Privacy-preserving Search over Encrypted Personal Health Record in Multi-Source Cloud

M Hema Latha1, A Suhasini Snigdha2, B Vijaya Lakshmi3, K Padmavathi4, M Satish5.
1Department of IT, Lakireddy Bali Reddy College of Engineering, Mylavaram-521230, India,
2,3,4,5MCA Department, Lakireddy Bali Reddy College of Engineering, Mylavaram-521230, India.
lathamunnangi@gmail1, ssnigdha6630@gmail.com2, vijayalakshmiботла24@gmail.com3, padmavathikolusu1496@gmail.com4, marrelasathish@gmail.com5

Abstract. The biggest domain which needs a cloud platform in this smart world is the medical domain, where hospitals have a huge amount of data regarding the patient's records. When these personal health records are stored in the cloud, the system needs some encryption mechanism to secure the records. In this paper, an enhanced Attribute-Based Encryption (ABE). The health records of the patient are decrypted with the help of private keys. In this model, the patient doesn't need any certificate for authorization, so the model uses the details of the patients as attributes, and these details are encrypted. A threshold value is defined as $t$, if the number of encrypted attributes is matched with the threshold value then only the data owner can decrypt the data. A hierarchical mechanism helps the model to have smooth access to the data. The traditional algorithms are good at handling a small amount of data but this algorithm provides security as well as scalable property. The enhanced mechanism includes the unsupervised machine learning approach to label the patients based on their disease. In the cloud, a group of clusters are formed to separate the patients and allocate the doctors based on their disease treatment. The process of labeling is major drawback in case of attribute based encryption mechanism.

Keywords. Personal Health Records, Authority, Enhanced ABE, Cloud Security, Encryption, Decryption.

1. Introduction

Medical Encryption is the process of encoding the patient's details, which are gathered as features into an unreadable format. These are accessible by only data owners and other authorized persons. For better working with the cloud environment, the system should provide fine and quick access to data. This can be provided by using the ABE algorithm. The attribute-based encryption can be performed in three ways as shown in fig 1.

KP-ABE is popular in the distributed environment where it has more untrusted resources. It acts as black-box testing by checking the functional specifications of the cloud. CP-ABE exchanges the data with the cloud without worrying about accessing the policies associated with the execution...
environment. In Non-Moton systems, it uses AND, NOT, and OR structures, it applies encryption only on the selected features. The encrypted attributes are known as “Positive Attributes” and normal attributes are considered as “Negative Attributes”. Nowadays, cyber crimes are increased even in the health care industry data because hackers attack patients’ data and try to steal their personal information, and create fake profiles for earning money through health insurance or for misusing their data for selling their organs without their knowledge. So more attention is required to provide security for the health domain.

![Fig. 1. Classification of ABE](image)

2. Related Work

In [1] Yao, X et al provided a solution to another critical issue in the cloud storage systems in accessing the data controller to various authorities. This concept of multi authorities has more security breach attacks than without providing security to the data. A data access scheme known as “DAC-MACS”, has a drawback because it supports a double encryption mechanism in both ways which takes a huge amount of time, and as well as it doesn’t have any proper control mechanism for the second time encryption. The model also tries to find the solution for both data privacy, and user identity by designing a controlling mechanism known as “Anony”, but it is very weak at handling the common request from different users and is very difficult to handle bottleneck problems that occur during the communications between the servers and other devices. It is very bad at handling the decryption process, so the model suggested one more solution which is robustly known as “RAAC” in which the centralized authority is further divided into distributed administrators. The number of distributed administrators is dynamically decided based on the number of clusters formed in the network.

In [2] Qin et al proposed a framework for protecting the data in the cloud using the tree data structure along with preserving the semantics of the attributes. With the advancement of electronic media, there is a huge amount of data that has to be outsourced to the cloud. Medical systems which provide data storage on their databases have to be upgraded to the cloud for getting security and for providing more analytical reports on the patient's health status. This model majorly addresses the issues that arise due to the queries that affect the privacy of the patient. So, to solve this issue the model has proposed an encryption technique based on order preservation. This is done by constructing a binary search tree because of which the increase in the height of the tree can lead to an overhead problem. To solve the overhead problem, a multi-linear map data structure is implemented to reveal the order of preservation.

In [3] S Parthima et al studied various privacy-preserving for health records. While the era has been shifted from databases to third-party tools called “Cloud Environment”, it has to take care of integrity, authorized access, preservation, and confidentiality. Apart from cloud computing, it also studied about few fog computing environments integrated with blockchain technology, in these
In [4] Liang, P et al proposed a secure cloud healthcare data access using ABE (CP-ABE) ciphertext policy. In this system, only one central agency is responsible for the attributes of users and private keys (CA). In the olden techniques, the system checks the certification authorization every time the user wants to upload the file, this increases the overhead on the cloud, and the performance of the cloud decreases exponentially. This model solves this issue by giving authoritative power to various administrators, general it happens in the distributed systems by maintaining a hierarchy to access the data. So, the attributed-based encryption is converted into a Multiple Authority ABE mechanism and all the search queries can also be handled using the decentralized powers. There is a multitude of independence authorities in which attributes can be monitored and privately owned keys distributed. The model cannot be corrupted that much easier. But the limitation of this system it has taken care of the load and power consumption but it is not worried about the privacy of the patients. The second issue is that it needs one more secondary primary key, to pass through the second layer of authorities.

A contribute-based encoding (PPMA-ABE) scheme has been imported, which safeguards privacy. In their works, the central authority had been abolished, but the interaction between multiple authorities was required. A PHR sharing scheme for ABE multi-authority is efficiently proposed but in case of any intruder central authority, it has to face privacy leakage of patients. There have been several works in recent years to improve the situation. A multi-competence ABE (MA-CP-ABE) ciphertext policy scheme which uses an AND gate with latencies to access structure. The GID user’s data can be easily disclosed to anyone with security breach protocols.

In [5] Doulamis, N et al stated that to handle the heavy load operations in the cloud and to make it scalable, the system needs some enumeration tools for monitoring the networks. One of the tasks is to remove the start time from the shuffle by predicting the run-time of the task. As part of performance analysis and optimization, the task execution time prediction is usually presented. Use random forest performance model’s prediction work time to lower Spark Streaming configuration optimization costs. uses samples to record the convergence trend of the algorithm and an experimental technique to predict algorithm time with various iterative features. The impact on multiple performance metrics of the Spark configuration change through the execution-driven model. A performance approximation technique for the simulation by stochastic Markov model of the performance of multistage iterative applications, use small samples for the application duration to fit the linear model and aid in predicting large-scale duration. The system needs a complete prediction model to improve the performance either it can be in small scale, medium, or large scale environments. The hypotheses functions act as grossly model time (application length and complete time of task), which in the actual work execution procedure may ignore the random factor.

3. **Proposed Methodology**
This proposed algorithm combines KPABE and CPABE algorithms to provide security to health records. The major elements in this framework are data owners, who can access the data from anywhere, and who have all the privileges to the file he/she has uploaded. Data providers are the third party resource persons like nursing staff, doctors, or any other employee of the hospital who can...
upload the patient’s details; they can even edit the data for entering the continuous track record of the patient. This model also helps the data owners to visualize their health status periodically. Simultaneously, this model helps the doctors to view their patient's status even after they get discharged until their observation period so that this will help the doctors to treat them if anything gets serious or they change the medications if necessary.

This cloud environment for hospitals can help the patients without carrying their medical records when a patient wants to move or get treatment from other hospitals because the model supports both data sharing and collaboration. This collaboration is done even with the clinical centers where a patient gets the reports from the lab. The model even supports a search query mechanism to find the details of the patients along with the treated doctor details. The model designs a matrix with symptoms and patients attributes, this helps the doctor to recommend medicines or a second opinion for further treatment. The model generates hashtag indexes to every record and the centralized system maintains a metadata table so it can recover at later stages also. The systematic approach is described in fig 2.

![Fig. 2. Enhanced ABE Encryption and Decryption Process](image)

The algorithm implemented in this paper is discussed below step by step to ensure security.

### 3.1. Algorithm

Input: patient’s health record data, PHRD
Output: Encrypted health record data
Procedure: Key_Generation(PK,MK,AttList)
1. $r \leftarrow \text{random}()$
2. $\text{grp1} \leftarrow \text{PK} \times r$
3. $b_{\text{inv}} \leftarrow 1/\text{MK}$
4. for $i$ in $\text{AttList}$:
   a. $r_{\text{attr}} \leftarrow \text{random}()$
   b. $k_{\text{attr}} \leftarrow \text{grp1} \times \text{hash(grp1)}^{r_{\text{attr}}}$
   c. $m_{\text{attr}} \leftarrow \text{PK}^{r_{\text{attr}}}$

Procedure: Encrypt($\text{PK}, \text{MSG}, \text{P}_\text{str}$):
1. $\text{users}=[]$
2. for $i$ in $\text{num_cols}$
   a. $r_{\text{attr}} \leftarrow \text{random}()$
   b. $\text{users}.\text{append}(r_{\text{attr}})$
   c. $\text{sum} = \text{sum} + \text{row}[i] \times \text{users}[i]$
   d. $k_{\text{1_attr}} \leftarrow \text{PK} \times \text{users}[i]^{\text{sum}}$
3. print $k_{\text{1_attr}}$

4. Experimental Results

Fig 3 provides the credentials to access the cloud and store their patients in the framework for encryption to start.

![Login Credentials Screen](image1)

**Fig. 3.** Login Credentials Screen

Fig 4 provides the provision for uploading or download the patient’s files by the data owner either by encrypting or decrypting the patient’s record.

![File upload](image2)

**Fig. 4.** Patient’s record encryption or decryption process

Fig 5 shows the encryption process in which the patient details like name, age, pulse rate, blood pressure, and temperature are uploaded in the panel. When the user enter his/her details all the attributes are read and using a private key and master key random values are generated. Basically the
key generator assigns one of the key to the user based on the diseases he/she has because the TPA takes care of dividing the patients based on the symptoms and diseases. Every disease is identified as a group and has allocated with some master key. Now using this private key and master, the systems encodes all the attributes present in the record as follows.

![Encryption](image)

**Fig. 5. Patients Attributes Encryption Process**

Fig 6 shows the decryption process when the data owner or doctor wants to access the patient’s record. During the decryption process, the TPA has master key and patient or doctor who wants to access the file first sends a approval to the TPA, then auditor asks for the key exchange between both. Then the system starts a verification process, on successful completion of it, TPA allows users to access the file.

![Decryption](image)

**Fig. 6. Patients Attributes Decryption Process**

5. Conclusion
The patient’s health records need a security assessment to successfully deploy an organization EHR process into the cloud environment which supports BigData framework with ETL tools, and Apache tools to perform various activities on the health records. The proposed algorithm addresses a lot of security breach issues than the current systems. Here, entire data is stored in the remote server and no backup is provided in the client system. They provide continuous reports on the audit log to the data provider and data owner to prevent malicious attacks on the data. Since some of the patient’s attributes contains information in textual format, so this model has taken care of the semantics preservation because a single word change may make the interpretation wrong and it may lead to either wrong treatment or any serious issues.

References
[1] Yao X., Lin, Y., Liu, Q., & Zhang, J. (2018). Privacy-Preserving Search Over Encrypted Personal Health Record In Multi-Source Cloud. IEEE Access, 6, 3809–3823. doi:10.1109/access.2018.2793304
[2] Liang, J., Qin, Z., Xiao, S., Zhang, J., Yin, H., & Li, K. (2019). Privacy-preserving range query over multi-source electronic health records in public clouds. Journal of Parallel and Distributed Computing. doi:10.1016/j.jpdc.2019.08.011
[3] Prathima S., Priya C. (2020) Privacy Preserving and Security Management in Cloud-Based Electronic Health Records—A Survey. In: Peng SL., Son L., Suseendran G., Balaganesh D.
(eds) Intelligent Computing and Innovation on Data Science. Lecture Notes in Networks and Systems, vol 118. Springer, Singapore. https://doi.org/10.1007/978-981-15-3284-9_3

[4] Liang, P., Zhang, L., Kang, L., & Ren, J. (2019). Privacy-preserving decentralized ABE for secure sharing of personal health records in cloud storage. Journal of Information Security and Applications, 47, 258–266. doi:10.1016/j.jisa.2019.05.012

[5] Doulamis N.D., Doulamis, A.D., Kokkinos, P., et al.: Event detection in Twitter microblogging. IEEE Trans. Cybern. 46, 1–15(2015)

[6] Manekar, A., Pradeepini, G., Optimizing Cost and Maximizing Profit for Multi-Cloud-Based Big Data Computing by Deadline-Aware Optimize Resource Allocation, Studies in Computational Intelligence, 2021, 921, pp. 29–38.

[7] A.Mallikarjuna, B. KarunaSree, “ Security towards Flooding Attacks in Inter Domain Routing Object using Ad hoc Network” International Journal of Engineering and Advanced Technology (IJEAT), Volume-8 Issue-3, February 2019.

[8] MANEKAR , A.S. and PRADEEPINI, G., 2019. Experimenting cloud infrastructure for tomorrows big data analytics. International Journal of Innovative Technology and Exploring Engineering, 8(5), pp. 885-890.

[9] Srinivasa Reddy, K., Suneela, B., Inthiyaz, S.,Kumar, G.N.S., Mallikarjuna Reddy, A.” Texture filtration module under stabilization via random forest optimization methodology “ International Journal of Advanced Trends in Computer Science and Engineering, Volume 8, No.3, May - June 2019.

[10] Mayank Singh and Viranjay M. Srivastava, “Implementing Architecture of Fog Computing for Healthcare Systems based on IoT,” International Journal of Engineering and Advanced Technology (IJEAT), vol. 8, no. 4c, pp. 23-27, April 2019.

[11] K. Sudheer Reddy, G. P. S. Varma and S. S. S. Reddy, "Understanding the scope of web usage mining & applications of web data usage patterns," 2012 International Conference on Computing, Communication and Applications, 2012, pp. 1-5, doi: 10.1109/ICCCA.2012.6179230.

[12] Shekhar Singh, Mayank Singh and Sandhya Tarar, “Improved Facial Recognition based Authentication approach to Secure Big Data,” International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8, no. 6c, April 2019.

[13] C. R. T, G. Sirisha and A. M. Reddy, "Smart Healthcare Analysis and Therapy for Voice Disorder using Cloud and Edge Computing," 2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), Mangalore, India, 2018, pp. 103-106, doi: 10.1109/iCATccT44854.2018.9001280.