Maintaining Healthcare Data Security in the Cloud

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Jihong Min

Spring 2022

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Sean M. Ferguson, Department of Engineering and Society

Introduction

According to statistics from Purplesec, the healthcare industry has been plagued the most by data breaches in recent years. This has been further elevated recently due to COVID-19, as hospitals admit more people, and the amount of patient data increases. Healthcare is a data-intensive domain where immense amounts of data are created, stored, and accessed every day. For example, in everyday tasks like a patient undergoing some tests like computerized scans, the data is first viewable to the physician, then stored in the hospital database, which may later be accessed by a physician in another hospital within the same network (Hulme, 2021). This gives the data a greater chance to be exposed to cybercriminals.

To cybercriminals, patient data is a very valuable and attractive target. Nadrag (2021) reports that medical records can sell at up to $1000 each, while credit card numbers only sell for around $5 each and social security numbers for $1. There are several reasons for this price discrepancy. Credit card numbers can be easily canceled, whereas medical records contain a lot of immutable data, such as a patient's medical health history, as well as health insurance and contact information. According to Cherian (2022), another reason healthcare data is often targeted is because it is very vulnerable. Many healthcare providers are still stuck in the past, using outdated IT infrastructure such as Windows 7 and Windows Server 2008. These operating systems have been retired by Microsoft for a while now, and don't receive security patches or updates. To make things worse, Nadrag (2021) also reports that more than half of medical devices operate on legacy stems, and 83% of medical imaging devices are on outdated operating systems that also don't receive updates. Without secure systems, medical records can be easily stolen by today's hackers. And once they're are stolen, cybercriminals often connect with members of criminal networks on the dark web experienced in drug trafficking and money

laundering who are eager to buy medical records to support their criminal activities. This includes illegally obtaining prescription medications, filing bogus medical claims or simply stealing the patient's identity to open credit cards and fraudulent loans.

Fortunately, many healthcare providers are turning to utilizing cloud computing for their services. Currently over 83% of healthcare organizations are already using some form of cloud platform ("Definitive Guide," 2022). One reason for the acceleration of cloud services is that only cloud-based solutions offer healthcare providers and patients the access they require while maintaining compliance for the Health Insurance Portability and Accountability Act (HIPAA). Cloud use in the healthcare industry is bound to keep growing, and security challenges are something that must be continued to dealt with. There are many challenges to maintaining security in the cloud, as cyber criminals become more skilled and cloud providers must continue to abide by regulations and standards set by the government. This paper will explore cloud computing security specifically in the health industry using the framework of Social Construction of Technology (SCOT). Mainly, it will examine different security solutions and innovations stemming from the interpretively flexible space of threats in medical data.

Framework

The framework, SCOT, according to Pinch and Bijker (1984), focuses on engineers, technologists, and entrepreneurs as they work with relevant social groups to interactively construct new innovations, disrupt existing sociotechnical arrangements, and otherwise define and solve problems. The definition and solving of problems come from interpretive flexibility. Interpretive flexibility means that each technological artifacts has different meanings and interpretations for various groups, and can lead to different solutions to what they define as the problem.

SCOT could be seen in the discussion surrounding the adoption of cloud computing in many industries. Khan and Al-Yasiri (2016), did a study to identify cloud security threats to strengthen the cloud computing adoption framework. Their study included interviews with cloud developers and security experts to help them understand current and future security challenges with cloud computing. As today's businesses can decide whether to keep their IT infrastructure on-site, or on the cloud, the biggest factor to think about is security, as that is what customers and users want for their personal data, and what the government is striving for with different regulations like the Payment Card Industry Data Security Standard (PCI DSS), HIPAA, and more. However, when considering security, IT professionals in a business need to assess whether or not they can manage data on premises better than the cloud providers. Besides security, they also need to considers factors like cost, uptime, and scalability, which also needs to be considered by upper management. This is an ongoing discussion in many industries as businesses need to consider their customers' demands, government regulations, cloud provider options, and their own IT personnel when considering going forward with cloud adoption. As many healthcare providers have now adopted cloud computing for managing health records, they must find the best way to manage various security threats and keep data safe while addressing the problems brought up from the relevant social groups.

Healthcare Data Background

Healthcare providers generally store Electronic Medical Records (EMRs) that contain medical and clinical data related to a given patient. To support the management of EMRs, Health Information Systems (HIS) were designed with the capability of creating new EMR instances, storing them, and querying stored EMRs. However, with patient mobility becoming the norm in society nowadays, EMRs did not have enough information than the pre-existing paper records to

allow easy transfer (Garret and Seidman, 2011). Electronic Health Records (EHRs) were designed to allow patient medical history to move with the patient or be made available to multiple healthcare providers. EHRs have richer data structure than EMRs ("What is EHR," 2019). Also, with health records access needing to be more flexible and the rise in Personal Health Records (PHR) coming from personal health tracking devices, there had to be a HIS based on an ecosystem of solutions that is able to seamlessly exchange data among themselves and provide the abstraction of a single health data storage for any given patient. This has led to cloud computing rising as an appealing solution for Health Information Systems (HIS). Cloud computing has the ability to support real-time data sharing regardless of geographical locations, to provide resource elasticity as needed, and to handle big data to obtain useful insights from the analysis of big healthcare data for research and policy decision making ("Definitive Guide," 2022). Securing this ecosystem include using cryptographic primitives, such as those based on public key infrastructure and public clouds to ensure data confidentiality and privacy.

Blockchain Technology

Healthcare's progressive shift of data and services to the cloud is a trend being observed, partly due to convenience and savings. Convenience coming from the availability of complete patient medical history in real-time and savings coming from the economics of healthcare data management. However, there are limitations to using conventional cryptographic primitives and access control models to address security and privacy concerns in an increasingly cloud-based infrastructure. With the conventional approach, the searchability of the data is greatly limited. Healthcare providers have to decrypt the data prior to searching on the decrypted data, resulting in increases in time and costs for the data retrieval and diagnosis. There are also access control models that care used to regulate access to data based on pre-defined access policies that are

effective for external attacks, but ultimately ineffective against internal attackers as they are likely to have access to the data (Esposito et al., 2018). Patient data remains vulnerable. Esposito and a group of professors across many universities around the world noticed and addressed this problem by examining the potential of blockchain technology to protect healthcare data hosted within the cloud.

To address this limitation, utilizing blockchain has been a recent interest in the provision of securing healthcare data management in the cloud. Blockchain technology, although mainly used for transactions and cryptocurrencies, is able to build and open and distributed online database, which consists of a list of data structures that are linked with each other. These blocks are distributed among multiple nodes and are not centrally stored. Each block contains information such as production timestamp, the hash of the previous block, and the transaction data ("What is blockchain technology?"). In the context of its use in healthcare, it would have a patient's healthcare data and healthcare provider information. When new healthcare data for a particular patient is created, a new block is instantiated and distributed to all peers in the patient network. After a majority of peers have approved the new block, the system will insert it in the chain. This allows a global view of the patient's medical history in an efficient, verifiable, and permanent way. If a block is not approved, a fork in the chain is created and the block is defined as an orphan and does not belong to the main chain. Once the block is inserted into the chain, the data in any given block cannot be modified without modifying all subsequent blocks. This way, modification is easily detectable (Esposito et al., 2018). Overall, utilizing blockchain provides the capability to achieve decentralized consensus and consistency, and resilience to intentional and unintentional attacks.

Analysis

There are many benefits to utilizing a blockchain approach. First, there is no single point of failure where health records can be hacked. Next, blockchain data is complete, timely, accurate, and easily distributed. Lastly, changes to blockchain are visible to all members of the patient network, and all data insertions are immutable (Esposito et al., 2018). However, there are many challenges and unknowns in taking on this new approach. Blockchain was originally designed to record transaction data in the banking, which is relatively small and linear. Healthcare data, on the other hand, has the potential to contain large amounts of data, related to patient records and treatment documents. It is unclear, according to the study, how well blockchain storage can deal with this amount of data.

Another challenge is dealing with privacy protection laws. There are quite a number of regulations related to the healthcare industry including the Federal Regulations of American Insurance (HIPAA) and various state regulations. HIPAA covered entities must implement appropriate administrative, physical, and technical safeguards to ensure the confidentiality, integrity, and availability of electronically protected health data. HIPAA includes rules designed to allow all the patients to be able to access their own medical records, correct errors, and have the right to get information regarding how their personal information is used or shared ("What is HIPAA,", 2022). Also, Article 17 of the General data Protection Regulation in the EU allows individuals to be able to request personal data to be erased. The problem is that once data is stored in blockchain, it cannot be deleted or altered ("What is blockchain technology?"). Although this comes from the strong data integrity of blockchain, this would violate personal data privacy laws. Healthcare data is very sensitive, so there needs to be a way to use blockchain while also conforming to these laws. Also, as cyberattacks on healthcare organizations are on the rise, it is understandable that the Department of Health and Human Services' Office for Civil

Rights has increased enforcement of HIPAA rules. A potential solution to these challenges that the study mentions is off-chain storage of data, where the data is stored outside of the blockchain in a database, but the hashes are still in the blockchain (Esposito et al., 2018). This would allow healthcare data to be erasable when needed, while the immutable hashes on the blockchain are used for checking accuracy and authenticity of the off-chain data.

Blockchain technology, while there are challenges that need to be addressed, seems like a good potential solution for the limitations of traditional methods of security in the cloud. And if the cloud is to be continued to be used for storing health data, a way to overcome these challenges need to be found quickly.

Automation

As seen from the blockchain solution, compliance to HIPAA and many other government regulations becomes a problem that needs to be dealt with by healthcare providers' IT and cloud providers. To address this problem, Kim and Joshi's (2021) study explored the possibility of automating HIPAA regulations for Cloud Health IT services. This was through developing a semantically rich knowledge graph using semantic web technologies to represent HIPAA rules in a machine-processable format. In summary, the HIPAA knowledge graph incorporated relationships between HIPAA rules and stake holders and refined hierarchies of provisions to make it more universal. Next, the HIPAA Privacy and Security Rule was reviewed to make the ontology more semantically rich with added details. Semantic web technologies were used to develop the HIPAA ontology and reasoning with the policy expression. The web tools enable data to be annotated with machine understandable meta data, allowing the automation of their retrieval and their usage in the correct contexts (Kim & Joshi, 2021).

Analysis

As healthcare providers adopting cloud-based services to manage patient data increases, compliance with the rules and policies of HIPAA regulation becomes increasingly complex. This is due to the multi-tenancy characteristics of cloud architectural designs, which most on-premise systems do not have. Conflicting interests of healthcare providers are a significant obstacle to HIPAA compliance in cloud systems. To resolve this, healthcare and cloud providers must write a Business Associate Contract and specify compliance requirements, service levels, and legal liability (Kim & Joshi, 2021). Although the contract can clarify the responsibilities, the money spent on compliance becomes unreasonable due to the complexity. Currently, HIPAA rules are only available in text format and require significant human effort to implement into Health IT systems. With the relaxation in telehealth policy due to the COVID-19 pandemic, every change needs to be manually implemented in the IT system. Therefore, this policy and rules automation would facilitate the compliance process to help healthcare providers focus on protecting patients' data, and can efficiently access the big data for public health. The problem of HIPAA compliance can be seen as a problem stemming from the demands of the government and needs to be dealt with efficiently and accurately through automation like this study showed, so that more effort can be put into healthcare itself.

Mutual Authentication

As the role of telehealth increased due to the pandemic, telecare medicine information systems (TMISs) have become a good alternative for patients to get remote health services conveniently. TMISs have improved the healthcare process between patients and doctors by providing an efficient communicating platform over the internet, which overcomes some of the drawbacks of traditional health-care services such as the barrier between patients and doctors caused by the inconvenience of patients' time and locations (Kumar et al., 2018).

For healthcare providers that also provide telehealth services, it is also very important to preserve security and privacy of the patient medical data in TMISs. With healthcare providers, the cloud users, or health IT, store medical data in the cloud database to recapture the data securely. As this paper has covered, the cloud is not completely secure, so a protected and authenticated framework is required to prevent simple security attacks. In a network environment, the client, or patient authenticates the server and vice-versa. In this way, network users can be assured that they are doing business exclusive with legitimate entities and servers can be certain that all would-be users are attempting to gain access for legitimate purposes. With mutual authentication, a connection can only occur when the client trusts the server's digital certificate and the server trusts the client's certificate (Kumar et al., 2018). This process reduces the risk that an unsuspecting network user will inadvertently reveal security information to cybercriminals or insecure websites. A study by Kumar, Jangirala, and Ahmad (2018) proposed a new mutual authentication protocol that would help secure TMISs in a cloud computing environment. Mutual authentication in this case, is accomplished between the healthcare center and cloud server, patients and cloud server, doctor and cloud server, and patient and healthcare center. The presented framework would be more capable in terms of computation cost and be resilient in dealing with various security attacks.

Analysis

Telecare is a rising form of healthcare that is definitely here to stay. It has opened numerous possibilities and helped healthcare become accessible to more people who might live in an environment that does not have reachable in-person healthcare. However, new forms of healthcare will naturally open up new ways for attackers to get to the data. As this form is done completely over the cloud, it opens new challenges for security. In non-face-to-face interactions,

mutual trust needs to be achieved between the patients and healthcare professionals. Healthcare professionals need to know if they are talking with the right person, and patients need to make sure they are talking to a trusted professional. Multi-factor authentication for the patient alone is not enough in these situations. With mutual authentication, this trust can be achieved, allowing cloud computing systems to stay secure to safely provide healthcare through TMIS.

Cloud Provider Guidance

Examining these potential solutions to security challenges in the cloud, there are some things that may be more difficult to fix. This is the human aspect of security. Cloud computing is still a rising field in computer science, and still lacks in the number of experts in the field. This is especially worrying as cloud security is not just the responsibility of the cloud providers, but also the users or clients. This means that users retain control of the security they choose to implement to protect their own content, platform, applications, systems, and networks no differently than they would in an on-site data center. In this case, healthcare providers need to have IT personnel that are capable of implementing a secure plan utilizing the available tools offered by cloud providers (Connell, 2020). Also, there are potentially going to be new ways to secure systems coming up to overcome security challenges, as attackers come up with more sophisticated ways to breach systems. This paper alone covered blockchain, automation, and mutual authentication in the cloud.

To aid in this responsibility, cloud providers supply users with an extensive set of tools and detailed documentation that outlines best practices. More specifically, they provide security-specific tools and features across network security, configuration management, access control, and data encryption, making up the technical artifacts. (AWS, 2021). As most users may not be experts, they provide additional guidance through online resources, personnel, and partners.

As cloud providers are doing everything they can to make securing systems as easy as possible, it is up to IT professionals in every industry to make use of these tools. This can involve doing online training to learn of best practices, studying documentation, and obtaining certifications. Personnel shortages may take time to overcome, but early initiatives taken by cloud providers will help this challenge not last too long.

Conclusion:

Cloud computing's role in healthcare will continue to rise and cement its place as the primary place for data. So, dealing with the security challenges is something that needs attention. Security challenges are not something that will go away once some are dealt with, but will keep coming as attackers continue to come up with new ways to get through the walls surrounding patient data. Security becomes even harder as demands from different groups become greater. Patients always want secure systems protecting their personal data, and health IT and cloud providers must continue to develop more tools and strengthen their systems to keep up with the attackers. Also, government laws concerning personal data may continue to become stricter, making it even more difficult for healthcare and cloud providers to develop systems that will comply with them. And this security must also transition to new forms of healthcare like telehealth that provide for people who don't live in ideal places for accessible healthcare. Blockchain technology serves as potential new way to secure patient data in the cloud, automation serves as an efficient way to comply with personal data laws, and mutual authentication serves as a better way to ensure security in a rising form of healthcare in telecare. Lastly, cloud providers continue to develop new tools and ways of learning, so that IT workers in healthcare can better learn how to secure their systems. Cloud computing's stability in maintaining healthcare data will rely on the ability to continuously overcome various security

challenges, meet demands from various groups, and in the end, cover all areas of potential weaknesses in the system.

References

Amazon Web Services. (2021). *Security and Compliance*. AWS Whitepaper. Retrieved from https://docs.aws.amazon.com/whitepapers/latest/aws-overview/security-and-compliance.html.

Cherian, S. (2022, January 17). *Council post: Healthcare Data: The perfect storm*. Forbes. Retrieved March 10, 2022, from https://www.forbes.com/sites/forbestechcouncil/2022/01/14/healthcare-data-the-perfect-storm/

Connell, J. (2020, October 15). *Overcoming the cloud security skills shortage by encoding expertise*. Threat Stack. Retrieved March 10, 2022, from https://www.threatstack.com/blog/overcoming-the-cloud-security-skills-shortage-by-encoding-expertise

Definitive guide to cloud computing in Healthcare: TrueNorth. TrueNorth ITG. (2022, March 2). Retrieved March 10, 2022, from https://www.truenorthitg.com/cloud-computing-in-healthcare/#:~:text=Currently%20over%2083%25%20of%20healthcare,business%20ahead%20of%20industry%20trends

Esposito, C., De Santis, A., Tortora, G., Chang, H., & Choo, K.-K. R. (2018). Blockchain: A panacea for healthcare cloud-based data security and privacy? *IEEE Cloud Computing*, *5*(1), 31–37. https://doi.org/10.1109/mcc.2018.011791712

Garrett, P., & Seidman, J. (2011, August 26). EMR vs ehr – what is the difference? Health IT Buzz. Retrieved March 20, 2022, from https://www.healthit.gov/buzz-blog/electronic-health-and-medical-records/emr-vs-ehr-difference

Hulme, G. V. (2021, November 10). *Healthcare Cloud Security explained*. Healthcare Cloud Security Explained. Retrieved March 10, 2022, from https://businessinsights.bitdefender.com/healthcare-cloud-security-explained

Khan, N., & Al-Yasiri, A. (2016). Identifying cloud security threats to strengthen cloud computing adoption framework. Procedia Computer Science, 94, 485–490. https://doi.org/10.1016/j.procs.2016.08.075

Kim, D.-young, & Joshi, K. P. (2021). A semantically rich knowledge graph to automate HIPAA regulations for cloud health IT SERVICES. *2021 7th IEEE Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)*. https://doi.org/10.1109/bigdatasecurityhpscids52275.2021.00013

Kumar, V., Jangirala, S., & Ahmad, M. (2018). An efficient mutual authentication framework for healthcare system in cloud computing. *Journal of Medical Systems*, *42*(8). https://doi.org/10.1007/s10916-018-0987-5

Nadrag, P. (2021, January 26). *Industry voices-forget credit card numbers. medical records are the hottest items on the dark web*. Fierce Healthcare. Retrieved March 10, 2022, from https://www.fiercehealthcare.com/hospitals/industry-voices-forget-credit-card-numbers-medical-records-are-hottest-items-dark-web#:~:text=Cybersecurity%20firm%20Trustwave%20pegged%20the,as%20little%20as%20%241%20each

Pinch, T. J., & Bijker, W. E. (1984). The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. Social Studies of Science, 14(3), 399–441. http://www.jstor.org/stable/285355

What is an electronic health record (EHR)? HealthIT.gov. (2019, September 10). Retrieved March 10, 2022, from https://www.healthit.gov/faq/what-electronic-health-record-ehr

What is blockchain technology? - IBM Blockchain. IBM. (n.d.). Retrieved March 20, 2022, from https://www.ibm.com/topics/what-is-blockchain

What is HIPAA compliance? Compliancy Group. (2022, January 19). Retrieved March 14, 2022, from https://compliancy-group.com/what-is-hipaa-compliance/