© 2016 IJSRSET | Volume 2 | Issue 2 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Intelligent Data Aggregation and Merging Algorithms for Secured Storage of Medical Information in Cloud

A. Antonidoss*, D. Manjula,

Department of Computer Science and Engineering, College of Engineering Guindy, Anna University, Chennai, Tamilnadu, India

ABSTRACT

Data aggregation is playing major role in cloud environment for providing data security in terms of effective storage. Secure storage is a crucial task when stored the confidential data such as medical reports, official secrets and organization resources on cloud. This is time to propose a new technique for secured storage. For that purpose, we propose a new secured storage model using cryptography. This model uses a newly proposed data aggregation algorithm for forming effective group and a new data partitioning method for extracting the useful data which are grouped securely. Finally, a new data merging algorithm for segregating the original data which are partitioned in this model. The proposed model is useful for securing confidential medical data and also for making effective decisions over the diseases by medical expert systems. Experiments have been conducted in this research work for evaluating the efficiency of the proposed secured storage model by using UCI Repository medical datasets.

Keywords : Cloud Computing, Medical expert system, Secured Data Storage, Data Aggregation, Data Partitioning, Data Merging.

I. INTRODUCTION

In India, Diabetes is the major disease which is to be effectively managed by both patients and doctors to provide better health services. This can be achieved effectively by transforming the practices of healthcare planning and implementation using Mobile cloud computing. Diabetes is a serious disorder which is characterized by high levels of blood glucose in the human body. Moreover, it originates from the defects for humans in the insulin production, insulin usage or both and is a silent killer. The insulin hormone which is secreted by pancreatic beta cells is regulated by the intake of the glucose from the blood into most cells of human body (Alberti & Zimmer 1998). Health complications such as damage of heart including stroke, high blood pressure, eye problems with severe vision loss or blindness are led by the inability of the human body to generate sufficient insulin hormone leading to increased level of blood glucose (Rosenthal et al 2007, Amos et al 1997). In such a scenario, maintaining patient data using relational databases will require the user to use the data only from specific locations.

Cryptography is useful for providing data security during the transmission of intellectual data in networks. It provides the facility to encrypt the valuable data to any authenticated person (user) through wired or wireless network environments. The same data decrypts by the respective authorized user. For these data encryption and decryption process uses the keys such as private key, public key and group key. This encryption and decryption process is also called data hiding. Data partition and data merging also proceeded by using any encryption and decryption methods which are present in Cryptography.

In this paper, we propose a new secured storage model for effective storage in cloud. This model uses three newly proposed algorithms for data aggregation, data partition and data merging. Rest of this paper is organized as follows: Section 2 provides the literature survey. Section 3 explain in detail about the proposed model and their algorithms. Section 4 shows the experimental results and discussion. Section 5 gives conclusion and future works.

II. METHODS AND MATERIAL

A. Literature Survey

There are many works have been proposed in this direction by various researchers in the past. Among them, Nissim et al (2010) proposed a novel approach for achieving k-anonymity by partitioning the original dataset into several projections. They used rejoin and projections in their model for data partitioning and merging process. An identity-based data storage scheme called selective-identity model was proposed by Jinguang Hana et al (2013). This scheme is suitable to the cloud computing scenario and it supports both intra-domain and inter-domain queries. In their model, the access key is bound not only to the requester's identity but also to the requested ciphertext. The key can be computed by the owner independently without the help of the group head or they centralized authority. Moreover, their scheme is secured against collusion attacks.

Piotr K. Tysowski et al (2013) proposed a new key management system for providing security to data on emerging applications. Sudip Misra et al (2014) proposed a cloud-assisted Wireless Body Area Network (WBAN) based framework for aggregating data from Local Data Processing Units (LDPUs) of the patient records. They also introduced an algorithm called Optimal Channelization Algorithm (OCA) for channelizing data through dynamic gateway allocation.

Ji-Jiang Yang et al (2015) proposed a new privacy preserving model for medical data sharing in cloud. They used vertical partition for partitioning the medical data to achieve the confidentiality with various privacy constraints. Their model consists of vertical data partition, data merging and integrity checking by using plaintext and ciphertext technique. Their model uses the statistical analysis and cryptography methods for effective secured storage. Muthurajkumar et al (2015) introduced a new log management system for providing data security in cloud with temporal feature. A new mobile cloud database service solution is proposed by Lihui Lei et al (2015) for effective data storage and managing the mobile data on cloud databases. They also provide location dependent query processing features. However, security issues are not focused in their work. Hongwei Li et al (2015) developed a security model using effective encryption method which is suitable for secured storage and retrieval of encrypted data in cloud.

B. Proposed Work

In this paper, we propose a new secured storage model for effective data storage. This model uses three new algorithms namely Algorithm for Data Hiding through Aggregation (ADHA), Data Partitioning Algorithm (DPA) and Data Merging Algorithm (DMA) for effective data storage in cloud database.

i. Aggregation Scheme

In this section, we introduce the newly proposed aggregation algorithm for providing security using data hiding through aggregation, partitioning and encryption. This section explains in detail about the proposed data aggregation algorithm for first level security with sample input and output of this work.

Algorithm for Data Hiding through Aggregation (ADHA)

```
/* For each element, this algorithm finds the sum of all the elements of corresponding rows and columns in the original dataset */

Input   : Medical dataset
Output  : Aggregated data

Step 1: Read a [i,j] for i = 1 to n and j = 1 to n from Original Table (OT)
Step 2: Convert into Numeric valued Table (NT) of the data using look up table.
Step 2: For each element of the table NT
   (a) Find the sum of all the elements in its row.
   (b) Find the sum of all the elements in its columns excluding the current element.
Step 3: Store the aggregated values into the corresponding position in the secured table.
```

This proposed algorithm works on the basis of sum of columns and rows values of the respective data on table. The aggregated data will be replaced for the respective place of data on table. The original data is read from dataset and the numeric/nonnumeric data are converted into numeric data using ASCII values. Now, the aggregation is performed to provide first level data security.

For example, consider a table with six attributes such as Insurance ID, Patient ID, Patient Name, Date of Birth, Address and the medical information. Table 1 shows the Numeric Valued table for the original information which is converted into the numeric form.

Table 1 Original Table

	Column1	Column2	Column3	Column4	Column5	Column6
Row1	1	2	3	4	5	6
Row2	7	8	9	10	11	12
Row3	13	14	15	16	17	18

Here, the patient ID is used as the primary key. The elements of the first row are hidden using the following equation.

$$a_{11} = (a_{11} + a_{12} + a_{13} + \cdots + a_{1n}) + (a_{21} + a_{31} + a_{41} + \cdots + a_{n1})$$

$$a_{12} = (a_{11} + a_{12} + a_{13} \ldots + a_{1n}) + (a_{22} + a_{32} + \cdots + a_{n2})$$

.

.

.

$$a_{1n} = (a_{11} + a_{12} + a_{13} + \cdots + a_{1n}) + (a_{2n} + a_{3n} + \cdots + a_{nn})$$

Similarly, the elements of the second row are hidden using the following equations.

$$a_{21} = (a_{21} + a_{22} + a_{23} + \cdots + a_{2n}) + (a_{22} + a_{32} + a_{42} + \cdots + a_{n2})$$

$$a_{22} = (a_{21} + a_{22} + a_{23} \ldots + a_{2n}) + (a_{12} + a_{32} + \cdots + a_{n2})$$

.

.

.

$$a_{2n} = (a_{21} + a_{22} + a_{23} + \cdots + a_{2n}) + \left(a_{12} + a_{22} + \cdots + a_{(n-1)2}\right)$$

Similarly, all the elements of the original table are hidden using this method. For example, the first element of the original table is 1. The other elements in its row are 2,3,4,5 and 6. The summation is 1+2+3+4+5+6=21. The other elements in its column are 7 and 13. The summation of these values with the row sum value is 21+7+13=41. The first element of the secured table will become 41. In this way, all the elements of the table are hidden to provide the first level of security. Table 2 shows the values of the secured table.

Table 2 Secured Tables

	Column1	Column2	Column3	Column4	Column5	Column6
Row1	41	43	45	47	49	51
Row2	71	73	75	77	79	81
Row3	101	103	105	107	109	111

Here, the secured table is obtained here for providing security using the aggregation algorithm.

ii. Data partitioning

The second level of the proposed security model consists of two techniques such as data partition and data storage. This section provides the detailed functionality about this model. Hence, it discusses about the techniques used for data partition.

Data partition: The security module involves the partitioning agent to partition the original medical data table R into two tables with primary attributes and the sensitive attributes respectively. At the end of this process, these two tables are stored separately in the cloud.

This subsection provides the steps of the proposed algorithm. Let R be the original relation with attributes are $a_1, a_2, \ldots . a_n \epsilon A$. The original table R is represented as $R(\ a_1, a_2, \ldots . a_n)$.

This table is partitioned with primary attributes and sensitive attributes using projection but with primary key attributes to be present in both partitions. Now, the primary attributes $a_1, a_2, \ldots . a_n$ are marked as $PA_1, PA_2, \ldots . PA_m \epsilon A$ and the sensitive attributes are marked as $SA_1, SA_2, \ldots . SA_p \epsilon A$. Now, if a_1, a_2, a_5, a_6 are primary attributes then they are renamed as PA_1, PA_2, PA_3, PA_4 and a lookup table is made to see their original names. Similarly, if a_3, a_4, a_7, a_8 and a_9 are sensitive attributes then they are renamed as $SA_1, SA_2, SA_3, SA_4, SA_5$.

Now, the relation R is split into two relations R_1 and R_2 using projection. Now,

$$R_1 = \left(\prod_{a_1,a_2,a_5,a_6}(R)\right) \quad (1)$$

$$R_2 = \left(\prod_{a_3,\, a_4,a_7,a_8\,,a_9}(R)\right) \quad (2)$$

R can be reconstructed (merged) using join of R_1 and R_2 as follows:

$$R = R_1 \text{ Join } R_2 \quad (3)$$

Now, the elements of $R_1 = (\, PA_1, PA_2, \ldots.PA_m)$ are transformed into $HR_1 = (\, PHA_1, PHA_2, \ldots.PHA_m)$ using Caesar cipher. That is, $HR_1 = E(R_1(K_1))$ where K_1 is the key for Caesar cipher [17]. The formula is used for encryption is $C = P + K$.

The elements of $R_2 = (\, PS_1, PS_2, \ldots.PS_p)$ are transformed in to $HR_2 = E(R_2(K_2))$ where K_2 is a key matrix to apply the Hill Cipher [17] encryption.

Data Partition Algorithm (DPA)

Input: R= $\{\, a_1, a_2, \ldots.a_n \,\epsilon\, A\}$
Output: Hidden (Encrypted) data of R in two tables namely HR_1 and HR_2.

```
Step 1: Read the data from R.
Step 2: Divide the aggregated table obtains from R is
        into two tables ($HR_1$ and $HR_2$) based on primary
        and sensitive attributes using projection.
Step 3: Rename the attributes using a lookup table.
Step 4: The elements of $R_1$ are hidden (encrypted) by
        applying the Caesar Cipher (William Stallings
        2013).
Step 5: The elements of $R_2$ are hidden (encrypted) by
using Hill Cipher (William Stallings 2013).
Step 6: Stored the encrypted data of R in two tables
namely $HR_1$ and $HR_2$.
```

The proposed data partition algorithm is used to perform the first level encryption over the original table. During the process, the original table is also partitioned into two tables based on the availability of important attributes whether primary or sensitive in the original table. Finally, the encrypted table is stored in the database.

iii. Data Merging

Data merging: This component is utilized by the system to achieve the dataset-level medical data access. With the authorization of the data owner, the data are decrypted separately using the keys used in Caesar cipher and Hill cipher. Now, they are merged into a single table using Join operation. The joined table applies the reverse aggregating process to get the original values from aggregated values.

The decryption process can be done as follows:

Let R_1 and R_2 is the input data for this data merging process. First, R_1 is decrypted by using the following formulae.

$$R_1 = D\,[HR_1(K_1)] \quad (4)$$

where K_1 is the key used by the Caesar cipher.
Here, the formula for decryption is P = C – K. Similarly, the table R_2 is also decrypted by using the following formulae.

$$R_2 = D\,[HR_2(K_2^{-1})] \quad (5)$$

Here, K_2^{-1} is a matrix which is the inverse of the matrix K^{-1} found using the formula.

$$K_2^{-1} = \frac{1}{|K_1|} Adj\,(K_1) \quad (6)$$

Using R_1 and R_2, the original table R is obtained using join operation using the equation 3. Finally, the merged (decrypted) data can be obtained by using this data merging technique.

Data Merging Algorithm (DMA)

Input : Rows of HR_1 and HR_2
Output: Original (Decrypted) data of R.

Step 1: Read the data from HR_1 and HR_2.
Step 2: Decrypt HR_1data using the equation (4).
Step 3: Decrypt HR_2data using the equation (5).
Step 4: Apply the Join operation for data merging using equation 3.
Step 5: Display the Joined and decrypted data of R (Original) to the data owner.

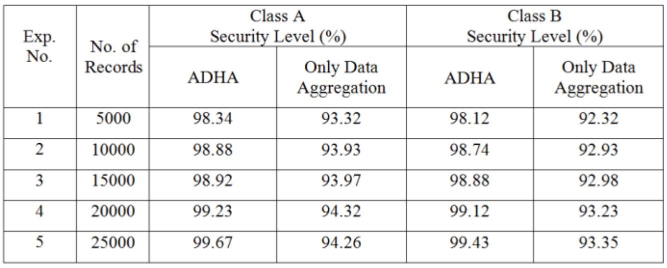
Here, the data are secured since the aggregation process is not revealed to the attackers. In addition, keys used for encryption are known only to the data owner.

III. RESULTS AND DISCUSSION

This research work has been implemented by using Java programming language. The various experiments have been conducted for evaluating the proposed model with different set of medical records. We have tested the proposed model with diabetic, cancer and heart disease affected patient original medical records which are collected from two levels of hospitals such as Class A and Class B.

Table 3 shows the performance of security level for the Class A hospital data and the Class B level hospital data. Here, we have considered the different number of records for carrying out the experiments such as 5000, 10000, 15000, 20000 and 25000.

Table 3 Security Level Analysis for the Class A & Class B level hospital data



Exp. No.	No. of Records	Class A Security Level (%)		Class B Security Level (%)	
		ADHA	Only Data Aggregation	ADHA	Only Data Aggregation
1	5000	98.34	93.32	98.12	92.32
2	10000	98.88	93.93	98.74	92.93
3	15000	98.92	93.97	98.88	92.98
4	20000	99.23	94.32	99.12	93.23
5	25000	99.67	94.26	99.43	93.35

From table 3, it can be observed that the performance of the proposed method which uses both encryption and aggregation is better than the existing model that uses only aggregation. Here, the performance is measured in terms of the percentage of accuracy. It is also observed that the proposed algorithm provides better results on both Class A level hospital data and Class B level hospital data.

Figure 1 shows the performance of the proposed secured storage model and the existing secured storage model (Muthurajkumar et al 2015).

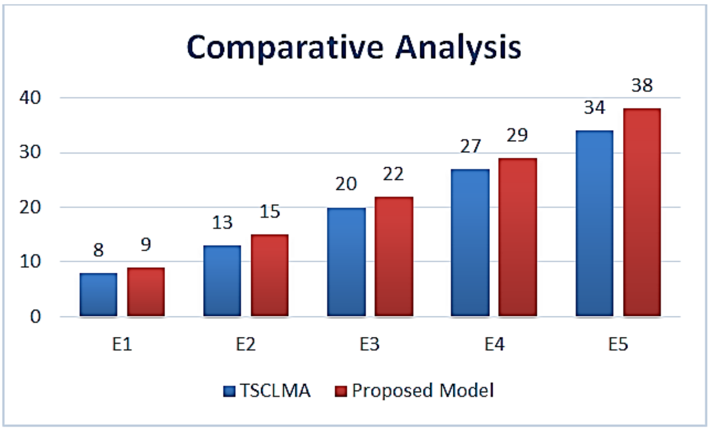


Figure 1. Comparative Analysis

From figure1, it can be observed that the performance of the proposed model is better than the existing model in terms of number of threats identified.

IV. CONCLUSION AND FUTURE WORKS

A new cryptography based secured storage model is proposed and implemented in this paper for effective data storage in cloud environment. This model uses three newly proposed algorithms namely Algorithm for Data Hiding based Aggregation (ADHA), Data Partitioning Algorithm (DPA) and Data Merging Algorithm (DMA) for effective storage. Experimental results show the better performance of the proposed model by using UCI Repository medical datasets. Future works in this direction could be the use of new intelligent agents for effective aggregation and merging process in cloud data storage.

V. REFERENCES

[1] Nissim Matatov, Lior Rokach, Oded Maimon, “Privacy-preserving data mining: A feature set partitioning approach”, Information Sciences, Vol. 180, pp. 2696–2720, 2010.

[2] Sudip Misra, Subarna Chatterjee, “Social choice considerations in cloud-assisted WBAN architecture for post-disaster healthcare: Data aggregation and channelization”, Information Sciences, Vol. 284, pp. 95–117, 2014.

[3] S Muthurajkumar, S Ganapathy, M Vijayalakshmi, A Kannan, “Secured Temporal Log Management Techniques for Cloud”, Procedia Computer Science, Vol. 46, pp.589-595, 2015.

[4] Piotr., K, Tysowski., M, & Anwarul Hasan. (2013). Hybrid Attribute- and Re-Encryption-BasedKey Management for Secure and Scalable Mobile Applications in Clouds. IEEE Transactions on Cloud Computing, 1(2), 172-189.

[5] Hongwei Li, Dongxiao Liu, Yuanshun Dai, & Tom H. Luan. (2015). Engineering Searchable Encryption of Mobile Cloud Networks: When QoE Meets QoP. IEEE Wireless Communications, 74-80.

[6] Lihui Lei, Sabyasachi Sengupta, Tarini Pattanaik, & Jerry Gao. (2015). MCloudDB: A Mobile Cloud Database Service Framework. 2015 3rd IEEE International Conference on Mobile Cloud Computing, Services, and Engineering, 6-15.

[7] Jinguang Hana, Willy Susiloa, & Yi Mu. (2013). Identity-based data storage in cloud computing. Future Generation Computer Systems, 29, 673–681.

[8] Ji-Jiang Yang, Jian-Qiang Li & Yu Niu. (2015). A hybrid solution for privacy preserving medical data sharing in the cloud environment. Future Generation Computer Systems, 43(44), 74–86.