ISBN 978-93-91286-40-8

KRISTU JYOTI COLLEGE OF MANAGEMENT & TECHNOLOGY
IQAC | Department of Computer Applications
RESEARCH HUB

CERTIFICATE
OF PRESENTATION

THIS CERTIFICATE IS PROUDLY PRESENTED TO

Merlin Mathew

OF SAINTGITS COLLEGE OF APPLIED SCIENCES, PATHAMUTTOM
FOR SUCCESSFULLY PRESENTING A PAPER AT THE **FIRST INTERNATIONAL CONFERENCE ON ADVANCE MODERN COMPUTING TRENDS AND TECHNOLOGY (ICAMCTT 2021)** ON 30TH & 31ST OF JULY 2021

Paper Title : Big Data Analytics in Cyber Security

REV. FR. JOSHY CHEERAMKUZHY CMI
Principal

SUSHEEL GEORGE JOSEPH
Conference Secretary

ROJI THOMAS
Conference Director

BINNY S
Conference Convenor

ISSN 2394-3777 (Print)
ISSN 2394-3785 (Online)
Available online at www.ijartet.com

International Journal of Advanced Research Trends in Engineering and Technology (IJARTET)
Vol. 8, Special Issue 1, August 2021

Big data Analytics in Cyber Security

PRATIBHA ANN JAYESH [1] , MERLIN MATHEW [2], ASHLY MATHEW [3]

SCHOLAR, BCA DEPARTMENT ,SAINTGITS COLLEGE OF APPLIED SCIENCES , PATHAMUTTOM , KOTTAYAM, INDIA [1]

SCHOLAR, BCA DEPARTMENT ,SAINTGITS COLLEGE OF APPLIED SCIENCES , PATHAMUTTOM , KOTTAYAM, INDIA [2]

SCHOLAR, BCA DEPARTMENT ,SAINTGITS COLLEGE OF APPLIED SCIENCES , PATHAMUTTOM , KOTTAYAM, INDIA [3]

Abstract — **Linkage of personal data of individuals possess a severe threat to privacy and civil rights. To overcome these challenges, we need new strategies and security solutions to improve security operations , detecting and analysis of threats and attacks of security. It is important to improve the techniques with the existing cyber security threats. The amount of data is increasing in the internet at an enormous rate. As the volume of data is high , cyber attacks will also increase in an exponential rate . So it is very necessary to interpret and visualize it, inorder to identify certain threats and attacking patterns. Big data analytics helps in tracking large set of user activities inorder to avoid various threats associated with it. This helps to avoid many data breaches**

In this paper we analyse the importance of big data and highlights how is it used as a tool in cyber security to support some security activities. Here we summarize a detailed review on Big Data Analytics in Cyber security field.

Keywords— Bigdata, Bigdata Analytics, Big data query Data analytics, Privacy, threat, Bigdata security, cyber security, visualisation , anomaly

INTRODUCTION

Over the past few years data is being produced rapidly from various application which had led to the production of enormous amounts of data also known as Big Data. Developments in internet services and other communication system in the past years have introduced the term big data, which involves amounts of data which is generated in different forms at a huge rate. The capability to process these bulk amount of data is through big data analytics. By utilizing data which are collected from networks, cloud, computers and other devices can help to detect system exposure and accordingly.

Big Data refers to a huge volume of data which incorporates both structured and unstructured data.

Big data is generally used to facilitate better customer services and provide increased security on customer data moreover using bigdata helps an organisation to make faster and highly informed business decisions.

Major concepts of Bigdata are generalised into 5.V's they are :

volume, velocity, variety ,veracity, value

- Volume is the amount of data that is been generated .
- Velocity can be defined as the rapidness of data from various orgins.
- Variety refers to the diversity or it can be defined as different types of available data such as structured, unstructured and semi-structured data.

- Veracity defines the accuracy of data. It is not at all about the quality of data but about its reliability.
- Value can be refered as the benefit that bigdata can provide.

Big data is defined as too complex to process and analyse. It enables many people to overcome the problems that are associated with small samples of data. There is a need to find new possibilities for accomadating these data because it is growing day after another in an exponential manner.

Two major modes of Big data analytics are : verification and identification.

In verification, data analyst already has an assumption about certain property of services that he wants to verify by means of data analytics

In identification, data analyst gather a large dataset potentially from multiple sources and tries to identify interesting facts hidden within the dataset

. Two key concepts of aggregation of datasets within the big data content must be defined –

The first is the aggregation of schematically identical datasets . For example joining the service access logs of two different online services that are saved on the same web server implementation. mostly wed to attain more information within an existing content.

Second type of aggregation is about linkage created from joining two datasets from disjunct contents, based on some key information shared in both datasets to be aggregated.

A key challenge of big data analytics consists in identifying linkage – a link can be a user email addresses, postal codes/combinations of IP addresses and timestamps

The identity of service plays a major role.

This linkage via user identify bear some very challenging pitfalls in the field of privacy.

Bigdata Analytics is the art of processing, storing, and gathering large data.

Big data mainly focuses on the detection of anomalies and attacks. It allows analysing structured and unstructured data like documents, images and videos which are used as digital evidence in computer forensic process.

When the count of the data increases , it is very difficult to secure it. Confidentiality is the most important side when we consider big data protection. Big data analytics is used as a tool for any data or all business, organization possibilities.

One of the important tools which improves the method processing is Hadoop . In this method they are managing the characteristics of huge volumes of enterprise data. Hadoop combination and revolution analytics giving gain advantages, to unmet the requirement of business for making of strategic decisions. Hadoop split and stores data in different devices and the copy of each dataset will be saved in each devices or in other words those enormous count of data are distributed into large data sets across hundreds of inexpensive servers with help of scalable storage platform are called Hadoop.

It is operated in parallel.

Cyber security is the process of protecting user's data from unauthorised access , attacks or damages. Cyber security has now gone beyond the traditional way . Big data has unfold new ways for cyber security sector.

Here is an overview of fields in cyber security where big data analytics can contribute :-

Forensic Analysis –

Forensic focuses on the analysis , preservation and interpretation of computer data. This field deals with a large dataset, we use various conceptual models for forensic analysis inorder to remove reduntant data .By applying visualisation technique we can reduce the time and improve the effectiveness to find suspicious files.

Big data solutions provide two essential approaches so that the analyst can make his search in abundant data easier. First one is an integrate information from different sources and second has customised visualisation tools.

Malware Detection -

we use bigdata for malware detection.

These are the methods for classifying , combining Bigdata analysis with machine learning, binary instrumentation and dynamic instruct flow analysis.

Security offence -

Security offence include cyber description threat hunting and attack detection.

Cyber desception-

Nowadays it is motivated to use artificial intelligence, game theory and big data to enhance cyber security strategies against attackers.

The main objective of the cyber description is to detect attacks.

Threat hunting-

It is an active defence searching .It is an iterative activity to check through hardware and detect threats in advance instead of waiting for attack alerts. By using big data solution , processing of large amount of information generated by logs can be handled .

Attack detection -

It is very important to detect attacks in the shortest time if possible. It will reduce the time between detection and attack response.

Even though big data enhances security, on the other hand Big data gives a great chance not only for the development of an organisation but also for cyber criminals because they have much more to achieve when they track such a huge volume of data.

EXISTING METHODS

There are various algorithms and analytics used to find out information. They are also applied based on the nature of the data. Some examples for this kind of algorithms are :

Apriori Algorithm and Naive Bayes Classifier Algorithm.

Aprori algorithm works on the principle of bringing frequent data variables, then extending them to larger as long as they are frequent in nature.

Naive Bayes Classifier Algorithm based on Bayes Theorm. It is a classification algorithm with assumptions of independence among predictors. This model is easy to build and work very well for large datasets.

Data mining is also an important process when it comes to big data analytics. It processes large, pre-existing data. It is used for find misure detection and also anomaly detection.

LITERATURE SURVEY

In the paper entitled ‘big data analytics technique in cyber security’ the authors mentioned what bigdata is and how it is useful for the development of an organisation.

Here the corresponding authors proposes the usage of Big Data Analytics for enterprise data which is the data generally shared by users of an organisation.

Their main objective is to access unstructured data from all extreme, and to convert processed data to structured form so that the process of accessing will be more easier. For the easier protection and storage of Big data many organization use tools like Hadoop which distribute and stores the huge data efficiently by using the method of parallel processing. This method is an efficient and best method for Big Data Analytics because it is less expensive since the datas are distributed to inexpensive servers and it is less time consuming.

Here big data is described in a way that it increases data processing efficiency. Here various authors enumerate the major differences between traditional and Bigdata Analytics. This technique is divided into Batch processing and stream processing. In this paper various authors mentioned the desire to build different platforms to store and analyse data. The process is partially enriched and partially illustrative .

In the paper entitled “Special Issue on Big Data Applications in Cyber Security and Threat Inteligence - Part 2” [by kim- Kwang Raymond Choo, Senior Member, IEEE, Mauro Conti, Senior Member, IEEE and Ali De. Dehghantanha , Senior Member, IEEE] focuses on big data applications and threat intelligence. They also shows various research topics on big data for future research which includes anomaly detection for big data, big forensic data provenance, analysis of big data for cyber intelligence, advanced persistent threats detection, big data analytical technique for cyber defence, big data forensic data management and reduction.

In the paper entitled ‘Special issue on Big data applications in Cyber Security and threat Intelligence part 1’ - [by Kim Kwang Raymond Choo, Senior Member, IEEE, Mauro Conti, Senior Member, IEEE, and Ali Dehghantanhe, Senior Member IEEE]focuses on importance of big data analytical techniques to overcome cyber security threats. They shows various technique to interpret, mine and visualise big data from different sources so that it can be applied in cyber forensic, cyber security and threat intelligence.

In the paper entitled "Challenges of Privacy protection in Big Data Analytics"[by Merko Jensen.] Shows various challenges related to big data analytics on privacy .He proposed that data erosion in terms of privacy and user's rights may due to the upcoming trend in big data analytics. He proposed various fields of research on privacy in big data analytics. The most challenging part of privacy in big data analytics is that to provide transparency of personal data of the individuals with respect to type of processing . It is always necessary to process information bound to an individual. Informed consent means that there are many types of big analytics based on complex data algorithm, so each Individual must be given an explanation of all these algorithms so that they can understand what is happening there ,this is a big challenge to data analysis.

An individual decides to revoke the consent for processing personal data later. This is similar to getting a person used among various data collectors and data analysts that is not easier to stop processing on these datas and to delete it. This has become a highly challenging issue. There are various types of attacks such as targeted identification attacks, correlation attacks and arbitary identification attacks. Most threatening type of attack is targeted identification attack. It is to identify some more details of an individual. Inorder to create more unique database entries we link a dataset of uniform data values to other sources . Correlation attacks consist of this kind of linking form datasets. There datasets contain more information per User ID. This helps in analysing more on individual.

Arbitary identification attacks shows failures of a set of anonymized data. This type of attack link atleast to one entry of the dataset to identify a human individual.

A threat to big data analytics is if the information gathered is valid or not. Various types of results can be formed. It will depend on the type of query used by a big data analyst.

Results from different big data query sometimes become a completely wrong final statement. A lot of threats to privacy can also arise from economic consideration in such data trading economic issues of the big data, paradigm is considered to be the fourth category of threats. So threats can be caused due to intentional attacks. It can also caused due to

false data processing methodology or caused by interaction with concerned individuals. So field of privacy in big data faces a lot of challenges.

In the paper entitiled ‘Big data and analytics ‘ the authors enumerate about the rapid growth of data. Contribution of smart devices, such as smartphones hand held computers, wireless networks and social media generating more data over past few years.

In social media domains such as facebook, more than 30 million users are updating posting and sharing their images and video per minute .

Like in instagram , also 300 million instagram users share more than 60-million photos everyday.

More than 100 hours of video are uploaded in every minute. This huge enormous data is Big Data and there is a need to protect and secure these data from & unauthorized access.

This Big data allows new possibilities in technolgy as well as in research field.

In the paper entitiled ‘ Big data analytics for cyber security ‘ explains about the spontaneous growth of the internet has resulted in the exponential increase of the number of cyber attacks. Many organisations tried many popular cyber security to prevent these attacks. Also, the intoduction of Big Data made internet with enormous amount of data . To regale this issue, many researches are now focusing on Security Analytics, which is one of the important application of Big Data Analytics techniques to cybersecurity. This paper provides a survey on the art of Security Analytics which including its states such as its description, , trends, technology and tools.

In the paper entitled "Challenges of Privacy protection Big Data Analytics" presented challenges to privacy of Individuals. The paper discusses about various set of challenges that may threaten privacy of individuals. Another threat with respect to privacy in big data analytics is the ability to perform "re-identification attacks", also validity of the result gathered is also a threat. Another threat covers the economic issues of big data paradigm.

In the paper entitled "Research about New Media Security Technology base on Big Data Era” [by Zheng-wu Lu, Communication University of China, Beying] proposed that high-precision, robust, lightweight and identification and understanding of technology is very important.

It will be the direction of future research. Big data based on cloud computing technology will become a major trend. Difficulty of the new media big is because recognition and understanding of new media content is difficult.

To create a healthy innovative new media environment , we need to research how we can safely provide, consume data and dig information faithfully from these datas.

In the paper entitled “An Insight into Big Data Analytics - Methods and Application

[by Dr. Manjula Sanjay and Alamma 13. H Department of Master of Computer Applications, Dayananda Sagar Academy of Technology of Management, Banglore, India] shows that generation of analytical software like Hadoop or other analytical database can be done through commodity hardware. They shows how traditional data analytics differ from big data analytics now They described about three methods of data analytics and various applications of big data on business, social and scientific applications.

In the paper entitled "Security. Analytics : Big Data Analytics for Cyber security"[by Dr..Tariq Muhammed and Uzma Afzal] proposed that malicious and suspicious patterns can be identified by network managers particularly in the surveillance of real-time network streams. They shows the survey on the art of security Analytics. Also the authors proposed that cyber application of analytics will become an imminent part in cybersecurity in the future. They mentioned different types of big data sources for analytics solution.

In the paper entitled.,”"Big Data Aanlytics Techniques A survey" by [Poonam Vashist and Vishal Gupta] proposed that big data consist of structured, semi-structured and unstructured data. They shows the methods. to analyse the audio, video and text. They shows different challenges

faced by researches while performing big data analysis They also discussed various big data analytics methods and techniques.

CONCLUSION

This paper contains a detailed review on Big Data Analytics in Cyber Security sector . Big data is a new alternative to improve security operations. It has the ability process voluminous data in different format in short time. It is applied to monitor operations and detection of anomalies. Moreover it is used in protective strategies such as threat hunting on cyber deception. It can also detect attack patterns by processing immense data from heterogeneous source.

Big Analytics is often used in cyber security lots of reasons. It facilitate the working of an organization more easier by increasing security with the use of various algorithms and techniques.

The main objective of Big Data analytics is to generate a safe environment for users to protect their data from unauthorised access attacks.

REFERENCE

1] Kim -Kwang Raymond C+ hoo ,Mauro Conti, Ali Dehghantanha

"special issue on Big Data Application in Cyber Security and threat intelligence – part 1"

IEEE transaction on Big Data , July – September (2019)

2] Kim -Kwang Raymond Choo, Mauro Conti, Ali Dehghantanha

" Special Issue on Big Data Application in Cyber Security and threat intelligence – part 2"

IEEE Transaction on Big Data , October – December (2019)

3] Fontugne R Mazel I and Fuhada K. Hashdoop "A MapReduce framework for network anomaly detection "IEEE conference on work shops (2014)4] Meiko Jensen "Challenges of Privacy Protection in Big Data Analytics"

IEEE International Congress on Big Data (2013)

5] Aviral Apurva, Pranshu Ranakoti, Saurav Yadav, Shashank Tomer, Nihar Ranjan Roy

"Redefining Cyber Security with Big Data Analytics" (2017) International Conference on Computing and communication technologies for Smart Nation (I c3TSN).

6] Poonam Vashisht , Vishal Gupta

" Big Data Analytics Techniques: A survey (2015) International conference on Green Computing and Internet of things (ICGI0T)

7] Dr. Tariq Muhammed, Uzma Afzal

" Security Analytics Big Data Analytics for Cyber Security (2013) 2nd National Conference on Information Assurance (NCIA)

8] Zheng - Wu Lu "Research about New Media Security Technology bare on Big Data Era"

(2016) IEEE 14th Inernational Conference on Dependable/ Automatic and Secure Computing, 14th international conference on Pervasive Intelligence and computing, 2nd international conference on Big Data Intelligence and computing cyber Security and Technology Congress

9] Danda B Rawat "Cyber Security in Big Data era:

From securing " Big Data to Data Driven Security"

IEEE

10] Neha Srivasta , prof. Umesh Chandra Jaiswal

" Big Data Analytics Technique in Cyber Security-

A Review" proceedings of third international conference on Computing Methodolgies and Communication (ICCMC 2019)