INTI JOURNAL | Vol.2024:25
eISSN:2600-7320

Cybersecurity in Big Data Era: From Securing Big Data to Data-Driven Security

Manjunatha V [1*], Shreedhara N Hegde [1], Nur Fatin Liyana Binti Mohd Rosely[2]

[1] DSATM Kanakapura Road, UDAYAPURA, Bangalore Karnataka 560082
[2]Faculty of Data Science and Information Technology, INTI International University 71800 Nilai Negeri Sembilan, Malaysia

***Email:** manjunathvenkatesh478@gmail.com

Abstract

In the age of information, the proverb "knowledge is power" has been shown to be true. admission to, which ultimately leads to knowledge acquisition. The relevance of the ability to glean knowledge from vast amounts of facts has increased. To describe the process of distributing, storing, and gathering enormous amounts of data for future analysis, researchers coined the term "big data analytics" (BDA). Data is generated at an alarming rate. The Internet of Things (IoT), the net's explosive expansion, and other technological advancements are the main forces behind this long-term growth. Since the information generated reflects the environment in which it is formed, the use of information gleaned from systems to understand the inner workings of those systems. The goal of protecting assets has been developed into a crucial component of cybersecurity. Additionally, big data now has the status of a high-value target due to the growing value of data. Current cybersecurity research in relation to big data has been reported here to explore big data security and its potential use as a cybersecurity tool. This gives trends, open research projects, and challenges along with a summary of current studies in the form of tables. In addition to current advancements and unanswered questions in this area of active research, this research work also provides readers a more thorough understanding of safety in the big data era.

Keywords

Big Data Security, Big Data Driven Security, IDS/IPS, Data Analytics

Introduction

Over the past 15 years, data has increased dramatically in many applications, bringing in the massive info age. West makes the point that big data has specific special qualities that can be applied to a range of situations. Using life-size data to identify threats or attacks is one example. Alvin Toffler once said, "As our technological powers increase, so do the side effects and potential hazards," which very well sums up the culture we live in today.

Submission: 4 May 2024; **Acceptance:** 6 August 2024

Originally, hacking was linked to the act of publicly vandalising property. Hackers engage in hacking activities for the sake of entertainment and gaining recognition. Contemporary attacks, on the other hand, are characterized by a greater level of intentionality and motivation. Countries level accusations against each other regarding cyber intrusions. Moreover, there has been a substantial surge in industrial espionage, posing potential risks. Individuals or organizations may originate from nation-states or competitive corporations with the intention of acquiring information or gaining an advantage over their competitors to enhance their own position. Emphasized significant the key concerns that need to be resolved in relation to data security and privacy are confidentiality, privacy, and reliability. The combination of several access control rules and the implementation of restrictions in large-scale data sources have been shown to provide challenges in ensuring data confidentiality. Examples of cyber responsibilities include user verification, access control, and user tracking. Have been identified as crucial in detecting and thwarting threats. The author asserted that contemporary technology, such as encryption, might offer both sanctuary and solitude.

The survey conducted by Abdulla and Zain presents a hierarchical classification of big data security and privacy, emphasizing the difficulties and methods involved, and examining its applications in many fields (A. A. A. A. Abdulla and J. M. Zain, 2023). The work by Cárdenas explores the importance of big data in the field of security, highlighting its capacity to detect and prevent threats (A. A. Cárdenas, et al., 2018).

Chen mentioned in his study to focus on the security and privacy challenges of big data in smart city environments, proposing a framework for protecting sensitive information (Z. Chen, et al, 2018). The study centers on the security and privacy obstacles posed by big data in smart city settings, presenting a framework for safeguarding confidential data. Meanwhile, Gupta and Chaudhary examine how big data analytics is used in cybersecurity, specifically focusing on its contribution to threat intelligence, vulnerability assessment, and incident response (H. Gupta and N. R. Chaudhary, 2018).

Jain and Mao expressed in their article to explore the synergy between artificial intelligence and big data in the context of cybersecurity, emphasizing their combined potential for enhancing security measures. The essay delves into the correlation between artificial intelligence and big data in the realm of cybersecurity, highlighting their collective capacity to improve security measures.

Kambatla also provides a comprehensive analysis of big data security, examining the difficulties, strategies, and prospects in this dynamic area of study (K. Kambatla et al., 2019). Marchetti mentioned to explore the cybersecurity concerns that are unique to the era of big data (M. Marchetti, et al., 2020). It offers valuable insights into the weaknesses and dangers that come with processing massive amounts of data. Figure 1 shows the big data as a security solution and security attacks that are unique to big data enabled systems.

.

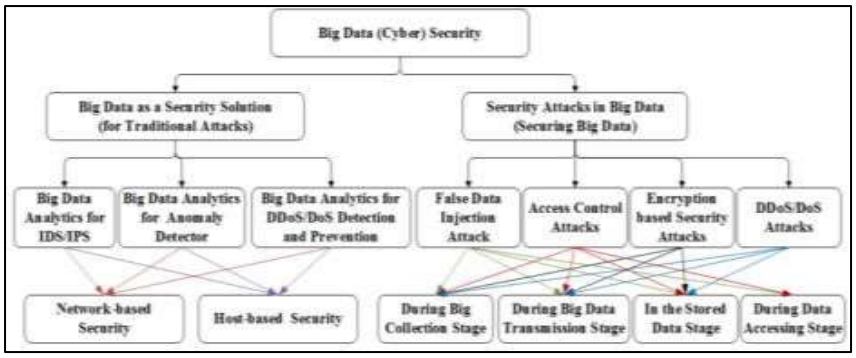

Figure 1. Big data (analytics) as a security solution and security attacks that are unique in a typical big data enabled systems

Prominent security firms agreed to exchange information with the aim of extracting intelligence from the shared data, known as the SecIntel Exchange. The company aimed to offer reliable safety equipment to its customers, and to achieve this, they sought to get as much knowledge as possible from the increasing hazards that emerged daily. They recognized the necessity of collaborating for the benefit of the collective. Considering the proliferation of adaptable malware and other developing threats, it became imperative for them to acquire extensive information about these risks to comprehensively comprehend the nature of the challenges they were confronted with and devise strategies to mitigate them. The effectiveness of older antivirus categorization systems was diminishing. Securing data gets increasingly challenging as it expands to a massive scale. It is important to investigate the security implications linked to big data and cloud computing. It was observed that the majority of companies transfer their extensive data databases to the cloud. However, the internet of things still presents a multitude of concerns.

Both cryptography and access control are focused on ensuring anonymity and security. The primary distinction lies in the fact that encryption primarily focusses on ensuring data confidentiality. Data might be acquired by either a dependable or unscrupulous entity. Encryption guarantees that only authorized and trustworthy users possess a copy of the data. On the other hand, access control seeks to restrict data access. Trusted parties often encounter data restrictions. Consequently, encoding methods must be more efficient than access control mechanisms. Decryption significantly restricts data confidentiality.

Methodology

Developed a framework to protect a combination of similar and dissimilar data, aiming to overcome limitations posed by various encryption techniques, such as the challenge of exchanging keys and the inefficiency in handling large-scale data activities. BigCrypt employs a probabilistic approach to address the PGP (Pretty Good Privacy) problem. Big Crypt employs a symmetric key to encode communication, which is subsequently encrypted and associated with the text using a readily available receiver's key. Subsequently, the email is sent.

Upon reception, the symmetrical key is acquired, asymmetrically encrypted, and utilised to decipher the main communication. The proposed model underwent successful testing on a physical device, as well as on the web and an online server.

The study in proposed a hybrid approach-based framework for large data Access security and confidentiality that creates and enforces privacy rules to contain privacy requirements in a system for control of access. Gao and colleagues presented a large amount of information by data online safety administration system. Cloud computing has increased the total amount of data on the internet. As a result, there were substantial thefts of information and losses. As a result, it was necessary to provide the necessary degree of safety. They conducted big data research to that end, assessing current big info ecosystem. Gupta et al. suggested a large data security compliance methodology. Initially, the approach to big data platforms benefits from safety and accessibility control. Confidentiality and access management were the key solutions for huge privacy concerns. Researchers, on the other hand, have explored different ways that may or may not entail some type of encryption. Because of the nature of massive data, it is difficult to preserve everything. Some academics have attempted to identify the most crucial bits of huge data to secure only those components.

Verizon's Information Breach Crimes Conclusion states that threats might arise from various sources. Cracking was employed in 62% of the instances. The incidents of cyber-attacks account for 51% of the total, while malware attacks constitute 43%. Human errors comprised 14% of the whole amount. Consequently, to execute a successful attack, a perpetrator typically requires the presence of another individual. In such circumstances, individuals, rather than technology, are the target of an attack. The predominant forms of these attacks include email-based phishing and fraudulent activities. A recent study revealed that 52% of successful email attacks prompt individuals to click within one hour, while 30% elicit clicks within a few minutes. This work reports the role of big data in such instances of targeted attacks.

Two enquiries were conducted to acquire further data for Enron Email phishing and fraudulent activities. A recent study revealed that 52% of successful email attacks prompt individuals to click within one hour, while 30% elicit clicks within a few minutes. The researchers examined the role of big data in these specific situations, where data collecting was employed in the initial study. The second study had freshman participants who were tasked with analyzing how email scams circumvented safety measures depending on user behavior. The collected data was further analyzed using Enronic software, and then subjected to email topic classification. The researchers found that phishers or attackers can employ big data analytics to analyze the actions of email users and subsequently develop phishing emails that pose security risks based on the obtained information.

The research's intention was to create a framework for addressing security issues in email communication. The company provided a system that used big data to detect junk and phishing emails by utilizing a global honeynet. The analysis approach involved gathering data from several sources, including pcap files, honeynet logs, blacklisted sites, and social networks. The framework

utilised Hadoop and Spark to handle the diverse data acquired in Hadoop Distributed File System (HDFS). Nevertheless, this system lacks the capability to perform real-time analysis on big datasets. Modern persistent threats are a form of sophisticated and meticulously orchestrated attacks.

Advanced Persistent Threats (APTs) provide a significant challenge in terms of identification. To address this difficulty, the use of big data analysis can be employed to effectively identify and counter APTs. These methodologies can play a crucial role in detecting risks at an early stage, especially when thorough analysis of patterns across different types of data is employed. Due to the proliferation of a significant number of Advanced Persistent Threat (APT) attacks targeting enterprises, a comprehensive APT security plan has been established. The proposed framework integrates advanced and multidimensional defense mechanisms. To counter Advanced Persistent Threat (APT) attacks, the system must be categorized. Information is categorized based on its degree of confidentiality. Assessing the characteristics of data is an essential element of ensuring data security. Researchers investigated the application of real-time big data analytics (BDA) and the associated risks involved in protecting substantial volumes of data. They emphasizes that effective protection of big data should prioritize its vast volume, rapid velocity, and wide-ranging variety (R. K. Naha et al., 2018; J. Nagy, et al., 2018; Tao, Q. et al., 2019).

Big Data:

Addressing big data security is crucial at the application, operating system, and network levels. Implementing the conventional security strategy becomes challenging when dealing with large volumes of often changing data. Consequently, Singh suggests employing machine learning techniques to safeguard huge datasets, with a particular focus on supervised rather than unsupervised learning. The study presented a secure method for safeguarding privacy in mobile big data using the dot product. The secure dots item was originally employed for data extraction. to help defend against attacks on statistical analysis.

The use of anonymized personal identification matching is gaining popularity in the field of big data. The preliminary investigation specifically focused on determining the suitability of mobile big data. Nevertheless, there remains untapped potential for further advancement. The study examined the potential of employing data anonymization to facilitate the integration of encrypted information. This strategy ensures the preservation of privacy during the collection and merging of data, while also enabling the sharing of data between several parties with the involvement of a third party. The collective result outlined in this report does not infringe upon the entity's temporal autonomy.

Moreover, the proposed method allows for the storage of different information from various entities at many third-party facilities while maintaining the confidentiality of the data owners' identities. The anonymized data can be securely linked within a reasonable timeframe. The experiments demonstrate that by employing the optimized secure merging technique, it is

possible to merge 100,000 data entries in approximately 1.4 seconds. To address the inquiry regarding the management of system categorization in a sanctuary.

Bertino examined the issues surrounding the protection of vast amounts of data and the concerns related to solitude, including confidentiality, privacy, and reliability. A multitude of restricted access methods were implemented, and control measures were placed on extensive data resources to address the difficulties identified in safeguarding information. Ensuring user identification, access control, and user monitoring are essential for detecting and preventing threats. The author suggests that contemporary advancements, such as data encryption, can provide both anonymity and security (E. Bertino, 2023).

Results and Discussions

The proposed secure method is for safeguarding privacy in mobile big data by utilizing the dot product. The privacy-preserving dot technique has long been used in data anonymization to protect against statistical analysis attacks.

Since data is anonymized, the practice of doing private identity comparison is gaining popularity in the field of big data. The research was a preliminary investigation focused on determining the suitability of mobility big data. Nevertheless, there remains untapped potential for further advancement. The research also examined the potential of employing data anonymization to facilitate the integration of encrypted information. This approach ensures the preservation of privacy during the collection and merging of data and enables the sharing of data between several parties without the involvement of any third party.

The merging outcome given in this study does not infringe upon the privacy of the individual. Moreover, the proposed method allows for the storage of separate datasets from various entities at multiple third-party facilities while maintaining the confidentiality of the data owners' identities. The de-identified data can be securely linked within a reasonable timeframe. The experiments demonstrated that by employing the optimized secure merging technique.

The study examined the issues and difficulties arising from the overwhelming influx of big data, which is characterized by its immense volume, rapid velocity, and diverse variety, making it incompatible with conventional database systems. The NIST risk management framework was employed to illustrate the security issues presented by the handling of large-scale data. The NIST SP800-30 framework is a guide for implementing data risk management. The study investigated quality assurance methods for the protection of large-scale data in security applications. The interest in quality assurance arises from a lack of confidence in the outcomes of big data applications. The risks connected with big data analytics arise from a deficiency in quality assurance.

Conclusion

This study examines the latest research that utilizes big data in the field of safety. The task was divided into two distinct components. The initial segment of the study concentrated on

investigating the application of big data for the purpose of protecting. The second portion focusses on doing research related to the security of large-scale data. The recent advancements in the application of BDA as a defensive tool were investigated. The significance of neural networks in this domain, along with the obstacles that deep comprehension must overcome before it can be integrated as a crucial element of the cybersecurity toolkit were also discussed.

In addition to recent literature regarding strategies for securing huge datasets, the principal study subjects for ensuring the safety of huge data are decryption and access control systems, as the focus is typically on maintaining data secrecy. In addition to cryptography and physical control, other approaches to protecting large data, which involve the use of technology to safeguard the other components of the CIA trinity were discussed. Furthermore, this work also anticipate advancements in the secrecy and security of big data, along with the challenges that will arise alongside these gains.

Acknowledgement

The researcher did not receive any funding for this study, and the results have not been published in any other sources.

References

Abdulla, A. A. A. A., & Zain, J. M. (2023). Big data security and privacy: A survey and a layered taxonomy. *Big Data Research, 32*, 100383. https://doi.org/10.1109/ISCIT.2016.7751634

Bertino, E. (2023). Privacy in the era of 5G, IoT, big data, and machine learning. *IEEE Security & Privacy, 21*(1), 91–92. https://doi.org/10.1109/MSEC.2022.3221171

Cárdenas, A. A., Ferrante, J., & Masri, A. M. (2018). Big data in security. *IEEE Security & Privacy, 13*(6), 8–11.

Chen, Z., Xiang, Y., Zhou, Y., & Wu, D. (2018). Big data security and privacy protection in smart city. *China Communications, 15*(2), 124–143. https://doi.org/10.1007/s11432-017-9229-2

Gupta, H., & Chaudhary, N. R. (2018). Big data analytics for cybersecurity: A review. *Wireless Personal Communications, 101*(1), 345–368. https://doi.org/10.1109/ICSCAN.2018.8541263

Haryadi, E., Yuliandari, D., Abdussomad, A., Wijayanti, D., Amelia, M., & Syafrianto, S. (2021). Maintaining the continuity of the company's operation using the NIST framework for SME. *Jurnal Khatulistiwa Informatika, 7*(1), 74–78. https://doi.org/10.31294/jtk.v4i2

Jain, A. K., & Mao, J. (2018). Artificial intelligence and big data in cybersecurity. *IEEE Cloud Computing, 5*(1), 26–31. https://doi.org/10.1109/MCC.2017.4721440

Kambatla, K., Kollios, G., & Srivastava, J. (2019). Big data security. In *2019 IEEE International Conference on Big Data (Big Data)* (pp. 1–1). IEEE.

Marchetti, M., Tesoriero, R., & Ronchetti, F. (2020). Cybersecurity threats in the big data era. In *Cybersecurity Threats in the Big Data Era* (pp. 1–12). Springer. https://doi.org/10.1109/TSC.2019.2907247

Naha, R. K., Garg, S., Georgakopoulos, D., Jayaraman, P. P., Gao, L., Xiang, Y., & Ranjan, R. (2018). Fog computing: Survey of trends, architectures, requirements, and research directions. *IEEE Access, 6*, 47980–48009. https://doi.org/10.1109/ACCESS.2018.2866491

Nagy, J., Oláh, J., Erdei, E., Máté, D., & Popp, J. (2018). The role and impact of Industry 4.0 and the Internet of Things on the business strategy of the value chain—The case of Hungary. *Sustainability*, *10*(10), 3491. https://doi.org/10.3390/su10103491

Tao, F., Qi, Q., Wang, L., & Nee, A. Y. C. (2019). Digital twins and cyber–physical systems toward smart manufacturing and Industry 4.0: Correlation and comparison. *Engineering*, *5*(4), 653–661. https://doi.org/10.1016/j.eng.2019.01.014