International Journal of Emerging Trends & Technology in Computer Science (IJETTCS)
Web Site: www.ijettcs.org Email: editor@ijettcs.org
Volume 6, Issue 2, March - April 2017 **ISSN 2278-6856**

ANALYZING AND PERFORMING PRIVACY PRESERVING DATA MINING ON ELECTRONIC HEALTH RECORDS

Mrs. G Jayalakshmi (Ph.D)[1], V.N.V. Srikanth[2], V. Ramya Devi[3], T. Kalyan Babu[4], V. Rajitha[5]

[1]Assistant Professor, Department of IT,
VR Siddhartha Engineering College, Vijayawada, India

[2]Department of IT,
VR Siddhartha Engineering College, Vijayawada, India

[3,4,5]Department of IT
VR Siddhartha Engineering College, Vijayawada, India

Abstract

Medical records that store all the services offered by a hospital must accurately contain all of the patient's medical history. These medical records can be stored easily with the help of a Hospital Management System which stores patient/doctor details and regains these details automatically and easily. In this system, the data will be entered in electronic format by the staff in the hospital and registered patients of the hospital. While projecting these details, protecting the privacy of patients is of utmost importance. These kind of patients' medical records should not be released directly to the public as it may potentially reveal sensitive information of individual patients. This paper aims to explain the use of Privacy Preserving Data Mining (PPDM) techniques like Anonymization, Suppression and Data Hiding on different fields in electronic health records (EHR), for the data to be more secure while projecting it to public. This paper shows how new primitives in attribute-based security can be used to build a secure and privacy-preserving EHR system that facilitates patients to share their data among various healthcare providers in a flexible, dynamic and scalable mode.

Keywords: privacy, preserving, hospital, medical data, data mining, health records

1. INTRODUCTION

Last few years have seen exceptional development in applicability of Computer Science in everyday exercises. Associations, communities and individuals show an expanded pattern of storing their information electronically. The large sets of data gathered can be utilized for analyzing patterns in market, individual or society. Data mining activities help to extract knowledge from this huge amount of data. In case of hospitals and medical records of patients, the sensitive information of the patients/doctors may be disclosed to general public which creates many ethical and privacy issues. Due to this reason, many patients do not like to share their information publicly and hence only provide partial data to the management of the hospital. This creates problems when there is an emergency situation and the management needs complete details of the patient for contacting them. PPDM has gained its importance due to this very reason viz., to address the privacy concerns of the patients.

Privacy plays a vital role in Electronic Health Record systems. EHRs store patient data, their medical history, prescriptions and other useful data about the patients. EHRs are very helpful during healthcare research and other investigation purposes. The data from EHRs should be projected in such a way that the complete identity of the patient should not be revealed. Hence, in order to protect their privacy, PPDM techniques have to be applied on the patients' details first and then it should be published to external sources.

1.1 Electronic health record

An Electronic Health Record is an electronic variant of a patient's medicinal history that is kept up by the supplier over time, and may incorporate the greater part of the key regulatory clinical information important to that persons care under a specific supplier, including that persons details, prescriptions, past medical history, bills, laboratory data, blood reports and other user specific data [1]. The EHR computerizes access to data and can possibly streamline the clinician's work process. The EHR likewise can bolster other care-related exercises directly or indirectly through different interfaces, including proof based choice support, quality administration, and outcomes.

EHR is next generation of health care systems that strengthen the bond between patients and doctors. EHR must be robust, highly secure, contain real time patient based information. An EHR system should have features like:

- Health information of each patient, i.e., all the details of their medical history.
- Possibility for the betterment of healthcare sector.
- Secure real time access to globally available data of any individual by authorized users only.

- Provide accurate and useful information at any point of time about any individual.

1.2 Hospital Management System

Hospital Management System (HMS) is a complete, unified information system that is designed to manage major operations of a hospital such as legal, financial, medical, administrative and managerial services [7].

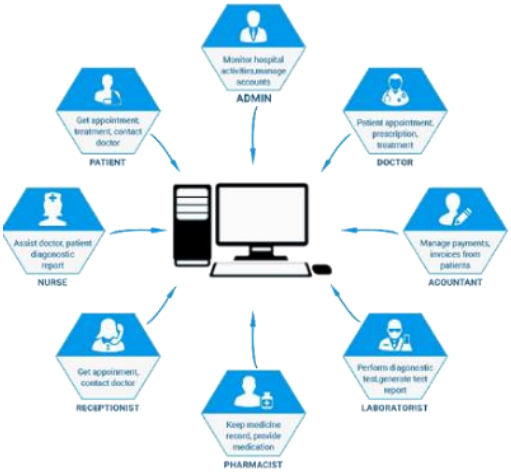

Figure 1 Components of a Hospital Management System

In manual way of managing the operations of a hospital, it is very difficult to retrieve and to find particular information, like to find out about the patient's history etc. For this the user has to go through various registers. This results in inconvenience and wastage of time. Lot of storage space is also necessary to store all these registers. Changes to any part of information is a hectic task and takes a lot of time and effort. Hospital Management System solves all these problems as it stores all the data electronically. This paper explains about the Hospital Management System developed and maintained for "Samatha Multi-Specialty Hospital based in Mangalagiri, Andhra Pradesh" which previously used to store all the details manually in records and registers. It is a dependable, affordable and efficient system that plays a vital role in the success of a medical centre.

However, medical records are highly personal documents and there are many privacy issues regarding this kind of data such as the extent of third party access to this data and accuracy of the data. Thus, privacy need to be applied such that no information is lost while doing so. There are many methods like anonymization, randomization, attribute removal, aggregation on numerical values, suppression etc. that are applied on these data sets to provide privacy.

1.3. Privacy Preserving Data Mining

Data mining is the process of extracting useful information by analyzing huge amounts of data and finding hidden patterns in the data. But before data mining tasks are carried out, several methods must be applied to protect privacy of individuals. Privacy Preserving Data Mining (PPDM) is the technique of mining appropriate results without the loss of individual's privacy [8]. PPDM techniques like anonymization, suppression, data hiding are used in this work to better protect the privacy of the individuals involved in this HMS. These PPDM techniques can be applied while the data is being extracted or after the mining results are obtained.

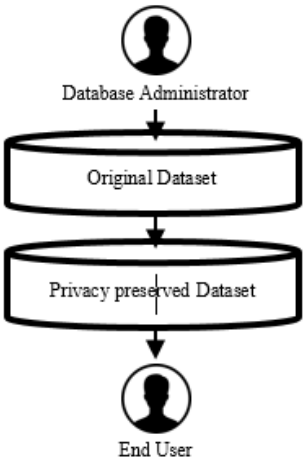

Figure 2 Privacy preserving data publishing

2. RELATED WORK

A detailed literature survey shows that these PPDM techniques are mainly of two types and these techniques are applied at two different levels. These are described as follows:

- Preserving the data at the time of mining process.
- Preserving the private data after the results are obtained and before the results are displayed.

Several algorithms have been used for preserving the privacy of the individual data over the past few years and each of these algorithms have both advantages and disadvantages. These algorithms have been described through the following table:

Table 1: Literature Survey

Each of these systems have one or more disadvantages which are described through the remarks section.

Year	Authors	Paper	Algorithm	Remarks
2013	A. A. AlShwaier and A. Z. Emam	Data Privacy On E-Health Care System	Condensation	Requires to ensure confidentiality constraint
2014	Xu, Yang, Tinghuai Ma, Meili Tang, and Wei Tian	A survey of privacy preserving data publishing using generalization and suppression	Generalization and suppression	Doesn't support partial display of data
2014	A. Hood, C. M. Fung	Privacy-Preserving Medical Reports publishing for Cluster Analysis	Machine Learning by cluster analysis	Training set should be very large for machine learning
2015	Tomar, Divya, and Sonali Agarwal	Hybrid Feature Selection Based Weighted Least Squares Twin Support Vector Machine Approach	Feature selection based least squares	Data is insecure as it is stored in cloud

3. PROPOSED SYSTEM

Information is to be extracted from the medical reports like Electronic Health Records (EHR) and before publishing this data, it has to be anonymized according to the needs of the user. If original data is directly published without applying PPDM techniques may cause threat to individual privacy.

The proposed software system is a Hospital Management System (HMS) with PPDM techniques applied to the data present in the system. Such a system can be used in any Hospital, Medical Institute, Clinic or any dispensary to get the information about the patients and doctors and this information can be stored for future usage. Patient's details like name, email, address, phone, gender, date of birth, age and blood group are entered electronically by the staff in the hospital such as a receptionist. Whenever a new patient comes up, his information is freshly stored. Each patient will be provided with a unique user name and password with which he can login to the system and view all his medication history, appointments, prescriptions, bills and reports. These details are however, not visible to unauthorized users and even doctors can only see anonymized data about the patients. PPDM techniques such as Anonymization, Suppression and Data hiding are used in this process to preserve sensitive details of the patient with the help of "ruby" programming language. After applying these techniques, the output will be completely privacy preserved data of the patients and doctors in the hospital.

3.1 Anonymization

Anonymization is one of the PPDM techniques in which the structure of actual data is maintained while the privacy is not lost [9]. In anonymization, the private part of the data is replaced by a specific value such as a "*" or "-" or any other symbol or character. The remaining part of the data will be as it was in actual data. In this HMS, this technique is applied on the phone number attribute and it looks as follows:

Table 2: Anonymization

Phone Number	Anonymized data
9618036096	96180*****
9124721412	91247*****
7238036095	72380*****

Here we have set the confidentiality constraint to be 50% and hence 50% of the total characters i.e., 5 characters have been replaced by the "*" symbol as per user's requirement.

3.2 Suppression

As the name suggests, in this technique the original data is suppressed to limited form. The end user cannot access the original data and can only access the suppressed data. For example, in case of an attribute like address, the entire is not necessary for all the end users. Only the city, which is the last word in the address is essential. Hence, the original addressed is suppressed to show only the name of the city. The suppressed data in this HMS after the application of suppression technique is as follows:

Table 3: Suppression

Address	Suppressed Data
S/O VENKATESWARA RAO, DOLASNAGE	DOLASNAGE
D/O NATA RAJU, NARASARAOPETA.	NARSARAOPETA
S/O.SUBANI, MANGLAGIRI	MANGALAGIRI

3.3 Data Hiding

In the technique of data hiding, entire attribute fields are hidden from the end user which are considered private. In this HMS, the name and department of the doctor are

sufficient for the patient. Hence all other details of doctor are hidden.

Table 4: Data Hiding

Name	City	Depart-ment	Visible Data	
Bindu	VJA	Pediatrix	Bindu	Pediatrix
Srikanth	VJA	Dental	Srikanth	Dental
Prasad	VJA	ENT	Prasad	ENT

4. RESULTS

This HMS stores patient's records, doctor's records and tests the proposed approach with this data. The dataset contains the name of the patient, email, address, phone number, gender, date of birth, age and blood group. The above techniques have been applied on this data for getting secure data ensuring patient and doctor privacy.

Figure 3: Login Screen of HMS

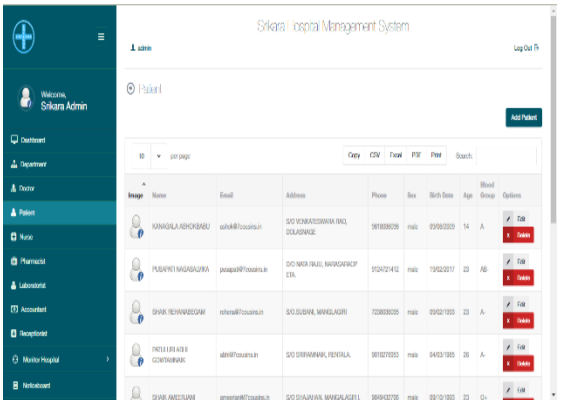

Figure 4: Details of all patients, with no anonymization and suppression, from Admin Panel

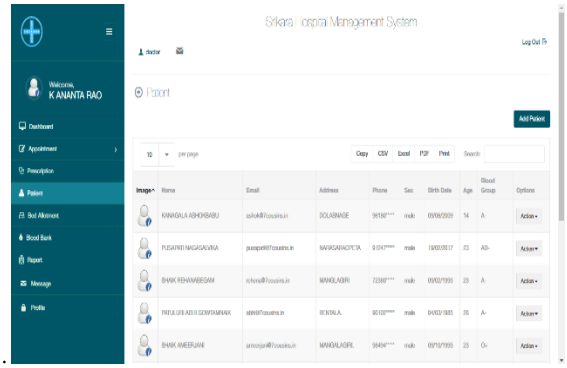

Figure 5: Details of patients who applied for appointment, with anonymization and suppression, from Doctor Panel

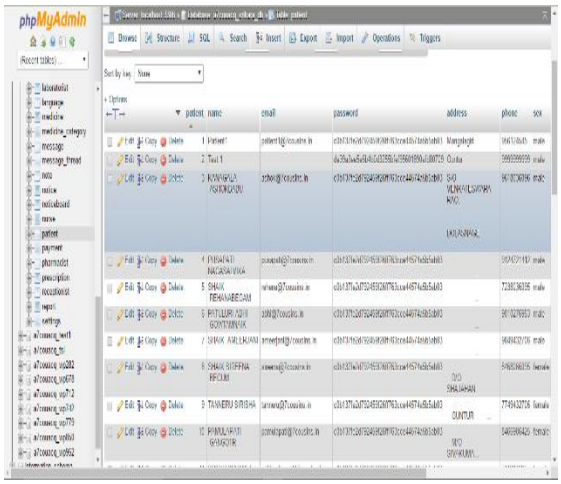

Figure 6: Details of all patients in backend database

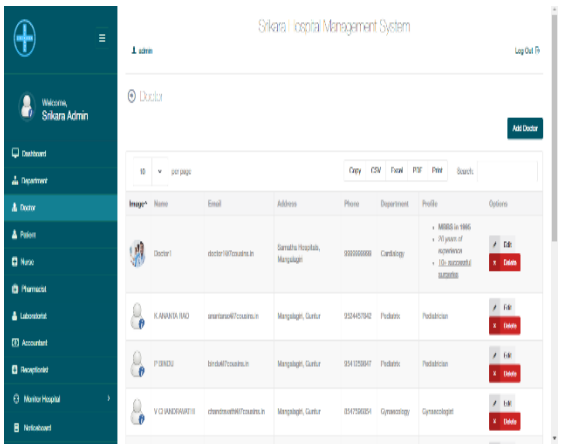

Figure 7: Details of all doctors, without data hiding, from Admin Panel

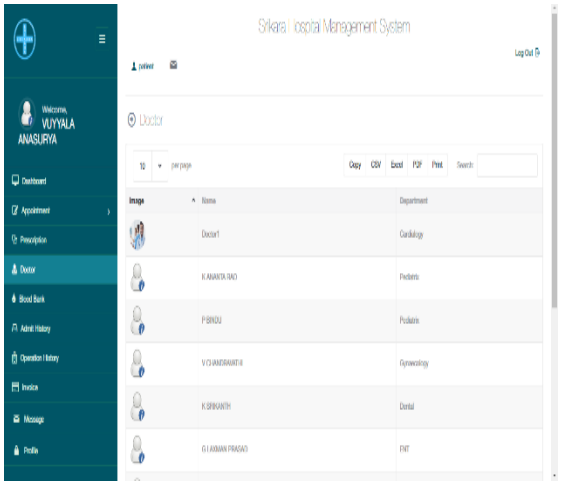

Figure 8: Details of all doctors, with data hiding, from Patient Panel

5. CONCLUSION

A HMS has been built for a multi-specialty hospital. It has the problem of exposing patients' private data. This problem was overcome by applying 3 PPDM techniques on the data preserved in this system to protect the privacy of the patients of the hospital. This system can also be used in dispensaries, clinics, medical institutions and many

other places. It can be extended to other hospitals with vast data whose privacy needs to be preserved using privacy preserving data mining techniques.

References

[1] Anvita Srivastava and Gaurav Srivastava, "Privacy Preserving Data Mining in Electronic Health Record using K anonymity and Decision Tree", International Journal of Computer Science & Engineering Technology (IJCSET), (2003).

A. Dr. D. Aruna Kumari and B. Ch.Mounika, "PRIVACY PRESERVING DATA MINING IN HEALTH CARE APPLICATIONS", International Journal of Advanced Computer Technology (IJACT), (2007).

[2] P. Shyja Rose, J. Visumathi and H. Haripriya, "Research Paper on Privacy Preservation by Data Anonymization in Public Cloud for Hospital Management on Big data", International Journal of Advanced Computer Technology (IJACT), (2016).

[3] Brian Martin, "Suppressing research data: methods, context, accountability, and responses", Accountability in Research, Vol. 6, 1999, pp. 333-372.

[4] Somy.M.S, Gayatri.K.S, Ashwini.B, "Privacy Preserving Health Data Mining", IJCST Vol. 6, Issue 4, Oct - Dec 2015.

[5] Alpa Shah, Ravi Gulati "Privacy Preserving Data Mining: Techniques, Classification and Implications" Accountability in Research International Journal of Computer Applications (0975 – 8887) (2016).

[6] Premkumar Balaraman, Kalpana Kosalram "E - Hospital Management & Hospital Information Systems" I.J. Information Engineering and Electronic Business, 2013, 1, 50-58 (2013).

[7] Agrawal, Rakesh, and Ramakrishnan Srikant. "Privacy-preserving data mining." In ACM Sigmod Record, vol. 29, no. 2, pp. 439-450. ACM, (2000).

[8] Aruna Kumari, Y. Vineela, T. Mohan Krishna and B. Sai Kumar, "Analyzing and Performing Privacy Preserving Data Mining on Medical Databases", Indian Journal of Science and Technology, Vol 9(17), (2016)

[9] N. Zhang, "Privacy-Preserving Data Mining", Texas A&M University, pp.19-25, 2006.