International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 11, 2017 ISSN: 2348-6600
Reference ID: IJCS-209 PAGE NO: 1307-1311
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

BIG DATA ANALYTICS IN CYBER CRIME DETECTION

M.SATHYAPRIYA[1]
Assistant Professor of Computer Science
Department of Computer Applications,
Gobi Arts & Science College,
Gobichettipalayam.
sathymuthusamy@gmail.com

K.J.PRAVEEN KUMAR[2]
Assistant Professor of Computer Science
Department of Computer Science,
Gobi Arts & Science College,
Gobichettipalayam.praveen.kka@gmail.com

ABSTRACT- The recent advances in data collection and computational statistics coupled with increases in computer processing power, along with the plunging costs of storage are making technologies to effectively analyze large sets of heterogeneous data. Applying big data technologies to an ever growing number and variety of internal and external data sources, businesses and institutions can discover hidden correlations between data items, and extract actionable insights needed for innovation and economic growth. This paper presented a review about how big data can be used to fight against cyber crime.

Keywords: Big data, Big data analytics, cyber crime

INTRODUCTION

The volume and variety of data produced by and about individuals, things or the interactions between them have exploded over the last few years. Such data can be replicated at low cost and is typically stored in searchable databases which are publicly (or at least easily) accessible over the Internet. According to recent IBM estimates, 2.5 billion Gigabytes of data are created everyday around the globe, and the creation rate is growing continuously. Cyber crime is gaining pace with the increasing threats due to online frauds and unethical hacking. Big Data can reduce the processing time of large volumes of data in the distributed computing environment. It also can predict potential cyber security breaches, help stop cyber attacks, and facilitate post-breach digital forensic analysis. This paper introduces Big Data applications in detecting cyber crime.

BIG DATA

The term Big Data refers to large scale information management and analysis technologies that exceed the capability of traditional data processing technologies. Big Data is differentiated from traditional technologies in three ways: the amount of data (volume), types of structured and

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 11, 2017 ISSN: 2348-6600
Reference ID: IJCS-209 PAGE NO: 1307-1311
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

unstructured data (variety),the rate of data generation and transmission (velocity) and data accuracy and consistent(veracity).

Volume: Big data implies enormous volumes of data. The data is generated by machines, networks and human interaction on systems like social media the volume of data to be analyzed is massive. Yet, Inderpal states that the volume of data is not as much the problem as other V's like veracity.

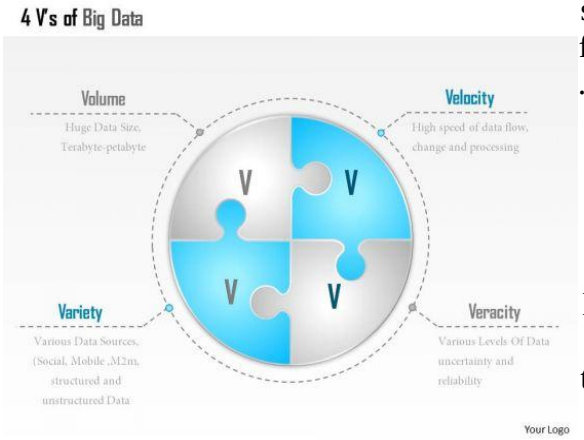
Variety: Variety refers to the many sources and types of data both structured and unstructured. We used to store data from sources like spreadsheets and databases. Now data comes in the form of emails, photos, videos, monitoring devices, PDFs, audio, etc. This variety of unstructured data creates problems for storage, mining and analyzing data.

Velocity: Big Data Velocity deals with the pace at which data flows in from sources like business processes, machines, networks and human interaction with things like social media sites, mobile devices, etc. The flow of data is massive and continuous. This real-time data can help researchers and businesses make valuable decisions that provide strategic competitive advantages for the companies.

Veracity: Big Data Veracity refers to the biases, noise and abnormality in data. Is the data that is being stored, and mined meaningful to the problem being analyzed. In scoping out your big data strategy you need to have your team and partners work to help keep your data clean and processes to keep 'dirty data' from accumulating in your systems.

BIG DATA ANALYTICS

Big Data analytics the process of analyzing and mining Big Data can produce operational and business knowledge at an unprecedented scale and speci ficity . The need to anal yze and lever age trend data colle cted by businesses is one of the main drivers for Big Data analysis tools.



The technological advances in storage, processing, and analysis of Big Data include (a) the rapidly decreasing cost of storage and CPU power in recent years; (b) the flexibility and cost effectiveness of datacenters and cloud computing for elastic computation and storage; and (c) the development of new frameworks such as Hadoop, which allow users to take advantage of these distributed computing systems storing large quantities of data through flexible parallel processing. These advances have created several differences between traditional analytics and Big Data analytics.

Big Data Sources for Security Analytics

The Data sources are to be taken into considerations so that
the security analysis can be more secure.

1 System-Based Data i.e. IP Locations,

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 11, 2017 ISSN: 2348-6600
Reference ID: IJCS-209 PAGE NO: 1307-1311
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

Keyboard typing or Mouse click stream patterns etc.

2. Mobile-Based Data i.e. GPS Locations, Network Locations etc. Time and Location of physical excess of network.
3. Travel Data i.e. travel patterns, sources, destinations etc.
4. Data from external unauthorized sources.
5. Credential Data i.e. user name & password.
6. OTP i.e. One Time Passwords, which are used for online access.
7. Digital Certificates, used for the authentication.
8. Biometric Identification Data i.e. fingerprints, iris, speech recognitions.
9. Social Media Data i.e. Facebook, Google Drive, Twitter etc.

BIG DATA ANALYTICS TO FIGHT CYBER CRIME

Cybercrime is any kind of crime that can be done in, with, or against networks and computer systems. Big data analytics may be the key to fighting cyber crime. Using big data to combat cyber crime, is becoming a decisive strategy for businesses willing to stay secure. With security risks becoming larger, from structured and unstructured data inside the network servers to smart phones, businesses need to be extremely alert due to tremendous increase in cyber threats. Several organizations are leveraging big data analytics for supporting their business processes. However, there are only few organizations that have realized the potential benefits of analytics towards ensuring information security. When data mobility is at a high level, there are highly increased risks, especially when data is transferred to another country with a different regulatory framework. Data driven information security dates back to bank fraud detection and anomaly based intrusion detection systems. Fraud detection is one of the most visible uses for Big Data analytics. Credit card companies have conducted fraud detection for decades. However, the custom-built infrastructure to mine Big Data for fraud detection was not economical to adapt for other fraud detection uses. Off-the-shelf Big Data tools and techniques are now bringing attention to analytics for fraud detection in healthcare, insurance, and other fields.

Analyzing logs, network packets, and system events for forensics and intrusion detection has traditionally been a significant problem; however, traditional technologies fail to provide the tools to support long term, large scale analytics for several reasons:

1. Storing and retaining a large quantity of data was not economically feasible. As a result, most event logs and other recorded computer activity were deleted after a fixed retention period (e.g., 60 days).
2. Performing analytics and complex queries on large, structured data sets was inefficient because traditional tools did not leverage Big Data technologies.
3. Traditional tools were not designed to analyze and manage unstructured data. As a result, traditional tools had rigid, defined schemas. Big Data tools can query data in flexible formats.
4. Big Data systems use cluster computing infrastructures. As a result, the systems are more

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 11, 2017 ISSN: 2348-6600
Reference ID: IJCS-209 PAGE NO: 1307-1311
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

reliable and available, and provide guarantees that queries on the systems are processed to completion.

A. ANALYTICS TECHNIQUES FOR DETECTING FRAUD

Detection and prevention are two ways to counter fraud. Fraud detection systems recognize attempts to fraud, while fraud prevention systems prevent it from occurring. While it is logical to use both in unison, prevention systems lead to hackers changing their strategies, which affects detection ability. Similarly, the existence of a detection system makes hackers devise novel ways to access confidential data, which weakens the system's own detection abilities. According to Bart Baesens, an expert in fraud analytics, and author of a book on fraud detection, techniques to detect or reduce fraudulent activities include descriptive, predictive and social network analytics.

Descriptive analytics tracks behavior that is unusual or deviates from the norm, using techniques such as association rules, clustering, and peer group analysis. Predictive analytics uses historical data sets containing real fraudulent transactions to create fraud detection models that can be subsequently used to detect fraud in real-time data. Over time, the models have to continue to learn as new types of cyber-attacks are discovered. Techniques used to analyze fraudulent data sets include neural networks, random forests, and linear / logistic regression.

Companies are increasingly turning to Social Network Analysis (SNA) to combat fraud. SNA helps detect fraud patterns across functions and products lines, beating earlier limitations caused by departments working in silos. It's advanced visual and analytics capabilities enables firms to detect and prevent fraud through online or traditional business channel.

CONCLUSION

This article speaks about how big data is helpful in cyber-crime detection and more often it says about how the things can be managed and become easy when the analysis part becomes strong while analyzing complex data sets and variety of data. It usually becomes a compulsion to improve the techniques that can be embedded in order to avoid/ prevent cyber-attacks and cybercrimes as well.

REFERENCES

1.http://www-01.ibm.com/software/data/bigdata/
2. http://gartner.com/it-glossary/big-data/
3. Ž. Spalević, Cyber security as a global challenge today, Singidunum Journal of Applied Sciences, 2014, pp. 687-692.
4. M. K.Kakhani, S. Kakhani and S. R.Biradar, Research issues in big data analytics, International Journal of Application or Innovation in Engineering & Management, 2(8) (2015), pp.228-232.
5. A. Gandomi and M. Haider, Beyond the hype: Big data concepts, methods, and analytics, International Journal of Information Management, 35(2) (2015), pp.137-144.
6. T. K. Das and P. M. Kumar, Big data analytics: A framework for unstructured data analysis, International Journal of Engineering and Technology, 5(1) (2013), pp.153-156.

International Journal of Computer Science
Scholarly Peer Reviewed Research Journal - PRESS - OPEN ACCESS
ISSN: 2348-6600
http://www.ijcsjournal.com Volume 5, Issue 1, No 11, 2017 ISSN: 2348-6600
Reference ID: IJCS-209 PAGE NO: 1307-1311
Sri Vasavi College, Erode Self-Finance Wing 3rd February 2017
National Conference on Computer and Communication NCCC'17
http://www.srivasavi.ac.in/ nccc2017@gmail.com

7. R. Magoulas and B. Lorica, "Introduction to Big Data", Release 2.0 (Sebastopol O'Reilly Media), Feb, 2009.
8. P. Breuer, L. Forlna, J. Moulton, "Beyond the hype: Capturing value from big data and advanced analytics", Perspectives on retail and consumer goods, Springer 2013.
9."Big Data and Analytics for National Security" [Online]. Available: web.stanford.edu/group/mmds/slides2012/s-fahey.pdf
10. R. J. Miller, "Big Data Curation," DIMACS Big Data Integration Workshop, 2013.