Big data security and privacy issues in healthcare

Nanthealth

Harsh Kupwade Patil and Ravi Seshadri
Nanthealth
Dallas, US
E-mail: hkupwade@nanthealth.com

Abstract—With the ever-increasing cost for healthcare and increased health insurance premiums, there is a need for proactive healthcare and wellness. In addition, the new wave of digitizing medical records has seen a paradigm shift in the healthcare industry. As a result, the healthcare industry is witnessing an increase in sheer volume of data in terms of complexity, diversity and timeliness. As healthcare experts look for every possible way to lower costs while improving care process, delivery and management, big data emerges as a plausible solution with the promise to transform the healthcare industry. This paradigm shift from reactive to proactive healthcare can result in an overall decrease in healthcare costs and eventually lead to economic growth. While the healthcare industry harnesses the power of big data, security and privacy issues are at the focal point as emerging threats and vulnerabilities continue to grow. In this paper, we present the state-of-the-art security and privacy issues in big data as applied to healthcare industry.

Keywords; healthcare; big data security; privacy; security analytics

I. INTRODUCTION

The new wave of digitizing medical records has seen a paradigm shift in the healthcare industry. As a result, healthcare industry is witnessing an increase in sheer volume of data in terms of complexity, diversity and timeliness. The term “big data” refers to the agglomeration of large and complex data sets, which exceeds existing computational, storage and communication capabilities of conventional methods or systems. In healthcare, several factors provide the necessary impetus to harness the power of big data. For example, in the last two decades, healthcare costs have increased at an alarming rate and healthcare expenses are now estimated at 17.6 percent of GDP. As healthcare experts look for every possible way to lower costs while improving care process, delivery and management, big data emerges as a plausible solution with the promise to transform the healthcare industry. The McKinsey Global Institute estimates a $100 billion increase in profits annually, if big data strategies are leveraged to the fullest potential [1]. For instance, harnessing the power of big data analysis and genomic research with real-time access to patient records could allow doctors to make informed decisions on treatments. Furthermore, big data will compel insurers to reassess their predictive models.

With the increasing cost for healthcare services and increased health insurance premiums, there is a need for proactive healthcare management and wellness. This shift from reactive to proactive healthcare can result in improved quality of care, decrease in healthcare costs, and eventually lead to economic growth. In recent times, technological breakthroughs have played a significant role in empowering proactive healthcare. For instance, real-time remote monitoring of vital signs through embedded sensors (attached to patients) allows health care providers to be alerted in case of an anomaly. Furthermore, healthcare digitization with integrated analytics is one of the next big waves in healthcare Information Technology (IT) with Electronic Health Records (EHRs) being a crucial building block for this vision. With the introduction of EHR incentive programs [2], healthcare organizations recognized EHR’s value proposition to facilitate better access to complete, accurate and sharable healthcare data, that eventually lead to improved patient care.

As healthcare industry explores myriad ways of applying big data analysis from diagnosis, to treatment, to population health management, and eventually capital and strategic planning, the opportunities are endless. Furthermore, as healthcare leaders move from a volume-based to a value-based business model (value refers to the association between quality of care and costs), data will play a pivotal role in the transition [3]. As the healthcare industry witnesses large volumes of data, the first step will involve governance and linking accurate and actionable data in real-time. In this age of connectivity, integrating health systems with large amounts of clinical, financial, genomic, social and environmental data will be crucial for real-time analytics and patient care. The goal is to understand population health for disease control and predictive analysis. For instance, predictive analysis can help understand aggravating health conditions and could prevent adverse health events from occurring (e.g. chronic diseases such as diabetes). Hence, collecting, linking and analyzing multi-dimensional data in real-time becomes imperative. A logical next step in a patient-centric model would be a new all-inclusive scale for measuring the health and wellness of a patient by including, but not limiting to clinical, physical, social, psychological, environmental and genomic data pertaining to a patient. Fig. 1 shows a need for a real-time
holistic model for healthcare, with an emphasis on parameters from different domains affecting the condition of a patient. For example, a patient’s vital signs can be normal, but his/her psychological and environmental factors can have dire consequences, (factors not considered as part of the prognosis).

Figure 1. Real-time holistic model for healthcare

The explosion of the Internet of Things (IoT) and its ability to provide real-time monitoring and expedited access to care is one of the driving factors for its adoption in healthcare. Gartner estimates 26 billion IoT devices will be functional by 2020 and the amount of traffic generated by such devices will be large enough to place it in the category of big data [4]. Several definitions for IoT exist but currently the focus is primarily on low-cost, low-powered resource constrained (storage, computation and bandwidth) devices [5]. In addition, with the introduction of Body Sensor Networks (BSN) and their direct application to healthcare [6], care providers will be able to monitor vital parameters, medication effectiveness, and predict an epidemic. Body sensors generate massive data, and linking such healthcare data from disparate resource-constrained networks will be crucial for driving healthcare analytics. Hence, healthcare providers have enormous opportunities to revolutionize healthcare by harnessing the power of big data. Nevertheless, such gains will be realized only if security and patient privacy are at the core of any product design and development.

The past decade has seen a steady increase in security breaches in healthcare IT. In 2013, Kaiser Permanente (one of the largest non-profit healthcare providers in US) notified its 49,000 patients that their health information had been compromised due to theft of an unencrypted USB flash drive containing patient records [7]. In 2012, Verizon’s data breach investigation report stated that its forensic investigation and security division compiled data from 47,000 reported security incidents and found 621 confirmed data breaches [8]. Furthermore, a study on patient privacy and data security showed that 94% of hospitals had at least one security breach in the past two years [9]. In most cases, the attacks were from an insider rather than external. In addition, the study stated that the external attacks originated from China, US and Eastern Europe (Romania recording the highest number of external attacks).

With the ever-changing risk environment and introduction of new emerging threats and vulnerabilities, security violations are expected to grow in the coming years. Moreover, the Affordable Care Act will lead to more enrollments for health insurance [10], making it an attractive focal point for hackers and opening a floodgate of healthcare breaches in the coming years. Security breaches of EHR can risk patient privacy and violate the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health (HITECH) Act in the United States [11], [12]. Hence, EHR security must be a high priority to ensure patient safety.

II. SECURITY AND PRIVACY IN HEALTHCARE

Adoption of big data in healthcare significantly increases security and patient privacy concerns. At the outset, patient information is stored in data centers with varying levels of security. Moreover, most healthcare data centers have HIPAA certification, but that certification does not guarantee patient record safety. The reason being, HIPAA is more focused on ensuring security policies and procedures than on implementing them. Furthermore, the inflow of large data sets from diverse sources places an extra burden on storage, processing and communication. Fig. 2 portrays a big data healthcare cloud that hosts clinical, financial, social, genomic, physical and psychological data pertaining to patients.
Traditional security solutions cannot be directly applied to large and inherently diverse data sets. With the increase in popularity of healthcare cloud solutions, complexity in securing massive distributed Software as a Service (SaaS) solutions increases with varying data sources and formats. Hence, big data governance is necessary prior to exposing data to analytics.

A. Data governance

As the healthcare industry moves towards a value-based business model leveraging healthcare analytics, data governance will be the first step in regulating and managing healthcare data. The goal is to have a common data representation that encompasses industry standards (e.g. LOINC, ICD, SNOMED, CPT, etc.) and local and regional standards. Currently, data generated by BSN is diverse in nature and would require normalization, standardization and governance prior to analysis.

B. Real-time security analytics

Analyzing security risks and predicting threat sources in real-time is of utmost need in the burgeoning healthcare industry. At present, healthcare industry is witnessing a deluge of sophisticated attacks ranging from Distributed Denial of Service (DDoS) to stealthy malware. Furthermore, social engineering attacks are on the rise and the risks associated with such attacks are difficult to predict without considering human cognitive behavior. Cognitive bias, for example, can come into play, especially in the case of elderly patients. “Cognitive bias is a pattern of deviation in judgment, whereby influences about other people and situations may be drawn in an illogical manner” [13]. For example, a man-in-the-middle attack can be effected perhaps by coaxing an elderly patient to accept a digital X.509 certificate. Such scenarios must be taken into account when designing an end-to-end authentication solution.

In the IoT environment, implementing security in resource-constrained networks has been a challenge and will continue to grow more complex with the increase in the number of IoT devices [14]. For instance, conventional symmetric and asymmetric key distribution and revocation schemes cannot be extended to a billion IoT devices. Hence, new scalable key management solutions leading to seamless inter-operability between disparate networks (e.g. IoT and legacy IP networks) is crucial for IoT’s integration of big data in a cloud environment.

As healthcare industry leverages on emerging big data technologies to make better-informed decisions, security analytics will be at the core of any design for the cloud based SaaS solution hosting Protected Health Information (PHI). Additionally, real-time security intelligence will steer new directions in risk management. Consequently, healthcare IT providers can monitor risks in real-time and take preemptive measures before affecting the healthcare business.

C. Privacy-preserving analytics

Invasion of patient privacy is a growing concern in the domain of big data analytics. An incident reported in the Forbes magazine raises an alarm over patient privacy [15]. 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. For instance, data anonymization prior to analytics could protect patient identity. Furthermore, privacy- preserving encryption schemes that allow running prediction algorithms on encrypted data while protecting the identity of a patient is essential for driving healthcare analytics. As the industry leverages on IoT devices to transmit vitals to healthcare clouds, there is a need for processing and analyzing data in an ad-hoc decentralized manner. However, performing resource-exhausting operations (required for analytics) while preserving privacy is a challenge in a resource-constrained environment. Additionally, as healthcare analytics gains popularity, new privacy laws need to be drafted to protect patient privacy. For instance, “informed consent” from patients is required prior to performing any analytics on patient data, and new laws need to be drafted to clearly illustrate all processes involved in performing big data analytics on patient data.

III. CONCLUSION

As big data transforms healthcare, security and patient privacy is paramount in driving such technologies. As healthcare clouds with big data become prominent, hosting companies will be more reluctant to share massive healthcare data for centralized processing. Hence, we envision distributed processing across disparate clouds and leveraging on collective intelligence. Secure patient data management is inevitable as healthcare clouds aggregate and link large amounts of data from disparate networks. Additionally, secure and privacy preserving real-time analytics will propel proactive healthcare and wellness. In
this paper, we review some of the security and privacy issues in healthcare and foresee a need for technological breakthroughs in computational, storage and communication capabilities to meet the growing demand of securing healthcare data.

IV. REFERENCES


