Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 70 (2015) 448 – 454

Procedia Computer Science

4th International Conference on Eco-friendly Computing and Communication Systems, ICECCS 2015

Preserving Privacy of Patients based on Re-identification Risk

Himanshu Taneja*, Kapil†, Ashutosh Kumar Singh‡

Department of Computer Applications, National Institute of Technology Kurukshetra-136118, India

Abstract

Electronic Medical Records (EMR) holds the medical data of the patients in an electronic form. It appears as a massive Big Data that needs to be stored in elastic clouds which are generally handled by a third party. The cloud service provider acts as a third party here and have the access to all the EMRs. Also, most of the times, the patient information needs to be shared with other research analysts and medical professionals for research or expert opinion. This raises serious concerns regarding the privacy of the patient's data.

In this paper, an approach based on reducing the re-identification risk is proposed to preserve the privacy of the EMRs. The proposed solution is based on *k*-Anonymity, *l*-Diversity, *t*-Closeness & δ-Presence and is implemented through ARX Anonymization tool. We have implemented the solution on randomly generated medical dataset based on extension of publicly available EMRs. The result shows that the re-identification risks are reduced to 2.33% from 100%.

Peer-review under responsibility of the Organizing Committee of ICECCS 2015

Keywords: Big Data; Cloud Computing; Electronic Medical Record; Electronic Health Record; HealthCare; *k*-Anonymity; *l*-Diversity; *t*-Closeness; δ-Presence

1. Introduction

The growth of Information Technology has resulted in the advancement of Healthcare services across the globe. The complexity of diagnosis and treatment of the disease has been reduced. The use of technology helps in disease prevention, its spreading and also has been able to provide a feasible solution to human impairments. The healthcare industry consists of patients' treatment, medication, laboratory testing, diagnostic tests, and other services offered by the medical institutes. The medical industry is listed in one of the top 14th industry of the globe and is able to

* himanshu.taneja24@gmail.com
† kapil@nitkkr.ac.in
‡ ashutosh@nitkkr.ac.in

Peer-review under responsibility of the Organizing Committee of ICECCS 2015
doi:10.1016/j.procs.2015.10.073

contribute a significant factor in improving the economy of a nation. The quality of healthcare services can be improved when the diagnosis of the problem, its solution and monitoring of the treatment is done in an effective and efficient manner[4]. It largely depends upon how the information is gathered, stored, managed and accessed in the information systems[15].

The healthcare institutes are continuously working to improve their services being offered and finding different mechanisms to automate the manual processes in an optimum manner. The requirement of providing better services in a rapid and cost-effective manner fits well with the cloud computing domain. There are many medical institutes which have already deployed cloud computing for their organizations and medical facilities are being provided based out of cloud[21]. The invention of cloud computing technologies and big data helped in sharing of the resources across different organizations in an easy and cost effective manner where a lot of development time is saved. The cloud services are generally provided by a third party which takes care of everything related to cloud and the resources stored in it[24].

As the data resides at some other place, belonging to the third party, this situation raises serious concerns such as privacy and security despite providing many benefits as discussed above[16]. The cloud acts as a library of shared resources for holding the EMR and different parties can access this shared common data. The healthcare institute is not aware about where their data is stored, how it is stored and the other tenants who can access the data. They are left with no option other than trusting the cloud service provider[6]. Medical information is taken as the most confidential information as it directly contains the personal data of the patients. It has become uttermost concern to preserve the confidentiality of the patient's data despite the fact that this data needs to be shared with other medical bodies, in case it is required. Though, the platforms providing the cloud based data are increasing, it has also resulted in the privacy and security concerns of the medical data stored in the cloud. As a result, many organizations are reluctant to deploy cloud computing solutions for their organization[9, 18]. It is important in the medical domain to take opinions of other practitioners for providing better diagnosis and treatment facilities. The medical data needs to be shared with other advanced labs for analyzing the results and prescribed medication. Considering the fact that this information is extremely personal to the patient, it has become important to decide which part of the information needs to be shared with others so that the privacy of the patients remain intact. Therefore, sharing of the medical data is considered as a critical activity in the healthcare organizations because it also needs to focus on improving the quality of services being provided by the institute. It is vital for the improvement of healthcare services[11]. The patient's EMR can be modified, corrupted, or misused if shared with different entities[22]. Different researchers and industry experts are working to find an effective solution for this problem. Due to the diversification of the resources used and lack of standards followed in the Healthcare industry, this problem has been further complicated as one solution does not fit well for all the issues which appear in a diversified environment.

We have proposed an approach based on the combination of *k*-Anonymity, *l*-Diversity, *t*-Closeness & δ-Presence with the help of ARX Anonymization tool for solving the privacy concern that handles the problem in an effective manner. Different parameters of the anonymity techniques are used to measure the re-identification risks of the data. This will help in identifying the best values for the parameters so that the risk can be minimized and information utility can be maximized. One cannot deny the sharing of records, but the access to the information being shared can be controlled with the help of different technologies. With its help, even if an intruder becomes successful in gaining the unauthorized access to the medical records, he would not be able to decipher any information out of this data. So, the data will not be of any use to them. As a result, the privacy of patient's data can be preserved. This paper is organized as follows: Section 2 describes the state-of-art work done based on different anonymization techniques. Section 3 demonstrates the proposed work, background details and implementation. The experimental results are shown in section 4. Finally, section 5 concludes the result.

2. Related Work

The research and ideas implemented in this paper are motivated by the state-of-art work done by different researchers based on the following areas:

2.1 Anonymization Approaches

Anonymization refers to hiding the specific details of an individual. Many studies are based on the anonymization techniques to preserve the privacy while accessing sensitive records. The concern of sharing