Digital Evidence Analytics Using NLP and Security Algorithms

[1]Dr. Swapnaja Ubale, [2]Miss. Ankita Wankhade, [3]Mr. Shubham Shinde, [4]Miss. Sharayu Kachi

[1] Assistant Professor, [2,3,4]UG Student, [1,2,3,4]I.T.Departmet, Zeal College of Engineering and Research, Pune, Maharashtra, India.

Abstract: Digital forensics has numerous application areas amongst digital evidence is one of the fields of forensic investigation which has vital importance. Many issues arise while dealing with network evidence. It becomes troublesome to gather network evidence as the network is volatile. Often, such a piece of information could vary according to time or maybe situated on a remote server that requires an authorization to get access. In this paper an approach is presented called Evidence Gathering to collect network evidence. Particularly, web pages, photos, chats or videos would be taken into consideration as an origin for obtaining data. Such an approach is adequate to intellectuals as well as non-intellectuals because the user is considered throughout the method called evidence acquisition. The information obtained from the remote source such as network packets and any information collected by the user is automatically acquired throughout this process. Trusted Third Party (TTP) acts as a digital lawyer not only to ensure the collected evidence but also to the acquisition process.

Keywords: Live Network Evidence (LNE), Digital Investigations, Digital Forensics,
Trusted Third Party (TTP).

Introduction:

Live Network Evidence (LNE) [1] is formed using information which is accessible through the web. The contribution of a Trusted Third Party (TTP) [1] during the process of acquisition, for instance, a lawyer is a conventional way to generate a trustworthy LNE. In such a situation, the lawyer ought to verify by papers, pictures, videos and info collected by an investigator through that of the examined web-based services. Once the evidence gathered, it is digitally approved, time stamped moreover given to the investigator. The one who generates a document including leading information, that is accessible via laypeople besides any guidance of specialized experts or superior tools known as a collector. Ultimately, execution as well as the assessment regarding a fully developed model to obtain LNE [1], named Live Network Evidence Acquisition (LINES) [1], is exhibited. The gathered evidence comprises robustness furthermore its authenticity could be checked once the process of acquisition is complete. A common example of LNE is nothing but data or an information stored on the internet in the form of web page. It is assumed that digital evidence resides on a digital device. Traditional warehouse forensics believes the same device is accessible to the investigator for the investigation. In contrast, an LNE has been described as group of information passing via the web. The client-server model is responsible for communication on computer networks. The information which is inquired by the clients via an appropriate protocol is saved on a web controlled by servers. The main focus is on the acquisition of LNE. First, they are saved into the repository of the webserver which will later be broadcasted across the web like a series of packets once it receives the request of a client. Primary information which is stored on the web server may be modified by the several components with the transmission path throughout this process. Frequently the information given to the user can be altered and immaterial as compared to the original information. There can be a possibility that the server response is replaced by the attacker before forwarding it to the client. The role of an analyst comes into the picture to solve this issue by capturing the imgrating information from the various components with a

communication medium. It enables connecting this information at various levels to enhance efficiency as well as the security of the gathered evidence.

Related work

A. Live Network Evidence (LNE):

The concept of LNE [1] presented to date is quite common. The need of this secton arises to clear the goal regarding the work. Therefore it is necessary to give an exact description concerning LNE [1] to assess whole problems associated with its acquisition and administration. An LNE [1] can be represented as a digital information which contains properties such as: (1) the system comprising geographically inaccessible evidence (2) the destination information can gain access by requesting to the web.

B.The aquisition of LNE:

To analyze and resolve the technical difficulties associated with the LNE acquisition [1] several mechanisms for digital researches have been introduced in the past decade. These are generally termed as Network Forensic Tools (NFTs) which can be further divided into local tools and remote tools. The local tools involve software running on the workstation of an investigator, whereas remote tools consist of devices executed as a third party service, accessible by the investigator via the web. In each of two instances, a social third party could be hired to authenticate the services offered by the investigator throughout the process of acquisition.

C. Modern approach for LNE acquisition:

Here, we introduce the modern method to obtain LNE [1]. Although this method may be adopted to collect information from a general system, for ease we concentrate on the acquisition of information from the world wide web (www) [1], which uses Hyper Text Transfer Protocol (HTTP) as a transmission protocol.

Motivation:

The goal of this project is to get data from the Trusted Third Party (TTP) [1] who gathers data on account of the investigator applying the Live Network Evidence (LNE) [1] process. The investigator authenticates a relationship with Trusted Third Party (TTP) [1]. This relationship is secure to preserve isolation.

System Architecture:

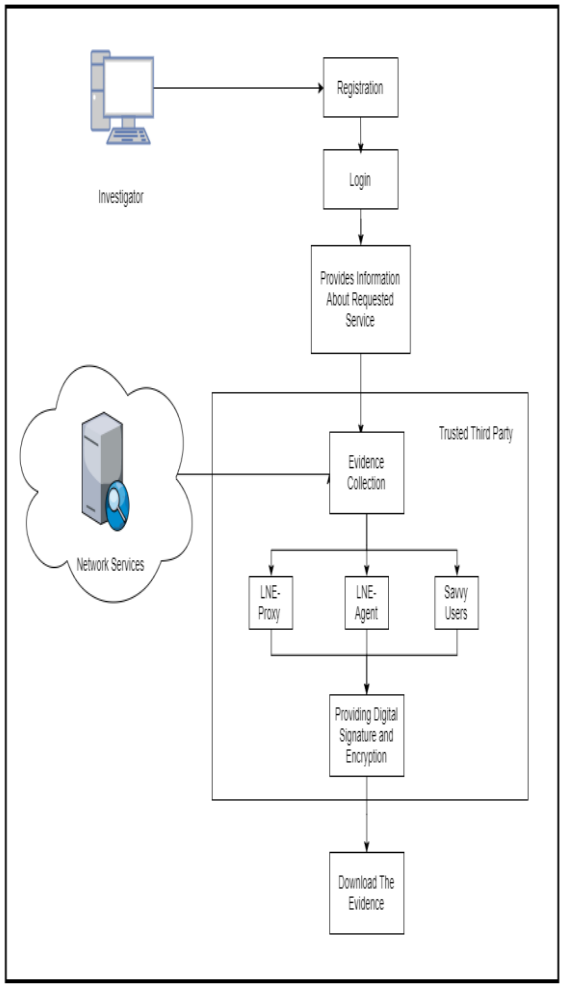

Fig 1. System Overview

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 we will provide the Digital Signature for accessing

evidence file from cloud. After entering digital signature, we can download file in zip format.

Algorithms:

A. Natural Language Processing (NLP):

Natural Language Processing (NLP) provides way for computers to examine, understand and obtain meaning from human language in an efficient manner. It allows extraction of all kinds of text properties. In our project, we used the NLP algorithm to categorize comments into different categories such as harmful, very harmful, neutral, positive and very positive.

pseudo code -

Input: Comments in the form of words, sentences or paragraphs.

Output: Prediction of comments such as harmful, very harmful etc.

Step1: String sentiment1 ="";

Properties props = new Properties();

Step2:

```
props.setProperty("annotators","tokenize, ssplit,
parse, sentiment");

StanfordCoreNLP pipeline = new
StanfordCoreNLP(props);

int mainSentiment = 0;
```

Step3:

```
if (line != null && line.length() > 0)
{
        int longest = 0;
        Annotation annotation =
pipeline.process(line);
        for(CoreMap sentence :
annotation.get(CoreAnnotations.SentencesAnnotatio
n.class))
        {
                Tree tree =
        sentence.get(SentimentCoreAnnotations.Senti
        mentAnnotatedTree.class);
        int sentiment =
edu.stanford.nlp.neural.rnn.RNNCoreAnnotations.get
PredictedClass(tree);
        String partText = sentence.toString();
        if (partText.length() > longest)
        {
                mainSentiment = sentiment;
                longest = partText.length();
        }
}
}
```

Consider a scenario in which user posted a comment, “You are stupid”.By using the following steps NLP will produce a result, “very harmful” because of negative comments.

Step 1: Read some text from the user.

Step 2: tokenize, ssplit, parse and sentiment annotators are used for classifying comments into either harmful, very harmful, positive, very positive or neutral.

Step 3:

Step 3.1: tokenize and ssplit

To tokenize the text tokenize annotator is used. It splits the text according to the language’s rule. After performing this step, the output will be -

Input - You are stupid

Output –

You
are
stupid

Step 3.2: parse

Parse annotator will represent each word in the form of a tree structure according to a noun, adjective, adverb, etc. This is shown in the following fig.

Mathematical Model

Let, S be the System Such that,

Where S = {I ,F,O ,DD ,NDD ,Success , Failure}

Where,

I= Input, F=function, O=Output,DD=Deterministic Data, NDD=Non Deterministic Data}

I= {IS,IIS,IR,IC,IP,IUA}

Where,

IS ->User

IIS->post

IR->comment related to their topic

IUA->user name and password

Input:

I=. Set of input i.e., Tweets.

Function:

F1=Save log4 file Function (This function is used for save login user and password)

F2=Track user info Function (This function is used for detecting)

F3= Get user info from log file Function (This function is used for finding getting data)

Output:

Get anomaly list

O1=Success Case (It is the case when all the inputs are given by system are entered correctly)

O2=Failure Case (It is the case when the input does not match the validation Criteria)

Conclusion:

A concept called the acquisition of Live Network Evidence (LNE) [1] from online services is built upon the Trusted Third Party (TTP) [1]. TTP obtains the data on account of the investigator. There are three different modes in which evidence would be collected. A piece of evidence gathered by TTP is reliable and its efficiency could be monitord later. The evidence is more vivid and its completeness, as well as reliability, could be assured.

References:

[1] Aniello Castiglione, Giuseppe Cattaneo, Giancarlo De Maio, Alfredo De Santis, Gianluca Roscigno, "A Novel Methodology to Acquire Live Big Data Evidence from the Cloud", IEEE TRANSCATIONS ON BIG DATA, VOL.X,NO.Y, JULY-SEPTEMBER 2017

[2] Ben Blakeley, Chris Cooney, Ali Dehghantanha, Rob Aspin, "Cloud Storage Forensic: hubiC as a Case-Study", 2015 IEEE 7th International Conference on Cloud Computing Technology and Science

[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.