International Journal of Scientific and Technical Advancements
ISSN: 2454-1532

Big Data Analysis and Digital Forensic

Tariq Hussain Sheikh
Lecturer, GDC Poonch
tariqsheakh2000@gmail.com

Rachna Gupta
GCET Jammu

Abstract: Due to the emergence of internet and the futuristic internet of things the data is enhancing day by day. It is not only difficult to stock the data in proper manner but also provide security and authentication to it. The huge amount of data is produced form applications like internet, social networks, bio-informatics, sensors, weather forecasting etc. Processing if this gigantic amount of data using database system is impossible. The voluminous of data, challenges the research scholars for preventing the cyber criminals for doing their business. The paper entitles to explore the challenges and opportunities for digital investigator in digital forensic analysis due to big data.

Keywords: big data, digital forensic, forensic analysis, gigantic.

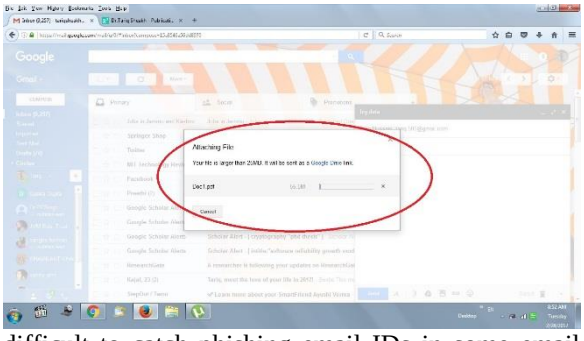
A.INTRODUCTION

There is no hard and fast rule about precisely what size a database needs to be in order for the data inside of it to be considered "big." [12] Instead, what typically defines big data is the need for new techniques and tools in order to be able to process it[12]. In order to use big data, we need programs which span multiple physical and/or virtual machines working together in concert in order to process all of the data in a reasonable span of time.

Getting programs on multiple machines to work together in an efficient way, so that each program knows which components of the data to process, and then being able to put the results from all of the machines together to make sense of a large pool of data takes special programming techniques[12]. Since it is typically much faster for programs to access data stored locally instead of over a network, the distribution of data across a cluster and how those machines are networked together are also important considerations which must be made when thinking about big data problems [12].

Nobody will disagree that the existing world is going through the big Data cohort [10]. Every person is engaged in various kinds of deeds on internet either by using social media or by online transactions, online shopping or by any other thousands of events. Due to all these facts they are inadvertently producing huge amount of data every second by their active impact on internet[1]. Moreover, the data generated by various services such as Facebook, WhatsApp, online shopping websites, etc is also increasing the volume of big Data to the larger extent.

The extreme addiction of people on internet, subsequently, increased the rate of cybercrimes to the larger extent and, and due to this fact the job of Digital Forensic investigations has become more puzzling because they are supposed to crack out the potential evidences from the puddle of Big Data. However, the Big Data presents contests but it can also be utilized as the breaks by the forensic investigators [2]. Challenge in terms of complications in identifying evidences from the pool of organized and unorganized data and the mistrusts behind the crime. For example, it is



difficult to catch phishing email IDs in some email server as there is not any specific filtered data present on the server. On the other hand Big Data also presents

Tariq Hussain Sheikh and Rachana Gupta, "Big data analytics and digital forensic," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 2, pp. 207-210, 2018.

opportunities such as connecting distinct data sets to identify some criminals or criminal activities.

Fig.1screenshot showing Big data doc.larger than 255mb.

B.BIG DATA [10]

Big Data is so gigantic in volume that it cannot be measured in terms of gigabytes or terabytes instead it is as large as petabytes or zettabytes [3]. In addition to this the volume is still increasing on quicker rate with every second. The Big Data is a blend of structured as well as unstructured data. Big Data is characterized by the five Vs which are variety, velocity volume, veracity and value [4].

B.1.TOOLS USED FOR BIG DATA [11].Perchance the most persuasive and established tool for analyzing big data is known as Apache Hadoop. Apache Hadoop is a context for storing and processing data in a large scale, and it is completely open source. Hadoop can run on commodity hardware, making it easy to use with an existing data center, or even to conduct analysis in the cloud. Hadoop is broken into four main parts:

- The Hadoop Distributed File System (HDFS), which is a distributed file system designed for very high aggregate bandwidth;
- YARN, a platform for managing Hadoop's resources and scheduling programs which will run on the Hadoop infrastructure;
- MapReduce, as described above, a model for doing big data processing;
- And a common set of libraries for other modules to use.

C. FORENSIC ANALYSIS [10] Digital Forensic is a branch of Applied Science which deals with identification, collection, organization, preservation and presentation and presentation of evidence data which is permissible is court room [5].Recently Network forensics has been advanced from digital forensics, which deals with assortment of confirmations from Internet or local intranets [6]. Digital or Network Forensics helps the security and forensics, which deals with collection of evidences collected from internet. This type of forensic analysis also deals with the cloud and other dispersed environment. The process of Digital Forensic comprises of following main sub-processes [7]:

- Identification
- Collection
- Organization
- Preservation
- Presentations

Fig.2 showing process of digital forensic.

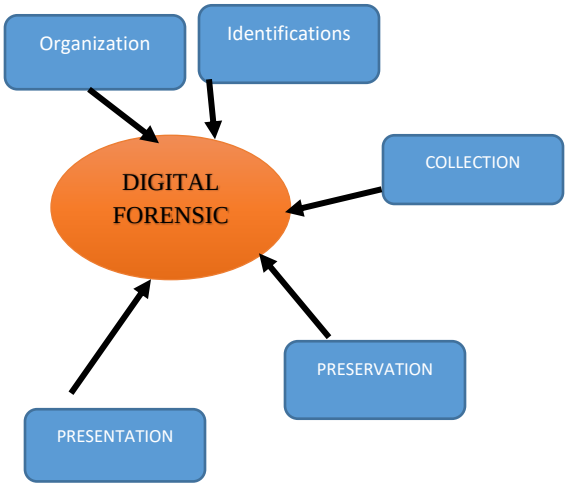


FIG.2 Process of Digital Forensic.

D.BIG DATA AS CHALLENGES FOR FORENSIC ANALYSIS [10]

Big Data is massive data of diversified varieties generated on very high velocity Traditional digital forensic tools are not proficient to handle big data in order to identify and analyze the evidences effortlessly [8]. Some potential challenges of Big Data Forensic are:

D.1 IDENTIFICATION [10]

Finding exact evidence from the Big Data is a hard-hitting job. Because penetrating a meaningful piece of

Tariq Hussain Sheikh and Rachana Gupta, "Big data analytics and digital forensic," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 2, pp. 207-210, 2018.

information from gigantic volume of data to reach some deduction regarding certain incident is not stress-free and that too when data to be analyzed is snowballing at high velocity.

D.2 COLLECTION [10]

The major trial which may be faced during this step is the collection of specious or insignificant data. For example if some errors were occurred during the identification phase then they will surely be spread in this phase. In addition to this the volume and velocity of the evidence data may always pose the problem of collection in front of the investigators.

D.3 ORGANISATION [10]

Organization of the Big Data evidence in appropriate manner is the biggest challenge confronted by forensic investigators. Due to the internet characteristics of Big Data i.e. Variety, volume velocity and veracity, it is not possible to evaluate and consolidate evidences physically in most of the cases effective data mining tools are required to organize the evidences. But the data mining tools capable of handling big data evidences are not accessible which sometimes deteriorate the situation even more.

D.4 Preservation [10]

Forensic investigator often aspect the problematic, while conserving the evidences to uphold the security and integrity for the forthcoming use, due to the same fact- inaccessibility of apposite tools.

D.5 Presentation [10]

In this step the investigator formulates final testimony on the basis of the accessible evidences which may become sometimes while dealing with big data. In addition to this the partial gen of judges on big data may also deteriorate the situation. It may be difficult to elucidate the scrutiny of big data evidences in front of them.

E.BIG DATA AS OPPORTUNITIES FOR FORENSIC ANALYSIS [10]

E.1To correlate distinct criminal data set.

All the criminal cases which are already resolved by the investigators can be stockpiled in big data tools along with their specifics. These datasets can be mines in the future at any point of time to associate distinct criminal data sets to recognize either crime or criminal Moreover, these datasets can also be filtered to decide upon the evidences for any illegitimate case.

E.2 Identification of Cyber Criminals [10]**.**

Big data spawned from various social websites, online transactions and other online cyber criminals. The modus operandi and criminal arrays of cybercriminal can be sheltered in some big data tools which can be mined or referred in forthcoming to identify or collect evidences in certain cases.

E.3 To identify mental state of a criminal. [10]

The mental state and crime configurations of diverse criminals can be deposited in the big data tool which may be referred in the future to adopt upon the mental case and cruelty of the crime.

E.4 Identification of Phishing Email. [10]

The phishing emails which have already been acknowledged by the cyber investigators can be made accessible on the cloud. Internet Service Providers, cyber investigators or universal operators can refer that list to identify and block those phishing emails to shield their very private and delicate data. Moreover an alarm system may also be fused with these lists which notify the users when they are retrieving these illegal emails.

E.5 ALERTS ON ACCESSING FAKE SOCIAL MEDIA ACCOUNTS [10]**.**

Bogus social media versions may be made accessible online. This may help users guarding themselves from criminal minded people who can damage them by sharing or posting their very private data online. To protect general users from these accounts a vigilant system may be intended which acquaint them by displaying warning messages against such accounts.

E.6 Using loT devices in forensic investigation [10]**.**

There are a hefty number of electrify loT devices which are being used for various services. These devices are furnished with sensors and intelligent programs and are expert to perform vital services. These devices may be consumed efficiently by

Tariq Hussain Sheikh and Rachana Gupta, "Big data analytics and digital forensic," *International Journal of Scientific and Technical Advancements*, Volume 4, Issue 2, pp. 207-210, 2018.

providing them with intellectual programs to identify and criminal cases. Moreover these can be used to identify and collect evidences against certain criminal cases.

F. CONCLUSIONS AND FUTURE WORK

The role of big data is offering new encounters and occasions in front of digital forensic investigators .tt stresses upon the prerequisite of novel tools that are well trained to identify, collect, preserved and analyses big data evidences in protected manner. In addition to this the tools should be adept of to avert the data from tempering to uphold the integrity of the evidence future use. The new technique and training personnel are also required by the digital forensic investigators to deal with the challenges presented by the big data. The future prospects require new algorithmic approach for solving the complexity and challenges of forensic analyzer to investigate and report the evidence to the court.

G. REFERENCES

[1] B.Davis, "How much data we create daily".http://goo.gl/a0lmFT,2013.

[2] Analytics, Big Data. "Big Data analytics for security" [2013].

[3] Zikopoulos, Paul and Chris Eaton. Understanding big data: Analysis for [4] enterprise class Hadoop and streaming data McGraw-Hill Osborne Media,2011.

[5] B.Marr, "Why only one of the 5vs of big data really matters", http://goo.gl/azsnse,2015.

[6] Casey, Eoghan. Digital evidence and computer crime: Forensic science, computers, and the internet. Academic press,2011.

[7] Maukkamala, Srinivas, and Andrew H. Sung. "Identifying significant features for network forensic analysis using artificial intelligent techniques." International Journal of digital evidences 1.4 [2003]:1-17.

[8] E. Casey, " Digital Evidences and Computer Crime", Elsevier Inc.,2011.

[9] A Guarino. "Digital forensics as a big data challenges." ISSE 2013 Securing Electronic Business Processes. Springer Fachmedien Wiesbaden,2013.197-203.

[10] Big data –challenges and opportunities in digital forensic by Sapna Sexana and Neha Kishore.

[11]www.digitalocean.com/community/tutorials/an-introduction-to-big-data-concepts-and-terminology.

[12] https://opensource.com/resources/big-data