An Evolution of Forensic Data Analysis and Methodologies

S. Brintha Rajakumari

Department of CSE,
B. I. S. T. (Bharath University), Chennai, India

Abstract: Every day huge amount of data is generated from various sources. For the past few years, the amount of data that an organization uses and stores has been increasing rapidly. Big data is the mixture of any type of data. This paper seeks to provide basic information regarding the evolution of big data analytic and methodology of forensic analytic model. This paper presented a survey of the area and an introduction into the vast and complicated area known as forensic data analytics.

Key words: Big Data - Computer Forensic.

INTRODUCTION

Big Data is a term for a collection of data sets which are so large and complex which includes climate sensors devices, social media web sites, digital image pictures and videos, purchase transaction records and cellphone GPS signals. The term Big Data refers to large-scale information management and analysis technologies that exceed the capability of traditional data processing technologies [1-3].

Big Data is differentiated from traditional technologies in three ways: the amount of data -volume, the rate of data generation and transmission -velocity, and the types of structured and unstructured data -variety [4].

Computer forensics requires applying computer science to answer legal questions. Arranging events chronologically is a good way of telling a clear, brief story. As valued as date and time based information often is to a case, none of the leading Forensic tools offer usable date and time oriented tools. Log2timeline is an excellent tool for extracting date and time based information from digital evidence [2].

Forensic Data Analytics as a topic and its adoption within the industry had long been overdue. With the advent of technology and increasing incidents of fraud, there has been a significant rise in adoption of Forensic Data Analytics. Due to this, company appointed auditors and independent directors are now seeking to implement proactive fraud-prevention solutions and are avoiding post incident remediation processes.

Forensic Data Analytics is a science used to proactively seek opportunities to prevent and detect fraud, waste and abuse by leveraging information in corporate data assets. It enables identification of meaningful patterns and correlations in existing historic data to predict future events and assess the reasons for various fraudulent activities. Such insightful predictive information is generally “invisible,” but provides a platform on which organizations can take business decisions related to fraud, disputes and misconduct.

Big Data Analytics: Analysis is based on a large population of transactions instead of sample. It is a process of examining large amount of data of a variety of types of uncover hidden patterns, unknown corrections and other useful information. Act of transforming data with the aim of extracting useful information and facilitating the achievement of factual conclusions. Characterization of Big –Data is Volume, Velocity and Variety (V3) [1].
Computer Forensics: Computer forensics is a new and fast growing field that involves carefully collecting and examining electronic evidence that not only assesses the damage to a computer as a result of an electronic attack, but also to recover lost information from such a system to prosecute a criminal. With the growing importance of computer security today and the seriousness of cyber crime, it is important for computer professionals to understand the technology that is used in computer forensics [5].

Computer forensics involves the preservation, identification, extraction, documentation and interpretation of computer data [6]. The three main steps in any computer forensic investigation are acquiring, authenticating, and analyzing of the data. Acquiring the data mainly involves creating a bit-by-bit copy of the hard drive. Authentication is the ensuring that the copy used to perform the investigation is an exact replica of the contents of the original hard drive by comparing the checksums of the copy and the original. Analysis of the data is the most important part of the investigation since this is where incriminating evidence may be found.

Forensic Analysis is the use of controlled and documented analytical and investigative techniques to identify, collect, examine and preserve digital information. Recognizing the fragile nature of digital data, and the legal and regulatory requirements to properly preserve electronically stored information (ESI) during forensic investigations, Secure State maintains standards relating to protecting ESI against manipulation or destruction [7].

Forensic data analysis is as much an art as it is science. Each case presents unique challenges, and the facts can often be hidden in plain sight. The objective of computer forensic data analysis is to determine the facts, as recorded on computer systems and electronic media, in an efficient and non-biased manner. It involves following the chain of evidence as it unfolds.

MATERIALS AND METHODS

Data Identification: It is mapping of Electronically Stored Information and paper documents, Identification of structured and unstructured data and also Identify relevant thirdparty data.

Forensic Collections: It can collect data using forensic preservation standards, Maintain chain of custody and Perform data integrity check to validate completeness.

Data Fusion: It use temporal and entity keys to integrate structured and unstructured data and Superimpose data sets to derive context.

Forensic Analytics: It apply rules-based detection on 100% of transaction data to identify anomalies (fraud, terrorism threats, etc.), Develop statistically based models to identify previously unknown patterns and Optimize anomaly detection rule sets through a feedback loop [8].

Evolution of Forensic Data Analytics
Big Data: data coming into the organization have reached unprecedented levels. About 2.5 Exabyte’s of data are created each day, and that number is doubling every 40 months. Without big data analytics, companies are blind and deaf, meandering aimlessly like a deer on freeway.

Issues in Managing Big Data: Big data requires high performance analytics to process billions of rows of data with hundreds of millions of data combinations. The traditional data warehousing techniques may not be able to identify anomalies in the existing data set thus preventing proactive fraud management. Torture the data, and it will confess to anything.

Key Risk Events: Recent scams in the limelight: In the recent times, India has been hit with multi billion value scams associated with the Anti Money Laundering, Bribery and Corruption, Procurement fraud and collusion in bidding process, Accounting misstatement and Absence of forensic evidence is not evidence of forensic absence.

Forensic Data Analytics: Adoption of forensic data analytics: The associated risks could have been mitigated if key stakeholders would have paid attention to anomalies at an earlier stage. This could have been possible if existing data assets were analyzed from forensic perspective to avoid wrongful or criminal deception intended to result in financial or personal gain.

Types of Computer Forensic Data Analysis: Disk Forensics: The process of acquiring and analyzing the data stored on physical storage media (computer hard drive, cell phones, PDAs, removable media, etc.). Disk forensics includes both the recovery of hidden and deleted data and also file identification, the process of identifying who created a file or message [9].
Network Forensics: The process of examining network traffic, transaction logs, real-time monitoring, using sniffers and tracing.

Internet Forensics: The process of piecing together where and when a user has been on the internet or internal company network. This is used to determine whether inappropriate Internet content access and downloading was accidental or not. It is also used to determine if sensitive information was emailed inappropriately using a personal email account.

Email Forensics: The study of source and content of electronic mail as evidence. It includes the process of identifying the actual sender, recipient, date, time and location and email originated from. Email has become a significant issue for individuals and organizations. Harassment, discrimination or unauthorized activity violating company policy can be identified via email forensics.

Forensic data analysis requires specialized skills and training. Computer Forensic Services expert forensic examiners include federally trained analysts who specialize in and have developed extensive techniques for processing computer evidence [10].

CONCLUSION

From this survey one can conclude the cost of inaccurate analytics in forensic investigations is very high, fusing structured and unstructured data brings better insights in forensic investigation. This paper presented an overview of forensic methodologies and evolution [11-16].

REFERENCES

4. www.cloudsecurityalliance.org/research/big-data,