Journal of Engineering Technology Vol. 11: 9-17, 2023
ISSN 2231-8798© 2021 UniKLBMI

Big Data Security Issues: A Review

Syafik Azhar[1], Noor Hidayah Mohd Yunus[1], Hasya Nabilah Mohd Yusof[1], Nur Aliah Hussin[1], Nurizzah Zafirah Norhisham[1], & Aini Juhaida Zainuddin[1]
[1]Advanced Telecommunication Technology, Communication Technology Section
Universiti Kuala Lumpur British Malaysian Institute
Batu 8, Jalan Sungai Pusu
53100 Gombak Selangor, Malaysia.

Corresponding email: noorhidayahm@unikl.edu.my

Abstract: Terms such as security and privacy are significant when discussing big data topics. Security for big data can be defined as the tools and techniques needed to protect both data and analytical operations. The primary goal behind big data security is to protect valuable data from attacks, stolen data, and other destructive actions. For cloud-based businesses, big data security concerns are multifaceted. These dangers include theft of data stored online, malware, and DDoS attacks that may bring a system to a halt. These risks may have severe financial consequences for the company, including losses, legal expenses and penalties. The use of cloud technologies for storing and managing substantial amounts of data and its associated applications brought a new element to the challenge concerning data security and privacy across substantial amounts of data. Data validation must be performed to ensure the stored data is effectively managed. Multiple layers of encrypted protection are recommended to ensure protection for users in the event of any worst-case scenario, especially a security data breach.

Keywords: Cloud technologies, electricity meters, malware, encryption code

1.0 INTRODUCTION

Big data is a term that refers to data sets that are so large and complex that they cannot be processed by ordinary data processing programs. It refers to a large amount of data, organized and unstructured. Big data is evident in finance and business, where large amounts of stock exchange data, banking, online, and on-site purchases pass through computerized systems every day and are then collected and stored for inventory management, user behavior, and market activity.

Big data security refers to the process of protecting digital data against being corrupted because of a cyberattack or a breach of data security. A data breach is defined as illegal access to data, often to read or copy information. Customer data, credit card numbers, and trade secrets are all examples of confidential information that may be obtained via data breaches. A cyberattack is significantly more violent than a physical attack. An effort by a hacker to fully disable or damage a computer system or computer network. For example, in 2017, hackers attacked a petrochemical plant in Saudi Arabia. However, there are controls over critical safety shutdown systems that have been designed to prevent explosive events from occurring. [1-2]. Malicious software, referred to as Triton or Trisis, was detected in 2017 and has the capability of running unauthorized applications [2-4]. Malware may also analyse and trace the control system, providing reconnaissance and issuing commands in the process [1-6].

The accessibility of big data and its usage for various private and privacy-sensitive activities necessitates that data privacy and security are both key requirements. For instance, the availability of different datasets that can be integrated and evaluated quickly simplifies the collection of sensitive information. The widespread collection of data from a variety of sources and technologies, such as smart electricity meters, smartphones, and computers, exacerbates the issue of data privacy. Such ubiquitous data gathering often tries to capture personally identifiable information, such as personal habits. Additionally, since data often contains intellectual property (IP) as well as other highly valuable information to firms, attacks are increasingly focused on stolen data and penetration. These attacks arise not just from outside, nevertheless also from within the organization.

In this era of big data, the Internet of Things (IoT) and cloud computing technologies are widely used to manage the exponential growth of data in every company, academic institution, and business sector. Big data has rapidly evolved into a trendy topic that has gained widespread interest from professionals in this field worldwide. Big data faces significant issues in terms of security and privacy when working with massive and

heterogeneous data encounters every day. Numerous individuals, including doctors, scientists, government institutions and corporate figures have shared data on a massive scale. Failure to implement proper inspections is one of the worst big data security risks confronting businesses. These inspections should be carried out regularly to discover security vulnerabilities that could expose the business to risk. This vital work is frequently put off or overlooked because of a lack of time, money, clarity regarding security requirements and qualified personnel.

Since data is often utilized to make essential decisions, data reliability is a significant necessity. Data must be protected against unauthorized change. Data should be accurate, comprehensive and current. Comprehensive data trustworthiness methods are challenging to implement because they need the integration of a variety of methodologies, including semantic integrity, digital signatures and data quality procedures. Assuming data reliability may involve intelligent control over data management processes, which has privacy implications.

2.0 GENERAL HISTORY

In the 1980s, the security of mass data became a major issue, at a time when computer clubs were starting to grow and viruses were becoming more common [7]. The first virus is an error in an algorithm that is an error in a program that has the power to reproduce itself. Once a virus is discovered, it is often used as a joke or as proof of one's programming ability, depending on the situation. The public's obsession with viruses, especially dangerous ones, continues to grow. During the first meeting of the Chaos Computer Group in 1985, which is now considered the largest hacker club in Europe, a German computer engineer named Ralf Berger made a keynote speech [8]. He encouraged others to investigate this new element of computer programming.

In 1986, the Brain computer virus was created, which targeted floppy discs and was the world's first purposefully harmful computer virus ever created [9-10]. Developed by two brothers, Amjad and Basit Farooq Alvi, who claimed to be worried about their software being cloned and spread elsewhere worldwide, the virus element was a result of their efforts to combat this. The brain is a virus that operates on IBM PC systems, modifying a floppy disc by replacing the boot sector with one that contains the virus. The virus will cause the disc drive to run slower and prevent seven kilobytes of RAM from being used. A new law was developed in the same year, which is Computer Fraud and Abuse, however, no law included for computer viruses.

During the 1990s the emergence of viruses and hackers wreaked havoc at an alarming rate [10-11]. Due to these situations, the modern version of big data security evolved. To prevent unwanted intrusion into computer systems, precautions were taken, and computer workers received warnings and memos on how to identify viruses. These efforts included the creation of separate backups, which ensured that even if the data on a computer was damaged, it was still accessible at a different location. A common technique of storing backup data has swiftly become software. To keep hackers out, passwords and encryption became more common.

As reported by [10], [12-13], in 1998, two people, a sixteen-year-olds from California and an 18-year-old tutor from Israel hacked into the United States government's computer network. Computer systems under the authority of the Department of Defense were hacked, the computer systems under the control of the government, the army and the private sector were taken away. Investigators first thought that Iraqi hackers were behind the attack since it was carried out with the aid of a computer virus. The Department of Defense was very concerned and required assistance from the National Security Agency (NASA), the Federal Bureau of Investigation (FBI), the Central Intelligence Agency and the United States Department of Justice (DOJ). The effort to apprehend the offenders was termed 'Solar Sunrise' and after the attacks, the defense department took extreme measures to avoid such instances from occurring in the future.

Harmful online conduct was transformed into a profitable crime syndicate throughout the first decade of the 21st century, with the majority of its members motivated primarily by monetary benefit. It was in August 2003 when the Sobig Worm, a computer worm, infected millions of Microsoft Windows computers linked to the Internet [14].

Sobig is a computer worm that multiplies on its own and also a Trojan horse that steals information from the machine or computer and pretends to be something other than malware [14-15]. To attack users, the Sobig worm disguises itself as a normal email message with a harmless subject line, such as "Thank you!" and an attachment designed to attract the user's curiosity and urge them to open the message.

Then, in 2004, there was the infamous "MyDoom" virus to deal with. MyDoom is known as the most malicious worm in the world [13], [16]. MyDoom was founded in January 2004 and has since become notorious. It is sent as an email attachment and becomes active as soon as it is opened. A predetermined termination date of February 12, 2004, was included in the initial package. The end date is actually of little significance, in part because the worm creates a backdoor that enables the worm's developer to get access to the computer at any time, and in part because modern online criminals have begun to use the worm's functionality. MyDoom virus family, which

includes the Sobig malware, has caused significantly more harm compared to other malware.

3.0 RESEARCH FINDINGS

Due to the rapid growth of the information technology sector, the term big data has become common in the current period. Data is generated from a variety of sources, including social media, digital video, corporate records, and photos, resulting in a rapid proliferation of data [17-21]. A challenging task is to manage this massive volume of data, which is referred to as big data. Hacking big data poses a severe hazard because malicious malware might be installed in operating systems and applications. In this paper, previous methods to recover big data from cyber-attacks are reviewed.

S.Padmapriya, N.Partheeban, N.Kamal, A.Suresh, S.Arun, (2019) [22] proposed a set of SQL injection codes to ensure that hackers and attackers cannot access user information stored in databases. Prefixes and suffixes are added to the original data to make it harder for hackers to decipher. This information is recorded on a database table, while the original user information is maintained on a separate table. Because the data is stored in a new table with the changed database instead of the original database, security is improved. When hackers gain access to the data, they will only find altered databases that are no longer in use, rather than the original enhanced database.

Malicious software is designed to harm the operating system or sensitive data without the user's permission. Malware refers to computer viruses, potentially harmful programs and other applications that might damage a computer. Internet hackers utilize malicious software to harm a variety of individuals and organizations throughout the world. Many harmful actions have been reported on the internet, involving innovative cyberattacks produced by unknown variations that hide their behavior in order to avoid detection.

Sitalakshmi Venkatraman and Ramanathan Venkatraman, (2019) [23] studied malicious features that use a four-step process, the first of which is to uncover malware. The binary executable is unpacked in step two. API call extraction is done in step three, while API call mapping and statistical feature analysis is done in step four. Polymorphism and metamorphism seem to be the most common obfuscation methods used by hackers to avoid signature-based detection. By manually unpacking the code and examining the application programming interface (API) calls, software tools can be used to resolve the issue.

Anupama Jha, Meenu Dave and Supriya Madan, (2017) [24] discuss data mining techniques used to offer an overview of big data privacy. Big data is a new term that describes innovative techniques and tools for analyzing large volumes of complex data sets collected from various sources and at varying rates. In the field of big data analysis, data mining techniques prove to be very useful. The ability to extract meaningful information from massive databases is known as big data analytics.

4.0 BASIC PRINCIPLE OF OPERATION

Organizations can use security measures to secure their big data analytics tools in a variety of ways. One of the most frequently used security technologies is encryption, which is a basic technology that can be very effective. External factors, such as hackers, cannot access encrypted data if the hacker does not have the key to decrypt the data. Furthermore, data encryption ensures that information is fully protected at input and output points.

Another important element for big data security is the establishment of a strong firewall. Firewalls are excellent at filtering traffic that enters and exits servers. The creation of powerful filters that block out any third parties or unknown data sources may help organizations avoid attacks before attacks occur. Finally, managing who gets root access to business intelligence (BI) tools and analytics platforms is another important aspect of data security. By implementing a layered access strategy, users may limit the probability of an attack occurring on the network.

Data masking is another important element of getting big data. Masking certain areas of data can help keep it safe from exposure to malicious outside sources, as well as from staff who might be tempted to use the information. For example, the first twelve digits of a credit card number may be masked in the database to protect the cardholder's identity. Furthermore, another way of big data security is data deletion. There are situations where data that is no longer active or in use must be removed from all systems and disposed of accordingly. For example, when a customer has requested the deactivation of a personal account, the customer's information should be permanently removed from the system. The final element of getting big data is data durability. In case an organization's data is accidentally deleted, damaged or stolen during a data breach, a backup copy of the information can be used to restore the information.

There are many reasons for the importance of following the best practices for big data security. It enhances the protection of non-relational record scores in a variety of ways. The best practices help contribute to the implementation of endpoint security. Next, ensures the security of transactions and records of data storage activities. Rely on big data cryptography and use custom-made solutions to meet users' needs. Implement security monitoring and compliance procedures in real-time. It is also improving the flow of information and the availability of information. Allows for more efficient resource sharing while also increasing the efficiency and resilience of the system. Finally, preventing illegal access secures and improves the overall performance and security of the organization.

5.0 TECHNIQUES USED

This chapter discusses the history of big data, data mining and encryption, as well as a process for selecting attributes to protect the value of big data.

5.1 Data Mining Solutions

A data mining solution is an analysis service that includes one or more data mining projects [25]. A data mining solution can be developed on multidimensional data for instance an existing cube, exclusively related data, like tables and views in a data warehouse or text files, Excel workbooks and other external data sources. This service is an instance that establishes the solution that must be constructed to accommodate multidimensional objects and data mining objects.

Thus, this process could also help create data mining objects within a database that has already existed. Usually, this will produce an outcome as such that the creator has developed a cube and wants to do data mining using the cube as the origin. When moving and saving cube-based models, the cube must also be replaced or duplicated. This is the preferred method for developing data mining models as its process and querying are faster compared to relational data origin. In addition, it could also move freely and save models between servers using the 'Export' and 'Import' commands.

The data mining objects are generated in the selected analysis services instance, in a database with the same name as the solution file, when the solutions are deployed, thus providing an overview of the steps to use the Data Mining Wizard to generate data mining solutions [26-28]. Data sources are created and viewed using relational data, text files, and other sources to create a mining structure. Models created from online analytical processing (OLAP) data may be stored as a data mining dimension, or the data set plus the models can be saved as a new cube.

5.2 Secure the encryption tools

Encryption is a method of protecting digital data that uses one or more mathematical techniques and passwords or "keys" to decrypt the data [29-30]. The cryptographic process transforms the data using algorithms that make the original data unreadable. For example, this technique can convert plaintext to ciphertext, which is an alternative version of plaintext. When an authorized user needs access to the data, the binary key can be used to decrypt the data. The ciphertext is converted to plain text and allows authorized users to view the original data.

Individuals or business companies should use encryption to protect sensitive information from hackers. To prevent stolen or fraud of personal information, for example, websites that send credit cards or bank account numbers should always encrypt sensitive information. Cryptography is a mathematical study and application of cryptography. Absolutely, a good encryption code is determined by the length of the encryption security key. In the second half of the 20th century, web developers used 40-bit or 56-bit encryption [25], [31]. This is key with 240 possible permutations. However, by the end of the century, hackers could use brute force attacks to crack these keys. Therefore, the normal encryption length for web browsers is 128 bits [31-33].

Advanced Encryption Standard AES is a data encryption system developed by the National Institute of Standards and Technology in the United States in 2001. The block size of AES is 128 bits, while the key lengths are 128, 192, and 256 bits. These implies that the data is encrypted and decrypted using the same key. The encryption and decryption procedures of asymmetric-key methods employ distinct keys. Although 128-bit encryption is now the industry norm, most banks, armies, and governments still employ 256-bit encryption.

Encryption may be performed in a variety of methods. Symmetric Encrypted Cryptography is the first method [33-34]. The source encrypts the raw message, sends the encrypted message to the recipient, and then uses the same private key to decrypt the message at the destination. Although the following is one of the most basic instances of symmetric encryption, there are many more complicated alternatives available for increased security. This system has the advantage of being simple to install with no operational overhead. However, the system has challenges with shared key security and scalability.

5.3 Access control

Access control is a way to verify an identity to clarify and be authorized to access the company's data. At the most basic level, access control is a selective restriction of data access throughout the employee identification process [35]. Authentication and authorization are two important components to ensure the safety of users which could avoid any misuse of valuable information or data. Authentication is a process to identify that someone is the authentic person interacting with. Authentication alone will never be enough when protecting any kind of data. An additional layer of approval is required to evaluate whether users are allowed access to the data or complete the transaction they are trying to complete. There would be no data security without the two authentication and authorization components, especially in the process of securing any data service companies.

Organizations need to determine which model to use based on the type and confidentiality of the data they are processing. Previous access models included voluntary access control (DAC) and mandatory access control (MAC) are the ones used for this method. The most popular model today is role-based access control (RBAC),

while the latest format is attribute-based access control (ABAC) [35-37].

5.3.1 Discretionary Access Control (DAC)

DAC is a form of access control system that assigns permissions to users based on user-provided rules [38]. DAC principles state that subjects can control who has access to their objects. The DAC paradigm uses access control lists (ACLs) and feature tables. The `Subject` row and the` Object` column form the skill table. The operating system security kernel checks the table to check whether access is granted. The security kernel ensures that unauthorized changes do not occur when the program only has given access to read the file. Fig. 1 shows the operating systems of DAC in Microsoft Windows file systems.

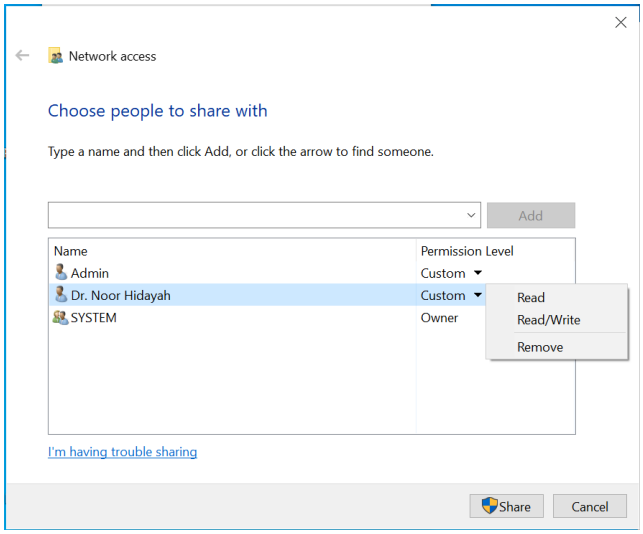

Fig. 1: DAC display

5.3.2 Role-Based Access Control (RBAC)

RBAC, also known as non-discretionary access control, is used when system administrators need to grant permissions based on the responsibility of the organization rather than the individual user accounts in the organization [39]. These allows the company to deal with the "minimum privilege" idea. Access is related to the employment of people, and give only the access when needed. Fig. 2 shows the example of RBAC links model for inter-system for each group proposed by Li Y et al (2022) [40]. Each group has individual file permissions, and every person is assigned to groups based on the work role.

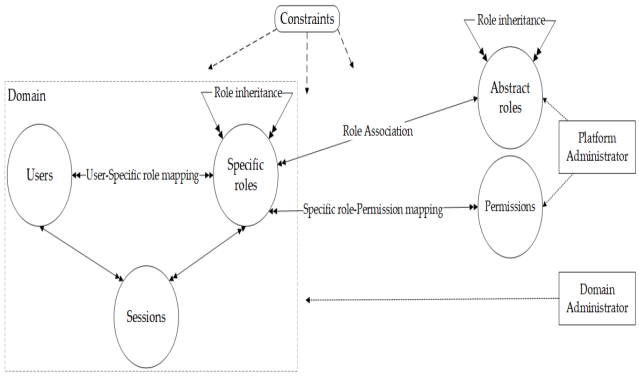

Fig. 2: RBAC link

5.3.3 Mandatory Access Control (MAC)

MAC is the toughest of all access control systems. Governments often use MAC designs and finishes [41]. Manage access to files/resources using a hierarchical method. The settings established by the system administrator control access to resource objects in the MAC environment. This indicate that the operating system controls access to resource objects based on parameters set by the system administrator. Users cannot customize access control for resources. MACs use security tags to identify resource elements on the system.

The security labels are related to two pieces of information which have a classification (high, medium, low) and category of the department or project that provides the information. Fig. 3 shows an example of the defined level for the classification. Classification and category characteristics are also linked to each user account. In case both properties match, the system grants the user access to the corresponding article. When a user has a high ranking then does not belong to the item category, the user will not be able to access the item. MACs are the most secure access control because articles and account names are constantly updated. However, the access system requires a great deal of preparation and advanced system management.

Fig. 3: MAC of defined classification levels of Windows Vista-8 display

5.3.4 Attribute-Based Access Control (ABAC)

All ABAC resources and all users are assigned several features. Use comparisons of user characteristics such as time of day, location, and location to make choices about resource access using these dynamic techniques. It is important for organizations to choose the best model based on data confidentiality and business data access needs. Organizations handling other categories of sensitive information, such as personally identifiable information (PII), Health Insurance Portability and Accountability Act (HIPAA) and unmanaged information (CUI), are at risk of exposure [42-43].

5.4 Intrusion Detection and Prevention System (IDPS)

One of the most important devices in an enterprise's overall security strategy is the intrusion detection and prevention system (IDPS) [44]. With too much data for a single analyst to find signs of an intrusion, IDPS can help alert people to incidents and investigate and prioritize efforts to identify them. IDPS also serves as a valuable auditing tool. IDPS should be established to identify security breaches caused by computer misconfiguration. This is the technical backbone of frustration and resistance technology. Additionally, certain attacks persist until patches or mitigations are available. IDPS may be able to detect traffic indicating new attacks when a command is issued or the attack activity is typically abnormal.

Intrusion detection and prevention implementation can also help identify malicious behavior when it occurs in the network. Many Companies offer powerful threat detection systems, which can immediately alert network administrators and security professionals of suspicious behavior or signs of an attack. An attacker can use the Telnet interface of the router used by the company to access the Internet and launch a counterattack. This interface has been maintained for use by maintenance companies who are given regular maintenance and updates to diagnose problems. Hundreds of authentication attempts per minute on a Telnet interface from one or multiple IP addresses should allow a properly installed and configured IDS or IPS to trigger an alarm.

IDS and IPS technologies can also detect malicious network activity based on differences in “normal" traffic patterns on the network. An "abnormal" based IDS or IPS is called this form of detection. The advantage of using IDS or IPS to identify suspicious traffic is that the system already has the basis for how the network works and can be used to compare traffic patterns. When unusual network traffic is detected, it could also contact the administrator for additional investigation.

6.0 METHODS AND APPLICATIONS

Within big data analytics, data mining techniques are used to do intelligent behavior mining on access restrictions, authentication and incident logs. Data mining technologies look for patterns in unstructured data and can be used with various data types, such as solely relational data and data from other external sources. It may not be possible to precisely specify the data that users can access in a large data environment. Adopting risk-adaptive access controls based on statistical approaches and information theory might be appropriate in this situation. However, in this technique for recognizing and analyzing risks of the big data is a significant challenge.

The benefit of using encryption tools is that they can help protect data from breaches, regardless of whether the data is at rest or in transit. This protection keeps personal data safe from hackers, fraudsters, spammers, internet service providers and even government entities. Wherever encryption is used, it adds an extra step to the data retrieval and transfer process. Additionally, encryption requires the application of highly advanced mathematical processes to each and every (literal) bit of data, which puts additional pressure on the processor system.

The advantage of an access control system is that it can guard against data theft by restricting access. This is crucial when gaining access to a piece of personal account information enables the owner’s identity to be stolen or manipulated. Numerous websites that require personal information to provide their services, particularly those that require a person's credit card information, are required to implement an access control mechanism to safeguard this information. However, this strategy has a drawback in that the system can be hacked. When a system is hacked, a hacker gains access to the data of numerous persons, depending on where the data is placed. Not only does hacking an access control system provide the hacker with information from a sole source, however it also helps the hacker to legitimately breach further control systems without being noticed. Despite advancements in security,

access control systems can still be tampered with and broken into.

An intrusion detection and prevention system (IDPS) is a form of intrusion detection system that uses IDS to assess the number and type of attacks. This data can be used to improve security measures or create more effective restrictions. Additionally, it can be studied to find vulnerabilities or configuration issues with network devices. After that, the metrics can be used to conduct future risk assessments. However, the downside of this strategy is that IP addresses can still be manipulated. Although IDSs read the information included in IP packets, network addresses can still be spoofed. When attackers use fake addresses, threats become more challenging to identify and assess.

The application of big data security applications can be seen as a technology that brings many benefits. However, nothing is perfect besides these techniques will bring pros and cons [25-46]. The advantages and disadvantages of each technique used in big data security can be summarized in Table 1:

Table 1: Advantages and disadvantages of each technique

Techniques/ methods	Advantages	Disadvantages
Data mining	Can be used on unstructured data and can built on multidimensional data	Complicated procedures.
Encryption tools	Protect information from data breaches	There are performance penalties
Access control	Theft prevention	Access control systems potential to be hacked
Intrusion Detection and Prevention	Assess and quantify attackers	IP address still can be fake

In terms of application in healthcare organizations, the challenge is in enabling the implementation of effective treatment record keeping, healthcare organizations store, store and transfer large amounts of data. However, protecting these records has been a challenging task for decades. Adding to the confusion, the healthcare business remains one of the most vulnerable to publicly disclosed data breaches. Attackers can use data mining techniques and processes to uncover sensitive information and make it public, resulting in a data breach. While implementing security measures is still a challenging endeavor, the stakes are constantly growing as new methods to circumvent security constraints are discovered and implemented. As a result, it is important for businesses to deploy healthcare database security solutions that secure critical assets while meeting regulatory requirements.

Big data security is also important in banking sector operations to protect consumer assets. As more individuals go cashless, transactions are conducted online or through physical credit scanners. In both cases, personally identifiable information (PII) may be transferred elsewhere and used for malicious purposes. This affects more than just buyers. Additionally, it caused severe damage to banks as they struggled to recover the data. When data is held hostage, banks may have to spend a lot of money to recover the data. As a result, their customers as well as other financial institutions lost faith in them.

Retail businesses are increasingly aware of the critical nature of consumer data in order to give an exceptional level of service. As businesses amass more data on customers, they discover that they may enhance their products via effective personalization. However, managing large groups of people's data comes with cost businesses must keep it secure and protect their customers' privacy or face a big problem.

7.0 CONCLUSION

Every business must guarantee that all large databases are secure and not compromised by security risks. It is important for business companies to protect their data from hackers to preserve their privacy as well as the interests and protection of the organization. Big data security has a favorable influence on the ability of healthcare organizations to preserve and deliver large amounts of data to provide good care. Many individuals have personal data in healthcare, which prompts hackers to target these organizations. However, to protect user data, big data security controls have grown increasingly complex in recent years.

Big data security benefits the banking industry by protecting not only the assets of each customer, but also bank management. Customers with assets can trust the bank without fear of being hacked. Customer data is essential in the retail industry to provide exceptional service. As a result, big data security is essential to guarantee that customer data and company data are always safe. Many methods or approaches were found throughout the study to prevent hackers from accessing user data, including data mining, encryption tools, access control and intrusion detection and prevention. When addressing data protection issues, each approach or strategy has its own set of advantages and disadvantages. However, the main goal of all approaches is to guarantee that hackers cannot access personal data information.

Acknowledgement

The authors wish to thanks UniKL BMI for the support given in facilitating and contributing to the success of this article review.

REFERENCES

[1] Goudarzi, A., Ghayoor, F., Waseem, M., Fahad, S., & Traore, I. (2022). A Survey on IoT-Enabled Smart Grids: Emerging, Applications, Challenges, and Outlook. Energies, 15(19), pp. 6984. Available: https://doi.org/10.3390/en15196984

[2] Hemsley, K., & Fisher, R. (2018). A history of cyber incidents and threats involving industrial control systems. In *Critical Infrastructure Protection XII: 12th IFIP WG 11.10 International Conference, ICCIP 2018, Arlington, VA, USA, March 12-14, 2018, Revised Selected Papers 12*, pp. 215-242. Springer International Publishing. Available: https://doi.org/10.1007/978-3-030-04537-1_12

[3] Reinhold, T., & Reuter, C. (2021). Toward a Cyber Weapons Assessment Model—Assessment of the Technical Features of Malicious Software. *IEEE Transactions on Technology and Society*, *3*(3), pp. 226-239. Available: doi: 10.1109/TTS.2021.3131817.

[4] Firoozjaei, M. D., Mahmoudyar, N., Baseri, Y., & Ghorbani, A. A. (2022). An evaluation framework for industrial control system cyber incidents. International Journal of Critical Infrastructure Protection, 36, pp. 100487. Available: https://doi.org/10.1016/j.ijcip.2021.100487

[5] Skopik, F., & Pahi, T. (2020). Under false flag: Using technical artifacts for cyber attack attribution. Cybersecurity, 3, pp. 1-20. Available: https://doi.org/10.1186/s42400-020-00048-4

[6] Bin Mohd Shukran, M. A., bin Mohd Yunus, M. S. F., Sham Shariff, W. S., Ariffin, M. S., & Maskat, K. (2014). Pixel Value Graphical Password Scheme: Identifying Design Features and Requirements. In Applied Mechanics and Materials, Vol. 548, pp. 1561-1565. Trans Tech Publications Ltd. Available: https://doi.org/10.4028/www.scientific.net/AMM.548-549.1561

[7] Dunn Cavelty, M. (2013). From cyber-bombs to political fallout: Threat representations with an impact in the cyber-security discourse. International Studies Review, 15(1), pp. 105-122. Available: https://doi.org/10.1111/misr.12023

[8] Wagenknecht, S., & Korn, M. (2016, February). Hacking as transgressive infrastructuring: Mobile phone networks and the German Chaos Computer Club. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, pp. 1104-1117. Available: https://doi.org/10.1145/2818048.2820027

[9] Parikka, J. (2007). Digital contagions: A media archaeology of computer viruses (Vol. 44). Peter Lang.

[10] Milošević, N. (2013). History of malware. arXiv preprint arXiv:1302.5392. Available: https://doi.org/10.48550/arXiv.1302.5392

[11] Szor, P. (2005). The art of computer virus research and defense: Art comp virus res defense _p1. Pearson Education.

[12] Denning, D. E. (2001). Activism, hacktivism, and cyberterrorism: The Internet as a tool for influencing foreign policy. Networks and netwars: The future of terror, crime, and militancy, pp. 239 - 288.

[13] Middleton, B. (2017). A history of cyber security attacks: 1980 to present. CRC Press.

[14] Hill, J. B., & Marion, N. E. (2016). Introduction to Cybercrime: Computer Crimes, Laws, and Policing in the 21st Century: Computer Crimes, Laws, and Policing in the 21st Century. ABC-CLIO.

[15] Bagchi, K., & Udo, G. (2003). An analysis of the growth of computer and Internet security breaches. Communications of the Association for Information Systems, 12(1), pp. 46. Available: https://doi.org/10.17705/1CAIS.01246

[16] A. Gupta, P. Kuppili, A. Akella and P. Barford, "An empirical study of malware evolution," 2009 First International Communication Systems and Networks and Workshops, Bangalore, India, 2009, pp. 1-10, doi: 10.1109/COMSNETS.2009.4808876.

[17] Suhaidi, M. I. A., & Yunus, N. H. M. (2021). Development of Blynk IoT-Based Air Quality Monitoring System. Journal of Engineering Technology, 9, pp. 63-68.

[18] Kamarudin, M. A., Yunus, N. H. M., Razak, M. R. A., Nadzir, M. S. M., & Alhasa, K. M. (2022). Development of Blynk IoT platform weather information monitoring system. In Advanced Materials and Engineering Technologies, pp. 295-305. Cham: Springer International Publishing. Available: https://doi.org/10.1007/978-3-030-92964-0_29

[19] Hadi, N. A. L. A., Yunus, N. H. M., & Nadzir, M. S. M. (2023). Development of a Wireless Solar Power Transmission for Battery Chargers. In Advancements in Materials Science and Technology Led by Women, pp. 199-207. Cham: Springer Nature Switzerland. Available: https://doi.org/10.1007/978-3-031-21959-7_14

[20] Yunus, N. H. M., Rafi, A. N. M., Hadi, N. A. L. A., Mazlan, M. A., & Sampe, J. (2022). A Review of Nanotechnology Applications in the Telecommunication Industry. Journal of Engineering Technology, 10(1), pp. 172-179.

[21] Yunus, N. H. M., Kamarudin, M. A., Karim, A. Z. A., Aziz, P. A., & Hamidon, F. Z. (2015). Development of Auto Solar Tracking Technique for Electricity Generation Systems. Available: http://localhost/xmlui/handle/123456789/10566

[22] Padmapriya, N., Parteeban, N., Kamal, N., Suresh, A., & Arun, S. (2019). Enhanced Cyber Security for Big Data Challenges. International Journal of Innovative Technology and Exploring Engineering, 8(10), pp. 3478-3481.

[23] Venkatraman, S., & Venkatraman, R. (2019). Big data security challenges and strategies. AIMS Math, 4(3), pp. 860-879.

[24] Jha, A., Dave, M., & Madan, S. (2017). Big data security and privacy: A review on issues, challenges and privacy preserving methods. International Journal of Computer Applications, 975, pp. 23-28.

[25] Wirth, R., & Hipp, J. (2000, April). CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, vol. 1, pp. 29-39.

[26] Pyle, D. (1999). Data preparation for data mining. morgan kaufmann.

[27] Badiozamany, S. (2010). Microsoft SQL server OLAP solution-A survey.

[28] Peña-Ayala, A. (2014). Educational data mining: A survey and a data mining-based analysis of recent works. Expert

systems with applications, 41(4), pp. 1432-1462. Available: https://doi.org/10.1016/j.eswa.2013.08.042

[29] Dhany, H. W., Izhari, F., Fahmi, H., Tulus, M., & Sutarman, M. (2017, October). Encryption and decryption using password based encryption, MD5, and DES. In International Conference on Public Policy, Social Computing and Development 2017, pp. 278-283. Atlantis Press. Available: doi: 10.2991/icoposdev-17.2018.57

[30] Yunus, M. S. F. M., Isa, M. R. M., Shukran, M. A. M., Wahab, N., Rahayu, S. B., & Fadzlah, A. F. A. (2023). Pixel Value Graphical Password Scheme: Analysis on Time Complexity performance of Clustering Algorithm for Passpix Segmentation. Journal of Engineering & Technological Sciences, 55(1), pp. 52-59.

[31] Lee, H. K., Malkin, T., & Nahum, E. (2007, October). Cryptographic strength of SSL/TLS servers: Current and recent practices. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, pp. 83-92.

[32] Oppliger, R. (2016). SSL and TLS: Theory and Practice. Artech House.

[33] Spenger, G. (2003). Authentication, Identification Techniques, and Secure Containers–Baseline Technologies. Digital Rights Management: Technological, Economic, Legal and Political Aspects, pp. 62-80. Available: https://doi.org/10.1007/10941270_5

[34] Bokhari, M. U., & Shallal, Q. M. (2016). A review on symmetric key encryption techniques in cryptography. International journal of computer applications, 147(10), pp. 43-48.

[35] Samarati, P., & de Vimercati, S. C. (2001). Access control: Policies, models, and mechanisms. In International school on foundations of security analysis and design Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 137-196. Available: https://doi.org/10.1007/3-540-45608-2_3

[36] Ameer, S., Benson, J., & Sandhu, R. (2022). An Attribute-Based Approach toward a Secured Smart-Home IoT Access Control and a Comparison with a Role-Based Approach. Information, 13(2), pp. 60. Available: https://doi.org/10.3390/info13020060

[37] Gupta, E., Sural, S., Vaidya, J., & Atluri, V. (2022). Enabling Attribute-based Access Control in NoSQL Databases. IEEE Transactions on Emerging Topics in Computing. Available: doi: 10.1109/TETC.2022.3193577.

[38] Zhou, Jinmeng, et al. (2022). Automatic Permission Check Analysis for Linux Kernel. IEEE Transactions on Dependable and Secure Computing, pp. 1849-1866 Available: doi: 10.1109/TDSC.2022.3165368.

[39] Rawal, B. S., Manogaran, G., & Peter, A. (2022). Implement and Manage Authorization Mechanisms. In Cybersecurity and Identity Access Management, pp. 167-172. Singapore: Springer Nature Singapore. Available: https://doi.org/10.1007/978-981-19-2658-7_12

[40] Li, Y., Du, Z., Fu, Y., & Liu, L. (2022). Role-Based Access Control Model for Inter-System Cross-Domain in Multi-Domain Environment. Applied Sciences, 12(24), pp. 13036. Available: https://doi.org/10.3390/app122413036

[41] Xiao, W., Kaneko, M., El Rachkidy, N., & Guitton, A. (2022). Integrating LoRa Collision Decoding and MAC Protocols for Enabling IoT Massive Connectivity. IEEE Internet of Things Magazine, 5(3), pp. 166-173. Available: DOI: 10.1109/IOTM.001.2200055

[42] Mills, J. L., & Harclerode, K. (2017). Privacy, mass intrusion, and the modern data breach. Fla. L. Rev., 69, pp. 771.

[43] Chiasera, A. (2012). Privacy elicitation and utilization in distributed data exchange systems (Doctoral dissertation, University of Trento).

[44] Patel, A., Taghavi, M., Bakhtiyari, K., & Júnior, J. C. (2013). An intrusion detection and prevention system in cloud computing: A systematic review. Journal of network and computer applications, 36(1), pp. 25-41. Available: https://doi.org/10.1016/j.jnca.2012.08.007

[45] Anshari, M., Alas, Y., & Guan, L. S. (2016). Developing online learning resources: Big data, social networks, and cloud computing to support pervasive knowledge. Education and Information Technologies, 21, pp. 1663-1677. Available: https://doi.org/10.1007/s10639-015-9407-3

[46] Allam, Z., & Dhunny, Z. A. (2019). On big data, artificial intelligence and smart cities. Cities, 89, pp. 80-91. Available: https://doi.org/10.1016/j.cities.2019.01.032