ISSN (Print) : 2320 – 9798
ISSN (Online): 2320 – 9801

International Journal of Innovative Research in Computer and Communication Engineering
Vol. 1, Issue 4, June 2013

Confidential Databases Updation Using Anonymization

S.Thirunavukkarasu[1] Dr.K.P.Kaliyamurthie[2]

Assistant Professor, Dept of IT, Bharath University, Chennai, TN, India[1]

Professor &Head, Dept of IT, Bharath University, Chennai, TN, India[2]

ABSTRACT:Various types of data, including demographics, clinical, and genomic information, are increasingly collected and stored in Electronic Medical Record (EMR) systems. These information are essential for the discovery of new drugs and therapies. But the sharing of this medical data leads to the disclosure of patient's privacy. The existing policy governing the medical records privacy is HIPAA. This act was passed by the US congress and includes policies about security, use and transfer of records. The Institutional Ethics committee of a medical centre is responsible for protecting the patient information during research, which gets a signed consent before the disclosure of patient medical information and Audit trial systems are used in present scenarios for data protection. The disadvantage is that the identity can be disclosed by violating the laws. Through this project, a framework is proposed which uses the anonymization technique for protecting patient privacy. Two protocols of anonymization namely suppression based and generalization based anonymization are used. Anonymization refers to converting the information into a non human readable form. The anonymized personal information and the non- anonymized treatment details are forwarded for clinical research purpose. The proposed system thus focuses on providing an innovative tool to hospital sector for maintaining privacy and confidentiality of patient details.

I.INTRODUCTION

The rapid advances in the computerization of medical data, the question of the protection of medical records privacy has begun to arise. Storing a large amount of sensitive information in databases could open the door to "invasion of privacy".

PRIVACY is a fundamental human right. A society in which there is a total lack of privacy would be intolerable; but then again a society in which there was a total privacy would be no society at all and a balance is needed between the two. Privacy is the right of people to make personal decisions regarding their own intimate matters, it is the right of people to lead their lives in a manner that is reasonably secluded from public scrutiny, and it is the right of people to be free from things such as unwarranted drug testing or electronic surveillance. Privacy issues relating to personal data arise from insecure electronic transmissions of data, data trails and logs of email messages, online transactions and the tracking of web pages visited. Privacy invasion issues arise from data matching (the process of cross checking of data from one source against another source such as tax and social security data) and personal profile extraction processes which use this data alone or in combination with other publicly available data. For example, data that include patient identification may be transmitted electronically to personnel of a health insurance payer to facilitate claim adjudication for reimbursement. Health insurers may also be required to use these data to comply with regulating organizations such as the National Committee for Quality Assurance (NCQA) Health Plan Employer Data and Information Set (HEDIS) measurements.

CONFIDENTIALITY is the process of protecting an individual's privacy. It pertains to treatment information that an individual has disclosed in a relationship of trust, with the expectation that this information will not be divulged to others without permission. The principle of confidentiality means keeping information given by or about an individual in the course of professional relationship secure and secret from others. It applies to all forms of transmissions oral, written, digital, manual or hardcopy records.

ANONYMIZATION is one of the best solutions to maintain privacy and confidentiality. Anonymization is the act of permanently and completely removing personal identifiers from data, such as converting personally identifiable information into aggregated data. Anonymized data is data that can no longer be associated with an individual in any manner. When EDC (Electronic Data Collection) systems were used, and any PHI was to leave the institution's custody that is to another hospital then anonymization is required which transforms information stored in the medical records into an unknown format. Anonymization allows the medical data to be shared in a way that preserves privacy and data utility. Medical data may include demographics that is date of birth, gender, race etc and other clinical text. If privacy is breached then there are severe consequences to patients such as emotional and economical embarrassment and difficulty to conduct statistically powered studies.

II. SYSTEM METHODOLOGY

It is today well understood that databases represent an important asset for many applications and thus their security is crucial. Data confidentiality is also essential. Consider the medical application, data collected by the history of patients over several years may represent a valuable asset that needs to be adequately protected. Such a requirement has motivated a large variety of approaches aiming at providing better data confidentiality. Data confidentiality is not the only requirement that needs to be addressed. Today there is an increased concern for privacy. The availability of huge numbers of databases recording a large variety of information about individuals makes it possible to discover information about specific individuals by simply correlating all the available databases. Although confidentiality and privacy are often used as synonyms, they are different concepts: data confidentiality is about the difficulty (or impossibility) by an unauthorized user to learn anything about data stored in the database. In manual system, the patient information is stored in case files and it is difficult to retrieve the data and also the privacy of the patient is only to some extent. So, the computerized system comes into picture. Here the medical records are electronically stored. And to maintain the privacy of the patient, some approaches are used.

2.1 DISADVANTAGES IN EXISTING SYSTEM

Many approaches are used to protect the privacy of personal information in existing system. One approach is to remove the personal data from the database by a third party. But this approach has a disadvantage that the third party should be trustworthy. The next approach is to protect the database by using passwords. But the demerit is that the password can be hacked easily. Once if the password is hacked, the entire privacy of the database is lost. We can also protect the privacy of the database by using audit trial systems, which is used to track who accesses and modifies the records in the database. It provides privacy and confidentiality to some extent only. The main drawback of the existing system is that the password can be hacked, redundancy of data may occur and this may lead to the inconsistency. The manual system is so time consuming. The online user cannot find correct medication because it can be updated by any person, so it is not reliable.

III.SYSTEM MODEL

Data confidentiality and privacy of user details is achieved to a great extent using anonymization techniques. Data confidentiality is about the difficulty (or impossibility) by an unauthorized user to learn anything about data stored in the database. Usually, confidentiality is achieved by enforcing an access policy, and possibly by using cryptographic tools. Personal information in the database is anonymized, so that privacy is maintained. The anonymization techniques are generalization based anonymization and suppression based anonymization. Generalization based anonymization replaces original values by more general ones. Suppression based anonymization is achieved by encrypting the data. Encryption is achieved by using RSA algorithm in which public and private key are generated. Both anonymization techniques cannot be applied for same data. By encrypting the data, the level of privacy is improved. High privacy and confidentiality is maintained in proposed system. It is efficient, reliable, avoids data redundancy and inconsistency. It is very user-friendly. Proposed system provides more security and integrity to data. Doctors will be benefited as before the doctor visits the hospital they will be able to know how many patients have put appointments. So that the doctor can manage his time before visiting the hospital.

3.1 SYSTEM ARCHITECTURE

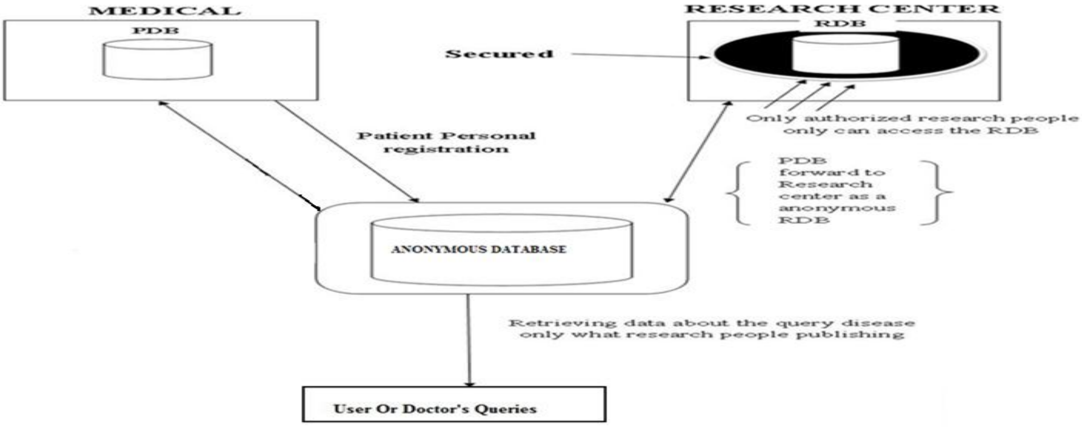
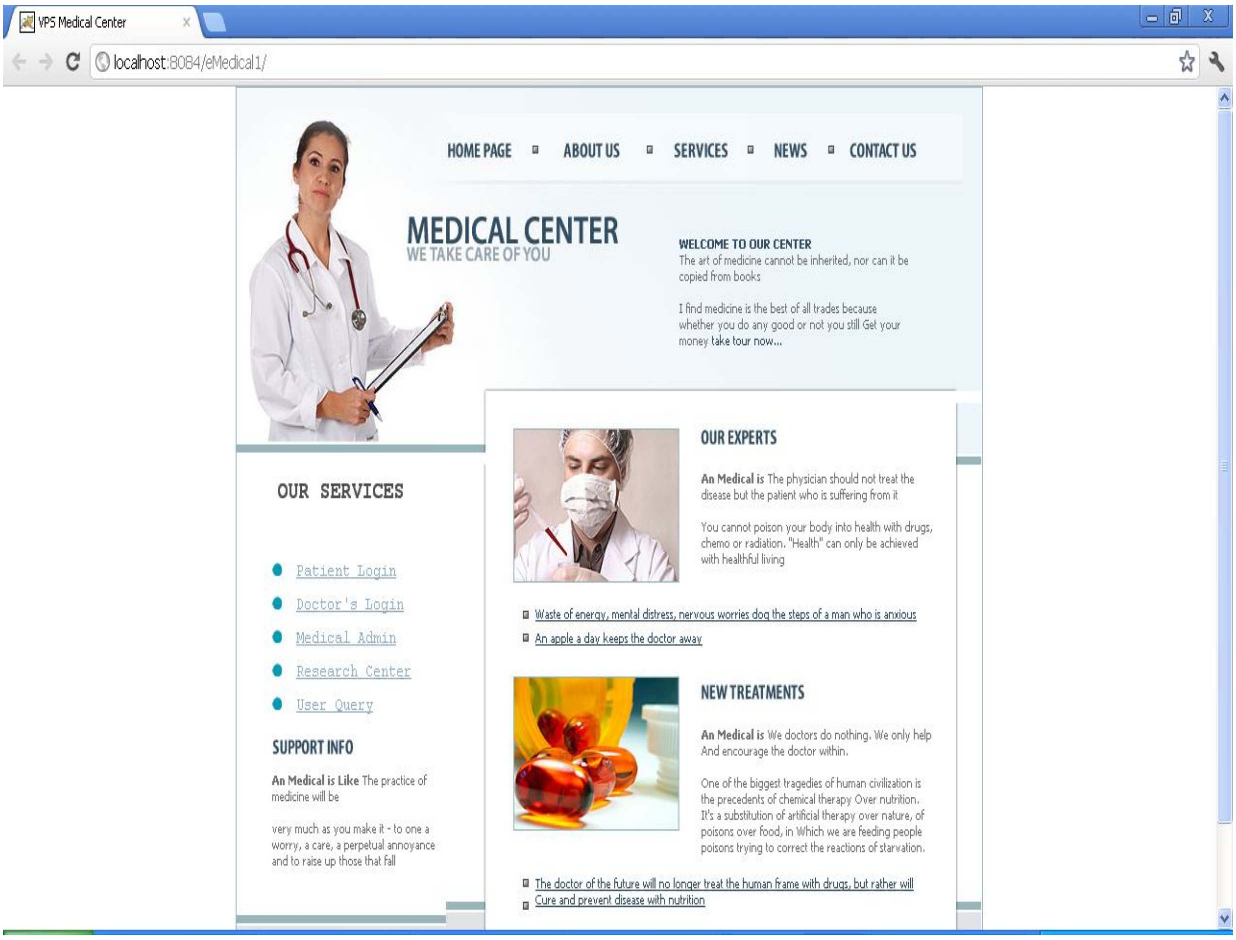

MEDICAL
PDB
Secured
RESEARCH CENTER
RDB
Only authorized research people only can access the RDB
Patient Personal registration
PDB forward to Research center as a anonymous RDB
ANONYMOUS DATABASE
Retrieving data about the query disease only what research people publishing
User Or Doctor's Queries

Figure 3.1 System Architecture

Medical database comprises of three tables namely patient table, doctor table and treatment table. Patient table stores the personal information provided by the patient while registration. This information is stored in an anonymized format. The doctor table contains the doctor's information, appointment details and researcher recommendation about any disease. The treatment and status details of the patient are stored in a table. Anonymization is done using two protocols namely suppression based anonymization and generalization based anonymization. The anonymized information does not reveal the patient's identity. Thus the privacy of the patient is preserved. Only the anonymized information is viewed by researchers. The treatment and status details of the patient are not anonymized and that can be viewed by researcher. Researcher updates their research results to doctor and user according to the level of requirements. The medication details about diseases are updated for doctor's reference. Any user can view the researcher's updates without need for registration.

IV. IMPLEMENTATION AND RESULTS

This chapter provides the screenshots of the admin module, patient module and doctor module which are designed using java server pages. describes the methodologies and processes adopted in the admin module, patient module, doctor module and researcher module along with their respective results.



WELCOME TO OUR CENTER
The art of medicine cannot be inherited, nor can it be copied from books

I find medicine is the best of all trades because whether you do any good or not you still Get your money take tour now...

An Medical is The physician should not treat the disease but the patient who is suffering from it

You cannot poison your body into health with drugs, chemo or radiation. "Health" can only be achieved with healthful living

- Waste of energy, mental distress, nervous worries dog the steps of a man who is anxious
- An apple a day keeps the doctor away

An Medical is We doctors do nothing. We only help And encourage the doctor within.

One of the biggest tragedies of human civilization is the precedents of chemical therapy Over nutrition. It's a substitution of artificial therapy over nature, of poisons over food, in Which we are feeding people poisons trying to correct the reactions of starvation.

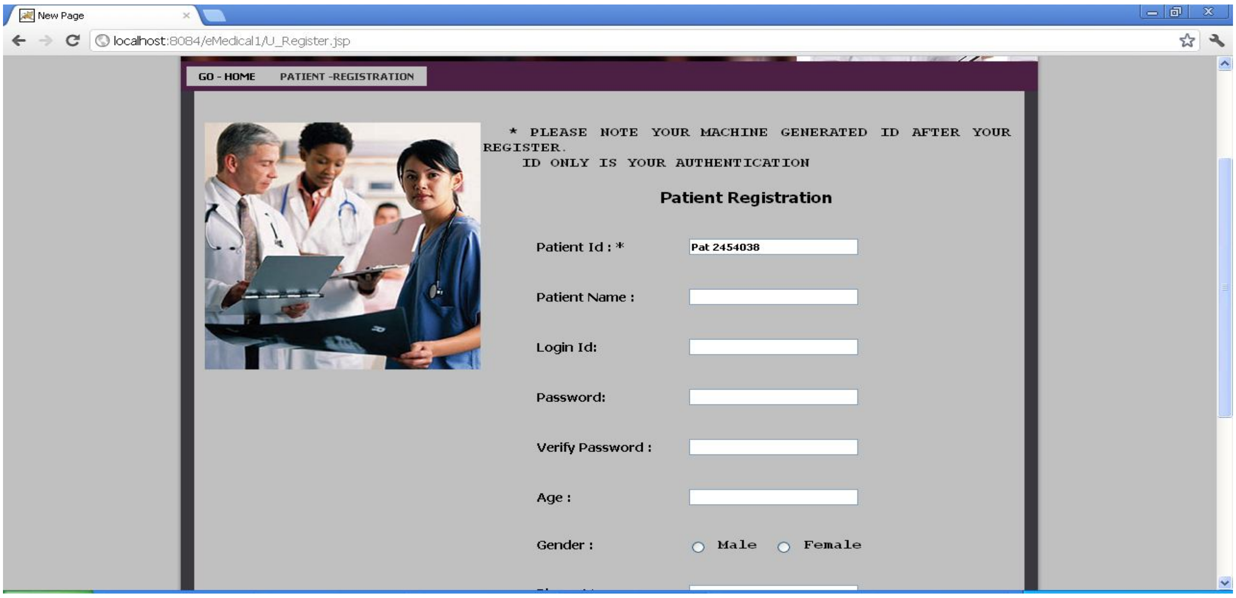
- The doctor of the future will no longer treat the human frame with drugs, but rather will Cure and prevent disease with nutrition

An Medical is Like The practice of medicine will be

very much as you make it - to one a worry, a care, a perpetual annoyance and to raise up those that fall

Figure 4.1 Home page

Figure 4.1 shows the home page which has hyperlinks to patient, doctor and admin login. User query link provides the user with research results.



* PLEASE NOTE YOUR MACHINE GENERATED ID AFTER YOUR REGISTER.
ID ONLY IS YOUR AUTHENTICATION

Patient Registration

Patient Id : * Pat 2454038

Patient Name :

Login Id:

Password:

Verify Password :

Age :

Gender : ○ Male ○ Female

Figure 4.2 Patient Registration

4.1PERFORMANCE EVALUATION

The proposed system is improved from the existing system in terms of

- Memory requirements
- Level of privacy
- Efficiency

In the proposed system, memory requirements has reduced, as a single copy of the information is only maintained. A single copy of the information is only accessed by both the hospital and research centre. The research centre views the anonymized information from which the identity of the patients cannot be revealed to the researchers. Anonymization allows medical data to be shared in a way that preserves privacy. In the contrast, the existing system has two copies of the same information. One for the hospital which is the original copy provided by the patients and the other for the research centre which has the same information with the personal details stripped off. Thus in the existing system two copies of the same information is maintained thereby causing data redundancy which is eliminated in the proposed system. In the proposed system, the information viewed by the researchers are anonymized that is in an unknown format. This increases the level of privacy of the patients and makes them feel respected. The memory requirements and level of privacy is improved in the proposed system compared to existing system. Thus the efficiency of the proposed system is also improved considerably.

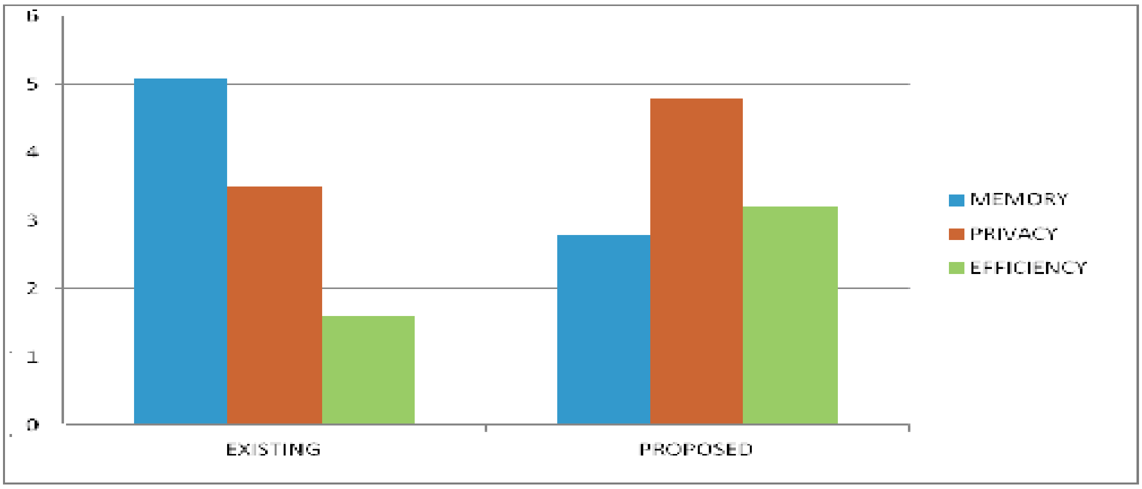


Figure 4.3 Performance Evaluation

V. CONCLUSION

We have used two secure protocols namely generalization based and suppression based anonymization for preserving privacy of the patient. Thus the privacy and confidentiality of the patient's personal information is improved. There is no redundancy of data in the proposed system. Memory requirements are reduced compared to the existing system.

REFERENCES

1. A. Meyerson, R. Williams, "On the Complexity of Optimal K-anonymity," In Proc. of ACM Symposium on Principles of Database Systems (PODS), Paris, France, 2004.
2. Bin Zhou, Yi Han, Jian Pei, Bin Jiang, Yufei Tao, Tan Jia, "Continuous Privacy preserving publishing of data streams ",2009.
3. Dawn Xiaodong Song, David Wagner and Adrian Perrig, "Practical Techniques for searches on encrypted data", 2000.
4. D. Boneh, G. di Crescenzo, R. Ostrowsky, and G. Persiano, "Public Key Encryption with Keyword Search," Proc. Eurocrypt Conf., 2004.
5. E. Bertino, R. Sandhu. "Database security - Concepts, approaches and challenges". IEEE Transactions on Dependable and Secure Computing, 2(1), 2005.
6. G.Agrawal,et al, "Anonymizing Tables", International Conference on Database Theory,Edin burg, Scotland, 2005.
7. H. Hacigumus, B. Iyer, C. Li, S. Mehrotra." Executing SQL over encrypted data in the Database service-provider model". In Proc. of ACM SIGMOD Int'l Conf. on Management of Data, Madison, Wisconsin, USA, 2002.
8. Alberto Trombetta, Wei Jiang, Elisa Bertino and Lorenzo Bossi, "Privacy-Preserving Updates to Anonymous and Confidential Databases", 2011.
9. A. Trombetta, E. Bertino. Private updates to anonymous databases. In Proc. Int'l Conf. on Data Engineering, Atlanta, Georgia, US, 2006.