing the evidence is physically inaccessible the target information can be accessed through a network request.

B.THE ACQUISITION OF LNE:

A number of tools for digital investigations have been proposed in the last years to try and solve the technical issues related to the LNE acquisition. These are generically called Network Forensic Tools (NFTs). For sake of clarity, programmer make a distinction between local and remote tools. The first category includes software operating on the investigator's workstation, while the second category comprises tools implemented as a third-party service, which can be accessed by the investigator through the network. In both cases, a human third-party can be employed in order to validate the operations performed by the investigator during the acquisition process.

C. A NEW METHODOLOGY FOR THE LNE ACQUISITION:

In this section programmer present a novel methodology to collect LNE. Although the methodology can be used to acquire information from a generic service on a common network, for sake of simplicity programmer focus on the acquisition of information from the World Wide Web (WWW), whose communication protocol is HTTP.

Motivation:

The motive of the project is to obtain information from trusted third party which collects information on behalf of investigator by using Live Network Evidence method. The investigator establishes the connection with TTP. This connection is secured to maintain privacy.

System Architecture:

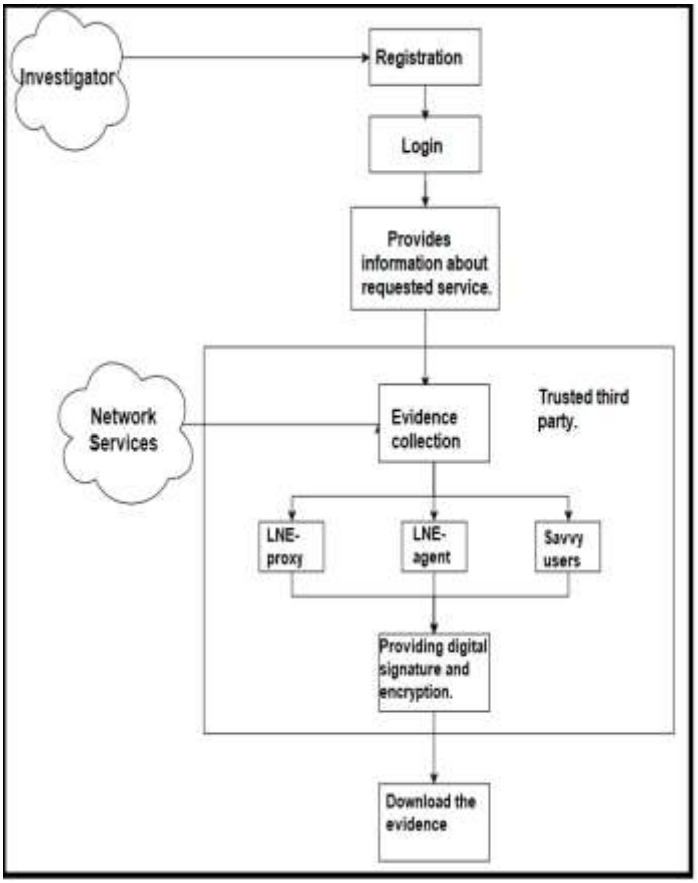

Fig 1. System Overview

Figure shows the flow of our System. The system consists of two phases: The Support System phase and Evidence Collection phase. In Support System phase, user enters username and password that will be store in Log4j file. User can upload the post and delete the post. In Evidence Collection mode there are three sub phases: LNE- Proxy, LNE Agent and Savvy Users. In LNE-Proxy mode Information related networks, servers, etc. is collected. In LNE-Agent mode the information collected by number of different sources is then put under the process of finding correlations. And in Savvy Users investigator is provided some investigation tools to collect furthermore information on his/her own. Generated m1,m2,m3 text file will be uploaded on cloud in encrypted format and after that programmer will provide the Digital Signature for accessing evidence file from cloud. After entering digital signature, programmer can download file in zip format.

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.

References:

[1] 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].
[2] 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
[3] "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
[4] 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.
[5] 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.
[6] B. Blakeley, C. Cooney, A. Dehghantanha, and R. Aspin, "Cloud Storage Forensic: hubiC as a Case-Study," in 7th IEEE International Conference on Cloud Computing Technology and Science, CloudCom 2015, Vancouver, BC, Canada, November 30 - Dec. 3, 2015. IEEE Computer Society, 2015, pp. 536–541. [Online]. Available: http://dx.doi.org/10.1109/CloudCom.2015. 4
[7] S. Nepal, R. Ranjan, and K. R. Choo, "Trustworthy Processing of Healthcare Big Data in Hybrid Clouds," IEEE Cloud Computing, vol. 2, no. 2, pp. 78–84, 2015. [Online]. Available: http://dx.doi.org/10.1109/MCC.2015.36
[8] L. Zhao, L. Chen, R. Ranjan, K. R. Choo, and J. He, "Geographical information system parallelization for spatial big data processing: a review," Cluster Computing, vol. 19, no. 1, pp. 139–152, 2016. [Online]. Available: http://dx.doi.org/10.1007/s10586-015-0512-2
[9] D. Quick, .-a. Martini, Ben, a. Choo, Kim-Kwang Raymond, and EBSCOhost, Cloud Storage Forensics. Amsterdam Boston Elsevier/Syngress, 2014. [Online]. Available: http://site.ebrary.com/id/ 10810980
[10] N. H. Ab Rahman, N. D. W. Cahyani, and K.-K. R. Choo, "Cloud incident handling and forensic-by-design: cloud storage as a case study," Concurrency and Computation: Practice and Experience, pp. n/a–n/a, 2016. [Online]. Available: http://dx.doi.org/10.1002/cpe.3868
[11] N. D. W. Cahyani, B. Martini, K.-K. R. Choo, and A. M. N. Al-Azhar, "Forensic data acquisition from cloud-of-things devices: windows smartphones as a case study," Concurrency and Computation: Practice and Experience, pp. n/a–n/a, 2016. [Online]. Available: http://dx.doi.org/10.1002/cpe.3855
[12] B. Martini and K.-K. R. Choo, "Distributed filesystem forensics: XtreemFS as a case study," Digital Investigation, vol. 11, no. 4, pp. 295– 313, 2014. [Online]. Available: http://www.sciencedirect.com/science/ article/pii/S1742287614000942