Tec Empresarial
P-ISSN: 1659-2395; E-ISSN: 1659-3359

"A REVIEW OF SECURITY AND PRIVACY PRESERVATION OF HEALTHCARE DATA USING BIG DATA ANALYTICS TECHNIQUES."

Ms. Bhavika Milind Gharat
PhD Research Scholar
Computer Engineering
Surendranagar University
gharatbhavi@gmail.com

Dr. Ishaan Tamhankar
Assistant Professor
Research Guide
Surendranagar University
prof.ishaantamhankar@gmail.com

Abstract:

Incredible amounts of data is being generated by various organizations like hospitals, banks, e-commerce, retail and supply chain, etc. by virtue of digital technology. Not only humans but machines also contribute to data in the form of closed-circuit television streaming, web site logs, etc. Tons of data is generated every minute by social media and smart phones. The voluminous data generated from the various sources can be processed and analyzed to support decision making. However, data analytics is prone to privacy violations. It is in this context that this paper aims to present the state-of-the-art security and privacy issues in big data as applied to healthcare industry and discuss some available data privacy, data security, users' accessing mechanisms and strategies.

Keywords:

Big Data, Privacy and Security, Healthcare, Privacy preserving, Security Lifecycle.

Introduction

The current trend toward digitizing healthcare workflows and moving to electronic patient records has seen a paradigm shift in the healthcare industry. The quantity of clinical data that are available electronically will be then dramatically increased in terms of complexity, diversity, and timeliness, resulting what is known as big data. Driven by mandatory requirements and the potential to improve care, save lives and lower costs, big data hold the promise of supporting a wide range of unprecedented opportunities and use cases, including these key examples: clinical decision support, health insurance, disease surveillance, population health management, adverse events monitoring, and treatment optimization for diseases affecting multiple organ systems. [1,2]

Even though the adoption of big data technologies in healthcare sector carries many benefits and promises, it also raises some barriers and challenges. Indeed, the concerns over the sensitive information security and privacy are increased year by year because of several growing trends in healthcare, such as clinician mobility and wireless networking, health information exchange, cloud computing and so on. Moreover, healthcare organizations found that a reactive, bottom-up, technology-centric approach to determining security and privacy requirements is not adequate to protect the organization and its patients [3]

To prevent breaches of sensitive information and other types of security incidents, a proactive, preventive approach and measures must be taken by every healthcare organization with attention to future security and privacy needs.

In this paper, we will discuss some successful and interesting related works. We will also present risks to the security of health data and discuss some newer technologies and redressal of these risks using new techniques. Then, we will focus on the privacy issue in healthcare, and mention various laws and regulations established by different regulatory bodies as well as some feasible methods and techniques used to ensure the patient’s privacy.

Related Work

Seamless integration of highly diverse big data technologies in healthcare can facilitate faster and safer throughput of patients, enable to gain deeper insights into the clinical and organizational processes, create greater efficiencies and help improve patient flow, safety, quality of care and the overall patient experience meanwhile contain costs.

Such was the case with the UNC Health Care (UNCHC), which is a not-for-profit integrated healthcare system in North Carolina that has implemented a new system allowing clinicians to rapidly access and analyze unstructured patient data using natural-language processing. In fact, UNCHC has accessed and analyzed huge quantities of unstructured content contained in patient medical records to extract insights and predictors of readmission risk for timely intervention and providing safer care for high-risk patients and reduce re-admissions. [4]

Another example in United States is the Indiana Health Information Exchange, which is a non-profit organization, provides a secure and robust technology network of health information linking more than 90 hospitals, community health clinics, rehabilitation centers and other healthcare providers in Indiana. It allows medical information to follow the patient hosted in one doctor office or only in a hospital system. [5]

One more example, is Kaiser Permanente medical network based in California, which has more than 9 million members, estimated to manage large volumes of data ranging from 26,5 Petabytes to 44 Petabytes. [6]

Big data analytics is used also in Canada, e.g. the infant hospital of Toronto. This hospital succeeded to improve the outcomes for newborns prone to serious hospital infections. In Europe this time, exactly in Italy, the Italian Medicines agency collects and analyzes a large amount of clinical data concerning expensive new medicines as part of a national profitability program. Based on the results, it may reassess the medicines prices and market access terms. [7]

After Europe, Canada, Australia, Russia, and Latin America, Sophia Genetics, global leader in Data-Driven Medicine, announced at the recent 2017 Annual Meeting of the American College of Medical Genetics and Genomics (ACMG) that its artificial intelligence has been adopted by African hospitals to advance patient care across the continent.[8]

Big data security in healthcare

Healthcare organizations store, maintain and transmit huge amounts of data to support the delivery of efficient and proper care. Nevertheless, securing these data has been a daunting requirement for decades. Complicating matters, the healthcare industry continues to be one of the most susceptible to publicly disclosed data breaches. In fact, attackers can use data mining methods and procedures to find out sensitive data and release it to public and thus data breach happens. While implementing security measures remains a complex process, the stakes are continually raised as the ways to defeat security controls become more sophisticated. As a result, it is crucial that organizations implement healthcare data security solutions that will protect important assets while also satisfying healthcare compliance mandates.

Technologies in use

Various technologies are in use for protecting the security and privacy of healthcare data. Most widely used technologies are:

1) Authentication: Authentication is the act of establishing or confirming claims made by or about the subject are true and authentic. It serves a vital function within any organization: securing access to corporate networks, protecting the identities of users, and ensuring that a user is who he claims to be. Most cryptographic protocols include some form of endpoint authentication specifically to prevent man-in-the-middle (MITM) attacks. For instance,11 Transport Layer Security (TLS) and its predecessor, Secure Sockets Layer (SSL), are cryptographic protocols that provide security for communications over networks such as the Internet. TLS and SSL encrypt the segments of network connections at the Transport Layer end-to-end. Several versions of the protocols are in widespread use in applications like web browsing, electronic mail, Internet faxing, instant messaging and voice-over (VoIP). [9]

2) Encryption: Data encryption is an efficient means of preventing unauthorized access of sensitive data. Its solutions protect and maintain ownership of data throughout its lifecycle — from the data center to the endpoint (including mobile devices used by physicians, clinicians, and administrators) and into the cloud. Encryption is useful to avoid exposure to breaches such as packet sniffing and theft of storage devices. [20,21] Healthcare organizations or providers must ensure that encryption scheme is efficient, easy to use by both patients and healthcare

professionals, and easily extensible to include new electronic health records. Furthermore, the number of keys hold by each party should be minimized. [22]

3) Data Masking: Masking replaces sensitive data elements with an unidentifiable value but is not truly an encryption technique so the original value cannot be returned from the masked value. It uses a strategy of de-identifying the data sets or masking personal identifiers such as name, social security number and suppressing or generalizing quasi- identifiers like data-of-birth and zip-codes. Thus, data masking is one of the most popular approaches to live data anonymization. k-anonymity first proposed by Swaney and Sarmatia [12,13] protects against identity disclosure but failed to protect against attribute disclosure. Truta et al. [14] have presented p-sensitive anonymity that protects against both identity and attribute disclosure. Other anonymization methods fall into the classes of adding noise to the data, swapping cells within columns, and replacing groups of k records with k copies of a single representative. These methods have a common problem of difficulty in anonymizing high dimensional data sets [15, 16].

4) Access Control: Once authenticated, the users can enter an information system, but their access will still be governed by an access control policy which is typically based on the privilege and right of each practitioner authorized by patient or a trusted third party. It is then, a powerful and flexible mechanism to grant permissions for users. It provides sophisticated authorization controls to ensure that users can perform only the activities for which they have permissions, such as data access, job submission, cluster administration, etc. Several solutions have been proposed to address the security and access control concerns. Role-Based Access Control (RBAC) [17] and Attribute-Based Access Control (ABAC) [18,19] are the most popular models for EHR. RBAC and ABAC have shown some limitations when they are used alone in medical system.

Big data privacy in healthcare

Recent years have seen the emergence of advanced persistent threats, targeted attacks against information systems, whose main purpose is to smuggle recoverable data by the attacker. Therefore, invasion of patient privacy is considered as a growing concern in the domain of big data analytics, which make organizations in challenge to address these different complementary and critical problems. In fact, data security governs access to data throughout the data lifecycle while data privacy sets this access based on privacy policies and laws which determine, for example, who can view personal data, financial, medical or confidential information. An incident reported in the Forbes magazine raises an alarm over patient privacy [10]. In the report, it mentioned that Target Corporation sent baby care coupons to a teen-age girl unbeknown to her parents. This incident impels big data to consider privacy for analytics and developers should be able to verify that their applications conform to privacy agreements and that sensitive information is kept private regardless of changes in the applications and/or privacy regulations.

Privacy of medical data is then an important factor which must be seriously considered.

Data protection laws

More than ever, it is crucial that healthcare organizations manage and safeguard personal information and address their risks and legal responsibilities in relation to processing personal data, to address the growing thicket of applicable data protection legislation. Different countries have different policies and laws for data privacy. Data protection regulations and laws in some of the countries along with salient features are listed in the Table below.

Country	Law	Salient Features
U.S.A	HIPAA Act Patient Safety and Quality Improvement Act (PSQIA) HITECH Act	Requires the establishment of national standards for electronic health care transactions. Gives the right to privacy to individuals from age 12 through 18. Signed disclosure from the affected before giving out any information on provided health care to anyone, including parents. Patient Safety Work Product must not be disclosed. Individual violating the confidentiality provisions is subject to a civil penalty. Protect security and privacy of electronic health information.
EU	Data Protection Directive	Protect people's fundamental rights and freedoms and in particular their right to privacy with respect to the processing of personal data.
Canada	Personal Information Protection and Electronic Documents Act ('PIPEDA')	Individual is given the right to know the reasons for collection or use of personal information, so that organizations are required to protect this information in a reasonable and secure way.
UK	Data Protection Act (DPA)	Provides a way for individuals to control information about themselves. Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects.
India	IT Act and IT (Amendment) Act	Implement reasonable security practices for sensitive personal data or information. Provides for compensation to person affected by wrongful loss or wrongful gain. Provides for imprisonment and/or fine for a person who causes wrongful loss or wrongful gain by disclosing personal information of another person while providing services under the terms of lawful contract.

Russia	Russian Federal Law on Personal Data	Requires data operators to take “all the necessary organizational and technical measures required for protecting personal data against unlawful or accidental access”.
Morocco	The 09-08 act, dated on 18 February 2009	Protects the one’s privacy through the establishment of the CNDP authority by limiting the use of personal and sensitive data using the data controllers in any data processing operation.

Conclusion

Limitless opportunities are offered for big data to drive health research, knowledge discovery, clinical care, and personal health management. However, there are a number of obstacles and challenges that impede its true potential in the healthcare field, including technical challenges, privacy and security issues and skilled talent. Big data security and privacy are considering as the huge barrier for researchers in this field. In this paper, we have discussed some examples of successful related work across the world. Privacy and security issues in each phase of big data life cycle are also presented along with the advantages and disadvantages of existing privacy and security technologies in the context of big healthcare data.

In this context, as our future direction, perspectives will focus more on achieving effective solutions to the scalability problem of big data privacy and security in the era of healthcare. And to go further, we will try to solve the problem of reconciling security and privacy models by simulating diverse approaches using exploiting the MapReduce framework. To, ultimately, support decision making and planning strategies.

References:

[1] L. Sweeney, “Achieving k-anonymity privacy protection using generalization and suppression,” in international journal on uncertainty, fuzziness and knowledge-based systems, vol. 10, 2002, pp. 571 – 588.

[2] P. Sarmatia, “Protecting respondent’s identities in microdata release,” in IEEE transactions on knowledge and data engineering, vol. 13, 2001, pp. 1010 – 1027.

[3] Ritu Ratra, Preeti Gulia, Nasib Singh Gill, Jyoti Moy Chatterjee (2022). Big Data Privacy Preservation Using Principal Component Analysis and Random Projection in Healthcare, Mathematical Problems in Engineering, vol. 2022, Article ID 6402274, pp. 12.

[4] T. M. Truta and B. Vinay, “Privacy protection: p-sensitive k-anonymity property,” in Proceedings of 22nd International Conference on Data Engineering Workshops, 2006, p. 94.

[5] N. Spruill, “The confidentiality and analytic usefulness of masked business microdata,” in Proceedings on survey research methods, 1983, pp. 602–607.

[6] S. Chawla, C. Dwork, F. M. Sheny, A. Smith, and H. Wee, “Towards privacy in public databases,” in Proceedings on second theory of cryptography conference, 2005.

[7]. Science Applications International Corporation (SAIC). Role-Based Access Control (RBAC) Role Engineering Process Version 3.0. 11 May 2004.

[8] A. Mohan, D. M. Blough, An Attribute-Based Authorization Policy Framework with Dynamic Conflict Resolution, Proceedings of the 9th Symposium on Identity and Trust on the Internet, 2010.

[9] J. Shafer, S. Rixner, and A. L. Cox. The Hadoop Distributed File system: Balancing Portability and Performance. Proc. of 2010 IEEE Int. Symposium on Performance Analysis of Systems & Software (ISPASS), March 2010, White Plain, NY, pp. 122-133.

[10] N. Somu, A. Ganga, and V. S. Sriram, "Authentication Service in Hadoop Using one Time Pad," Indian Journal of Science and Technology, vol. 7, pp. 56-62, 2014.

[11] C. Yang, W. Lin, and M. Liu, "A Novel Triple Encryption Scheme for Hadoop-Based Cloud Data Security," in Emerging Intelligent Data and Web Technologies (EIDWT), 2013 Fourth International Conference on, 2013, pp. 437-442.

[12] Shahzad, Aamir &amp; Kayani, Haroon &amp; Malik, A. (2023). Big Data Security, Privacy Protection, Tools and Applications. Pakistan Journal of Science. Vol75 No 2. 353-372. 10.57041/pjs.v75i02.850.

[13] Batko, K., Ślęzak, A. (2022). The use of Big Data Analytics in healthcare. J Big Data 9, pp.3. https://doi.org/10.1186/s40537-021-00553-4.

[14] Tilahun, Tewodrose &amp; Tsegaye, Solomon. (2022). The Security Challenges of Big Data Analytics: A Systematic Literature Review. Asian Journal of Research in Computer Science. 184-197. 10.9734/ajrcos/2022/v14i4303.

[15] Baig, M. I., Shuib, L., &amp; Yadegaridehkordi, E. (2020). Big data in education: a state of the art, limitations, and future research directions. International Journal of Educational Technology in Higher Education, 17(1), 1-23.

[16] Butpheng, C., Yeh, K.-H., &amp; Xiong, H. (2020). Security and Privacy in IoT- Cloud-Based e-Health Systems—A Comprehensive Review. Symmetry, 12(7), 1191. https://doi.org/10.3390/sym12071191.

[17] R. Ratra and P. Gulia (2020). “Experimental evaluation of open-source data mining tools (WEKA and orange),” International Journal of Engineering Trends and Technology, vol. 68, no. 8, pp. 30–35.

[18] Abouelmehdi, Karim &amp; Beni Hssane, Abderrahim &amp; Khaloufi, Hayat &amp; Saadi, Mostafa. (2017). Big data security and privacy in healthcare: A Review. Procedia Computer Science. 113. 73-80.10.1016/j.procs.2017.08.292.

[19] Abouelmehdi, K., Beni-Hessane, A. &amp; Khaloufi, H. Big healthcare data: preserving security and privacy. J Big Data 5, 1 (2018). https://doi.org/10.1186/s40537-017-0110-7.

[20] Kaur, P., Sharma, M., &amp; Mittal, M. (2018). Big Data and Machine Learning Based Secure Healthcare Framework. Procedia Computer Science, 132, 1049-1059. https://doi.org/10.1016/j.procs.2018.05.020.

[21] Mahmood, Tariq &amp; Afzal, Uzma. (2013). Security Analytics: Big Data Analytics for cybersecurity: A review of trends, techniques and tools. 129-134. 10.1109/NCIA.2013.6725337.

[22] H. Zhou and Q. Wen, "Data Security Accessing for HDFS Based on Attribute-Group in Cloud Computing," in International Conference on Logistics Engineering, Management and Computer Science (LEMCS 2014), 2014.