Available online www.jsaer.com

***Journal of Scientific and Engineering Research*, 2021, 8(9):6-13**

Review Article

ISSN: 2394-2630
CODEN(USA): JSERBR

Big Data in Cybersecurity: A Primer

Matthew N. O. Sadiku, Guddi K. Suman, Sarhan M. Musa

Roy G. Perry College of Engineering, Prairie View A&M University, Prairie View, TX 77446
Email: sadiku@ieee.org; guddikarna@gmail.com; smmusa@pvamu.edu

Abstract Big data refers to mining usable information from the massive amounts of data. It is becoming a focal point of cybersecurity. Cybersecurity has become a big data problem due to the size and complexity of the data and due to the fact that sophistication of threats has increased dramatically. While businesses and government agencies take advantage of big data analytics to improve operations, cyber criminals are mining the same data for unethical reasons. Traditional protection tools used for data mining and cyber-attack prevention are insufficient for several companies. Modern cybersecurity solutions are mostly driven by big data. Intelligent big data analytics allows data specialists to develop a predictive model. This paper is a primer on big data in cybersecurity.

Keywords big data, data analytics, cybersecurity, cybersecurity analytics

Introduction

Today, we live in cyber world where everything is digital and data is king. With over 4.57 billion people using the Internet in 2020, the amount of data being generated has exceeded 2.5 quintillion bytes per day [1]. This explosion of data volumes has created opportunities as well as challenges. Data is becoming our greatest asset. Countless sensitive data are processed by businesses across the globe on daily basis. The technology of processing, storage, and analysis of large data is known as big data.

The concept of big data reflects the appearance in recent years of datasets containing gigantic volumes of unstructured or disparate information. As shown in Figure 1, big data may be regarded as a huge volume of complex structured or unstructured data [2]. The term big data was popularized in 1990 by John Mashey. The rapid growth of big data has made it a prime target for hackers. Hackers are becoming smarter day by day and they are more innovative in exploiting the vulnerable data of individuals, organizations, and governments. Data security is now more important than ever. Big data has become an important issue in our society, science, health, engineering, medicine, finance, business, entertainment, sports, and agriculture. It is now showing promise within the cybersecurity industry.

Cyber security is the process of protecting computer networks from cyber-attacks or unintended, unauthorized access. It is the need of the hour. Big data powers the cybersecurity world. It is now applied extensively across the cybersecurity industry. They are being used to help cybersecurity process data quickly and spot suspicious activity. The rise of big data and the evolution of cybersecurity are intertwined. Big data tools are helping the fight against hackers. Big data is enabling companies to react to specific attacks automatically even before hackers are able to cause damage [3].

Growth and democratization have game-changing effects on cybersecurity. The growth in the amount and variety of data has led to a concurrent growth in the infrastructure that generates and supports that data. For companies to extract maximum value from their data, it must be democratized and made available at every level of the enterprise [4].

Review on Big Data

Big data (BD) refers to a collection of data that cannot be captured, managed, and processed by conventional software tools. It is a relatively new technology that can help many industries, including government. The three main sources of big data are machines, people, and companies. Big data can be described with 42 Vs [5]. The first five Vs are volume, velocity, variety, veracity, and value [6].

- *Volume*: This refers to the size of the data being generated both inside and outside organizations and is increasing annually. Some regard big data as Data over one petabyte in volume.
- *Velocity*: This depicts the unprecedented speed at which data are generated by Internet users, mobile users, social media, etc. Data are generated and processed in a fast way to extract useful, relevant information. Big data could be analyzed in real time, and it has movement and velocity.
- *Variety*: This refers to the data types since big data may originate from heterogeneous sources and is in different formats (e.g., videos, images, audio, text, logs). BD comprises of structured, semi-structured or unstructured data.
- *Veracity*: By this, we mean the truthfulness of data, i.e. weather the data comes from a reputable, trustworthy, authentic, and accountable source. It suggests the inconsistency in the quality of different sources of big data. The data may not be 100% correct.
- *Value*: This is the most important aspect of the big data. It is the desired outcome of big data processing. It refers to the process of discovering hidden values from large datasets. It denotes the value derived from the analysis of the existing data. If one cannot extract some business value from the data, there is no use managing and storing it.

On this basis, small data can be regarded as having low volume, low velocity, low variety, low veracity, and low value. Additional five Vs have been added [7]:

- *Validity:* This refers to the accuracy and correctness of data. It also indicates how up to date it is.
- *Viability:* This identifies the relevancy of data for each use case. Relevancy of data is required to maintain the desired and accurate outcome through analytical and predictive measures.
- *Volatility:* Since data are generated and change at a rapid rate, volatility determines how quickly data change.
- *Vulnerability:* The vulnerability of data is essential because privacy and security are of utmost importance for personal data.
- *Visualization:* Data needs to be presented unambiguously and attractively to the user. Proper visualization of large and complex clinical reports helps in finding valuable insights.

Figure 2 shows the 10V's of big data. In addition, the 10V's above, some suggest the following 5V's: Venue, Variability, Vocabulary, Vagueness, and Validity) [8]. The future of big data will bring more Vs.

Big Data Analytics

Big data sets can be staggering in size so that its analysis is daunting. Every day, data is growing bigger and bigger and big data analysis (BDA) has become a requirement for gaining invaluable insights into data such that companies could gain significant profits in the global market. Big data analytics can leverage the gap within structured and unstructured data sources. Once the big data is ready for analysis, we use advanced software programs such as Hadoop, MapReduce, MongoDB, Spark, Cassandra, Apache Storm, and NoSQL databases [9]. Big data analytics refers to how we can extract, validate, translate, and utilize big data as a new currency of information transactions. It is an emerging field that is aimed at creating empirical predictions. Data-driven organizations use analytics to guide decisions at all levels [10].

Big data analytics tools are used by analysts, researchers, and engineers to process massive amounts of dirty data and extract the gold information from it. It basically learns from data to predict the way individuals will behave in the future. Big data analytics tools are the first line of defense to provide integrated security threat prediction, detection, and deterrence. The objective of using big data analytics is to discover relevant information (such as consumer preferences, market trends, etc.) that can help a business make informed

decisions. Companies are placed at an increasing risk, and they need help from big data analysis to cope. It will improve the detection of many types of attacks and threats. Here are some things that data analytics can do to combat cyber threats [11]:

- It can identify anomalies in how a device is behaving.
- It can identify anomalies in employee and contractor behavior.
- It can detect anomalies in the network such as new threats without known signatures.
- It can analyze data to assess network vulnerabilities and risks.

What is Cybersecurity?

By nature, cyberspace or the Internet is difficult to secure. Intruders exploit the vulnerabilities to steal information and money and perpetrate crimes. The crimes include child pornography, banking and financial fraud, and intellectual property violations. They may also include accessing government and defense confidential information, tampering with commercially sensitive data, and targeting supply chains. Companies are constantly bombarded from all types of sources: criminal syndicates, cyber vandals, intruders, and disgruntled insiders/employees [12]. Figure 3 illustrates some cyber-security threats [13].

Cyber security is the process of protecting computer networks from cyber-attacks or unintended unauthorized access. It refers to a set of technologies and practices designed to protect computer networks and data from damage or unauthorized access. It is vital because governments, companies, and military organizations collect, process, and store a lot of data. Cybersecurity takes different forms including military, law enforcement, judicial, commerce, infrastructure, interior, intelligence, and information systems. Figure 4 shows the elements of cyber security [2]. The main objective of cybersecurity is to protect individuals, organizations, member states and their digital assets from criminal organizations, attackers, and others. The cybersecurity community often reacts to attacks after they have occurred. As shown in Figure 5, cybersecurity involves multiple issues related to people, process, and technology [14]. Cyber security is a major challenge for many companies due to constantly advancing threats.

A typical cyber-attack is an attempt by adversaries or cybercriminals to gain access to and modify their target's computer system or network. Cyber-attacks are becoming more frequent, sophisticated, dangerous, and destructive. They are threatening the operation of businesses, banks, companies, and government networks. They vary from illegal crime of individual citizen (hacking) to actions of groups (terrorists).

Cyber security blunders one must avoid include the following [15]:

1. Not backing up data
2. Not implementing the layered security approach
3. Not investing in security
4. Poor device management
5. Over-reliance on antivirus software
6. No culture of cybersecurity awareness.
7. Weak Passwords

Here is what a cybersecurity professional can do to protect their company [16]:

- Protecting your company’s data begins with understanding it first and then using advanced analytics to discover trends and keep pace with cybercriminals. In cyber security environment, knowing is half the battle.
- Take appropriate steps to secure systems and infrastructure.
- Required and enforce two-factor authentication for all log-in attempts without exceptions.
- Apply strong encryption is a widely-discussed solution to network security issues.
- Be certain that cloud service providers are doing all they can to protect information.
- Run regular anti-malware scans.
- Big data analytics is a must-have component of any effective cyber security solution.

Big Data Cybersecurity Analytics

The traditional tools used for stopping cyber-attacks are a lot more reactive than proactive and they gave rise to many false alarms. These tools are not sufficient enough for many businesses that handle big data. Increasingly sophisticated attack and offensive tools used by cyber criminals show that traditional approaches to mitigate cyber threats are becoming ineffective. Big Data has been creating a profound paradigm shift in addressing the growing cybercrime threats. Businesses and government organizations have experienced better cybersecurity because of big data and analytics. Big data analytics tools provide cybersecurity professionals the power to investigate different kinds of information from varied sources and react in real time. These tools can also be used gather information and counteract cyber-attacks [17].

It is expedient to follow the best practices for big data security for the following reasons [18]:

- It boosts the security of non-relational data scores.
- It helps to implement endpoint security.
- It ensures the safety of transactions and data storage logs.
- It relies on big data cryptography.
- It uses customized solutions.
- It involves practicing real-time security monitoring and compliance.
- It enhances communication and availability of information.
- It allows for convenient resource sharing.
- It increases systems efficiency and robustness.
- It avoids unauthorized access thus protecting and enhancing the performance and security of the organization.

Big data cyber security is a complex migration that requires a massive change for an organization and cannot easily integrate into an enterprise overnight. Big data cybersecurity analytics is substantially different from other types of security analytics. It includes functions such as technologies for integrating dissimilar data types and advanced data visualization applications.

The US government has developed a framework of cyber security standards and best practices. This is intended to prevent well-funded hackers and cyber criminals around the globe from interfering with critical infrastructure and from taking advantage of the massive amounts of military intelligence, trade secrets, financial data, etc.

Big data can provide cybersecurity capabilities in four key areas [19]:

1. The ability to ingest application data.
2. The ability to capture, store and process high volumes of any kind of security and security telemetry data at scale.
3. Perform universal processing of the data (transformation, enrichment, forensic analysis on the data).
4. Long term information storage.

When big data analytics is used along with machine learning, a whole new possibility for cybersecurity results. Big data analytics solutions, backed by machine learning, ensure businesses that their data processes can be kept secure even when hacked [20]. However, machine learning benefits cybersecurity efforts just as much as it helps hackers. Big data benefits both the world of business and the underworld of hackers and cyber criminals. Cyber criminals use big data to monitor their processes, improve their own efficiency, and learn more about breached databases and compromised information systems [21]

Applications

Figure 6 shows some cybersecurity applications for big data analytic [22]. The various uses of big data and analytics in cybersecurity include the following:

- *Smart Cities:* Traditional cities are being transformed to smart cities, with a large number of smart devices connected for different applications. These connected devices produce a large amount of data and bring any security vulnerabilities. The smart city incorporates data into city planning and hopes to achieve spans across various departments including water, energy, communication, housing, and mobility, all with the intention of improving people's quality of life [23].

- *Construction Industry:* The construction industry is one of the leading industries impacted by data security incidents. Cybersecurity is an urgent issue facing the construction sector today. The threat actors often seek to extort money, and the construction industry is a lucrative target. Construction contractors should consider their unique vulnerabilities and take measures to minimize their exposure in the event of a successful cyber-attack. Cybercriminals will continue to target the construction industry as companies adopt new technology in the office and at the worksite. Contractors who fall victim to an attack may face severe financial, reputational, and legal consequences. Addressing cyber-attacks in contracts is a relatively new practice [24].
- *Healthcare*: Cybersecurity is a matter of high concern for the healthcare industry. FDA is encouraging hospitals to be vigilant about any cyber issues in their devices. Medical device manufacturers are now incorporating better security strategies in new devices to secure data. Big data analytic tools can help improve the cybersecurity of sensitive medical devices and avoid jeopardizing the patient's health. Modern medical devices should have a software solution that provides data protection and countermeasures [25].
- *Automation:* A lot of cyber-attacks are due to the ignorance of employees and employee related breaches, also known as inside jobs. Some employees are unaware of cyber threats; hence they are easy targets for attackers. Big data analytics can help monitor the large set of activities of systems/users in order to keep threats away and minimize data breaches. Cybersecurity professionals can automate their processes to speedily combat data breaches when a cyber-attack occurs [26].

Benefits

The benefits of using big data analytics for cybersecurity are many. Almost any business can use big data for better cybersecurity. Big data analytics plays a big role in mitigating cybersecurity breaches caused by business employees, especially unauthorized staff. With big data, cybersecurity professionals are enabled to stay ahead of possible threats and help prevent attacks from happening. Big data in cybersecurity improves the detection of cyber threats with a more sophisticated approach. Businesses can use the tools to correlate information from multiple sources, detect vulnerabilities, and identify a breach as it occurs.

Challenges

As far as cyber security is concerned, big data is both an opportunity and a threat for businesses or organizations. Today, cyber security in big data environment faces multiple challenges and requires diligence and smarter technologies. Although big data analytics is a valuable new addition to the cybersecurity toolkit, using it to its full potential is an uphill task. There is an ongoing debate over the ethics of sharing the private information. The security of connected devices and the storage of the associated data are essential for the privacy and safety. As businesses are benefiting from the big data and analytics, they should be aware of cybersecurity threats.

Other challenges in securing big data in a cyber-security environment include [14]:

- It is difficult for security software to protect new toolsets or new technologies.
- Security tools do not have the same impact on data output from multiple analytical tools to multiple locations.
- Big data administrators generally mine the data without permission or notification.
- The size of big data installation is in terabytes to petabytes which are too big to handle.
- If there is no regular update of security done by a big data owner then there is risk of data loss and exposure.
- Big data security experts need to continuously update their knowledge regarding cleanup and removal of malware and threats.
- When faced with the volume of data to be processed and analyzed to prevent cyber-attacks, most businesses view cybersecurity as a major challenge.
- Most companies find it hard to keep up with emerging threats.
- There is a growing cybersecurity skills gap.

- The issue of protecting sensitive and personal information is important.
- Companies must be responsible for protecting data rights and ownership.

Conclusion

We are in the age of big data and cybersecurity. Big data analytics plays a critical role in strengthening cybersecurity capabilities. Protecting your company's data requires investing in the right tools and staying several steps ahead of cybercriminals. Big data cybersecurity tools make it possible to uncover patterns as they emerge and react in real time, preventing attacks as they occur. In the coming years, companies will continue to enhance big data cybersecurity with machine learning tools [27].

Cyber security is a lucrative job field. The demand for cybersecurity experts (such as cyber security engineers, cyber security analysts, cyber security managers, cyber security consultants) has increased in recent years. As cybersecurity threats become more common and serious, security concerns will take center stage and the demand for cybersecurity professionals will continue to rise. Cybersecurity engineer is the most in-demand job in the cyber security industry [28]. More information about big data in cybersecurity can be found in the books in [29-33] and the following related journals:

- *Journal of Big Data*
- *Big Data Research*
- *Big Data & Society*

References

[1]. M. M. Alana, "Big data in cyber security: A survey of applications and future trends," *Journal of Reliable Intelligent Environments,* vol.7, 2021, pp. 85–114.

[2]. A. M. AlMadahkah, "Big data in computer cyber security systems," *International Journal of Computer Science and Network Security*, vol.16, no. 4, April 2016, pp. 56-65.

[3]. R. Ijaz, "The challenges and opportunities of big data in cyber-security," https://www.datacenterknowledge.com/security/challenges-and-opportunities-big-data-cybersecurity

[4]. "When big data and cybersecurity collide," August 2018, https://www.cio.com/article/3295836/when-big-data-and-cybersecurity-collide.html

[5]. "The 42 V's of big data and data science," https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html

[6]. M. N. O. Sadiku, M. Tembely, and S. M. Musa, "Big data: An introduction for engineers," *Journal of Scientific and Engineering Research*, vol. 3, no. 2, 2016, pp. 106-108.

[7]. P. K. D. Pramanik, S. Pal, and M. Mukhopadhyay, "Healthcare big data: A comprehensive overview," in N. Bouchemal (ed.), *Intelligent Systems for Healthcare Management and Delivery*. IGI Global, chapter 4, 2019, pp. 72-100.

[8]. J. Moorthy et al., "Big data: Prospects and challenges," *The Journal for Decision Makers*, vol. 40, no. 1, 2015, pp. 74–96.
https://www.grandviewresearch.com/industry-analysis/industrial-wireless-sensor-networks-iwsn-market

[9]. M. N. O. Sadiku, J. Foreman, and S. M. Musa, "Big data analytics: A primer," *International Journal of Technologies and Management Research,* vol. 5, no. 9, September 2018, pp. 44-49.

[10]. C. M. M. Kotteti, M. N. O. Sadiku, and S. M. Musa, "Big data analytics," *Invention Journal of Research Technology in Engineering & Management*, vol. 2, no. 10, Oct. 2018, pp. 2455-3689.

[11]. E. Morris, "Big data and cyber security - A double whammy," May 2018, https://www.datasciencecentral.com/profiles/blogs/big-data-and-cyber-security-a-double-whammy-for-hackers

[12]. M. N. O. Sadiku, S. Alam, and S. M. Musa, "A primer on cyber security," *International Journal of Advances in Scientific Research and Engineering,* vol. 3, no. 8, Sept. 2017, pp. 71-74.

[13]. M. A. Rassam, M. A. Maarof, and A. Zainal, "Big data analytics adoption for cybersecurity: A review of current solutions, requirements, challenges and trends,"

Journal of Information Assurance and Security, vol. 11, 2017, pp. 124-145.
[14]. "Eliminating the complexity in cybersecurity with artificial intelligence," https://www.wipro.com/cybersecurity/eliminating-the-complexity-in-cybersecurity-with-artificial-intelligence/
[15]. S. Biswas, "7 Cybersecurity blunders in big data," https://pentestmag.com/7-cybersecurity-blunders-in-big-data/
[16]. M. James, "Strengthen your cybersecurity posture: 20 steps to take in 2020," https://www.smartdatacollective.com/strengthen-cybersecurity-posture-20-steps-to-take-in-2020/
[17]. "Cybersecurity and big data in business," December 2018, Unknown Source.
[18]. "Big data security – implementation, use cases and issues," https://techvidvan.com/tutorials/big-data-security/
[19]. "Cybersecurity – the killer app for big data," April 2016, https://www.vamsitalkstech.com/cybersecurity/cybersecurity-the-killer-app-for-big-data-34/
[20]. K. Lynch, "Using big data analytics in cyber security," May 2019, https://www.dataminingapps.com/2019/05/using-big-data-analytics-in-cyber-security/
[21]. "For cybersecurity, big data offers advantages and challenges," February 2018, https://www.villanovau.com/resources/bi/for-cyber-security-big-data-offers-advantages-challenges/
[22]. R. O. Andrade et al., "Application of big data analytic in cyber security," *Proceedings of International Conference on Advances on Applied Cognitive Computing*, 2019, pp. 26-32.
[23]. R. Doku and D. B. Rawat, "Chapter 8 - Big data in cybersecurity for smart city applications," *Smart Cities Cybersecurity and Privacy*, 2019, pp. 103-112.
[24]. T. Lawhorn, J. P. Vogel, and T. Fandrey, "Cybersecurity for the construction industry: Limiting liability for data breaches," July 2018, https://www.constructionexec.com/article/cybersecurity-for-the-construction-industry-limiting-liability-for-data-breaches
[25]. R. Kh, "Importance of data-driven cybersecurity for medical device," https://www.smartdatacollective.com/importance-of-data-driven-cybersecurity-for-medical-device-companies/
[26]. R. Narasimhan, "Big data analytics for cybersecurity & threat intelligence," November 2020, https://www.linkedin.com/pulse/big-data-analytics-cybersecurity-threat-intelligence-ram-narasimhan
[27]. E. Baez, "Big data cybersecurity: Why it matters and how it helps," April 2021, https://www.scalyr.com/blog/big-data-cybersecurity/
[28]. R. Kh, "Demand for data-savvy cybersecurity professionals grows in 2021," https://www.google.com/search?q=Demand+for+Data-Savvy+Cybersecurity+Professionals+Grows+In+2021&rlz=1C1CHBF_enUS910US910&oq=Demand+for+Data-Savvy+Cybersecurity+Professionals+Grows+In+2021&aqs=chrome..69i57j69i60l2.2166j0j7&sourceid=chrome&ie=UTF-8
[29]. A. E. Hassanien and M. Elhoseny (eds.), *Cybersecurity and Secure Information Systems: Challenges and Solutions in Smart Environments.* Springer 2019.
[30]. O. Savas and J. Deng, *Big Data Analytics in Cybersecurity.* New York: Auerbach Publications, 2017.
[31]. Y. Maleh et al., (eds.), *Machine Intelligence and Big Data Analytics for Cybersecurity Applications.* Springer 2021.
[32]. A. Felkner et al., *Cybersecurity Research Analysis Report for Europe and Japan: Cybersecurity and Privacy Dialogue between Europe and Japan*. Springer, 2021.
[33]. S. Sakr and A. Y. Zomaya (eds.), *Encyclopedia of Big Data Technologies.* Springer, 2019.

About the Authors

Matthew N.O. Sadiku is a professor emeritus in the Department of Electrical and Computer Engineering at Prairie View A&M University, Prairie View, Texas. He is the author of several books and papers. His areas of research interest include computational electromagnetics and computer networks. He is a fellow of IEEE.

Guddi K. Suman is currently working towards a PhD in Electrical and Computer Engineering at Prairie View A&M University, Prairie View, TX. Her areas of research interest include space radiation effects on electronics, nanomaterials, characterization of nanomaterials and semiconductors, thin film nanofabrication, and nanosensors. She worked as an intern at Pacific Northwest National Laboratory, Richland, WA in 2021. She is a student member of IEEE.

Sarhan M. Musa is a professor in the Department of Electrical and Computer Engineering at Prairie View A&M University, Texas. He has been the director of Prairie View Networking Academy, Texas, since 2004. He is an LTD Sprint and Boeing Welliver Fellow. His areas of research interest include computational electromagnetics and computer networks.