A Survey on Big Data Privacy and Security Issues in Healthcare Information System

K. Pragash and Dr. J. Jayabharathy

Received 14 September 2017; Accepted 15 October 2017; Available online 30 October 2017

Address For Correspondence:
K. Pragash, Assistant Professor, Computer Science and Engineering Department, Sri Ganesh College of Engineering and Technology, Podicherry.
E-mail: Kpragashx43@gmail.com

ABSTRACT

Enormous amount of data is being generated every day from big data. This paper explores the impact of big data Security issues in Healthcare information. As healthcare providers generate their medical records, clinical systems, and medical imaging, protecting the privacy of patient information and securing it becomes increasingly challenging. Widespread use of digital data in healthcare industry has provided potentially immeasurable benefits by instance access to patient information practically from anywhere in the world this put the data vulnerable to security threats and risk the privacy of patients. Healthcare organizations are facing more security threats that increase the risks of inappropriate access to patient information, inability to access the information, and impaired integrity of the information. This paper presents a survey on privacy and security of Medical data and details about the health information systems (HIS), privacy and security of health data due to digitization using new techniques to improve the quality of healthcare.

KEYWORDS: Big Data, Healthcare information system, privacy and security.

INTRODUCTION

Technology has advanced a great deal during the past forty years from the legacy systems based on network and hierarchical models to relational and object database systems. Database systems can also now be accessed by the web and data management services have been implemented as web services. Due to the explosion of web-based services, unstructured data management and social media and mobile computing, the amount of data to be handled has increased rapidly on daily basis from terabytes (TB) to petabytes (PB) and zetabytes (ZB) in just two decades. Such huge amounts of complex data have come to be known as Big Data [1]. ‘Big Data’ refers to novel ways in which organizations, including government and medical healthcare data which combine diverse digital datasets and then use statistics and other data mining techniques to extract from them both hidden information and other related information [2]. ‘Big Data’ is best understood as a more powerful version of knowledge discovery in databases or data mining, which has been defined as ‘the nontrivial extraction of implicit, previously unknown, and potentially useful information from data’ [3].

All the business that deals with healthcare data either stored or transmitted is required to act with these regulations. To prevent the misuse of data and to minimize the risks of security breaches, some controls can be used. Traditional security measures including firewall, antivirus and intrusion detection and prevention systems are no longer providing the required levels of granularity, protection and enforcement required for compliance with security and privacy regulations. American Recovery and Reinvestment Act (ARRA) of 2009 is a new driver of automation in healthcare especially in the adoption of electronic health records (EHR). Requirements are
expected to include new technologies like newer and advanced encryption techniques, data masking and access control policies to add new dimensions to security and privacy in healthcare.

This paper is organized as follows: Section II describes HIS and need for big data analytics in health care is discussed in section III. Section IV will give detail about Big data security and privacy issues and Section V and VI will give hints about existing system and future work and the conclusion is given in Section VII.

**Health Information Systems (HIS):**

The application of information processing involving both computer hardware and software that deals with storage, management, transmission, retrieval and sharing of information related to the health detail of individuals or the activities of organisations that work within the health division is termed as Health Information Systems (HIS). It includes district level routine information systems, disease surveillance systems and also includes laboratory information systems, hospital patient administration systems (PAS) and human resource management information systems (HRMIS). The basic goal of health information systems is to “integrate data collection, processing, reporting and use of information for improving health service effectiveness and efficiency through better management at all levels of health services” [4]. Therefore, its main objective is to produce appropriate and quality information to support decision making process in explicit manner [5].

**A. Six Components of Health Information System:**

In order to assess health information systems and sustainably improve them using a standard framework, Health Metrics Network (HMN), which operates as a network of global, regional and country partners, was established in 2005 under the aegis of World Health Organization (WHO). The HMN Framework divides the components of HIS into three categories: inputs, processes, and outputs. The Framework suggests a three-phase process for implementing a stronger HIS.

The three processes used by HIS are: indicators, data sources, and data management. Six components of HIS mapped to the three categories are: Inputs refer to HIS resources that include legislative, regulatory and planning frameworks to ensure a fully functional health information system and involve personnel, logistics support information and communication technology (ICT). Processes refer to the selection of indicators and data sources for collection and management of data. Indicators reflect changes over time in health information and include determinants of health, health system inputs, outputs and outcomes and health status. Determinants of health are socioeconomic, environmental, behavioural, demographic and genetic determinants. Input indicators include policy, organisation, health resources, financial resources, health infrastructure, equipment and supplies. Health service availability, information availability and quality are the output indicators.

Data sources can be divided into population based data sources and institution based data sources. Population based health information data resources include census, civil registration and population surveys. Institution based data resources include resource records, service records and individual records. Data management handles collection and storage of data (patient medical records), ensuring quality and flow, processing, compilation and analysis of data. Outputs deal with production, dissemination and use of information. Information products deal with transformation of data to produce information which can be integrated with other information and becomes evidence that is used by decision makers. Dissemination and use include use of information for decision making.

The data storage component of HIS mentioned in above section is important in ensuring proper storage and accessibility of patient medical records. Ensuring the proper storage not only leads to accuracy, timeliness, completeness and reliability but also helps in the analysis of disease trends, assessment of quality care and ultimately the equitable distribution of resources. With the introduction of HIS and digitisation of health data, an increasing number of hospitals and peripheral health facilities are moving towards the transformation of paper-based data into digitized data using data-to-information cycle (recording, reporting, aggregating, storing, analysis and using)

**Need for Big Data Analytics in Healthcare:**

To improve the quality of healthcare by considering the following:

1) **Providing patient centric services:**

If there is no pre stored data available means patient has to undergone all the tests from the beginning to know the type of the disease, where as if the data is available doctor can easily identify the diseases with the type of symptoms and this will help the patient to get best medicine with less cost and in effective manner. To provide faster relief to the patients by providing evidence based medicine-detecting diseases at the earlier stages based on the clinical data available, minimizing drug doses to avoid side effect and providing efficient medicine based on genetic makeup. This helps in reducing readmission rates thereby reducing cost for the patients.
2) Detecting spreading diseases earlier:
   In past decades we came to explore many diseases like dengue and Ebola etc within certain perimeter range this could be predicted earlier. Predicting the viral diseases earlier before spreading based on the live analysis. This can be identified by analysing the social logs of the patients suffering from a disease in a particular geo location; this helps the healthcare professionals to advise the victims by taking necessary preventive measures.

3) Monitoring the hospital’s quality:
   Many Healthcare centres have emerged within short period of time, in that many does not have proper internal facility and adequate equipments, and this should be regularly monitored by whether the hospitals are setup according to the norms setup by Indian medical council. This periodical check-up helps government in taking necessary measures against disqualifying hospitals.

4) Improving the treatment methods:
   Customized patient treatment monitoring the effect of medication continuously and based on the analysis dosages of medications can be changed for faster relief. Monitoring patient vital signs to provide proactive care to patients. Making an analysis on the data generated by the patients who already suffered from the same symptoms, helps doctor to provide effective medicines to new patients.

Big Data: Security and Privacy issues:
   Big Data is being tremendously used with its vast amount of datasets through different data sources. It is also giving rise to the security and privacy concerns in different domains particularly in healthcare data. As Big Data includes data and information that will be used for different purposes, the security and privacy of an individual is at risk. Here in this section we briefly discuss about the security and privacy issues in various domains of Big Data which are as follow:

5) Healthcare:
   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 healthcare, several factors provide the necessary force to harness the power of big data. 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. 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 any problem or difficult situation[6].

6) Security and Privacy issues in Healthcare:
   As healthcare industry is growing, so are the security and privacy concerns with it. Following information need to be consider when we talk about security and privacy issues related to it:
   a. Big Data is a collection of large and complex datasets and getting adopted in the healthcare significantly, security and privacy issues in healthcare becomes necessary to deal with. Most healthcare data canters are HIPPA certified, though this certification does not guarantee patient’s record safety as HIPPA is more focused on ensuring security policies and procedures than implementing them. [HIPPA (Health Insurance Portability and Accountability Act): It is the federal Health Insurance Portability and Accountability Act of 1996. The primary goal of the law is to make it easier for people to keep health insurance, protect the confidentiality and security of health information and help the healthcare industry control administrative costs.]  
   b. A study on patient privacy and data security showed that 94% of hospitals had at least one security breach in the past two years. In most cases, the attacks were from an insider rather than external [7].
   In HIS, security should be the top priority from day one. Patients’ data should be protected with comprehensive physical security, data encryption, user authentication, and application security as well as the latest standard setting security practices and certifications, and secure point-to-point data replication for data backup. These security issues have been extensively investigated for cloud computing in general [8, 9, and 10]. A major challenge to healthcare cloud is the security threats including tampering or leakage of sensitive patient’s data on the cloud, loss of privacy of patient’s information, and the unauthorized use of this information. Hence, a number of security requirements should be satisfied by healthcare cloud computing systems. The main security and privacy requirements for healthcare clouds are discussed below:

   • Authentication:
   In a healthcare, both healthcare information offered by CSPs and identities of users (HPs, practitioners, and patients) should be verified at the entry of every access using user names and passwords assigned to users by CSPs (Content security policy).
• **Authorization:**
  
  Is an essential security requirement that is used to control access priorities, permissions and resource ownerships of the users. Each user is granted privileges based on his account. Patient can allow or deny sharing their information with other healthcare practitioners. To implement patient consent in a healthcare system, patient may grant rights to users on the basis of a role or attributes held by the respective user.

• **Non-repudiation:**
  
  Implies that one party of a transaction cannot deny having received a transaction nor can the other party deny having sent a transaction. In a healthcare system, technologies such as digital signatures, timestamps, confirmation receipt, and encryption can be used to establish authenticity and non-repudiation for patients and practitioners.

• **Integrity and Confidentiality:**
  
  Integrity means preserving the accuracy and consistency of data. In the healthcare system, it refers to the fact that EHRs (Electronic Health Record) have not been tampered by unauthorized use. Confidentiality is defined by the International Organization for Standardization (ISO) in ISO-17799 as "ensuring that information is accessible only to those authorized to have access". Confidentiality and integrity can be achieved by access control and encryption techniques in EHR systems.

• **Availability:**
  
  For any EHR system to serve its purpose, the information must be available when it is needed. High availability systems aim to remain available at all times, preventing service disruptions due to power outages, hardware failures, and system upgrades. Ensuring availability also involves preventing denial-of-service (DoS) attacks.

**Existing Solutions:**

The issues of data access, storage, and analysis are not unique to the medical arena. These problems have been looked in a number of areas, from financial services to internet shopping, and technical solutions exist which can be applied to health care to increase privacy and security in a multi-user setting:

**A) Access control on Role-based:**

one of the most challenging problems in managing large networks is the complexity of security administration [11]. Role based access control, or role based security, is the dominant model for advanced access control. It results in the reduction of the complexity and cost of security administration in large networked applications. An example of role based access control for health care is in [12].

**B) Encryption:**

Encryption can be used to ensure the security of the data and help prevent eavesdropping and skimming. Encryption can be accomplished in hardware as well as in software. In order to ensure the highest level of security, it is best if both forms of encryption are used. Different symmetric and asymmetric key algorithms can be used to provide encryption in software [13]. In sensor networks, Tiny Sec [14] is specifically designed to provide encryption and authentication capabilities. Tiny Sec is already employed by some medical sensor systems such as the Kansas State University/University of Alabama in Huntsville WBAN.

**C) Validation Mechanisms:**

Authentication mechanisms can be used to ensure the data is coming from the person/entity it is claiming to be from [13]. There has been a number of authentication algorithms developed such as passwords, digital signatures, and challenge response authentication protocol. There are methods designed for sensor networks that are more energy efficient, such as the hash function in Tiny Sec, that can be used for authentication

**Area to be improved to increase security:**

While there are methods that can be employed to aid in security and privacy of medical data with these new technologies, there are still areas that can be improved upon.

**A) Define clear attributes for function based access:**

Clear rules for the role-based access need to be defined so that these systems can be put in place. These can be dynamic rules or static rules depending on what is appropriate.
B) Policy development:

New policy needs to be created that can deal with across state jurisdiction. While the HIPAA (Health Insurance Portability and Accountability Act) 1996 Privacy Rules provide some groundwork, more needs to be done to create clear rules that users can rely upon. The move toward EPRs and the increasing amount of medical data that will be gathered due to remote sensor networks creates the ability to transfer large amounts of data quickly. This necessitates a comprehensive set of regulations that protect a user’s privacy and security independent of which state the data is located in. In the current setting many patients are not sure about their privacy rights regarding medical data and are ill-informed. As more medical data becomes electronic and can be easily transmitted, this will magnify the confusion of users unless clear guidelines are defined.

C) Rules on patient’s privacy at home:

Can the patient have full control over how much of the data is sent to the central monitoring station, or does the patient only have partial control? Guidelines need to be drawn which will regulate what sensor data collection entails and who will have control over it.

D) Data mining rules and technological measures:

These include not only who has the right to analyze what type of data, but also include the rules on the collected data. The appropriate technical methods for ensuring these rules then need to be put in place if some form of automation is possible.

Conclusion:

Technology is enabling medical health records to be put in the electronic format, and making them available to the users via the Internet. But this technology and the connected devices expose healthcare data to increased security and privacy risks. In this paper we have detail discussed the healthcare system, big data and privacy and security issues that arise when integrating this new technology of big data into the traditional health care system. We explored some of the existing solutions that can be employed and the open research questions that need to be answered before the widespread use of the new technology is possible with minimal security and privacy risks.

REFERENCES