Information Fusion Research (2024) Volume 2 Issue 1
doi: 10.59429/ifr.v2i1.6381

RESEARCH ARTICLE

Constructing a cybersecurity big data and data presentation solution

Cheryl Ann Alexander[1,*], Lidong Wang[2]

[1] *Institute for IT Innovation and Smart Health, Mississippi, USA*

[2] *Institute for Systems Engineering Research, Mississippi State University, Mississippi, USA*

***Corresponding author:** Cheryl Ann Alexander, cheryl.alexander@techhealthsolutions.org

ABSTRACT

This paper introduces the requirements of Big Data analytics for cybersecurity, and the challenges of Big Data analytics in cybersecurity are highlighted. Methods and technologies used for the cybersecurity of big data are presented. The cybersecurity of big data in healthcare is introduced. Deep learning (DL) has been used in big data and cybersecurity. The applications of DL in cybersecurity and healthcare are summarized, including examples of tasks and specific DL method(s) for each task. The cybersecurity of big data in a medical center is presented as a case study. In today's highly technological healthcare environment, big data cybersecurity in healthcare is primarily through data encryption for both data at rest and data in motion. The healthcare industry utilizes encryption to protect the sensitive information included in the clinical chart so that only authorized users and recipients are able to view and read the data. Cybercriminals are typically unable to read the data because they lack the decryption key.

Keywords: cybersecurity; cyberattack; information infusion; big data; big data analytics; deep learning (DL); healthcare

1. Introduction

Today, business applications run on mobile services live in the cloud and constantly evolve, which introduces various types of cyberattacks. However, cybersecurity technology such as firewalls, anti-virus and malware tools, and email filters are not guaranteed to protect enterprises from malicious actors. While technology goes far in business protection, policies and procedures surrounding accessing, dealing with breaches, and securing data need to be part of a robust security plan. Asset management becomes a critical need in the business environment. Risk assessment is necessary to have vigorous risk management and governance of the asset management environment. Data security becomes a high-stake need for enterprises and protection procedures should be a priority for the IT security team. Awareness of malicious threats and training for all personnel who may be involved in access control and maintenance is necessary for data security, especially for sensitive data. Detecting incidents such as anomalies and events is a part of continuous security monitoring and the detection of malicious threats and malicious actors. Therefore, responding with a plan

ARTICLE INFO
Received: 11 June 2024 | Accepted: 3 July 2024 | Available online: 12 July 2024

CITATION
Alexander CA, Wang LD. Constructing a cybersecurity big data and data presentation solution. *Information Fusion Research* 2024; 2(1): 6381. doi: 10.59429/ifr.v2i1.6381

becomes essential to response planning and mitigation. Communication among managers, the security IT team, and staff plays an important part in making improvements and analysis of malicious threats. Recovering normal operations is a priority for any improvements in communication and detection of further recovery planning.

One of the most sensitive industries is healthcare. Not only is patient data used for monitoring patient health over time, but it also drives healthcare research for the entire industry. However, Big Data analytics and patient privacy have become hot-button topics as data breaches become all too common in healthcare as patient data becomes an invaluable attraction to malicious actors. Hacking/IT events dominated the healthcare industry in 2022 accounting for data breaches and leaking health information. However, the healthcare industry is not the only industry being affected by data breaches. Numerous industries now report sensitive data and have a substantial risk for malicious actors. Industry experts and enterprises have faced even greater cyber threats since the pandemic as the threat of exposing sensitive data becomes even more attractive to malicious actors. Therefore, many industry leaders have become even more interested in Big Data analytics to prevent data hacks. Big data and privacy issues are at the forefront of the security team's priorities as cyber-criminals threaten patient privacy and data security in healthcare organizations.

A system of big data for cybersecurity employs big data technologies such as Hadoop and Spark to capture/collect, store, and analyze security event data with a big volume for detecting cyberattacks. A framework was developed to quantify the impacts of procedures (duplicate data removal, features selection, signature-based detection of intrusion/attacks, and alert ranking) on the accuracy and response time regarding three factors (security data, machine learning model used, and the execution mode of a system)[1].

It is necessary to develop a co-productive approach to data collection and sharing that overcomes the paradox of cyber data sharing. A common standard for data collection should also be developed. There are the following specifics in a data collection system: 1) where and when to collect the data; 2) load data dynamically and store the collected data; 3) manage and control the data during the collection process; 4) be flexible and scalable regarding the amount and bandwidth of the data; 5) be efficient and stable, not interfering with the data during the collection, and avoiding computationally intensive operations; 6) no noise into the environment, without affecting the data quality; 7) learn and adapt to changes in the environment of the data being generated; 8) prevent data loss to ensure the integrity of the data; 9) prevent data leakage and verify the integrity and authenticity of the data; 10) protect user's privacy during the collection process; and 11) export the data to an external database or other system[2].

A big data sharing model was developed based on blockchain and the smart contract to ensure the safe circulation of data resources. A de-centered storage mode for big data allowed block data to be saved on the network by a node and exchanged in a peer-to-peer manner. Blockchain has the potential to improve the cybersecurity of healthcare big data [3,4]. Blockchain can help secure the management of big health data such as access control, privacy, etc.

The purpose of this paper is to deal with how big data is used in the implementation of effective data-

driven cybersecurity solutions, the requirements of Big Data analytics for cybersecurity, and methods and technologies used for the cybersecurity of big data, especially big data in healthcare and a medical center. The subsequent sections of the paper are organized as follows: the second section introduces the requirements of Big Data analytics for cybersecurity and challenges; the third section presents methods and technologies used for the cybersecurity of big data; the fourth section deals with the cybersecurity of big data in healthcare; the fifth section presents the cybersecurity of big data in a medical center; and the sixth section is the conclusion.

2. Requirements of big data analytics for cybersecurity and challenges

Table 1 [5] shows the requirements of Big Data analytics for cybersecurity. There are challenges in Big Data analytics although Big Data analytics has the potential to provide desirable solutions in many applications. Challenges of Big Data analytics in cybersecurity are as follows [5]:

- Analysis of stream data.
- Real-time analysis.
- Handling unstructured data.
- Data provenance—uncertain of the trustworthiness of data sources, authenticity and integrity needed for the data.
- Adaptiveness—difficult when using machine learning (ML), considering together with real-time and streaming data processing, data privacy, security, provenance, etc.
- Data privacy.
- Visual analytics—challenges when real-time analysis of stream data, tracking the dynamicity of events, and providing security analysts with clues using dashboards.

Table 1. Requisites of big data analytics for cybersecurity.

Requisites	Details
Handling data from multi-sources	Active directory files, firewall logs, SQL server logs, operating systems events logs, NetFlow data, IDS (intrusion detection system) data, SIEM (security information and event management) data, threat intelligence data, etc.
Handling data of various types	Data (structured, unstructured, semi- structured) from various sources such as blogs, e-mails, social networks activities, threat feeds, etc.
Management of large-scale data	Data with large volumes are collected; cloud computing as well as cluster and grid computing help store, process, retrieve data in a timely manner, and draw useful insights.
High-performance infrastructure technologies	Big data modeling, cloud computing, stream processing, MapReduce, grid computing, distributed computing, large-scale distributed systems, etc.
Visualization	Visualized connections between events, signatures, locations, devices, and IPs reveal data patterns, anomalies, and intrusions.

3. Methods and technologies used for the cybersecurity of big data

Big Data analytics can be used as a powerful tool to help protect the security of enterprises. Big data itself is also vulnerable to cyberattacks. The cybersecurity of big data has been a significant issue. **Figure 1** [6] shows approaches to securing big data. **Figure 2**[6] shows trends, open research challenges, and problems in the cybersecurity of big data.

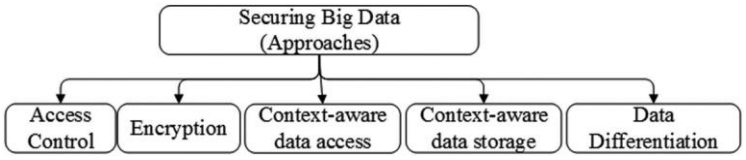

Figure 1. Securing the big data.

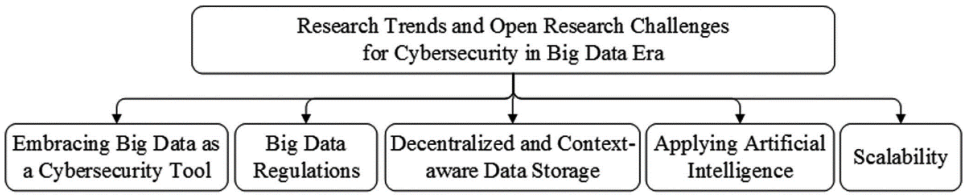

Figure 2. Typical trends, open research challenges and problems.

Attribute-based access control (ABAC) is an important method. It is different from other methods of access control (AC), such as list-based AC. ABAC provides a flexible and dynamic solution. **Table 2** [7] shows the analysis of ABAC approaches in big dataset processing (BDP). In the table, examples of AC methods, description of AC methods, security observations, and use case applications are provided.

Table 2. The analysis of ABAC approaches in BDP.

Examples of AC Methods	Description of AC Methods	Security Observations	Use Case Applications
Resource-based	Permissions are directly assigned to the data.	A strengthened method due to adding controls on the data processing resource.	Fine grain dataset control
Object-tagged, rule-based	Using meta-data tags as a part of the decision process of AC	Guaranteeing the integrity of attributes critical to trustworthiness	Dynamic control of datasets
Provenance, history-based	Using the data (meta-data) provenance as a part of the decision process of AC	Supports the execution of a privacy-preserving program	Sanitization needed for combining and processing datasets
Task-role based	Tasks are related to specific roles. Security policies are implemented according to a user's role and the tasks of the role.	A strengthened method due to adding controls to tasks	Fine-grain user control
Role-based, time-bound	Enabling a just-in-time access model; enabling a time-limited and	The risk of unauthorized access is restricted by time	Public Wi-Fi access

	temporary privilege for a user (assigned to a role).		
Content-sensitivity based	The sensitivity score of data is updated according to the data provenance; access to the data is allowed or denied according to a user's access right to the sensitivity score.	Complex (technically), depending on accurate inference.	Need for the inference of the access right
Relationship-based	A user's access to data is based on the relationship (e.g., friends).	A distributed control to a broad set of users results in inconsistencies.	Social networks (online)
Semantic & ontology role-based	Addressing mismatches in the attribute definition related to data from various sources by using semantics & ontologies as the basis for access control.	Complex (technically), depending on accurate inference.	The integration of dataset controls across various systems

Deep learning (DL) has been used in various areas such as big data and cybersecurity. There are various methods and algorithms in DL. The choice of a DL method or algorithm depends on the specific application and available resources (such as data, computers, and software). Examples of DL tasks and methods in cybersecurity and healthcare are shown in **Table 3**[8].

Table 3. DL applications in cybersecurity and healthcare: examples of tasks and methods.

Areas	Examples of Tasks	Examples of Methods
Cybersecurity	Security incidents and fraud analysis	Self-organizing map (SOM) based
	Intrusion detection & classification	Deep belief network (DBN) based
	Detection of network intrusions	Autoencoder (AE) and support vector machine (SVM) based
	Detection of DoS attacks	Restricted Boltzmann machine (RBM) based
	Suspicious flow detection	Hybrid deep-learning-based
	Android malware detection	AE and convolutional neural network (CNN) based
	Zero-day malware detection	Autoencoders and generative adversarial network (GAN) based
Healthcare	Malicious behaviors identification	Recurrent neural network (RNN) based
	Regular health factors analysis	CNN-based
	Cancer classification	Transfer learning based
	COVID-19 detection	CNN-LSTM (long short-term memory) based

4. Cybersecurity of big data in healthcare

A fundamental practice for big data security implementation is data encryption. Data is encrypted at rest and in motion. The healthcare industry uses encryption to protect sensitive data in transit and at rest. Authorized users and preferred recipients should be the only individuals able to read the data. Encrypted files are unreadable to cyber-criminals who do not have decryption keys. Healthcare providers choose their encryption methods based on the organizational workflow.

The healthcare industry must reinforce data privacy and data security. Only those users who require access to sensitive patient information should be allowed to access protected health data. Access control, such as passwords and permissions is an additional protective measure. User authentication ensures that only authorized staff have access to private patient data. These access restrictions limit the number of individuals who can access sensitive patient data. Multi-factor authentication can also be used to validate and secure the identity of users with two or more verification methods. Security teams can swiftly and easily detect unauthorized users and reveal the root cause of privacy leakages in healthcare by restricting and tracking access to protected health information.

Clinics, medical centers, and other healthcare organizations should take a proactive approach to patient privacy. Furthermore, an incident response plan should outline clear roles and responsibilities, regular risk management, and implementation of cybersecurity frameworks (CSFs). These are risk management guidelines that help reduce cybersecurity risks and maintain data management. CSFs work as road maps to secure IT systems and help uncover, counter, and avoid cyber threats and ultimately the consequences of a cyberattack.

Using data management services in healthcare is a key digital transformation resulting in the increased use of Internet of Medical Things (IoMT) devices in hospitals because of the increased use of technology. Enabling remote patient monitoring, these devices advance patient data access, etc. However, this can present a variety of novel privacy issues. Therefore, to ensure and establish big data security for healthcare organizations, ensuring and updating IoMT devices can ensure a complete authentication process.

Mobile and connected devices are used in patient care by both providers and other clinicians. IoMT devices, mobile phones or tablets, and apps are regularly used in the hospital by providers, other clinical staff, and administrators and create another source of vulnerabilities. Cyberattackers can retrieve information or passwords, or even the smartphone or tablet itself, eavesdrop, reconfigure them, or hack a connected device.

Novel digital technology has radically transformed the healthcare ecosystem. Recent global pandemics have accelerated data and processes that have caused global change, leaving the ability of healthcare organizations to protect vitally sensitive PHI and patient privacy questionable. Almost every hospital and clinic handles patient information in a variety of digital systems. Medical data is available to providers such as nurse practitioners, physicians, pharmacists, etc. in the form of EHRs and other software that works with medical data, a very tempting target for cybercriminals. Damage from Ransomware is an ever-increasing risk, and more and more cybercriminals are attacking medical infrastructure. Healthcare providers must understand and

become aware of how to protect patient privacy and data from malicious actors.

Cybersecurity for the healthcare ecosystem presents a unique challenge due to the nature of patient medical data. If a bank card is stolen, it can always be blocked, and a new one issued. However, when laboratory tests or disease information is stolen, leaked, or manipulated in any way, it is impossible to "cancel" it. Additionally, a patient's health and possibly their life can become endangered if the digital clinical chart fails. There are multiple digital complexes and networks in any clinic or hospital, HVAC systems, IoMT devices, and e-prescribing and decision support systems, infusion pumps, that can be threatened by malicious actors. Providers and associated businesses must balance the protection of patient privacy with providing a high level of care and compliance with HIPAA, GDPR, and other applicable regulatory standards. Cybercriminals are quick to take advantage of the fact that it is hard to implement security measures.

Healthcare organizations can prioritize cybersecurity protection to secure PHI. Securing PHI can be fulfilled by protecting devices, data, digital systems, and networks from attacks. Personnel training regarding cybersecurity such as securing PHI is important. Sometimes the most important method of protecting privacy and patient data is simply by training the staff because the lack of cybersecurity knowledge and skills can contribute to major threats to PHI and the organization. Studies have shown that most healthcare employees do not have the cyb ersecurity expertise to prevent or deter attacks. Employees should be able to address the following:

- Recognition of phishing emails—some emails are aimed at specific individuals and are more effective.
- Back-up data: Valuable patient information can be deleted or damaged during cyberattacks therefore, employees should create an encrypted and controlled backup regularly.
- Using digital hygiene technology: Employees should create robust passwords, and be instructed to never click on the unknown, suspicious links, etc.

Healthcare must prioritize cyber-threat protection and prevention by using safety measures aimed at protecting and securing PHI through the protection of digital systems, networks, devices, etc. A lack of IT security knowledge poses a major threat to the global healthcare industry. A survey, conducted by IONOS Cloud found that many healthcare employees lack cybersecurity expertise or knowledge of how to protect patient data, making routing professional training vital to the protection of PHI.

Employee training is often one of the most significantly underestimated cybersecurity practices in the healthcare industry. Data privacy concerns typically arise due to a lack of appropriately trained staff. As healthcare continues to move towards remote work-at-home models, healthcare organizations now have employees who are targeted by malicious actors in their homes. Training staff in data privacy and protection and cybersecurity is essential in preventing home attacks by cybercriminals, including holding simulation drills and conducting regular testing for cyberattacks to find the weakest links. This can help prepare the staff with critical thinking skills and the required knowledge to make smart decisions when a data breach happens. A robust data security risk management program depends on developing a data privacy and protection-conscious

culture supported by trained employees.

5. Cybersecurity of big data in a medical center

Charleston Regional Medical Center is a big medical center in Jackson, Mississippi, USA. This healthcare organization produces big data both internal and external data every day. The clinical chart is the most definitive data that is created, collected, and necessary to protect in the healthcare organization. A clinical chart is created by computers with EMR software, mobile devices such as phones or tablets, and imaging and diagnostics software. The clinical chart contains medical information such as diagnoses, clinical notes, billing data, ICD-10 data, provider and nurse's notes, research data if applicable, and diagnostic and imaging data. This clinical data can be delivered and contained both externally and internally in today's technological environment. Patients can access parts of the clinical charts for hospital admissions and clinic visits with providers. This is both data at rest and data in motion, which is protected by several methods, but primarily through encryption to prevent others from accessing the valuable data.

Malicious actors are highly motivated to steal patient information to sell to various third parties. For example, cybercriminals can access the diagnoses and clinical notes, as well as demographic information to sell to third-party insurance companies who can then deny the patient insurance or use the information to deny patient claims. This can cause hardships for the patient or the patient's family. Passwords, biometrics, fingerprinting, and encryption are all viable methods that are used by healthcare facilities to protect patient information. Clinical staff often use biometrics and passwords to open clinical charts to add clinical notes or other information.

In addition to the hospital or medical facility, whether it is a hospital, clinic, or research facility, the patient chart, or parts of the patient chart can be accessed by the patient at home, billing agencies such as CMS, Medicaid, or other third-party insurance company, or vendors, or research agencies such as CMS, AHRQ, NIH, etc. Encryption is the primary method for protecting the data in motion, but passwords are also used. External access to the patient chart, or parts of the patient chart, can be accessed at home by the patient, by diagnosticians who must read ECGs, ECHO reports, X-rays, etc., or by providers who receive data from the patient via telemedicine. It makes no difference where the patient data comes from or where the data is transmitted to, it must be protected on the same level.

There are numerous cyber threats in healthcare. Protecting private and sensitive patient data in the medical center is threatened by the four most common threats including data breaches, DDoS attacks, ransomware attacks, and phishing. IT security teams must work alongside facility management to provide data protection. Best practice approaches for healthcare service security include more than just responding to threats when they occur. Security teams must be proactive when they are defending against cyberattacks.

Data control usage considerations are also necessary for preventing cyberattacks. In addition to medical centers, provider clinics must also monitor and control malicious activity by implementing and running systems that block or control unauthorized data activities, stop unauthorized emails from being shared, prevent

employees from sharing or copying to external sources, etc. A medical center must use appropriate technology when protecting patient privacy. HIPAA guidelines state that encryption for patient files must be predefined. The cryptographic technology selected should be reasonably necessary and appropriate to prevent malicious actors from accessing patient files.

Data security is key for any healthcare enterprise. Data security concentrates on the protection of data, networks, and computers used by healthcare professionals and other organizations that must deal with the hospital enterprise. HIPAA is a driving factor in the protection and security of healthcare data. These standards offer general directions about managing sensitive patient data. HIPAA outlines necessary tools or techniques that IT teams find necessary to protect all PHI and how to control who can access patient data. Healthcare data management faces significant challenges in protecting patient privacy including health information exchanges, hackers and rising development of "hacktivism", outdated technologies, cloud computing, and mobile technologies, user error when adopting technology, etc.

Assessing the risks of third-party vendors is an important task for the security team. Providers and covered entities transmit patient data daily to provide excellence in service and care and facilitate payment. If a data breach occurs in the medical center because of the negligence of a third-party vendor, the center could be held responsible for the data breach. Therefore, monitoring third-party vendors and other business partners is a critical process for the healthcare ecosystem.

6. Conclusion

Today's business environment is evolving rapidly with multiple apps running on mobile services live in the cloud and the need to regularly develop new methods by which big data is protected is continually expanding because as enterprises evolve, various and more invasive types of cyberattacks occur. Furthermore, cybersecurity expertise and modes such as firewalls, anti-virus and malware tools, and email filters are not certain to protect any business from cybercriminals. Technology reaches the farthest in business defense, policies and procedures surrounding accessing, and handling data breaches, and tightening big data security needs to be part of a complete and robust security strategy. Asset management is a critical necessity in any business environment. Risk assessment does not always ensure vigorous risk management and governance of the asset management circumstances. Big data security is a high-stakes must for corporations and organizations, and procedures ensuring protection should be a high priority for the IT security team. Awareness of malicious threats and training for all personnel who may be involved in access control and maintenance helps ensure big data security, especially when sensitive data is involved in healthcare and medical centers. The future research of our research team includes 1) Big Data analytics, artificial intelligence (AI)/machine learning (ML), and automation in cybersecurity; and 2) defensive and offensive cybersecurity solutions.

Acknowledgements

The authors would like to express thanks to Technology and Healthcare Solutions, USA for its help and

support.

Conflict of interest

The authors would like to announce that there is no conflict of interest.

References

1. Ullah, F., & Babar, M. A. (2019, December). Quantifying the Impact of Design Strategies for Big Data Cyber Security Analytics: An Empirical Investigation. In 2019 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT) (pp. 146-153). IEEE.
2. Atapour-Abarghouei, A., McGough, A. S., & Wall, D. S. (2020, December). Resolving the cybersecurity data sharing paradox to scale up cybersecurity via a co-production approach towards data sharing. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 3867-3876). IEEE.
3. Chen, J., & Xue, Y. (2017, June). Bootstrapping a blockchain based ecosystem for big data exchange. In 2017 IEEE international congress on big data (bigdata congress) (pp. 460-463). IEEE.
4. Alexander, C. A., & Wang, L. (2019, April). Cybersecurity, information assurance, and big data based on blockchain. In 2019 SoutheastCon (pp. 1-7). IEEE.
5. Rassam, M. A., Maarof, M., & Zainal, A. (2017). Big Data Analytics Adoption for Cybersecurity: A Review of Current Solutions, Requirements, Challenges and Trends. Journal of Information Assurance & Security, 12(4).
6. Rawat, D. B., Doku, R., & Garuba, M. (2019). Cybersecurity in big data era: From securing big data to data-driven security. IEEE Transactions on Services Computing, 14(6), 2055-2072.
7. Tall, A. M., & Zou, C. C. (2023). A Framework for Attribute-Based Access Control in Processing Big Data with Multiple Sensitivities. Applied Sciences, 13(2), 1183.
8. Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Computer Science, 2(6), 420.