ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Identity Based Encryption for Secure Data Access in Cloud Computing using Machine Learning Algorithms

[1]A.Hemlathadhevi and [2]Dr.Rajeshwari Mukesh

[1]*Department Of Computer Science and Engg, St. Peter's University, Tamil Nadu, Chennai*
[2]*Department Of Computer Science and Engg, Hindustan University, Tamil Nadu, padur.Chennai.*

ARTICLE INFO

Article history:
Received 3 October 2015
Accepted 31 October 2015

Keywords:
Cloud Storage; Medical Data; Machine Learning Technique; Identity Based Encryption; Privacy Preserving Mechanism; Homomorphic Encryption

ABSTRACT

Cloud computing will be the future big data storage center for medical data. It contains numerous huge datacenters which is typically heterogeneous and geographically distributed. However, secure data access in cloud computing is a big challenge. In this paper proposed a new method for secure medical data storage and transformation in cloud computing platform. The medical data are classified as sensitive or insensitive data using a machine learning technique and those data are encrypted using the Identity Based Encryption (IBE) method for secure data access. Since the cloud medical data are shared across multiple users, data privacy is very essential. Thus, the proposed system makes use of privacy-preserving mechanism for user verification. Finally, compare the proposed method with other cloud data encryption methods through simulation and comprehensive analysis. The experimental results show that the proposed data accesses approach on cloud is very secure and feasible.

To Cite This Article: A. Hemlathadhevi and Dr.Rajeshwari Mukesh., Identity Based Encryption for Secure Data Access in Cloud Computing using Machine Learning Algorithms. ***Aust. J. Basic & Appl. Sci.,*** *9(33): 72-77, 2015*

INTRODUCTION

Cloud computing improves its quality in different medical applications. The medical data sharing through the cloud platform has become very common nowadays. Transmission of medical data online makes the process of medical expert easier, capable of storing, sharing and accessing the medical data easier (Danan Thilakanathan, *et al*., 2014). The cloud computing renders the vast storage application which are managed by the medical experts from the remote place. Thus the cloud platform has gained some attention from the IT vendors. It acts as an exchange platform where the hospitals and other health organizations exchange their patients' medical records and also acts as a storage place for those medical records. The cloud computing technology on the medical record has the merits and also the demerits over the data security since the security and the reliability are considered as the issues.

Medical data storage in the cloud computing platform allows the users to maintain their details conveniently. Initially the user encrypts the personal details and sends to the cloud storage. The cloud platform rearranges the encrypted data without knowing the original data. Usually the encrypted data in the cloud is maintained by the third party (Abhishek Kumar Gupta and Kulvinder Singh Mann, 2014). Therefore the integrity, confidentiality, security of the medical data is becoming doubtful than the private storage system. Thus it is important to verify whether the encrypted data on the cloud can be accessed or modified by any unauthorized third party. Many probabilistic encryption techniques are used in the cloud platform where the third party is responsible for the cloud storage and the secret key is generated by the third party. But still they are experiencing another issue that the unauthorized user can also query the data from the cloud storage that leads to the leakage of the data to the outsiders.

The paper proposes a secured data access scheme that provides more security and integrity in storing and sharing the confidential medical data over cloud platform. This proposed scheme incorporates four different methods in cloud platform that enhances the security, confidentiality and the integrity of the personal data of the user. First the patients' medical data is classified into sensitive and insensitive data which is possible by the Support Vector Machine (SVM), a machine learning algorithm. Second method is an encryption technique called Identity Based Encryption (IBE) where the access permission is used not only for the data user but also for the data on the cloud. The data owner provides this data access permission which is usually given by the third party in the cloud. Third one is

Corresponding Author: A.Hemlathadhevi Department Of Computer Science and Engg, St. Peter's University, Tamil Nadu, Chennai
E-mail:hemlathadhevi@gmail.com

homomorphic encryption which is optionally needed when the data owner wants to modify the shared data in the cloud platform. It improves the security of the cloud platform where the modification of the encrypted data is done without being aware of the original data and the authorized user retrieve and decrypts the data. Finally the ring signature is generated which is a Privacy Preserving Mechanism that maintains the privacy of the sensitive data by user verification process.

The rest of this paper is structured as follows. Section 2 describes the related work of the proposed techniques. Secured medical data sharing on the cloud is explained in section 3. Section 4 describes the results and its discussion with suitable diagrams and finally the conclusion of this paper are presented in the section 5.

Related Work:

The data owner brought the sensitive data and the data security and the access control becomes the main issue here. The paper (Seenu Iropia and Vijayalakshmi) solves the issue by applying the data access policies according to the data qualities and allows the data owner to do the calculation for the access control without revealing the data content. Therefore the Key Policy Attribute Based Encryption (KP-ABE) is implemented to solve all these issues mentioned in the paper.

The author (F. Ozgur CATAK and M. Erdal BALABAN, 2013) proposed the Cloud SVM training mechanism along with the Map-Reduce technique in the cloud environment. The distributed cloud environment adapted the SVM algorithm to classify the datasets accordingly. Then the combine the support vectors to each node in the cloud. These two steps are repeated continuously until the optimal classifier for the entire cloud dataset is detected. Since it is difficult to detect the optimal data set from the huge data sets using the SVM algorithm on a single system, the data sets are divided into many iterations and the iteration results are combine and go for the few more iterations to get the optimal dataset set.

Communication in the cloud computing is a big issue in terms of data security and the access control. Three main features such as Key Confidentiality, Key Freshness and the Key Authentication are essential in the key based sharing data in the group that is dynamically changing. To attain these three features the author (Dharani. R, M. Narmatha, 2014) proposes the Dynamic Group key protocol based on Key Generation Center. This method distributes the keys to the group members and dismisses the key distribution to the members not in the group. In addition the anonymous access is controlled by the short group signature scheme.

The remotely stored data encryption is been the big issue in the field of data sharing the cloud environment. All on the proper basis, the homomorphic encryption is a better approach that improves the security of the application in which the sensitive data resides in. The author (Shashank Bajpai and Padmija Srivastava, 2014) takes the encrypted data as input to the encryption and modify the proper changes to the encrypted data without decrypting the data and the data content is unaware to the user while doing modification. This method of homomorphic encryption proves that the private data remote sharing is achieved without disturbing the data privacy.

The key generation center is usually used to the normal decryption of all the messages that belonging to the particular user. But if the data is shared between any two users, then the data should be kept secret to the outsiders other than the two users who shared the data. Hence the author (M. Saranya and R.Vasuki, 2015) suggest a new encryption scheme that solves this problem by issuing the escrow-free key that provides the two party communications between the data storing center and the key generation center. The security performance is controlled by using the Third Party Auditing (TPA) and the security analyses that manage the data in the distributed sharing system securely by using the RBAC.

The cloud service providers in the cloud environment restrict the data protection based on the CSP. Many approaches are available for the prevention of CSP where the user privacy is under risk with the untrusted CSP. These existing approaches bring the overhead in communication and complex key management process. Hence the paper (Hassan Takabi, 2014) privacy aware access controls system that has two level of users' data protection in the cloud environment. CSP enforced access control mechanism protects the user data and the third party service provider protects the data using the multiple layers of encryption.

An algorithm (Divya Meena, *et al*., 2014) is needed to share the private data securely over many data users. Each node in the network is assigned with the ID numbers which has the range of 1 to N. This is possible using the ID assignment technique and this makes the data to be shared securely more complex than before. All the nodes as per the ID assignment, allotted an ID using the central authority. This algorithm for sharing the private data securely is compared with the previous existing algorithms and proves that the proposed algorithm beats the existing algorithms.

Secured Data Sharing On Cloud:

A. ***Materials and Methods:***

Patients' data consisting of clinical, demographics and general information obtained from the database. Nearly more than 6000 patients' details have received who are under the treatment of chest and lung disease. These databases comprises of the patients details such as name, age, gender, family

history regarding the disease, laboratory test results, physical verification with the doctor, image testing results and so on. The severity of the disease is differentiated into acute, sub-acute and the chronic stages. A subset (some attributes) of this data base is shown in the table.1 (Mridu Gulati, 2011). This table shows the condition of the lungs based on the health criteria. These data are gathered under clear observation since 2007. Most of the patients wanted to know their medical status and get their prescription from remote place. So it is necessary to store the data in the cloud platform. Since the details of the patients are having both sensitive and non-sensitive information, security and confidentiality becomes challenging issues. The Identity based encryption with the homomorphic encryption is used to overcome the security issue and the ring based signature scheme is used to overcome the confidentiality issue.

Table 1: Database Sample.

Gender	
Male	Idiopathic Pulmonary Fibrosis, Pneumoconiosis
Female	Lymphangioleiomyomatosis, Connective Tissue Disease Associated ILD
Age	
20-40 years	Connective Tissue Disease Associated ILD, Lymphangioleimyomatosis, Sarcoidosis
Above 50 years	Cryptogenic Organizing Pneumonia, Idiopathic Pulmonary Fibrosis.
Family history	Basic symptoms that depends on the gene
Physical Examination	Clubbing, Extra-Pulmonary Signs, Lung Testing such as Wheeze or Crackles
Pulmonary Function Testing	Lung Volumes, Resting and Ambulatory Oxygen Saturation, Arterial Blood Gas
Biopsy	Surgical Lung Testing and Bronchoscopy Testing
Laboratory Testing	Hypersensitive Pneumonitis Panel, Routine Blood Test, etc.
Imaging	High Resolution CT Scan, Chest Radiography (X-Ray)

B. Data Classification:

The dataset consists of both the common and private data for each patient. Before the encryption and the sharing process starts, it is significant to separate sensitive (disease, severity of the disease, etc.) and insensitive data (name, address, etc.) and provide privacy to the sensitive data. Hence, the most efficient Support Vector Machine (SVM) (Afif, M.H, *et al*., 2012) is used for the classification of patients' medical record into sensitive and insensitive data. Consider the elements (name, age, disease, severity, stage, status, level, etc.) of the medical record as a set. Let xi be the set and the elements i= name, age, disease, severity, stage, status, level, etc. according to our system, we define two classes namely sensitive and insensitive. Taking the elements one by one, assigning the weight to the elements based on the sensitivity. Here if the weight is 1 then it belongs to the sensitive data such as disease, severity, etc. if the weight is -1 then that element belongs to the insensitive data such as name, address, age, etc.

Let y be the classification of two classes and is represented as

$$y_i = \begin{cases} 1 \; if \; x_i \; in \; sensitive \; class \\ -1 \; if \; x_i \; in \; insensitive \; class \end{cases}$$

C. Secure Transaction:

The classified data should get encrypted before storing it in the cloud. Thus the most significant encryption technique called Identity Based Encryption is used (Varsha S. Agme and Archana C. Lomte, 2014). The process of IBE is depicted in the figure 1.

Initially the data owner encrypts the medical records and sends to the cloud server. If the patient wants to know his medical record from the cloud server, he/she request the data owner for the re-encryption key (secret key). The data owner accepts the request and verifies the authorization of the data user who sends the request. If the data user is an authorized user, then the re-encryption key is generated and sends to the authorized data user. Once the re-encryption is key attained, then the data user can able to access his/her medical reports. The main advantage of this Identity Based Encryption (IBE) is that it can provide the details of the particular patients he/she requested. The other patients detail will not be rendered without the proper authorization.

In our proposed system, there is only one group **G** consisting of medical experts as data owner and the patients as data users. Let us see the working of Identity Based Encryption (IBE) in our system. As in the normal IBE, the process consist of four steps namely Set up, Key Generation, Encryption and Decryption. The description of these four steps implemented in our proposed system is as follows:

Setup:

The Public Key Generator (PKG) randomly selects the public parameters T such that $k \leftarrow L(T)/DS(T)$, where k is a random pseudo square, $DS(T)$ is a set of quadratic residues modulo$T$, $L(T)$is the set of elements with Jacobi symbol 1 modulo T and T=ab where the master secret key is a and b.

Australian Journal of Basic and Applied Sciences, 9(33) October 2015, Pages: 72-77

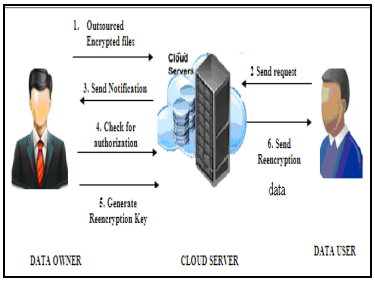

Fig. 1: Secured Cloud Data Storage Using Identity Based Encryption.

Key Generation:

The generation of common ID is done by the PKG and the hash function is calculated for each block of the data. The secret key for each block of the data is computed in this key generation. The hash function of the chosen common ID is computed as

$$S = H(ID) \quad (1)$$

And the secret key for each block of the data is $d_{ID} = t$

Where $t = \begin{cases} \sqrt{S} \text{ if } S \text{ is a square} \\ \sqrt{kS} \text{ if } S \text{ is not a square} \end{cases}$

Encryption:

The secret key is generated for each block of the data and the data has to be encrypted using the generated secret key. For example, consider a secret key of the block x and is computed as

$$d_x = (f_x^A + k^x S)/f_x \quad (2)$$

And the encryption of the data block is done using

$$C = Encrypt(T, ID_M, M) \quad (3)$$

Where M is the original message, T is the public parameters and ID_M is the identity of the whole message

Decryption:

The encrypted data stored in the cloud and the privacy preserving settings based on the following session are applied to encrypted message in the cloud platform. These privacy preserving settings are used for the purpose of allowing the specific users to read the corresponding file. This clearly says that the medical details of the particular patient will not be visible to the other patient even though they belong to the same group. The decryption of the encrypted message is done using the public parameters T, cipher text C, identity of the whole message ID_M and the most importantly the secret key of the specific file (medical report of the particular patient).

$$Decrypt(T, C, ID_M) = M \quad (4)$$

The medical data of a patient is not a constant one but have changes according to the health condition of the patients. Once the data encrypted and storing in the cloud, traditionally the future changes are updated by decrypting the data, make the changes, encrypt the data again and store in the cloud. But in addition to the IBE for the secure data access in the cloud, another encryption technique called homomorphic encryption is used. This method performs the changes on the encrypted data itself without decryption and without knowing the private key of the data. The private secret key is known only to the data owner.

D. ***Privacy Preserving:***

Privacy preserving is a concept of preserving the identity of the individual in a group. Consider the medical data set stored in the cloud is consisting of many patients' health details. The encrypted data in the cloud is divided into blocks and each block is signed using one of the group members (data users). While verifying the authorization of the data user in the IBE phase, the ring signature is used to verify the signature whether it belongs to the particular group. It will not determine the particular data user. Using the ring signature and known authorized users, the verifier cannot able to determine the signer's identity.

The generation of ring signature consists of three basic steps namely key generation, ring signature generation and the ring verify (Boyang Wang, *et al.*, 2014). In the key generation phase, all the users in the group generate their own private and public key.

For a user u_i, they randomly select the private key $x_i \xleftarrow{R} Z_p$ and calculate the public key using the private key as $w_i = g_2^{x_i}$.

During the ring signature phase, any one user in the group generates a signature on a block and a block identifier is generated using the public keys of all the group members and private key of the particular data user. Any two blocks can be differentiated between each other using this block identifier. The ring signature is computed as

$$\sigma_s = \left(\frac{\beta}{g^{(\prod_{i=s} w_i^{a_i})}}\right)^{1/x_s} \in G_1 \quad (5)$$

Where $R_s = H_1(id)g1^m \in G_1$ (6)

Where H_1 is the hash function, id is the block identifier, G_1 be a group.

Last step is ring verify, where a verifier checks whether the particular block is signed by any one of the group member. Using the number of users d, private and public keys, identifiers, block m, and a ring signature σ_s, the equation (2) is computed and then checks the equation (7).

$$e(\beta, g_2) = \prod_{i=1}^{d} e(\sigma_i, w_i) \quad (7)$$

Australian Journal of Basic and Applied Sciences, 9(33) October 2015, Pages: 72-77

If the equation (7) satisfies, then the given block is signed by the any one of the group members or else it is an unauthorized data access process. The major advantage of using the ring signature for the privacy preserving is that it does not need much space to store ring signature and dynamic operations are updated to the blocks without disturbing the privacy preserving process.

Result Analysis:

The important techniques used in the proposed system are evaluated in this section. The Support Vector Machine (SVM) is used for classifying the data as sensitive and insensitive. The figure 2 shows that the SVM classifier maintains the efficiency with consistent degree. From the experimental result, even though the database has more insensitive data than the sensitive data, the SVM maintains constant accuracy and that is shown in the figure 2. This figure also depicts that the SVM provide better performance even when the number of data to be classified is increased.

The Identity Based Encryption (IBE) in the proposed method is compared with the previous probabilistic encryption technique. In the probabilistic encryption, the full plain text or the partial plain text is easy to compute from the cipher text. Hence the IBE is performing better than the existing Probabilistic encryption. The figure 3 shows the comparison graph for both the probabilistic encryption and the Identity Based Encryption (IBE) that shows the IBE proves better efficiency than the existing encryption method.

The IBE alone shows the encryption efficiency better than the previously encrypting methods. But the modification of the data after encryption takes much time for decryption, data updating, and again encryption. So the homomorphic encryption is added with the IBE that reduces the time taken by the modification of data. The efficiency of the IBE and the IBE integrated with the homomorphic encryption is depicted in the figure 4.

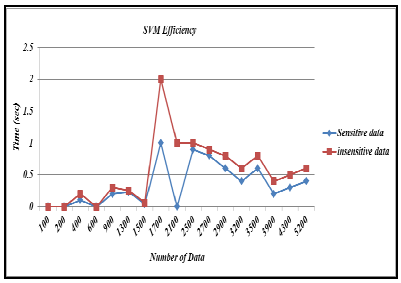

Fig. 2: SVM Classifier Efficiency.

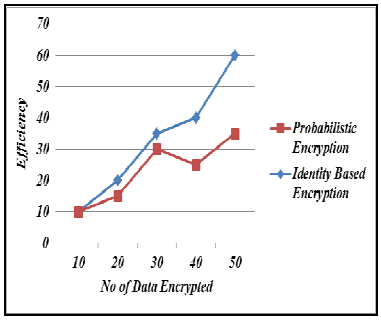

Fig. 3: Comparison of Ibe With Existing Probabilistic Encryption.

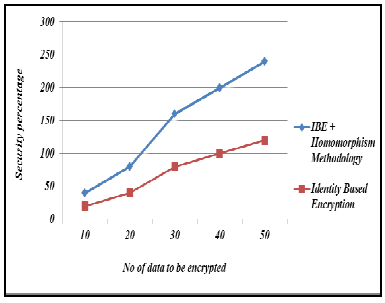

Fig. 4: Efficiency of Ibe with the Homomorphic Encryption.

Conclusion:

The paper proposed a secured medical data access on the cloud using the most elegant Identity Based Encryption which encrypts the data based on the user identity. The medical data is classified as sensitive and insensitive data for providing the security in an easier way. The modification on the data after stored in the cloud is done without decrypting it using the homomorphic encryption. The privacy of the medical data is preserved using the proposed ring signature that is known only to the group member to access the data. Finally the proposed system is compared with the existing probabilistic encryption on cloud and shows it is better than the existing method. The efficiency of Identity Based Encryption along with the Homomorphic encryption and the efficiency of the SVM classifier is shown.

REFERENCES

Danan Thilakanathan, Shiping Chen, Surya Nepal, Rafael A. Calvo, 2013. "Secure data sharing in the Cloud, Book Chapter on Security", Privacy, and Trust in Cloud Systems, by Springer.

Abhishek Kumar Gupta and Kulvinder Singh Mann, 2014. "Sharing of Medical Information on Cloud Platform-A Review", IOSR Journal of Computer Engineering (IOSR-JCE), 16(2): ISSN: 2278-0661.

Seenu Iropia and Vijayalakshmi, 2014. "Decentralized Access Control of Data Stored in Cloud Using Key Policy Attribute Based Encryption", International Journal of Inventions in Computer Science and Engineering, 1-2, ISSN (Online): 2348 – 3539, ISSN (Print): 2348 – 3431.

Ozgur Catak, F. and M. Erdal Balaban, 2013. "Cloud SVM: Training an SVM Classifier in Cloud Computing Systems", Springer Berlin Heidelberg, 7719, ISSN: 0302-9743, 57-68.

Dharani, R. and M. Narmatha, 2014. "Secured Data Sharing with Traceability in Cloud Environment", International Journal of Inventions in Computer Science and Engineering Volume 1, Issue 8, ISSN (Online): 2348 – 3539, ISSN (Print): 2348-3431.

Shashank Bajpai and Padmija Srivastava, 2014. "A Fully Homomorphic Encryption Implementation on Cloud Computing", International Journal of Information & Computation Technology, Volume 4, Issue 8, pp. 811-816, ISSN 0974-2239.

Saranya, M. and R. Vasuki, 2015. "Improving Data Security in KP-ABE with Third Party Auditing", International Journal of Inventions in Computer Science and Engineering, Volume 2, Issue 2, ISSN (Online): 2348 – 3539, ISSN (Print): 2348 – 3431.

Hassan Takabi, 2014. "Privacy Aware Access Control for Data Sharing in Cloud Computing Environments", proceeding of the second International Workshop on security in Cloud Compting, pp.27-34, ISBN: 4503-2805.

Divya Meena, M., A.R. Arunachalam and T. Nalini, 2014. "Confidential Data Sharing With Anonymous Id Assignment Using Central Authority", I-2. International Journal of Inventions in Computer Science and Engineering, ISSN (Online): 2348 – 3539, ISSN (Print): 2348 – 3431.

Afif, M.H. and A.R. Hedar, 2012. "Data Classification using Support Vector Machine Integrated with Scatter Search Method", Japan-Egypt Conference on Electronics, Communications and Computers (JEC-ECC), IEEE, pp: 168-172, ISBN: 4673-0485.

Varsha, S., Agme and Archana C. Lomte, 2014. "Cloud Data Storage Security Enhancement Using Identity Based Encryption", International Journal of Application or Innovation in Engineering & Management (IJAIEM), 3-4, ISSN 2319–4847.

Boyang Wang, Baochun Li, Hui Li, 2014. "Oruta: Privacy-Preserving Public Auditing for Shared Data in the Cloud", IEEE Transactions on Cloud Computing, 2(1).