Evidence Collection from Web Browsers using Data Mining Technique

Dayananda R B[1]Kavyashree Jayaprakash[2]Harish S[3]Gowtham A[4]Harsha S Gowda[5]

Abstract—*Increase in internet technology cyber-crime are being increased day by day and committed by attackers. Data recovery methods and practical frameworks are used for investigation. In cyber-crime huge amount log date is been collected, transaction of data leads to storage of large amount of data and analyze them.[1] It is difficult for Forensic investigators to find out the clue and analyze the data. In log files large amount data is generated in every of action so it is difficult for forensic investigators to find out clue and analyzing the data. This paper focuses on collecting the data from cyber system and web browsers, forensic analysis and remote system forensic which is to be used as evidence for detecting the suspect during the investigation[2]. Decision tree is one of the technique which can help for forensic investigation purpose so system can adopt a way by which using decision tree for generating, storing and analyzing the data retrieved from log files. This paper focuses on how decision tree can allow system to quickly, easily and inexpensively analysis of log data available in various file formats for file forensic analysis[3].*

Keywords—Data Collection; Log Data Collection; Digital Forensic Tool; Clustering

I. INTRODUCTION

Digital forensics is the branch of forensic science recovery and investigation for digital devices often done in computer crime. Digital forensic is a synonym for computer forensics and expanded to cover investigation done for storing the digital data. Digital forensics have a variety of applications. This is used to support hypothesis before criminal courts. In the form of technical investigation aspect is divided into several branches related with the digital device involved .The different types of forensic science are network forensic, forensic data analytics and mobile device forensics .Identification of evidence of a crime digital forensics has been used to attribute evidence to suspects. Investigation are much broader and scope than any other areas of forensic analysis involving complex times on hypothesis. The next few years computer crimes are increased and laws are passed to deal the issues of copyright privacy and child pornography.

Ex. Cyber-crime, cyber stocking , online predictors[5].

Recovery of data from digital devices is nothing but digital devices, Different Tools and applications are used for digital forensic but there are certain limitations .Technical challenges implement small scale of data mining in which decision tree can support for efficient classification of data[6].

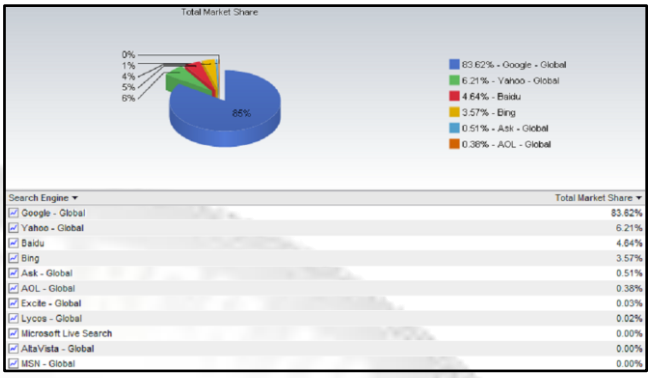

Fig.1: Global market share of search engines

Decision Tree (DT) technique is one among which can help for file forensic investigation purpose. Every system can use Decision Tree technique for generating, storing and analyzing large amount of data retrieved from log files which pose as evidence in file forensic analysis. This paper focuses on how Decision Tree can allow system to quickly, easily and inexpensively analysis of log data available in various file formats for file forensic analysis[4]. General methodology is being used for digital forensic analysis as follows

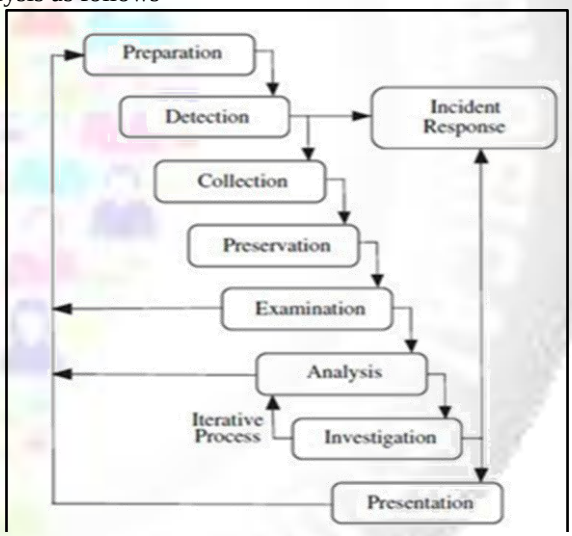

Fig.2: General methodology for digital forensic analysis

Preparation: In the preparation phase, the data is obtained and stored for backup in storage devices. A hash of all the trace data is preserved. A copy of the data will be analyzed and the original network traffic data which is not alter by hacker.

Detection: Once log file gathered, the next step is to recognize the presence of nature of an attack. If there is any suspicious activity, then type of an attack can be detected

Generation: generation of huge amount of data which requires amount of storage memory and the system must be able to handle different log data formats appropriately.

Examination: Once data get collected from different nodes it will get integrate into large dataset. This

large data set again get checked to see whether any exception is present.

Analysis: The indicators are classified and correlated to deduce important observations using the existing attack patterns. The attacks patterns are put together reconstructed and try to understand what the intention of attacker and methodology used is.

Investigation: The goal is to determine the path from a victim network or system through any intermediary systems and communication pathways, back to the point of attack beginning. The packet captures and statistics obtained are used for attribution of the attack.

Presentation: The conclusions are also presented using presentable format so that they can be easily grasped and make their decision.

Above Fig.2 shows, the general methodology to arrive at thevictim system. Among all steps, analysis step is an importantstep which helps to detect a crime activity and take a decision according to that.

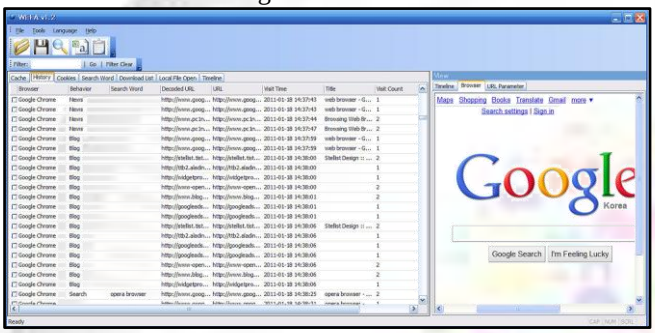

Fig.3: Web Browser Forensic Analyzer

II. LITERATURE SURVEY

It has also become important to extract information from these huge databases that might be valued to the owner of the data base. In network forensic analysis, it focused on the visual representation for network component for more easily analysis of evidence to the examiner[4].The visualization effect build to provide ease of understand the evidence or data. Forensic investigation process can be improved by using data mining technique. Collected evidences should be cross checked to each other it may cause investigator get some clues.[5]Different type of data mining strategies can be used for forensic analysis. K-means and a prior algorithm work well for clustering similar kind of data[6]. Proposed framework analyze huge amount of textual data using weka. Weka provides various machine learning concepts to analyze the collected data[7]. Framework using text mining technique. Using this text mining technique it's become easy to forensic investigator to find out particulars identity with the help of social networking sites. Text mining technique mine the data of particulars on different two or more social networking site. Different type of tools and methods are available to recognize the pattern of an attack. Handling complex and huge data for forensic investigation purpose become difficult task. Crime activity investigation involves memory forensic, network forensic, file forensic. So investigator must have to consider all this concepts while forensic investigation analysis[8].Oracle Express Edition (10g) software has been used to load the data for pre-process further. But there are certain limitations with this software's like Recuva is unable to recover or extract all files from flash drive. Hence the extracted data is not complete data through which crime activity get recognised[9].

III. DATA MINING TECHNIQUE TO DETECT AN ATTACK

Data are grouped together on the similarity scheme. Clustring is one of the data mining technique which make data instances into clusters of significant interest and evaluate the performance of the system. This proposed system implements clustering of data as normal users or attackers by using simple k-means algorithm[9].

K-means algorithm takes k, number of clusters is determined as an input parameter and partitions the given set of n objects into k clusters so that the resulting intra- cluster similarity is high while the inter-cluster similarity is low. This algorithm is used to minimize the means squared Euclidean distance of the data, from the center of their clusters. Decision tree has the capacity to classify the huge amount of data.

IV. PROPOSED SYSTEM

The proposed system proposes a stratergy which helps for the forensic investigators for the forensic investigation.The proposed system takes the log files as an input and used for forensic investigation and uses the data mining technique for this purpose. Using crime data mining system the nature of the attack is identified and alert administrator about similar attacks also our proposed system will helps to increase the security of the organization. The block diagram of the system consists of the following

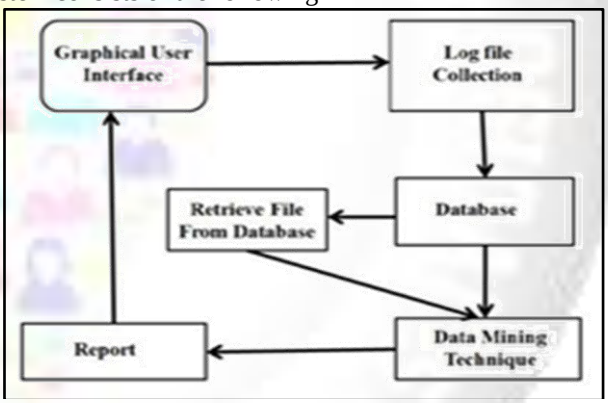

Fig.4: Block Diagram of proposed system

Graphical User Interface(GUI):It acts an interface between the system and the investigator. It provides an interface for proving the input for the desired output. The graphical interface makes the investigators easy for the forensic investigation and for finding the evidences from the system where crime occurs. This is also shows the result of processing data in presentable form.

Log file collection: In this the the log files are collected from the browser history. The collected data from this collection helps the investigator to collect the evidences related to crime. This data collected is stored in the database.

Database: Proposed system used the database to save browser history as evidence, provides ease to store data in graphical as well as in graphical format. The database provides flexibility to store the data.

Data Mining Technique: This technique is used to analyze the collected data. It consists of various modules

like association, characterization, cluster analysis, classification .The first step is to run undetermined clustering algorithm using K-Means. Then system uses the classification algorithm to verify the visualization pattern of the data instances and detect a type of an attack. Data pattern of attacker's data will get match with pattern in training dataset. This training data help to recognize the type of an attack.

Report: Finally system generates the report which shows the type of an attack has been occurred. Generated report shows the result of the analysis

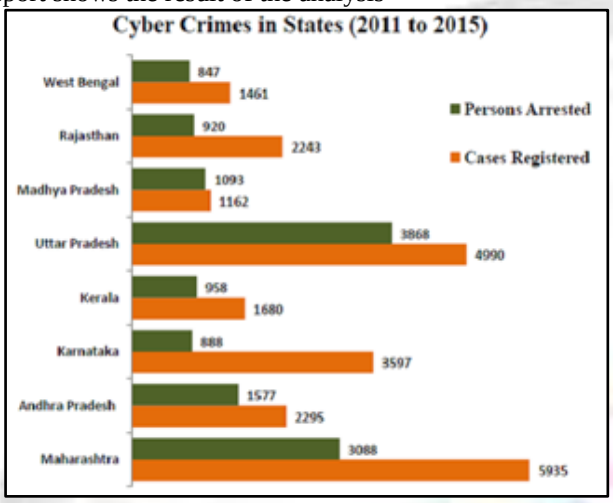

Fig.5: Cyber Crimes in states over the year 2011 to 2015[13]

V. CONCLUSION

In this paper cyber-crime data is collected with using of proposed methodology, the log file is captured and stored in database as evidence. The generated file and data is analysis with using digital forensic toolkit. Forensic toolkit is used to analysis of victim system where the attack is happened[15]. The physical memory data and logical memory data is analyze and find an evidence which help in crime investigation. Digital forensic tries to analysis network traffic data. To, implement file forensic analysis and network investigation analyzer in efficient way system applying data mining techniques which will help to forensics investigator to detect a crime activity. In system, the Decision Tree works as classifier used to analyse where an attack is happened and also the type of an attack.

REFERENCES

[1] Latesh G. Malik,"A Review on Data Generation for Digital Forensic Investigation using Datamining", IJCAT International Journal of ComputingandTechnology, Volume 1, Issue 3, April 2014.

[2] K. K. Sindhu, B. B. Meshram, " Digital Forensics and Cyber Crime Datamining, Journal of Information Security, May 2012.

[3] ChrysoulaTsochataridou, AviArampatzis, VasiliosKatos,"Improving Digital Forensics Through Data Mining", IMMM 2014, The Fourth International Conference on Advances in Information Mining and Management, September 2016.

[4] Daniel Compton, J.A. Hamilton," An Examination of the Techniques and Implications of the Crowd-sourced Collection of Forensic Data", IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, April 2014.

[5] SonalHonale , JayshreeBorkar," Framework for Live Digital Forensics using Data Mining", International Journal of Computer Trends and Technology (IJCTT) volume 22 Number 3, April 2015.

[6] Jooyoung Lee, SungkyungUn, and Dowon Hong, "Improving Performance in Digital Forensics", International Conference on Availability, Reliability and Security, 2009.

[7] Veena H Bhat, Prasanth G Rao, Abhilash R V,P DeepaShenoy,Venugopal K. R. "A Novel Data Generation Approach for Digital Forensic Application in Data Mining", Second International Conference on Machine Learning and Computing, 2010.

[8] Sebastian Schmerl, Michael Vogel, René Rietz, and HartmutKönig, "Explorative Visualization of Log Data to support Forensic Analysis and Signature Development", Fifth International Workshop on Systematic Approaches to Digital Forensic Engineering, 2010.

[9] FunminiyiOlajide, Nick Savage, Richard Trafford, "Forensic Memory Evidence of Windows Application", The 7th International Conference for Internet Technology and Secured Transactions (ICITST-2012).

[10] Jooyoung Lee, SungkyungUn, and Dowon Hong, "Improving Performance in Digital Forensics", International Conference on Availability, Reliability and Security, 2009.

[11] Veena H Bhat, Prasanth G Rao, Abhilash R V,P DeepaShenoy,Venugopal K. R. "A Novel Data Generation Approach for Digital Forensic Application in Data Mining", Second International Conference on Machine Learning and Computing, 2010.

[12] Sebastian Schmerl, Michael Vogel, René Rietz, and HartmutKönig, "Explorative Visualization of Log Data to support Forensic Analysis and Signature Development", Fifth International Workshop on Systematic Approaches to Digital Forensic Engineering, 2010.

[13] FunminiyiOlajide, Nick Savage, Richard Trafford, "Forensic Memory Evidence of Windows Application", The 7th International Conference for Internet Technology and Secured Transactions (ICITST-2012).

[14] Seung-hoon Kang, Juho Kim, "Network Forensic Analysis Using Visualization Effect", International Conference on Convergence and Hybrid Information Technology, 2008.

[15] Daniel Compton, J.A. Hamilton, "An Examination of the Techniques and Implications of the Crowd-sourced Collection of Forensic Data", IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, 2011.