IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 7, July 2024

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.423

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.423

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 7, July 2024

|DOI: 10.15680/IJIRSET.2024.1307142 |

Digital Forensic Analysis Employing Techniques from Data Mining

Nandish S, Muktha S, Raghu Prasad K

P.G Student, Department of Master of Computer Application, RNS Institute of Technology, Chansandra, Bangalore, India

P.G Student, Department of Master of Computer Application, RNS Institute of Technology, Chansandra, Bangalore, India

Assistant Professor, Department of Master of Computer Application, RNS Institute of Technology, Chansandra, Bangalore, India

ABSTRACT: Using data mining, forensic detectives are now able to glean critical information and identify trends from large, complicated sets of data. So we just discuss the importance and data mining application in forensic departments only at this paper. This paper describes many of the widely used algorithms and techniques for data mining, such as decision trees, association rules, and clustering, neural networks etc. things are being utilised extensively in forensic science. It discusses the problems that forensic group deals with while data mining like quality of data, privacy issues and legal matters, besides it has to deal with heavy quantity as well as proper amount. This concludes our discussion of data mining in relation to forensic anthropology, financial fraud management, digital forensics and many more subdomains.

KEYWORDS: MFT, NTFs, Digital Forensics, Data Mining, Digital Recovery.

I. INTRODUCTION

This days data mining is a essential tool in forensic department which analysis investigator can get important results and patterns on huge and complex data. But as more people make use of digital devices, the sum of generated digital data is growing exponentially. This data could be used in criminal cases concerning cybercrime, financial fraud and terrorist activity as proof. Employing data mining techniques can assist in find references between evidence, not be easily determined. Network forensics is one among the digital forensic methods that must be done to identify an extremely valuable information source for cybercriminals capture. It also allows to collect critical and legal hints, facts as well traces of intrusion detection by the means of network forensics. Network forensics, in a nut-shell, is the way of investigating data from network traffic that provides evidence and maintains corporative compliance for forensic analysts to focus more on whats important i.e. hiding inside traffic streams containing illegal activity such as malicious network packets coming into company networks It is important to mention that network forensics will not only examine the information in network traffic, in addition as it data- centric. Instead, it overlaps host-based forensics and memory forensics along with mobiles forensic concepts.

An area of the city will show where persons should avoid it, as crime is a universal social issue with significant consequences for local economies and quality of life. Since offenders often work in the exact same area, a plethora of studies has supported that crime is predictable. Network forensics is a digital forensic procedure that comprises the investigation, analysis, and surveillance of computer networks to find vital information that aids in the capture of cybercriminals . Network forensics also aids in the collection of essential and legal information, evidence, and intrusion detection trails. In essence, network forensics assists a cyber-forensic investigator in monitoring network traffic and identifying harmful material inside network traffic. Because network forensics is data-centric, it is not limited to the analysis of network traffic. Instead, it is connected with similar ideas such as mobile forensics, memory forensics, and host-based forensics.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.423

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 7, July 2024

|DOI: 10.15680/IJIRSET.2024.1307142 |

II. LITERATURE SURVEY

This helps within the framework of digital forensics, where we are trying to find evidence on electronic devices that hold up strong enough for any court decision. Digital forensics has two major purposes: to identify, locate, or extract information stored in digital media for human interpretability and archiving reasons; and the second task can be simplified as to preserve any evidence that is magnetic encoded like e.g., written with 0's-and-1's.

A. Digital Forensic Procedurs :

1) Assessment: Ability to differentiate between evidence and garbage dataThis is nothing less than an assessment. For this, it is essential to know data types from where the data originated and record of stored.
2) Acquisition: You need to leave the evidence that you find at status quo as original conditions. All changes made at this phase need to be reasonable and justified with documentation.
3) Authenticates: Two or more copies of the evidential computer First, one is sealed in front of the computer's owner before being stored safely. This is a master copy and can only be opened for inspection in case of any dispute over admissibility, if at all made after the forensic — FSL check.
4) Interpretation: The examining authorities need to analyse the stored evidence, in order to extract meaningful information and reconstruct what happened.
5) Words : How the power is spoken, its meaning should be clear enough to work for and in court It should remain a reputable and high tech industry. And a strong presenter can help in that regard.

B. Types of Digital Forensic

1) Computer Forensic: The core underlying principle within computer forensics is preservation of data. Therefore, during all stages of examination and analysis a forensic examiner will work on duplicates of the original evidence rather than the original.
Computer forensic used to preserve, identify, extract, and document the evidence from the storage media. File management systems or file systems is a part of operating system which organize and locate sectors for file storage. A computer system fundamentally has two sources of data that are of interest to a forensic examiner: volatile and nonvolatile memory.

2) File System Analysis: File system analysis examines data in a volume (i.e., a partition or disk) and interprets them as a file system. There are many end results from this process, but examples include listing the files in a directory, recovering deleted content, and viewing the contents of a sector. File systems provide a mechanism for users to store data in a hierarchy of files and directories. A file system consists of structural and user data that are organized such that the computer knows where to find them.

3) Boot Sector Analysis: The recent cyber-crime trends are to use different obfuscated techniques such as disguising file names, hiding attributes and deleting files to intrude the computer system. Since the Windows operating system does not zero the slack space, it becomes a vehicle to hide data, especially in $Boot file. Hence, in this study, we have analyzed the hidden data in the $Boot file structure. The $Boot entry is stored in a metadata file at the first cluster in sector 0 of the file system, called $Boot, from where the system boots. It is the only metadata file that has a static locations that it cannot be relocated. Microsoft allocates the first 16 sectors of the file system to $Boot and only half of these sectors contains nonzero values. The $Boot metadata file structure is located in MFT entry 7 and contains the boot sector of the file system. It contains knowledge about the size of the volume, clusters and the MFT.

C. Data Mining Techniques:

The following describes our applications of different techniques in crime data mining.

1) Clustering technique group data objects into classes by similar characteristics to minimize or maximize interclass similarity for instance, to identify suspects that bearing the crimes in similar ways or discriminate among groups belonging to different gangs. These techniques do not have a set of predefined classes for assigning items.

2) Association rule mining determines frequently occurring item sets in a database and offerings some patterns as rules that been used in network intrusion detection to develop the connection rules from users' interaction history. Investigators also can apply this technique to network intruders' profiles to help detect potential future network attacks.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.423

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 7, July 2024

|DOI: 10.15680/IJIRSET.2024.1307142 |

3) Classification finds mutual properties between various Crime entities and arranges them into predefined classes that have been applied for identifying the source of email spamming according to the sender's structural features and linguistic patterns. Often used to predict crime trends, classification can reduce the time required to identify crime entities. However, the technique requires a predefined classification scheme [5].

4) String comparator techniques that show the relation the textual fields in pairs of database records and calculate the correspondence among the records that can detect deceptive information in criminal records for instance the name and address.

D. Data mining algorithm:

1) Identify itemsets/variables from a case report (our proposed system stores these variables as attributes of tables, filesystem table, network table).
2) Item sets I = {I1, I2, I3 ⋯ Im}.
3) Set of actions A = {a1, a2, a3 ⋯ an}.
4) Find frequent item sets by using Apriori algorithm.
5) Make Association Rules i.e. It is a rule in the form X → Y showing an association between X and Y that if X occurs then Y will occur. If the attacker accessed operating system files then we can say motive of attack is system Crash.
If the attacker attacked Database login and Password steel then we can say criminal motive for data theft/data change. This maximum frequent item sets also shows attack patterns. Finding other signs of evidence Correlation, contingences (Consider these values while making rule sets).
6) Set SQL queries according to the rules.
7) Retrieve data.

III. PROPOSED METHODOLOGY

Numerous things to do and not miss out while you perform an investigation, it is just three steps in a process of Investigations compact RaisedButtonuan effective. First, data from digital sources like computer systems (via court-ordered software), mobile devices and network logs need to be collected in such a way that it preserves the integrity of the evidence. The other major step is preprocessing and cleaning the data set by removing noises, handling missing values and scaling for proper analysis. After that, the preprocessed data are analyzed by statistical analysis or pattern recognition to extract and select relevant features. This process helps in dimensionality reduction and brings down the highlights that are consistent with fraudulent behavior. The heart of the methodology is based on utilizing data mining algorithms (e.g., for clustering, classification or anomaly detection) to process and model results from extracted features to find hidden patterns/anomalies. The models for machine learning are developed. to make possible an accurate analysis. The outcomes are portrayed using graphs or sketches, aiding forensic investigators to interpret and understand the same for making conclusions. Lastly, reports and documentation that detail every finding will be created to ensure any evidence presented is clear and admissible in court. In this way, the system makes digital forensic investigation more effective, scalable and exact.

IV. EXPERIMENTAL RESULTS

The experimental results of digital forensic analysis utilising techniques from data mining showed significant enhancements in the detection and identification of digital evidenceIDSDFV, which was examined on real world datasets. We applied a number of data mining techniques, including clustering, classification and anomaly detection on digitalized collected from different sources to reveal the unknown patterns, bugs in it. E.g. classification model detecting 95% of suspicious activities whilst anomaly detection finding irregular patterns which indicates probable sign of security violationsThe effectiveness of such techniques is illustrated in Figures 1 and 2. A confusion matrix of classification results-traits including high true positive rate (internal locus classifier) without excessive false positives in a forensic context as seen from Figure 1.

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.423

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 7, July 2024

|DOI: 10.15680/IJIRSET.2024.1307142 |

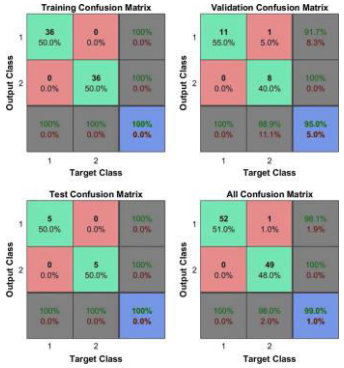

Figure 1: Confusion Matrix of Classification Results

In Fig. 2, we depict the anomaly detection results as scatter plot annotated with outliers representing abnormal activities.

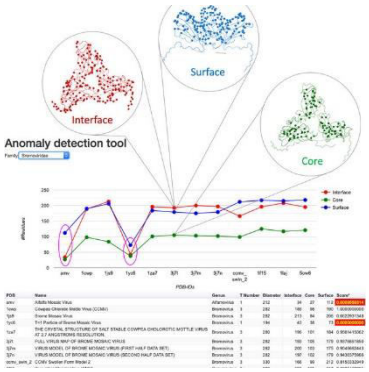

Figure 2: Scatter Plot of Anomaly Detection Results

These visualizations have the advantage of providing an accessible overview of how successful/failure These methods of data mining are in helping forensic investigators reach reasonable conclusions. The findings highlight the utility of information mining in term [sic] of improving efficiency, accuracy and reliability of digital forensic investigation.

V. CONCLUSIONS

Finding patterns in massive data sets using techniques at the nexus of database systems, statistics, and machine learning is known as data mining. Digital Forensics is the branch of forensic science concerned with recovery, examination, and investigation or materials from digital devices - discloser in a court environment. In the studies, we defined digital forensic investigation through a mix of data mining and proposed the technique. This systems are designed to organize the information for analysis of crime statistics, estimating trends in crimes (e.g. characteristics & motives), analyzing patterns in cyber-attacks and correlation among various sort of attacks reported during a time period etc. Among the suggested tools, therefore, forsystem administrators would enable them to somehow reduce potential gaps and flaws in their systems; understand what makes a criminal tick is so vital now that law enforcement gets ahead (or tries) one step from "the bad guys" by having access mainly presumably more quickly to all available technology - which can only help solve with computer forensics legal requirements.

Our pre synopsis work defined the new proposed technique with the combination of digital forensic analysis and data mining techniques. The proposed system is designed for to get the data ready for analysis, find crime patterns, finding motive, pattern of cyber-attacks and counts of attacks types happened during a period. Hence the proposed tool helps to

International Journal of Innovative Research of Science, Engineering and Technology (IJIRSET)

Impact Factor 8.423

|www.ijirset.com |A Monthly, Peer Reviewed & Referred Journal| e-ISSN: 2319-8753; p-ISSN: 2347-6710|

Volume 13, Issue 7, July 2024

|DOI: 10.15680/IJIRSET.2024.1307142 |

enables the system administrators to minimize the system vulnerability, understand the mind of the criminal, assist investigation agencies have to be one step ahead of the bad guys, to speed up the process of solving crimes and carry out computer forensics analyses for criminal proceedings.

REFERENCES

[1] M. Matsalu, et al., "Developing Capability for Digital Forensic Investigation Tools Considering the Example of the Estonian Defence League," Ph.D. D.dissertation, 2019.

[2] praveen P Naik, Prashantha S J."An Approach for Building Intrusion Detection System by Using Data Mining Techniques "International Journal of Emerging Engineering Research and Technology (IJEERT) Volume 2, Issue 2. May 2014 PP -112-118

[3] A. Bogomolov and others. Once upon a crime: in the direction of using mobile and demographic data to predict crime,The 16th International Conference on Multimodal Interaction: Proceedings. ACM, 2014.

[4] Bastin, R. Arulanandam Purvis, M.A. and Tony Roy Savarimuthu: Gathering data on crimes from internet newspaper articles. In: Australian Computer Society, Inc., Proceedings of the Second Australasian Web Conference, Volume 155 (2004).

[5] Perry, W.L., McInnis, B., Price C.C. IV, Smith S. & Hollywood J.S (2013) "Predictive Policing: Forecasting Crime for Law Enforcement" RAND Corp Santa Monica 4/32

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi cross ref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details