Security and Privacy of Health Data: A Review of the Challenges and Approaches

Kamya Eria
Faculty of Computing, Engineering & Technology
Asia Pacific University of Technology and Innovation
tp048982@mail.apu.edu.my

Abstract—Health data is generated on a daily basis amidst privacy and security threats. Because patients feel that their data is highly sensitive, they are not willing to disclose much about their health status. This is because of the privacy and security issues feared by the patients. As a result, patient data collected is left with a lot of missing values, outliers, noise and some inconsistencies. Those who willingly provide their sensitive data are left in suspense about the privacy and security of this data wherever its used. These security concerns are also a big threat to the medical institutions from hindering medical research to huge financial losses. This review focuses on the challenges and approaches to these security and privacy issues. Prevailing medical practices used in addressing these issues have also been discussed.

Key Words- Electronic Health Records, Medical Data, Privacy, Security.

1.0 Introduction

Data is known to be a vital asset for any organization for the insights derived from it (Sahi, Lai, & Li, 2016) . Different institutions generate data and make great use of this data. Health institutions as well gather data from their clients amidst many security and privacy challenges. As a result, a lot of data is left out by patients when approached for their health data. Such patients fear to have their health status privacy invaded. Security and privacy issues have always worried patients due to fear of stigma issues in their societies (Vayena, Salathé, Madoff, & Brownstein, 2015). Security and privacy issues have been accelerated by rapid technological developments in the health sector such as Electronic Health Records (EHRs) and other forms of digitization in information systems.

Due to these security and privacy threats, it has always necessitated organizations to employ robust and versatile data controls (Ehsan Rana, Kubbo, & Jayabalan, 2017). Information governance, data governance and data management have been given special attention to exercise maximum control and authority over the management and use of health data. This is aimed at restoring trust in clients about the safety, security and privacy of their data. Information governance refers to the management of organizational information through set procedures and policies. Data governance is all about exercising authority over the management and use of data as an organizational asset (Hall, 2017). Data management is concerned with acquiring, validating, storing, protecting and processing data to reveal useful information from it (Bärenfänger, Otto, & Gizanis, 2015). Having perfect control over data promotes its proper usage in medical research.

The remainder of this report is organized as follows: The next section looks at related works with emphasis on security and privacy of patient data. It is then followed by data governance and management practices and methods used to exercise control of medical data. Recommendations follow up after which conclusions are made.

2.0 Related works

In health institutions such as cancer hospitals, it is a very vital practice to keep record of whatever transpires during the institution's operations as they try to improve the health of patients (Ehsan Rana et al., 2017). As a result, a lot of patient data is always recorded. Such data turns out to be very invaluable when modern data analytical technologies are used, from getting actionable insights out of this data to as far as incredible medical innovations like personalized medication (Panahiazar, Taslimitehrani, Jadhav, & Pathak, 2014). As this data accumulates in the data warehouses, it has prompted practitioners to adopt cheaper means for easy, timely storage and access to this data (Sahi et al., 2016).

Recent technological trends such as cloud computing have made collection of this data easier and more efficient, to the extent of enabling real time data collection and analytics. This has been implemented through the use of EHRs and Electronic Medical Records (EMRs) (Jayabalan and O 'daniel, 2017; ISO, 2013). Moura and Serrão (2015) identified that because such data tends to be

unstructured and from various technological sources, the privacy and security of this data tends to be at stake.

However, it was made clear in the Universal Declaration of Human Rights by United Nations (1948) that every human being has a right to personal security and privacy. This was stated in the third and twelfth articles of the declaration respectively, thus giving every person an inherent right to live securely and without being vulnerable to their personal privacy invasion. It is therefore very important that as patient data availability is maximized, personal privacy that comes with this health data should as well be maximized. Jayabalan and O 'daniel (2017) added that since health data involves sensitive issues such as disease severity and past health status, it calls for maximum efforts to ensure privacy of this data. Filkins et al. (2016) also emphasized that the security of this health data should be ensured until it reaches the intended destination.

2.1 Security and Privacy

This section seeks to give a clear distinction between security and privacy. Privacy is the right for a person to have their personal secrets undisclosed (HIPAA, 2015). Nass and Levit, L. A. Gostin (2009) added that privacy varies from person to person, implying that what is regarded as highly private by one person may not be regarded the same by another person. Security refers to the measures and mechanisms put in place to ensure that privacy is observed (HIPAA, 2017). Khan, Sayed and Hoque (2016) added that security is concerned with protecting medical data from unintended access such as intruders, malwares and frauds. The authors also elaborated further on privacy saying its existence starts in any arrangement where personally identifiable and sensitive information of any form is collected and kept. Miller School of Medicine (2009) asserted that security is basically obtained through operational and technical controls to unauthorized access to patient data.

2.2 Medical Data

Khan, Sayed and Hoque (2016) described medical data as that which is collected for use in the process of medical diagnosis. However, in many occasions this data is collected to facilitate medical research (Nass & Levit, L. A. Gostin, 2009). Health institutions deal with data in both structured and unstructured formats. This data also comes in very large volumes. The reason behind varying formats is due to the proliferation of technology leading to new and advanced means of data collection (Jayabalan & O 'daniel, 2017). The authors also identified advanced means such as sensors and smart wearable devices responsible for the rapid inflow of health data. Global networking technologies such as the Internet of Things (IoT) and Cloud computing are the primary drivers for the unprecedented data amounts (Sahi, Lai and Li, 2016; Bertino, 2016). In fact, Ehsan Rana, Kubbo and Jayabalan (2017) acknowledge the affordability of cloud computing describing it as a cost effective way of data collection. This ultimately justifies the high rate of its adoption. Hospital Information Systems (HIS) have also been introduced purposely to boost the potential of data accessibility to both patients and medical workers (Mehraeen, Ayatollahi, & Ahmadi, 2016).

Medical data ranges from Personal Health Records (PHRs), EHRs and EMRs. PHRs are health data owned and managed by the patient himself (Sahi et al., 2016). The authors added that good ones of this kind contain precise information about the patient's historical data which can be provided to anyone authorized to access it. EHRs as confirmed by the same authors refer to a repository where information regarding a particular medical subject is kept. Liu et al. (2018) asserted that in EHRs, data is loaded onto a cloud system from which doctors and nurses can get access to it. Ehsan Rana, Kubbo and Jayabalan (2017) acknowledged that EHRs are very instrumental in helping medical practitioners to deliver fast and better health services. EMRs refer to the electronic version of paper records that health workers use during medical service provision.

2.3 Why Security and Privacy

The sensitivity that comes with medical data, given the fact that it is concerned with a person's health status should not be overlooked (Mehraeen et al., 2016). When patients choose to be served by a certain health institution, they deserve to receive back this trust by these institutions remaining trustworthy in securing their health data. Liu et al. (2018) added that confidentiality is one thing that patients are very much concerned about. Ehsan Rana, Kubbo and Jayabalan (2017) confirmed that over 75% of patients are always not contented with how their health data is shared with other health institutions involved in the provision of medical services to them. Confidentiality in this case refers to the obligations to those who collect and use medical data to ensure its security and privacy (Nass & Levit, L. A. Gostin, 2009).

Van Staa et al. (2016) also identified that patients are always not sure about who uses their data. They only come to appreciate this by the information drawn from this data. HealthIT.gov (2013)also added that ensuring privacy and security builds trust, confidence and customer loyalty in patients. It therefore facilitates research because patients are willing to disclose their health data once they are guaranteed about its security and privacy (Nass & Levit, L. A. Gostin, 2009). Furthermore, the authors asserted that breach of security and privacy commitments can cause irreparable harm to

individuals thus inflicting fear, esteem issues and loss of dignity. Such individuals can even shy away from public events and other social activities due to stigma issues.

2.4 Risks and Implications

Violating privacy and security requirements amounts to certain risks born to the health institution. Sometimes data violations and breaches can cause serious problems to the health institution leading to serious setbacks. Some of the highlighted risks that health institutions are exposed to in case of privacy and security problems are described below:

Data loss: Security attacks into health data can result into loss of data to these hackers or natural disasters. This data may not be recovered at all or recovered at high costs.

Criminal Penalties: Health institutions or individual medical workers exposing someone's privacy risk criminal penalties due to denial of patients from their inherent privacy rights (HealthIT.gov, 2013).

Civil Penalties: Health institutions risk civil actions from civil rights activists upon violations of the HIPAA rules (HealthIT.gov, 2013; Abouelmehdi, Beni-Hessane and Khaloufi, 2018).

Loss of Trust: Trust is built over a long period of time. However, once this trust is lost from a health institution's clients, it will take them another considerable while to rebuild it.

3.0 Approaches to Security and Privacy

Various approaches have been developed to protect the privacy and this is through effective security measures (Nass & Levit, L. A. Gostin, 2009). The authors also noted that an effective privacy protection attempt should have an effective security system. Sahi, Lai and Li (2016) identified cryptography as a widely accepted versatile approach to enforcing data security and privacy in cases where third parties are involved. Khan, Sayed and Hoque (2016) also added authentication and data masking on top of encryption. Filkins et al. (2016) also acknowledged these three methods but observed that of them is considered perfect. The authors advised that critical and thorough evaluation is the only way to make them good enough as well as a well-thought-out implementation process.

3.1 Authentication

This approach has long been used in enforcing security of data. It is a form of user verification which only responds to the user needs upon satisfaction. The traditional method is the use of passwords and usernames to verify the legitimacy of the user (Jayabalan and O 'daniel, 2017; Filkins et al., 2016). This is known as Single Factor Authentication and the simplest as well as inexpensive to implement. The authors however noted its vulnerability to easy cracks and phishing attacks especially to users who set weak passwords. In 2012, hackers are known to have exploited weak password of an administrator system (Filkins et al., 2016). As a result, 780,000 records of Medicaid patient data were breached.

The authors also called for the need to employ Multi Factor Authentication which combines two or more authentication elements such as smart cards, biometrics and security tokens. A Three Factor authentication system can offer greater security through combination of three different authentication elements(Jayabalan & O 'daniel, 2017). Such a strong authentication is however difficult to implement even though it is a more efficient security and privacy measure.

3.2 Data Masking

This approach is intended to hide the true identity of individuals whose information is included in health data thus preserving their privacy (Tucker et al., 2016). Abouelmehdi, Beni-Hessane and Khaloufi (2018) added that data masking de-identifies datasets by hiding one's unique identifiers such as names and social security number. It also uses generalization of quasi identifiers such as birth date and zip codes. Tucker et al., (2016) identified some examples of how data masking has been implemented:

- Diversity models have been developed to purposely transform data in ways that ensure that specific individuals cannot be identified within public databases. These provide a guarantee of a pre-specified level of anonymity based on non-uniqueness of records within the transformed data.
- Data reduction techniques such as generalization of the data can also be used to transform data. In this case, values are grouped into categories thereby masking of data since specific values or whole records are eliminated from the dataset.
- Data perturbation techniques can also be applied, which add random noise to the true values.
- Use of aggregate results, such as area level census data.
- Use of results from data analysis such as data mining through applications which hide certain data attributes rather than direct exposure to data.

3.3 Encryption

Abouelmehdi, Beni-Hessane and Khaloufi (2018) described this method as an efficient method of protecting sensitive data. It ensures data protection throughout the life cycle of data from the start to the end point. The authors asserted that providers of this method

should ensure that it has an efficient functional scheme which can simply be used by both patients and health workers. Several encryption algorithms have been developed such as RSA, Rijndael, AES, DES, 3DES, RC4 and RC6 (Abouelmehdi et al., 2018). The authors however added that selection of the appropriate algorithm to use is still a challenge.

3.4 Access control

With this method, authorized users can enter the information system upon authentication but still have to go through an access control system (Abouelmehdi et al., 2018). The authors further elaborated that the access control policy usually assesses the privileges and rights of practitioners to information which is authorized by patients or trusted third parties. Jayabalan and O 'daniel (2017) asserted that access control is a trusted mechanism in ensuring confidentiality and integrity of both data and resources since only authorized persons are granted access to information. The authors also identified the vitality of access controls in observing patient consent in information flows. In this case, patient consent should be incorporated in access control policies thus giving patients authority to regulate access to practitioners especially under normal situations.

3.5 Regulations about health privacy and security

Different international regulatory bodies have put up regulations to maintain the privacy and security of patient data. As a result the Health Insurance Portability and Accountability Act (HIPAA) privacy and security rules were established to ensure that health institutions abide by these set policies (Harsh, Patil, & Seshadri, 2015). The Privacy Preserving Data Publishing (PPDP) procedures were also established to preserve patient privacy in the course of publishing the results of medical research. PPDP recommends different measures to ensure confidentiality when patients' (Zaman & Obimbo, 2014) data are published such as anonymous publishing.

4.0 Challenges to privacy and security of medical data

Recently, challenges in health data privacy and security began to surface and hence attention has been given to them (Ehsan Rana, Kubbo and Jayabalan, 2017; Jayabalan and O'daniel, 2017). Both authors also identified that these challenges have long been in existence in financial institutions until recently that they also crossed to the health sector due to the high sensitivity of the data dealt with. Khan, Sayed and Hoque (2016) identified that the major challenge to medical data security and privacy is data sharing among practitioners while protecting personally identifiable information.

There are two major backups to the challenges encountered by hospitals as they pursue to ensure data security and privacy. These are: Collaboration and technological advancements. For instance, in many cases patients are referred to other hospitals probably for better medical services. The referral hospital may be a partner to the referring hospital or not. In both cases maximum cooperation is required in terms of patient data provision to facilitate faster response to the health problem at hand. Such cooperation presents a challenge because patient information is to be shared which exposes it to privacy and security issues especially at the receiving end.

Sometimes governments also need health information to conduct health research. In such a situation, cooperation is required from health institutions in terms of patient data which further presents these challenges (Nass & Levit, L. A. Gostin, 2009). Technology on the other hand presents these challenges through two ways: The first is about data hackers who use sophisticated technological tools to gain access to this data. The other is through the EHRs and other HIS that embrace cloud computing. Since these systems gained approval from many health practitioners (Ehsan Rana et al., 2017), information is exposed to various users in the network thus increasing the risk of privacy and security issues.

Having the major drivers to these challenges in mind, it now brings us to the actual challenges. Medical cooperation is driven by the motive to save lives in the shortest affordable time. It is during execution of this objective that formidable challenges of data breaches and confidentiality issues show up. Confidentiality being centered on trust and obligation, it remains a tricky challenge since patients need to trust health institutions who in turn are obliged to respect this trust.

4.1 Data breaches

These are simply data leakages during medical information flow and sharing. Because HIS have widely been adopted, many doctors and nurses get access to this data. Data leakages then become imminent in cases where weak authentication measures are used. Jayabalan and O 'daniel (2017) added that such measures include weak passwords by doctors who do not want to stress up with long (and most probably stronger) passwords due to the frequency of need to access medical data. These weak authentication measures make the HIS vulnerable to opportunists. Data hackers are also always on the lookout for such loop holes (Moura & Serrão, 2015). However, in many cases the strength of the password may not hinder them to hack into these systems because of technologically advanced tools at their disposal.

It should also be noted that there is a strong financial motive behind the actions of these hackers (Ehsan Rana,

Kubbo and Jayabalan, 2017; Khan, Sayed and Hoque, 2016). The authors further highlighted that over 89% of the data breaches are driven by financial motives which is reportedly 10 times financially valued than financial data in the black market. With such a driving force, data security and privacy is exposed to big risks which challenge health institutions to avoid the worst cases that can erupt from data breaches. Formidably, Sahi, Lai and Li (2016) asserted that the possibility for data breaches to occur is real and that they can happen either intentionally or accidentally. This is more especially with the use of cloud computing in data sharing.

5.0 Recommendations

Health institutions should observe some precautionary measures to avoid the worst out of a data security and privacy scenario.

Data back up: This method is used to create a separate version of the same data in order to recover data in case of data loss. This method responds to certain extreme situations such as natural disasters of earth quakes, floods and volcanos (HealthIT.gov, 2013). It also restores health data in case of data breaches. Health institutions should therefore back up their data to avoid total shut down of the HIS in case of any unprecedented scenario (Khan et al., 2016).

Extra care: Medical practitioners should be careful when sharing health information in order to minimize possible security and privacy violations. Khan, Sayed and Hoque (2016) added that workers should always sign out from systems to avoid giving room to any opportunist looking out for such a loop hole.

Routine checks: HIS should always be checked to spot out any likely system failures, hacks and other likely data security and privacy issues in advance (Khan et al., 2016). It should always be remembered that prevention is better than cure.

6.0 Conclusions

Health data such as the case for cancer patients is highly sensitive and deserves maximum security and privacy maintenance. Health institutions should therefore devote their best to avoid such embarrassing privacy and security incidences. Since HIS systems and other EHRs systems remain preferred in this technology era, it is highly advisable that data back-ups, routine checks and extra care be highly observed by these institutions in a bid to minimize the effects of an inevitable scenario.

References

Abouelmehdi, K., Beni-Hessane, A., & Khaloufi, H. (2018). Big healthcare data: preserving security and privacy. *Journal of Big Data*, *5*(1), pp.1–18, https://doi.org/10.1186/s40537-017-0110-7

Bärenfänger, R., Otto, B., & Gizanis, D. (2015). Business and Data Management Capabilities for the Digital Economy, *14*(May).

Bertino, E. (2016). Data Security and Privacy in the IoT. *Proceedings of the 19th International Conference on Extending Database Technology*, pp.1–3, https://doi.org/10.5441/002/edbt.2016.02.

Ehsan Rana, M., Kubbo, M., & Jayabalan, M. (2017). Privacy and Security Challenges Towards Cloud Based Access Control in Electronic Health Records. Retrieved from http://docsdrive.com/pdfs/medwelljournals/ajit/2017/27,pp.4-281.pdf.

Filkins, B. L., Kim, J. Y., Roberts, B., Armstrong, W., Miller, M. A., Hultner, M. L., … Steinhubl, S. R. (2016). Privacy and security in the era of digital health: What should translational researchers know and do about it? *American Journal of Translational Research*, *8*(3), pp.1560–1580.

Hall, J. (2017). Association for Information Systems AIS Electronic Library (AISeL) Data Governance at State Departments of Transportation Data Governance at State Departments of Transportation. Retrieved from http://aisel.aisnet.org/mwais2017%0Ahttp://aisel.aisnet.org/mwais2017/24.

Harsh, N., Patil, K., & Seshadri, R. (2015). Big data security and privacy issues in healthcare. Retrieved from https://pdfs.semanticscholar.org/6220/ac5410c1c8f30e8865221a8695f0f7f5034f.pdf.

HealthIT.gov. (2013). Guide to Privacy and Security of Health Information, (April), pp.27–40.

HIPAA. (2015). Privacy | HHS.gov. Retrieved March 4, 2018, from https://www.hhs.gov/hipaa/for-professionals/privacy/index.html.

HIPAA. (2017). The Security Rule | HHS.gov. Retrieved March 4, 2018, from https://www.hhs.gov/hipaa/for-professionals/security/index.html.

ISO. (2013). ISO/TS 14441:2013(en), Health informatics —

Security and privacy requirements of EHR systems for use in conformity assessment. Retrieved March 4, 2018, from https://www.iso.org/obp/ui/#iso:std:iso:ts:14441:ed-1:v1:en.

Jayabalan, M., & O 'daniel, T. (2017). Continuous and Transparent Access Control Framework for Electronic Health Records : A Preliminary Study. https://doi.org/10.1109/ICITISEE.2017.8285487.

Khan, S. I., Sayed, A., & Hoque, M. L. (2016). Digital Health Data: A Comprehensive Review of Privacy and Security Risks and Some Recommendations *. *Computer Science Journal of Moldova*, *24*(271), pp.273–292.

Liu, Y., Zhang, Y., Ling, J., & Liu, Z. (2018). Secure and fine-grained access control on e-healthcare records in mobile cloud computing. *Future Generation Computer Systems*, *78*, pp.1020–1026, https://doi.org/10.1016/j.future.2016.12.027.

Mehraeen, E., Ayatollahi, H., & Ahmadi, M. (2016). Health information security in hospitals: The application of security safeguards. *Acta Informatica Medica*, *24*(1), pp.47–50. https://doi.org/10.5455/aim.2016.24.

Miller School of Medicine. (2009). The difference between the privacy and security of health information. *University of Miami Health System*, *409*(10), pp.1–3. Retrieved from http://www.med.miami.edu/hipaa.

Moura, J., & Serrão, C. (2015). Security and Privacy Issues of Big Data. *Handbook of Research on Trends and Future Directions in Big Data and Web Intelligence*, *2*, pp.20–52, https://doi.org/10.4018/978-1-4666-8505-5.ch002.

Nass, S. J., & Levit, L. A. Gostin, O. L. (2009). *The value, importance, and oversight of health research*. *Beyond the HIPAA Privacy Rule: Enhancing Privacy, Improving Health Through Research*, https://doi.org/Damschrode.

Panahiazar, M., Taslimitehrani, V., Jadhav, A., & Pathak, J. (2014). Empowering Personalized Medicine with Big Data and Semantic Web Technology: Promises, Challenges, and Use Cases. *Proceedings : ... IEEE International Conference on Big Data. IEEE International Conference on Big Data*, *2014*, pp.790–795, https://doi.org/10.1109/BigData.2014.7004307.

Sahi, A., Lai, D., & Li, Y. (2016). Security and privacy preserving approaches in the eHealth clouds with disaster recovery plan. *Computers in Biology and Medicine*, *78*, pp.1–8, https://doi.org/10.1016/j.compbiomed.2016.09.003.

Tucker, K., Branson, J., Dilleen, M., Hollis, S., Loughlin, P., Nixon, M. J., & Williams, Z. (2016). Protecting patient privacy when sharing patient-level data from clinical trials. *BMC Medical Research Methodology*, *16*(Suppl 1), https://doi.org/10.1186/s12874-016-0169-4.

United Nations. (1948). Universal Declaration of Human Rights. *Universal Declaration of Human Rights*, pp.29–30, https://doi.org/10.1017/CBO9781107415324.004.

Van Staa, T. P., Goldacre, B., Buchan, I., & Smeeth, L. (2016). Big health data: The need to earn public trust. *BMJ (Online)*, *354*(July), pp.3636, https://doi.org/10.1136/bmj.i3636.

Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical Challenges of Big Data in Public Health. *PLOS Computational Biology*, *11*(2), e1003904, https://doi.org/10.1371/journal.pcbi.1003904.

Zaman, A. N. K., & Obimbo, C. (2014). Privacy Preserving Data Publishing: A Classification Perspective. *IJACSA) International Journal of Advanced Computer Science and Applications*, *5*(9), Retrieved from http://thesai.org/Downloads/Volume5No9/Paper_19-Privacy_Preserving_Data_Publishing_A_Classification.pdf.