ISSN: 0974-5823 Vol. 7 No. 6 June, 2022

International Journal of Mechanical Engineering

ANALYTICAL VIEW OF BIG DATA AND IN CYBER SECURITY

1.Dr. M. Sathyapriya

Assistant Professor, Department of Computer Science, Gobi Arts & Science College (Autonomous) Gobichettipalayam – 638453

2. Dr. K. J. Praveen Kumar

Assistant Professor , Department of Computer Science, Gobi Arts & Science College (Autonomous) Gobichettipalayam – 638453

3. Dr. T. K. Sumathi

Associate Professor, Department of Computer Science, Gobi Arts & Science College (Autonomous) Gobichettipalayam – 638453

4. Dr. K. Divya

Assistant Professor of Information Technology, Bharathidasan college of Arts and Science Ellispettai, Erode-638116

Abstract

Big data analytics in the field of cybersecurity refers to the capacity to compile vast quantities of digital information for the purposes of examining, visualising, and deducing insights that might make it feasible to anticipate and thwart cyberattacks. Together with advancements in security technology, this provides us with a more robust cyber defensive posture. They make it possible for companies to identify patterns of behaviour that are indicative of risks to a network. Within the scope of this paper, our primary focus is on the ways in which Big Data might enhance existing information security best practises.

I. INTRODUCTION

The term "Big Data" refers to data sets that are either extremely vast or extremely complicated, to the point that typical data set processing application software is either unable to cope with these huge or complex data sets or is unsuitable for the task. The primary distinction between traditional data analysis and big data is in the amount, velocity, and variety of the data. The terms "volume," "velocity," and "variation" relate, respectively, to the quantity of data that is being created, the pace with which that data is being generated, and the types of data that are being generated, including structured and non-structured data.

At present moment in time, big data is growing into a serious concern for research in practically every industry, but the field of cyber security is particularly affected by this trend. The key sources of this data are social networking sites and smart devices. Because this data also contains some important and sensitive data, such as bank account information, passwords, credit card details, and other details of a similar nature, it is very essential to keep this data secure. Because this data also contains some sensitive and important data, it is very important to keep this data secure. This rate of data creation gives rise to a number of different issues regarding the safety of the data that has been produced. This is due to the fact that it is of utmost significance to maintain the confidentiality of these records. In addition, improvements in big data analytics have led to the production of devices that are capable of collecting and utilising this data, making it simpler to breach the privacy of individuals. As a direct result of this, the development of controls to prevent the exploitation of big data is necessary concurrently with the establishment of tools for big data.

II. DEFINING AND ANALYTICS BIG DATA

The phrase "big data" refers to the vast amounts of information that are kept and transferred within a computer system.

Big Data is distinguished from conventional technology in three key respects, namely:

1. The volume of data, also known as the size of the datasets: the volume of datasets is an important factor, which refers to the total quantity of data that has been produced.
2. The pace at which data is produced as well as the rate at which it is sent (Velocity) A significant amount of complexity may be attributed to a number of critical aspects, including the behaviour, structure, and permutations of datasets.
3. The many types of data that are organised and unstructured (variety) Technologies, including tools and procedures that have been used previously, are a crucial factor to take into consideration while analysing big or sophisticated datasets.

III CYBER-SECURITY AND CYBER THREAT INTELLIGENCE

In this part, a fundamental foundation concerning cyber security and cyber threat intelligence is offered in order to build a platform for discussion in the subsequent subsections. These subsections will cover the following topics: We also investigate and have a conversation about the existing conventional methods to cyber security.

A. Terms and Definitions

When a vulnerability is discovered, countermeasures, strategies, and standards are implemented in order to protect a system or network inside an organisation or the Internet from harm. This definition is found in Section 1.

an attempt to gain access to system resources, services or information without proper authorization or an attempt at jeopardising system integrity are both examples of an attack.

According to the definition, an APT is "an adversary with advanced levels of experience and large resources that allow it to generate opportunities to achieve its objectives by exploiting numerous attack vectors (such as cyber, physical, and deception)". As such, an advanced persistent threat is an enemy capable of utilising a wide range of tactics to achieve its objectives.

An evidence-based knowledge about a present or upcoming danger or hazard to assets, as defined by Gartner in their vocabulary, may be used to influence decisions for the subject's reaction. Context, procedures, signs and consequences as well as practical suggestions are all part of this scientifically validated information.

ESG describes SIEM as "a platform designed to aggregate and correlate security events, logs and network traffic data for the objectives of securing and operating the enterprise." The acronym SIEM refers to the management of security-related data and events.

B. More Conventional Methods of Cyber security Assessment and Management

To safeguard and defend information technology systems, organisational networks and the Internet from cyber-security threats, a variety of security management techniques and processes were created and applied since cyber security was first coined. Figure 1 shows the 10 kinds of probable cyber security breaches that IDC has identified. A system's or network's security can be threatened by any of the threats listed in these categories, or variants of them. In addition to zero-day attacks, the most complex and long-term attacks include distributed denial of service (DDoS) and advanced persistent threats (APTs). These dangers need to be recognised as soon as possible and with precision.

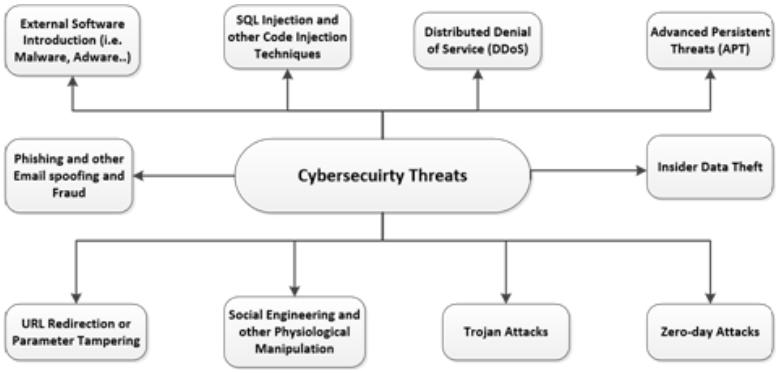

Figure 1: Cyber-security Threats

Throughout the course of information system or network security's history, a variety of defences against and approaches to mitigating the consequences of cyber-security threats, including diverse methods and techniques, have been suggested and developed. Figure 2 depicts the standard ways and processes for cyber-security management and analysis. These are the kinds of things that are often found in any organisation or in certain particular IT systems and may even be in use. Next this, the following paragraphs will provide a concise explanation of each strategy. It is essential to point out that the aforementioned methods are the most utilised of the many available methods, which will not be discussed in more detail in this assessment.

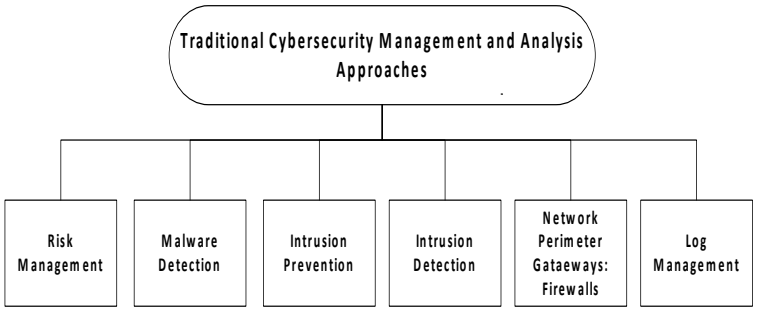

Figure 2: Traditional Cyber-security Management and Analysis Approaches

IV. TRENDS IN TECHNOLOGY

Along with analytics and cloud-based technology, big data is garnering a tremendous amount of interest from the corporate world, the media, and even the general public. These are the components of the present eco-system that was formed by the megatrends in technology.

Big data has emerged as a significant issue or overarching theme in the technology media, and it has also found its way into a number of agreements and internal audits. According to the results of the 2014 Global Forensic Data Analysis Survey conducted by EY, 72 percent of respondents feel that new big data technologies have the potential to play an important role in the prevention and detection of fraud. However, just a small percentage of respondents (approximately 7 percent) were aware of any specific big data technologies, and an even smaller percentage (about 2 percent) were actively utilising these technologies. There are technologies that fall under the category of forensic data analysis (FDA), and they may assist businesses in keeping up with the ever-growing volumes and speeds of data, as well as the rising complexities of their businesses.

Big Data is a wide concept that incorporates a variety of on-going technological breakthroughs and new trends. The following are the top ten developing technologies that are assisting users in coping with and managing Big Data in an efficient and cost-effective manner.

1. database organised in columns

Traditional, row-oriented databases are great for handling online transaction processing because of the rapid update speeds they offer, but they become less effective at handling queries as the number of data stored in them increases and as the data gets more unstructured.

2. Database without a schema, sometimes known as a NoSQL database This category includes a variety of database types, such as key-value storage and document stores, that concentrate on the storage and retrieval of large volumes of data that may be unstructured, semi-structured, or even structured. Examples of databases that fall into this category include the following:

3. Map Reduce

This is a way of thinking about programming that enables massively scalable task execution across thousands of computers or clusters of servers. Any implementation of MapReduce may be broken down into two tasks:

The "Map" job, which consists of converting an input dataset into a distinct collection of key/value pairs, or tuples, and returning the results. The "Reduce" task is one in which a number of the outputs from the "Map" task are merged to generate a smaller collection of tuples.

4. Hadoop

As a completely open source platform for the management of large amounts of data, Hadoop is often regarded as the most effective and widespread use of the map-reduce programming model. It is adaptable enough to operate with a number of different data sources at the same time. It may be used for a variety of purposes, but one of its primary uses is for managing enormous amounts of data that are in a state of continual flux, such as location-based information collected by weather or traffic sensors.

5. Hive

It is a bridge that has the features of SQL and enables typical BI apps to conduct queries against Hadoop clusters. It possesses these qualities since it was built using SQL. Although Facebook was the company that developed it in the beginning, it has been open source for quite some time now. It is a higher-level abstraction of the Hadoop framework that enables anybody to conduct queries against data that is stored in a Hadoop cluster as if they were manipulating a traditional data store. This is possible because it acts as a proxy for the data that is stored in the Hadoop cluster. Facebook was the company that first developed the framework.

6. Pig

Yahoo is responsible for the creation of PIG.

PIG is a bridge that is meant to bring Hadoop closer to the reality of developers and business users in a manner that is akin to that of Hive. PIG, on the other hand, is not a language that is similar to SQL in the way that Hive is. Instead, it is a language that is similar to Perl that enables users to execute queries on the data that is stored on a Hadoop cluster. Hive stands in stark contrast to this.

7. WibiData

Wibi data is a hybrid system that combines web analytics with Hadoop. It was developed on top of Hbase, which is a database layer that is built on top of Hadoop.

8. Sky Tree

It is a high-performance machine learning and data analytics platform with a specific focus on the handling of enormous volumes of data. The platform also includes a number of other features.

Because the scale of the data makes manual examination impractical, machine learning is a particularly significant component of big data.

V. THE DATA LIFE CYCLE FOR BIG DATA

The life of big data may be broken down into three stages.

1. The creation 2. The Process 3. The Generation of Output

There is a category of information that cannot be gathered, but until recently, this category of information was extremely infrequently put to good use (The position of a person at any given point in time, as well as the number of steps they take on a daily basis, are two examples that might be considered representative of the category as a whole.).

This information may now be recorded by more recent technologies for the purpose of analysis. Examples of these technologies include more modern sensors and software that has been particularly designed. Alterations in the fields of communication in the manner in which we communicate (for example, using social media rather than the telephone rather than texting or SMS rather than emailing or writing a letter) have also broadened our power to investigate topics such as consumer sentiment.

Processing

In light of the current circumstances, we are in possession of an extraordinarily huge volume of data that, for a number of reasons, has not been gathered and processed in the conventional manner. The fact that the cost of completing the processing is substantially more than the value insights that organisations can gain from its analysis is the fundamental explanation for this phenomenon.

Because of the extremely high costs associated with processing massive amounts of data, vast amounts of data are often left unprocessed.

However, thanks to recent advancements in technology, both the financial burden and the technical barrier associated with successful data processing have been significantly reduced. This has made it possible for businesses of any size to access the value that is hidden in a variety of data sources. For example, it is challenging for traditional relational databases to manage unstructured data because of the nature of the data.

The cloud is becoming an increasingly popular option for data storage among businesses of all sizes. With cloud computing, businesses are able to employ prebuilt big data solutions or rapidly create and deploy a powerful array of servers without incurring the significant expenditures that come along with owning physical infrastructure. Cloud computing also allows businesses to save time.

Output

It is neither simple nor inexpensive to capture or acquire data, store the data, and analyse the data; the data is of no use whatsoever until the information is pertinent, and it must also be easily accessible whenever it is required.

There are three essential enablers, which are as follows:

• Mobile – The proliferation of mobile networks has made the dissemination of information in real time significantly simpler.

• Visual and interactive: Thanks to technological advancements, even the most basic business users now have access to the tools necessary to analyse massive and intricate data sets.

• Human resource — There is a new generation of workers who are equipped with the knowledge to manage the complexities of big data and the capability to simplify the results for day-to-day use. These workers are referred to as "big data experts." This is a significant new turn of events..

VI. BIG DATA ANALYTICS FOR CYBER SECURITY

Analytics Conducted on Huge Amounts of Data Are Utilized in Fraud Detection

There are primarily two categories of methods that are utilised for detecting fraudulent activity, and these are statistical methods and artificial intelligence methods.

The following are some examples of methods for statistical data analysis:

1. Methods for the identification, validation, and repair of errors, as well as the completion of any missing or erroneous data Pre-processing procedures for the data
2. The computation of a variety of statistical parameters, including but not limited to averages, quintiles, performance metrics, probability distributions, and so on and so forth.
3. Models and probability distributions of the many different types of business operations, either in terms of the many parameters or the probability distributions.

4. Creating user profiles using computing.
5. An study of time-dependent data using a time-series format.
6. Utilizing clustering and classification to identify commonalities and correlations between different sets of data.
7. Employing matching algorithms in order to uncover irregularities in the behaviour of persons or transactions when compared to previously defined models and profiles. In addition, there is a demand for approaches that may reduce the number of false alerts, evaluate the potential dangers, and project into the future the behaviour of users or continuing transactions. Fraud management is a challenging endeavour that calls for a large amount of specialised knowledge.

The following are the primary AI strategies that are utilised for fraud management:

1. The use of data mining to organise, cluster, and segment the data, as well as to automatically discover relationships and rules within the data that may suggest interesting patterns, particularly those connected to fraud. 2. The use of data mining to automatically discover relationships and rules within the data that may suggest interesting patterns.

2. Fraud detection systems that rely on expert systems that are able to codify their knowledge in the form of rules.

3. Pattern recognition to uncover approximate classifications, clusters, or patterns of suspicious behaviour automatically (unsupervised) or to match inputs that have been submitted by the user. Either the utilisation of predetermined inputs or the application of unsupervised pattern recognition are both viable options for achieving this goal.

4. Methods of machine learning that can automatically recognise indicators of fraudulent behaviour.

5. Neural networks that are able to learn suspicious patterns from samples and can then utilise this knowledge to identify those patterns later.

2. Analytical methods based on large amounts of data are used to identify anomaly-based intrusions

Anomaly detection algorithms are quite straightforward to configure and carry out their duties on their own. Following the selection of several key performance indicators for an event and the subsequent establishment of thresholds. When a certain limit is breached, an alarm is triggered to notify more personnel to look into the situation. The choice of indicators that are going to be watched, the length of time that is going to be analysed, and the choices for the threshold value all have an impact on how successful this strategy is.

The configuration of anomaly detection algorithms is a straightforward process that does not require human interaction. The choice of parameters to be monitored, the analysis time, and the choices for the threshold value all have an impact on how successful this strategy is.

3. Supply Security Intelligence – They are able to cut down on the amount of time needed to correlate data for forensics purposes and produce a security reaction that can be put into action.

VII. CHALLENGES

1. It's possible that some businesses aren't data driven. They are apprehensive when it comes to analytics on big data because they do not comprehend the benefits of analytics.
2. Companies and other organisations could consider big data analytics to be a method for extracting value from data. However, the most important thing is to locate the appropriate use case in relation to the targeted business aim.
3. The users and the analytics team collaborate throughout the process of analytics, beginning with the determination of the scope and continuing through data extraction and delivery.
4. Because it is difficult to comprehend how the data may provide such results, managers might not be able to put their faith in the conclusions drawn from the analytics.
5. A scarcity of data scientists who have both the necessary education and relevant work experience.

6. Concerns over the safety of huge data.

CONCLUSION

The acquisition of intelligence that can be acted upon in real time is the primary objective of applying analytics to large amounts of data for the goal of enhancing security. There are three primary ways that Big Data may have a substantial impact on the business that you are currently in charge of running. It is able to assist you in the following areas:

1. Discover insights that were not previously visible. For instance, if you analyse customer survey data while also evaluating a high service cancellation rate, you might find a pattern or an underlying reason that wasn't obvious before and that you can eliminate in order to improve customer retention. This would allow you to keep more of your existing customers.

2. Make better decisions by expanding the types of information that are available to those who make decisions – If you take into consideration, for instance, a customer's social media profile, you can obtain a more accurate image of that consumer and their position in the globe. After that, you can put this information to use to improve how you respond to customers' service requests or to prioritise alerts regarding fraudulent activity.

3. Implement automated systems for corporate procedures. You may, for example, look at extensive stock trading statistics in order to find trends that lead to poorly performed trades. After identifying these patterns, you could then automate the process in such a way that specific steps are followed whenever the pattern appears again. One other illustration would be to look at the information that is available regarding the weather.

REFERENCES

A.P.H. de Gusmão, L. C. e Silva, M. M. Silva, T. Poleto, and A. P. C. S Costa, Information Security Risk Analysis Model Using Fuzzy Decision Theory. International Journal of Information Management, 2016. 36(1): p. 25-34.

A. Razaq, H. Tianfield, and P. Barrie. A Big Data Analytics based Approach to Anomaly Detection. In: 2016 IEEE/ACM 3rd International Conference on BigData Computing Applications and Technologies (BDCAT), 2016. IEEE.

B. Blakley, E. McDermott, and D. Geer, Information Security is Information Risk Management. In: Proceedings of the 2001 Workshop on New Security Paradigms2001, ACM: Cloudcroft, New Mexico. p. 97-104.

Big 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 – Ponemon Institute© Research Report Date: August 2016 [6] Richard A.Derrig,"Insurance Fraud", The Journal of Risk and Insurance",2002,Vol.69,No.3,271-287

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.

Bresfelean, Vasile Paul, Mihaela Bresfelean, Nicolae Ghisoiu, and Calin-Adrian Comes. 2007. "Data Mining Clustering Techniques in Academia." In ICEIS (2), pp. 407-410.

C. Alvaro. A, P.K. Manadhata, and S.P. Rajan, Big Data Analytics for Security. IEEE Security & Privacy, 2013. 11(6): p. 74-76.

C.C Lo and W.J. Chen, A hybrid Information Security Risk Assessment Procedure Considering Interdependences Between Controls. Expert Systems with Applications, 2012. 39(1): p. 247-257ublished by, www.ijert.org ICCCS - 2017 Conference Proceedings Volume 5,

Cyber-Security Definitions; National Initiative of Cyber-security Careers and Studies (NICCS), USA. https://niccs.us-cert.gov/glossary; Access date :31/03/2016.

Extending Security Intelligence with Big Data Solutions: Leverage Big Data Technologies to uncover Actionable Insights into Modern, Advanced Data Threats, IBM Software : Thought Leadership White Paper, 2013.

J. Hu and A.V. Vasilakos, Energy Big Data Analytics and Security: Challenges and Opportunities. IEEE Transactions on Smart Grid, 2016, 7(5): p. 2423-2436.

J. Oltsik, An-Analytics-based Approach to Cybersecurity, May 2015: Enterprise Strategy Group (EGS).

K. Gai, M. Qiu, and S.A. Elnagdy. A Novel Secure Big Data Cyber Incident Analytics Framework for Cloud-based Cyber-security Insurance. In: Big Data Security on Cloud, IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS).

Kasper Security Bulletin 2015: Kaspersky Overal statistics for 2015, in Kaspersky Corporation.

M. Marchetti,,F. Pierazzi, A. Guido and M. Colajanni, Countering Advanced Persistent Threats through Security Intelligence and Big Data Analytics. In: 2016 8th International Conference on Cyber Conflict (CyCon). 2016. IEEE.

Robert Eastman, Michael Versace, and Alan Webber, Big Data and Predictive Analytics: On the Cyber-security Front Line: White Paper, Feb 2015: International Data Corporation (IDC).

S.H. Ahn, N.U. Kim, and T.M. Chung. Big Data Analysis System Concept for Detecting Unknown Attacks. In: 16th International Conference on Advanced Communication Technology, 2014.

T. Mahmood and U. Afzal. Security Analytics: Big Data Analytics for Cyber-security: A Review Of Trends, Techniques and Tools. In: 2013 2nd National Conference on Information Assurance (NCIA), 2013. IEEE.