International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 14 (2015) pp 34969-34973

An Efficient Method to Encrypt Information in Large Database

V.S.Thiyagarajan
Research Scholar,
Department of Computer Science and Engineering,
Annamalai University, Annamalai Nagar – 608 002,
Tamilnadu
thiyagu.cse86@gmail.com

Dr. K.Venkatachalapathy,
Professor,
Department of Computer Science and Engineering
Annamalai University, Annamalai Nagar – 608 002,
Tamilnadu

Abstract- With the development of information technology and medical technology, developed countries have been establish organization to set standard for electronic medical records in response to new generation and information on the application, they gradually develop emerging medical information exchange mode, Personal Health Records (PHR). Here, a data dividing and integration approach for parallel privacy preserving clustering process will be developed. Initially, the input data are divided into subsets and then, subsets of data is given to parallel process to do the clustering process individually through probabilistic clustering process. Privacy protection is a crucial problem in
many medical signal processing applications. For this reason, particular attention has been given to the use of secure multiparty computation techniques for processing medical signals, whereby non-trusted parties are able to manipulate the signals although they are encrypted.

Keywords: Personal Health Records, parallel privacy preserving, clustering process, multilevel, data dividing.

1. INTRODUCTION

Clustering objects from various clusters. The literary works represents with a huge number of strategies for effective grouping of information. These strategies could be classified into nearest neighbour clustering, fuzzy clustering [1, 3], partition clustering, hierarchical clustering, artificial neural networks for clustering, statistical clustering algorithms, density-based clustering algorithm etc. In these techniques, hierarchical and partition clustering algorithms are two essential methodologies of expanding favour towards exploration groups. Various hierarchical clustering strategies can normally realize reasonable clustering outcomes. In spite of the fact that the hierarchical clustering procedure is regularly depicted as a superior quality clustering approach, this method does not contain any procurement for the rearrangement of data, which may have been crudely grouped at the initial stage. Moreover, the majority of the hierarchical clustering strategies is computationally accelerated and entails high memory storage.

PHR can integrate different kind of personal health records. With the Internet or portable device, PHR offers the integrity and accuracy personal health and medical records. Through electronic medical records, we can evaluate the quality of medical care, provide continued care to patients, promote the medical efficiency and increase the accuracy of medical diagnosis .Personal health record (PHR) is an emerging patient-centric model of health information exchange, which is often outsourced to be stored at a third party, such as cloud providers. [5, 19] However, there have been wide privacy concerns as personal health information could be exposed to those third party servers and to unauthorized parties. To assure the patients' control over access to their own PHRs, it is a promising method to encrypt the PHRs before outsourcing. Yet, issues such as risks of privacy exposure, scalability in key management, flexible access, and efficient user revocation, have remained the most important challenges toward achieving fine-grained, cryptographically enforced data access control.

An open source product called Java Simplified Encryption (JASYPT) allows you to replace clear text passwords in files with encrypted strings that are decrypted at run time. The following shows how this can be done. JASYPT can be integrated into the spring configuration framework so that property values are decrypted as the configuration file is loaded.

In this paper, we develop an algorithm for privacy preserving using Simple Encryption Algorithm called JASYPT and Attribute clustering algorithm. The main contribution of the work is to attain the privacy preserving and better clustering accuracy. Initially, the whole dataset is divided to small segments. The next step is to find the best sets of attributes combinations, which are attained through, attribute weighing process, which leads to attain the privacy preservation through horizontal grouping of attributes. The next is to apply the proposed attribute clustering algorithm for each segment, which produces the number of clusters for each segment. The next step is applying the Simple Encryption Algorithm on the clusters to attain the final clustered result.

2. RELATED WORK

Because of Advances in information and communications technology, electrical health records become a trend around worldwide, however the traditional medical records of EMR mainly provide information for the professional nurses in clinical medical use, not the health care and manage on patients' view[5,6]. Due to the higher percentage of self-consciousness and participation on patients, the concept of PHR continually defined

Chia-Hui Liu provided, a program about key management on Bilinear pairing, which perfectly switch in patient-centered Personal Health Records (PHR) in Cloud computing environment, and establish partial order to manage every user, in order to ensure every patient can manage and share their own medical record,[8,15] we design access control based on patients, also provide the solution of Multi-user access and lower the complexity of key management.

Ruoming Jin et al, have proposed a strategy called Fast and Exact K-means Clustering (FEKM). Individual or a least number of passes on the whole dataset was desired by the composed strategy and provably delivered the similar cluster focuses as reported by the first k-means algorithm. Now, the cluster focuses were balanced by taking one or more passes the whole datasets formerly, the planned strategy made primary cluster focuses by sampling. Likewise, a hypothetical investigation was given by them to demonstrate that the cluster focuses were equivalent as the one figured by the first k-means algorithm. The analytic results of genuine and manufactured datasets demonstrated that the planned strategy was performed better contrasted with K-means.

Fiza Abdul Rahim [17] has reviewed, identified and categorized related factors that influence information privacy concerns in EHR. The SLR technique had assisted in narrowing down the targeted information privacy articles in healthcare environment. The findings from this study may provide managerial guidance for healthcare organizations in several ways. The examined articles had emphasized the role of stakeholders in EHR, namely, [2] healthcare practitioners and patients. The process of information dissemination and being more computer literate would go further in developing effective information privacy policies. This in-progress study will proceed in evaluating the highlighted factors towards the design of information protection concerns framework for EHR.

This paper is mostly related to works in cryptographically enforced access control for outsourced data and attribute based encryption. To realize fine-grained access control, the traditional public key encryption (PKE)-based schemes [16, 17], either incur high key management overhead, or require encrypting multiple copies of a file using different users' keys. To improve upon the scalability of the above solutions, one-to-many encryption methods such as ABE can be used.

Tiancheng Li et al [4] revealed a unique procedure called slicing, which segments the data both on a level plane and vertically. They demonstrated that slicing conserves preferable information utility over simplification and might be employed for membership exposure assurance. An alternate vital preference of slicing is that it can deal with high-dimensional information. They demonstrated how slicing could be utilized for attribute revelation assurance and create an effective strategy for figuring the sliced information that comply with the l-diversity necessity. Their analyses affirm that slicing conserves preferred utility over simplification and is more powerful than bucketization in task assignments including the delicate quality. Their investigations likewise presented that slicing could be employed to avert membership exposure.

S. Patel et al [12, 14], have suggested a secrecy conserving distributed K-Means clustering of horizontally partitioned information that aids security in malevolent ill-disposed model. The essential development includes consumption of secret transferring system battered to code based zero knowledge identification plan. They employed secret sharing for secretly offering the data and code based distinguishing proof plan to provide help against malignant rivals.

3. METHODOLOGY

We have implemented a new technique of attribute-based clustering and securely retrieving a patient multi-level privacy-preserving cooperative authentication scheme realizing three levels of security and privacy requirement in distributed m-healthcare system. We are using the hybrid encryption algorithm so the level of privacy is more secure compare with existing system.

The proposed system is divided into four Modules:

A. Admin Data Injection.

B. Hybrid Encryption.

C. Attribute cluster.

D. Secure Decryption.

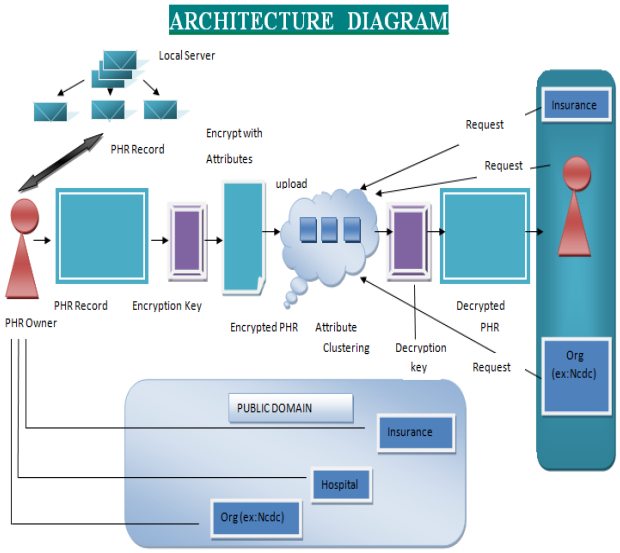

Fig 1.System Architecture.

A.ADMIN DATA INJECTION

In this phase we have to implement the data injection by the admin. Each and every data about the test result and lab reports are added by the admin, because responsibility of records are very important to ensure that the health data captured by a system or provided to any entity is true representation of the intended information and has not been modified in any way with paper-based medical records and prescriptions have also advanced to the Personal Health Records and the Electronic Health Records. The PHRs and EHRs, both are the electronic versions of patient health information. However, the PHRs are controlled by patients themselves; whereas, The EHRs are managed by the healthcare providers.

B.HYBRID ENCRYPTION

We are providing a hybrid Encryption to the system is shown in the figure 1. The cryptographic approaches commonly used in the e-Health cloud-based systems to protect data use encryption, such as Public Key Encryption

and Symmetric Key Encryption .However, there are some other key. The PKE technique requires two separate keys; one of the keys is private whereas the other is public. Solutions based on the PKE are secure but using the PKE alone seems computationally less efficient due to the slower operations and the larger key sizes. Therefore, the PKE is used in combination with the SKE where symmetric keys are used to encrypt the contents while public/private keys are used to secure the symmetric keys. Consequently, in this section we term the approaches that use the PKE in conjunction with symmetric cryptographic technique as hybrid approaches. The common public key algorithms use the (JASYPT) techniques for generating public/private parameters used for security services.

C.ATTRIBUTE CLUSTER

This module presents an attribute clustering method which is able to group attribute based on their interdependence so as to mine meaningful patterns from the health data. It can be used for grouping, selection, and classification. The partitioning of a relational table into attribute subgroups allows a small number of attributes within or across the groups to be selected for analysis. By clustering attributes, the search dimension of a data mining algorithm is reduced. The reduction of search dimension is especially important to data mining in expression data because such data typically consist of a huge number of attributes and a small number of expression profiles. Most data mining algorithms are typically developed and optimized to scale to the number of attributes. The situation becomes even worse when the number of attributes overwhelms the number of tuples, in which case, the likelihood of reporting patterns that are actually irrelevant due to chances becomes rather high.

D.SECURE DECRYPTION

The decryption operation requires a secret key that is derived from a master private key. The decryption keys are distributed by the patients to grants access over certain parts of the medical record. Moreover the approach provides an efficient mechanism for searchability of the encrypted data; it also assumes the presence of multiple trusted authorities in the PHR system. The trusted authorities ensure the enforcement of the sticky policies besides authorizing the users to get the decryption keys for read and write operations. The PKE is considered as less efficient in terms of computation whereas the ABE has a standing of costly decryption primitive because of bilinear computations. The data is encrypted using a content key and only the users having valid license are allowed to decrypt and use the content.

4. EXPERIMENTAL RESULTS

The final result of the work is characterized through the JSP Servlet page with a login page(i.e)Figure(2-7).The login page is accessed only by authorized user (phr owner) with a authorized entry of username and password. Then, the individual patient's detail injected through Admin. It contains Demographics, Emergency Contacts, Health Insurance Details, Laboratory Results and Health History.

Fig 2. Login Page

Fig 3. Demographics

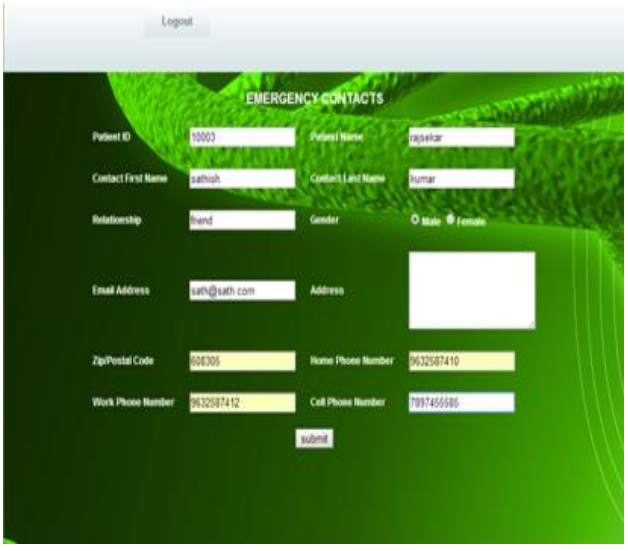

Fig 4.Emergency Contacts

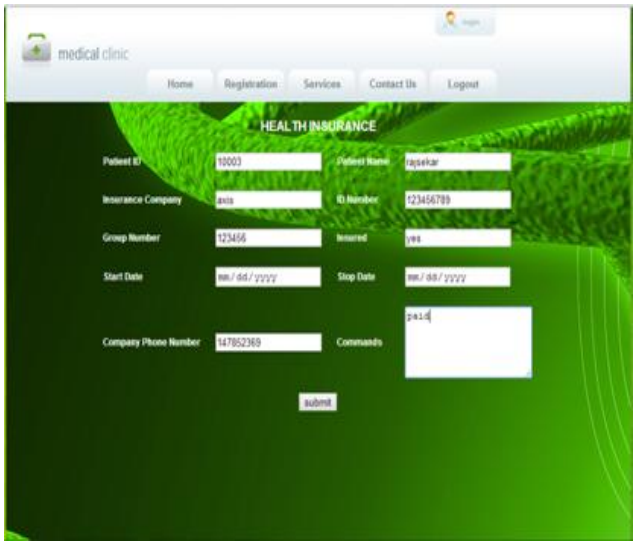

Fig 5. Insurance Details

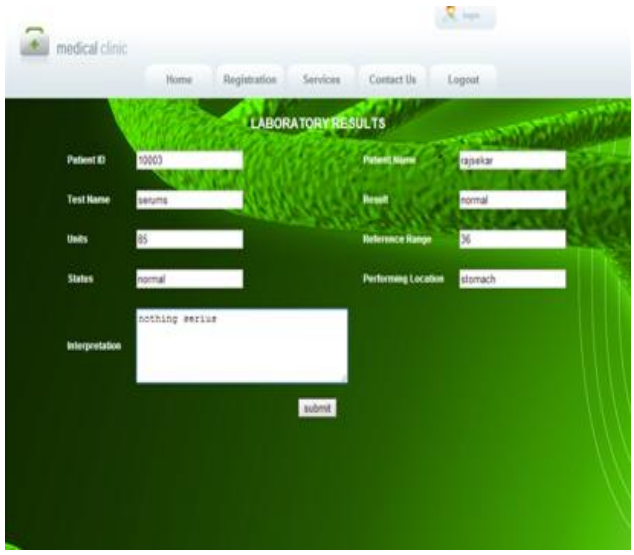

Fig 6. Lab Results

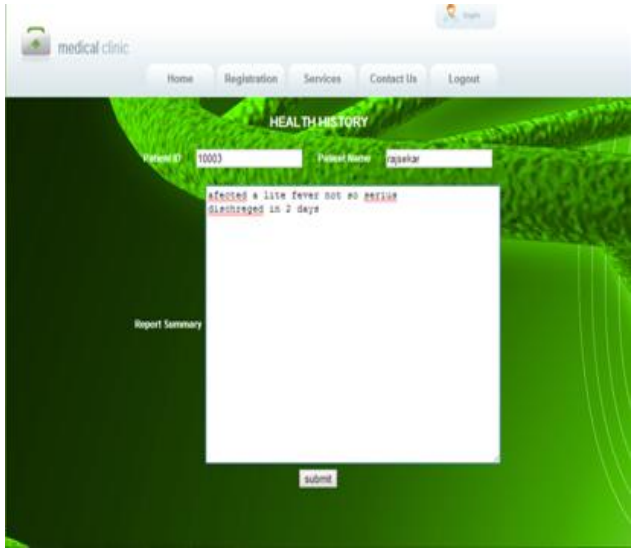

Fig 7. Health History of Patients

Next, it is to inject Physician personal and official information with valid proof numbers with valid E-mail ID to generate Secret key through mail. It is to decrypt and get the original information after encrypting.

Ex :(i.e Secretkey=UKJ6264476).

5. PERFORMANCE ANALYSIS

It represents the accuracy of proposed Attribute clustering and jasypt encryption algorithm with DES and AES for Encryption in clustering. By analyzing the below figure (8, 9), when the number of clusters increased, the accuracy of encryption is effective gradually from the three clustering technique as used for evaluation process. The performance clearly shows that the proposed Attribute clustering and jasypt encryption algorithm outperformed than the existing DES and AES for Encryption in clustering technique in terms of accuracy.

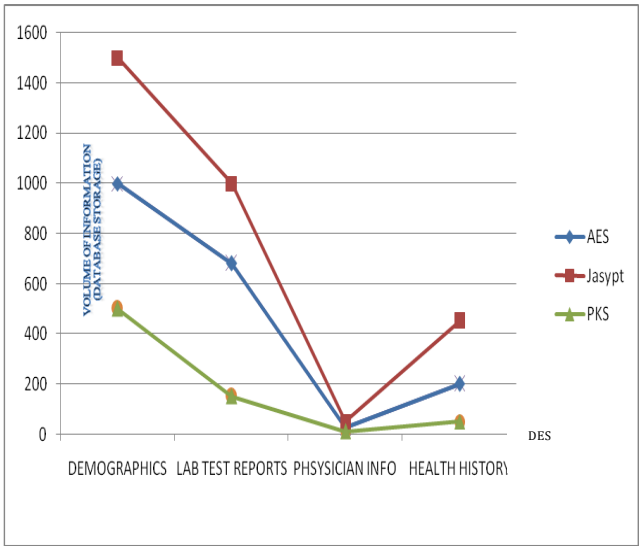

Fig 8. Database Storage of AES & DES Vs Jasypt

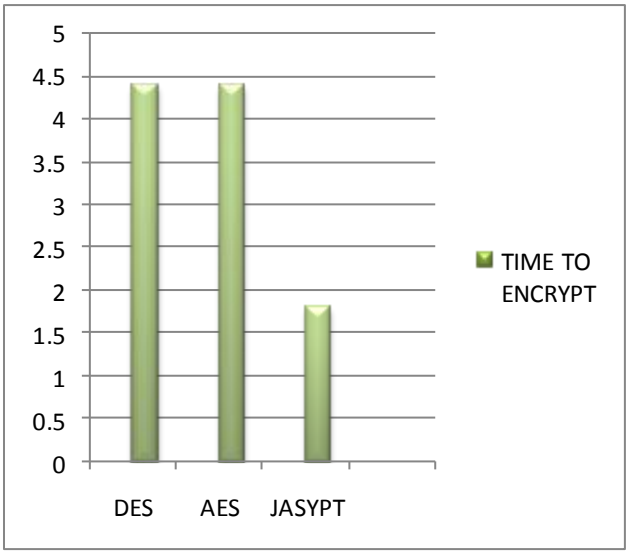

Fig 9.Time to Encrypt the information

6. CONCLUSION

Here ,I Conclude the work which is done on Hospital Patient's Repository by injecting data in Admin ,which stores hundreds of patient's information and their health history .They are need to be very secured. Simultaneously the work is being done on SQL server to protect from unauthorized agent. Because, patient's information are highly confidential. And add to a note that Data are being accessed frequently through JSP servelet, but viewed and added only through Web page document. Then it is being encrypted using JASYPT and AES Algorithm to make System secure .i.e. only authorized person can access using the known key either as sender/receiver. It is concentrated on Data security and authentication security as well as the clustering techniques is used to enhance the future application. so, it will provide privacy and preserving more than the existing system.

REFERENCES

[1] Narayanan, HemaAndal Jayaprakash, &Gunes, Mehmet Hadi. (2011). Ensuring access control in cloud provisioned healthcare systems. Paper presented at the Consumer Communications and Networking Conference (CCNC), 2011 IEEE.

[2] Andrew Kusiak and Matthew Smith, "Data mining in design of products and production systems", in proceedings of Annual Reviews in control, vol. 31, no. 1, pp. 147- 156, 2007.

[3] Bipul Roy, "Performance Analysis of Clustering in Privacy Preserving Data Mining", International Journal of Computer Applications & Information Technology, Vol. 5, no. II, May 2014.

[4] EshrefJanuzaj, Hans-Peter Kriegel and Martin Pfeifle, "Scalable Density-Based Distributed Clustering", Proceedings of 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), pp.231-244, 2004.

[5] M. Ester, H. Kriegel, J. Sander, and X. Xu, "A density-bsed algorithm for discovering clusters in large spatial databases with noise", In SIGKDD, pp. 226–231, 1996.

[6] Ng R. T., Han J.: "Efficient and Effective Clustering Methods for Spatial Data Mining", Proceedings 20th International Conference on Very Large Data Bases, pp.144-155, 1994.

[7] JinchaoJi , Wei Pang, Chunguang Zhou, Xiao Han, Zhe Wang, "A fuzzy k-prototype clustering algorithm for mixed numeric and categorical data", journal of Knowledge-Based Systems, vol. 30, pp. 129-135, 2012.

[8] Jinfei Liu, Li Xiong, Jun Luo, Joshua Zhexue Huang, "Privacy Preserving Distributed DBSCAN Clustering", transactions on data privacy, vol. 6, pp. 69–85, 2013.

[9] Josenildo Costa da Silva and Matthias Klusch, "Inference in Distributed Data Clustering", Engineering Applications of Artificial Intelligence, Vol.19, No.4, pp.363-369, 2005.

[10] Ron Wehrens and Lutgarde M.C. Buydens, "Model-Based Clustering for Image Segmentation and Large Datasets via Sampling", Journal of Classification, Vol. 21, pp.231-253, 2004.

[11] Ruoming Jin, AnjanGoswami and Gagan Agrawal, "Fast and Exact Out-of-Core and Distributed K-Means Clustering", Journal of Knowledge and Information System, Vol. 10, No.1, pp. 17-40, 2006.

[12] Sankita Patel, Viren Patel, Devesh Jinwala, "Privacy Preserving Distributed K-Means Clustering in Malicious Model Using Zero Knowledge Proof", Distributed Computing and Internet Technology, vol.7753, pp 420-431, 2013.

[13] G.Sheikholeslami, S. Chatterjee, and A. Zhang, "WaveCluster: A multi-resolution clustering approach for very large spatial databases", VLDB, pp. 428-439, 1998.

[14] Swagatam Das, Ajith Abraham, AmitKonar, "Automatic Clustering Using an Improved Differential Evolution Algorithm", IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems And Humans, Vol. 38, No. 1, 2008.

[15] Tiancheng Li, Ninghui Li, Jian Zhang, and Ian Molloy, "Slicing: A New Approach for Privacy Preserving Data Publishing", IEEE transactions on knowledge and data engineering, Vol. 24, No. 3, pp. 561-574, March 2012.