Big Data Analytics in Cyber Security

[1]Mrs.R.Kohila, Assistant Professor, Department of Cyber Security,Muthayammal Engineering College, Rasipuram - 637408, kohilamec@gmail.com[2]Mr.K.Boopathi, Student, Department of Cyber Security,Muthayammal Engineering College (Autonomous),Rasipuram - 637408, boopathibhoopathy@gmail.com[3]Mr.A.Dharunraj Student, Department of Cyber Security,Muthayammal Engineering College (Autonomous), Rasipuram – 637408, dharunraj0507@gmail.com[4]Mr.M.Mohanasabari Student, Department of Cyber Security,Muthayammal Engineering College (Autonomous), Rasipuram – 637408, mohanasabari052003@gmail.com

Abstract

The rapid growth of the Internet has brought with it an exponential increase in the type and frequency of cyberattacks. Many well-known cybersecurity solutions are in place to counteract these attacks. However, the generation of Big Data over computer networks is rapidly rendering these traditional solutions obsolete. the application of Big Data Analytics techniques to cybersecurity. Analytics can assist network managers particularly in the monitoring and surveillance of real-time network streams and real-time detection of both malicious and suspicious (outlying) patterns. Such a behaviour is envisioned to encompass and enhance all traditional security techniques.This paper presents a comprehensive survey on the state of the art of Security Analytics, i.e., its description, technology, trends, and tools. It hence aimsto convince the reader of the imminent application of analytics as an unparalleled cybersecurity solution in the near future.

Introduction

Big Data is a collection of data that is huge in volume, yet growing exponentially with time. It is a data with so large size and complexity that none of traditional data management tools can store it or process it efficiently. Big data is also a data but with huge size. Big data security analytics is simply a collection of security

Data sets so large and complex that it becomes difficult or impossible to process

using on-hand database management tools or traditional security data processing applications.

The technology known as Big Data is Bone of the most impactful innovations of the digital age. Patterns and correlations hidden in massive collections of data, revealed Aby powerful analytics, are informing planning and decision making acrossnearly every industry. In fact, within just the last decade, Big Data usage has grown to the point where it touches nearly every aspect of our lifestyles, shopping habits, and routine consumer choices.Here are some examples of Big Data applications that affect people every day.

- Transportation
- Advertising and Marketing
- Banking and Financial Services
- Government

- Media and Entertainment
- Meteorology
- Healthcare
- Cybersecurity
- Education

Types of Big Data:

- Structured data
- Unstructured data
- Semi-structured data

Structured data:

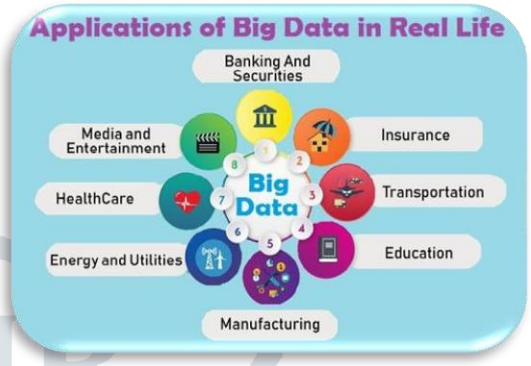
Structured data has certain predefined organizational properties and is present in structured or tabular schema, making it easier to analyse and sort. In addition, thanks to its predefined nature, each field is discrete and can be accessed separately or jointly along with data from other fields. This makes structured data extremely valuable, making it possible to collect data from various locations in the database quickly.

Unstructured data:

Unstructured data entails information with no predefined conceptual definitions and is not easily interpreted or analysed by standard databases or data models. Unstructured data accounts for the majority of big data and comprises information such as dates, numbers, and facts. Big data examples of this type include video and audio files, mobile activity, satellite imagery, and No-SQL databases, to name a few. Photos we upload on Facebook or Instagram and videos that we watch on YouTube or any other platform contribute to the growing pile of unstructured data.

Semi-structured data

Semi-structured data is a hybrid of structured and unstructured data. This means that it inherits a few characteristics of structured data but



nonetheless contains information that fails to have a definite structure and does not conform with relational databases or formal structures of data models. For instance, JSON and XML are typical examples of semi-structured data.

Characteristics of Big Data

Following are the big data core characteristics. Understanding the characteristics of big data is vital to know how it works and how you can use it. There are primarily seven characteristics of big data analytics:

1. Velocity

Volume refers to the amount of data that you have. We measure the volume of our data in Gigabytes, Zettabytes (ZB), and Yottabytes (YB). According to the industry trends, the volume of data will rise substantially in the coming years.

2. Volume

Velocity refers to the speed of data processing. High velocity is crucial for the performance of any big data process. It consists of the rate of change, activity bursts, and the linking of incoming data sets.

3. Value

Value refers to the benefits that your organization derives from the data. Does it match your organization's goals? Does it help your organization enhance itself? It's among the most important big data core characteristics.

4. Variety

Variety refers to the different types of big data. It is among the biggest issues faced by the big data industry as it affects performance. It's vital to manage the variety of your data properly by organizing it. Variety is the various types of data that

you gather from different kinds of sources.

5. Veracity

Veracity refers to the accuracy of your data.

It is among the most important Big Data characteristics as low veracity can greatly damage the accuracy of your results.

6. Validity

How valid and relevant is the data to be used for the intended purpose.

7. Volatility

Big data is constantly changing. The data you gathered from a source a day ago might be different from what you found today. This is called variability of data, and it affects your data homogenization.

8. Visualization

Visualization refers to showing your big data-generated insights through visual representations such as charts and graphs. It has become prevalent recently as big data professionals regularly share their insights with non-technical audiences.

BIG DATA ANALYTICS FOR CYBER SECURITY

1. Big Data Analytics Used in Fraud Detection

Techniques used for fraud detection fall into two primary classes: statistical techniques and artificial intelligence.

Examples of statistical data analysis techniques are:

1. Data pre-processing techniques for detection, validation, error correction, and filling up of missing or incorrect data.

2. Calculation of variousstatistical parameter such as averages, quintiles,

performance metrics, probability distributions, and so on.

3. Models and probability distributions of various business activities either in terms of various parameters or probability distributions.

4. Computing user profiles.

5. Time-series analysis of time-dependent data.

6. Clustering and classification to find patterns and associations among groups of data.

7. Matching algorithms to detect anomalies in the behaviour of transactions or users as compared to previously known models and profiles. Techniques are also needed to eliminate false alarms, estimate risks, and predict future of current transactions or users. Fraud management is a knowledgeintensive activity.

The main AI techniques used for fraud management include [AI]:

1. Data miningto classify, cluster, and segment the data and automatically find associations and rules in the data that may signify interesting patterns, including those related to fraud.

2. Expert systems to encode expertise for detecting fraud in the form of rules.

3. Pattern recognition to detect approximate classes, clusters, or patterns of suspicious behaviour either automatically (unsupervised) or to match giveninputs.

4. Machine learning techniques to automaticallyidentify characteristics of fraud.

5. Neural networks that can learn suspicious patternsfrom samples and used later to detect them.

2.Big Data Analytics Used to Detect Anomaly based Intrusion

Anomaly detection algorithms are very simple to set and functions automatically. Some key performance indicators are for an event chosen and then thresholds are set. If a threshold is exceeded, then the event is signalled for further investigation. The effectiveness of this method is influenced by the choice of indicators to be monitored, of the analysis period, and of the threshold value settings.Anomaly detection algorithms are very simple to set and functions without human intervention. The effectiveness of this method is influenced by the choice of parameters to be monitored, of the analysis period, and of the threshold value settings.

3. Provide Security Intelligence – They can reduce

The time taken to correlate data for forensics purpose andgenerate actionable security response.

CHALLENGES

1. Some organizations may not be data driven. They do not understand the benefits of analytics and hesitant regarding big data analytics.
2. Organizations may think of big data analytics as a way to create value from

data. But it is more about finding the right use case related to intended business objective.

3. Analytics team and the users work together in the various
4. phases of analytics process from scope definition to data extraction and delivery.
5. The management may not be able to trust the analytics outcome as it is difficult to understand how data can generate such outcomes.
6. Limited number of well trained and experienced data scientists.
7. Security issues of big data.

CONCLUSION

The availability of Big Data, low-cost commodity hardware, and new information management and analytic software have produced a unique moment in the history of data analysis. The convergence of these trends means that we have the capabilities required to analyse astonishing data sets quickly and cost-effectively for the first time in history. These capabilities are neither theoretical nor trivial. They represent a genuine leap forward and a clear opportunity to realize enormous gains in terms of efficiency, productivity, revenue, and profitability.

REFERENCES

CLOUD SECURITY ALLIANCE Big Data Analytics for Security IntelligenceBryant, Katz, Lazowska, 2008.Big Data Analytics for Detection of Frauds in

Matrimonial Websites Vemula Geeta et al | International Journal of Computer Science Engineering and Technology (IJCSET) | March 2015 | Vol 5, Issue 3, 57-61Big Data and Specific Analysis Methods for Insurance Fraud Detection Ana-Ramona BOLOGA, Razvan BOLOGA, Alexandra FLOREA University of Economic Studies, Bucharest, Romania.Big Data Cyber security Analytics Research Report – PonemonInstituteÂ© Research Report Date: August 2016Richard A.Derrig,Insurance Fraud, The Journal of Risk and Insurance,2002,Vol.69,No.3,271-287 Bresfelean, Vasile Paul, MihaelaBresfelean, NicolaeGhisoiu, and Calin-Adrian Comes. 2007. “Data Mining Clustering Techniques in Academia.” In ICEIS (2), pp. 407-410. Bresfelean, V. P., Bresfelean, M., Ghisoiu, N., & Comes, C. A. 2008. Determining students academic failure profile founded on data mining methods. In Information Technology Interfaces, IEEE, pp. 317-322.