International Journal on Recent and Innovation Trends in Computing and Communication
Volume: 5 Issue: 6

ISSN: 2321-8169
1422 – 1427

Implementation on Health Care Database Mining in Outsourced Database

Monali Gainkar[#1], Sonali Bodkhe[#2]
[#1,2]Department of Computer Science & Engineering
RashtrasantTukadojiMaharaj Nagpur University
Nagpur, Maharashtra, India
[#1]*monalimgainkar@gmail.com*,[#2]*sonali.mahure@gmail.com*

Abstract: Due to the EMR (Electronic Medical Record) system there will be a rapid growth in health data collection. As we have already discuss in previous review paper the different work of the health care data record for maintaining the privacy and security of health care most private data. Now in this paper we are going to implement sheltered and secretive data management structure that addresses both the sheltered and secretive issues in the managementor organization of medical datainoutsourceddatabases. Theproposed framework will assure the security of data by using semantically secure encryption schemes to keep data encrypted in outsourced databases. The framework also provides a differentially-private query or uncertainty interface that can support a number of SQL queries and complicated data mining responsibilities. We are using a multiparty algorithm for this purpose. So that all the purpose is to make a secure and private management system for medical data or record storage and accesses.

Keyword: *Outsourced database, data mining, multi-party algorithm, data encryption.*

*****

I. INTRODUCTION

Healthcare industries storean enormous amount of perceptive personal data, such as patient names, dates of birth, and personal medical records. Since healthcare data doubling every year, organizations need to provide in both hardware and software to store and manage large amount of data. Database outsourcing has gained significance in the past few years due to the appearance of the cloud computing. In Database-as-a- Service (DaaS), which is a sort of cloud computing services, the database administrator outsources both databases and querying services to a cloud server and clients issue queries over the database to the cloud server. In this context, privacy is a most important test and it is necessary to satisfy main privacy requirements of database owners and clients. In the budding cloud computing archetype, data owners become progressively more aggravated to outsource their complex data management systems from confined sites to the commercial public cloud for great elasticity and financial savings. For the contemplation of users' privacy, susceptible data have to be encrypted before outsourcing, which makes valuable data utilization a very tough task. In this domain, cloud computing is an effective solution for healthcare companies to handle huge amounts of medical records. However, healthcare organizations face two technical challenges.

First, data outsourcing exposes sensitive healthcare data to un-trusted cloud service providers. Unauthorized access to sensitive medical records can have a significant negative impact on healthcare services. To ensure the confidentiality of the medical data stored on the cloud, it should depend on semantically-secure encryption schemes. Using semantically- secure encryption schemes, it must be infeasible for a computationally-bounded adversary to derive significant information about a message when given only the cipher-text and the corresponding public key. In this regard, the challenge is how to ensure data confidentiality while allowing query execution over encrypted data.

Second, driven by mutual benefits and regulations, there is a demand for healthcare organizations to share patient data with various parties for making inquiries purposes. Healthcare organization may allow data analysts (e.g., researchers) to execute aggregate queries and perform some data analysis tasks (e.g., classification analysis) on the database. In this regard, the challenge is how to support aggregate queries or complex data mining tasks on encrypted data while preventinginference attacks.

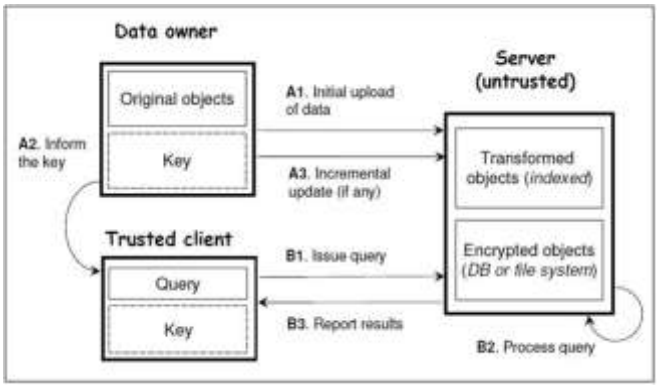

Figure 1.Architecture of proposed system

There have been a lot of research suggestions that independently address these two challenges. Most of the previous suggestions on secure outsourced databases suggest encrypting the data before moving it to the cloud. While encryption can provide data secrecy, it is of little use in deterring inference attacks. Similarly, there is a wide spread literature on private data analysis. However, all these proposals require access to unencrypted data to generate privacy-preserving answers and therefore do not satisfy the data confidentiality requirement. This reality demand privacy-enhancing technology that can simultaneously provide data confidentiality against an un-trusted database server, and prevent inference attacks from data analysts. Here proposing a general framework for secure and private data management in order to support effective data mining. The contributions of the approach are summarized as follows:

Based on real-life healthcare scenarios, first identify a new problem of secure and private data management of outsourced databases for data mining purposes.

• A new privacy-enhancing protocol that can provide data confidentiality against an un-trusted cloud server by using semantically-secure encryption schemes then extend the protocol to support aggregate queries or complex data mining tasks on encrypted data while preventing inference attacks.

• Taking decision tree learning as an example, show that it is possible to compute a classifier on the encrypted data. The computed classifier provides differential privacy guarantee to prevent an inference attack

II. RELATED WORK

By reviewing the previous work done by many people we are just taking into account there most precious work. One, in which the data owner and the data miner are two different entities, and another, in which the data is distributed among several parties who aim to jointly perform data mining on the unified corpus of data that they hold.

F. McSherry [1] has developed Data records that are protected from the data miner without enlightening original or secrete record information or data. Lindell and Pinkas [2] has the goal is to perform data mining while protecting the data records of each of the data owners from the other data owners and showed how to securely build an ID3 decision tree when the training set is distributed horizontally.

S.Barouti, D.Alhadidi and M. Debbabi [3] has presents protocols for executing keyword search and combined SQL queries that maintain the confidentiality of both the client and the database owner. N. Cao, Z. Yang, C. Wang, K. Ren and W. Lou [4] for the first time ever, define and solve the dilemma of privacy-preserving query over encrypted graph-structured data in cloud computing (PPGQ), and establish a set of authoritarian privacy necessities for such a secure cloud data deployment system to become a reality.

F.Chen and A.X. Liu [5] has reflect on a two-tiered sensor network structural design in which cargo space nodes gather data from in close proximity sensors and answer queries from the descend of the network. M. A. ALzain and E. Pardede [6] have suggested the design of a new reproduction suitable for NetDB2 architecture, known as NetDB2 Multi-Shares (NetDB2-MS). R. Mishra, D. P. Mishra [7] principally highlights some foremost security issues obtainable in current cloud computing surroundings. They aims at concurrently achieving data secrecy while still keeping the balancing relations intact in the cloud. His proposed system facilitates the data owner to assign most of computation intensive tasks to cloud servers without reveal data stuffing or user access right information.

III. PROBLEM DEFINATION

1) The data owner and the data miner are two different entities, and data is distributed among several parties whose aim is to jointly perform data mining on the unified corpus of data that they hold.
2) In the first setting, the goal is to protect the data records from the data miner. Hence, the data owner aims at anonymizing the data prior to its release.
3) The main approach in this context is to apply data perturbation.
4) Computation and communication costs versus the number of transactions *N* the perturbed data can be used to infer general trends in the data, without revealing original record information.
5) In the second setting, the goal is to perform data mining while protecting the data records of each of the data owners from the other data owners. This is a problem of secure multiparty computation.
6) The usual approach here is cryptographic rather than probabilistic.
7) The protocol that propose here computes a parameterized family of functions, which is threshold functions, in which the two extreme cases correspond to the problems of computing the union and intersection of private subsets is proposed on Healthcare system where human DNAs are consider. Those are in fact general-purpose protocols that can be used in other contexts as well.
8) Another problem while designing a healthcare data of secure multiparty computation that is the set inclusion problem.

IV. PROPOSED METHOD

From the above helpful and most accurate discursion on previous work done by a great people about

 ISSN: 2321-8169
1422 – 1427

outsourcing the data in encrypted form in cloud for some most important benefits and studying and grabbing idea related to database query processing help to build our system.

The aim is to propose a sheltered and confidential data management structure that addresses. The proposed framework makes certain the security of data by using semantically-secure encryption system to keep data encrypted in outsourced databases. In this scenario the data in the form of datasets are stored in database act as a container and datasets are health care data which is very sensitive data that the data owner or organization of data doesn't want to reveal to next un- authorized party while outsourcing data in the cloud.

The framework also provides a differentially-private query interface that can maintain a number of SQL queries and difficult data mining tasks. In this scenario client can query to the server and apart from all this server can copied only the information or sensitive data to client only as per query held by him not more than that information is provided by the server. Because the data is very sensitive and this data kept in protection of the server side.

The Framework will address both the security and privacy issues in the management of medical data in outsourced database. In the face of many benefits, data collected works and allotment have become a big concernas it intimidate individual privacy. The idea is to propose a secure and private data management framework that addresses both the security and privacy issues in the managementof medical data in outsourced database. The proposed framework ensures the security of data by using semantically-secure encryption schemes to keep data encrypted in outsourced databases. The framework also providesa differentially-private query interface that can support a number of SQL queries and complex data mining tasks.For Private access of Outsource data proposing a protocol which is based on two secure multi-party Algorithm

- One for computing the union (or intersection) of private subsets that each of the interacting players hold:
  In this a dataset of health care and perform classification our datasets by making unionof datasets or interaction of data sets and as a result of it the dataset weget is a common datasets and un-common datasets which is hold by a party or user which gone access thisdatabase.
- Another is a protocol that tests the inclusion of an element held by one player in a subset held by another:

After first step we are applying inclusion algorithm that can make a between the datasets held by one party with other one.

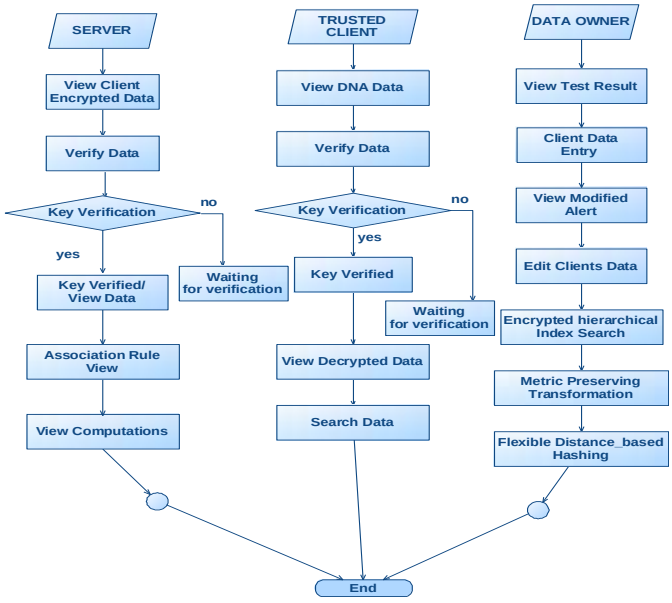

Figure 2. Flowchart of Proposed System

IV.1 Secure Multi-party algorithm:

- A set of parties with private inputs wish to compute some joint function of their inputs.
- Parties wish to preserve some security properties.
  E.g.,privacy and correctness.
- Security must be preserved in the face of adversarial behaviour by some of the participants, or by an external party.

Algorithm 1:

Range of the inputs known: (0, N)
Recap:V
A : public key e, private key dv
B: can access e, but not dvand

- Decd(Ence (x)) = x
- Decd (Ence (x)+y) = random looking thing (for someone who knows x,y,e but not d)

Step 1:

- A has i and B has j
- B generates a random x (of m bits)
- C = Ence (x)
- u =C- (j-1)
- Send u to A.

Step 2:

- A computes: for (t= 1 to N)
  ym = Decd (u+t)

 ISSN: 2321-8169
1422 – 1427

- Takes a prime p (of size about √m) and computes zi = yi mod p … for i = 1 to N
- p chosen such that |zm – zn | ≥ 2 for any m,n in [1 to N]

Step3:

- A send to B the following list: P, z1, z2, zi..... , (zi+1+1), (zi+2+1), …, (zN+1).
- Bob compares the jth entry of this list (excluding the prime p) with x mod p.
- If x mod p is = jth entry of the list implies i ≥ j.

IV.II RiJndael Algorithm

RiJandael Cipher is block cipher algorithm. It works iteratively. It uses 128 bit, 192 bit or 256 bit keys. Itencrypts 128 bit blocks. The purpose of selecting this algorithm as it provide security as well as efficient implementation both in hardware and software also it provide sufficient code length and memory utilization.

The 128 bit key is expanded as 32 bits words an array of 44 entries, for every round 4 distinct words are serve as a round key, key schedule relies on the S-box.

This algorithm consist of three layers

-Linear Diffusion

- Non Linear Diffusion

- Key Mixing

Algorithm 2:

State = X 1.

1. AddRoundKey (State, Key 0)
2. for r = 1 to (Nr - 1)
   a. SubBytes(State, S-box)
   b. ShiftRows (State)
   c. Mix Columns (State)
   d. AddRoundKey (State, Keyr)
   end for
1. SubBytes (State, S-box)
2. ShiftRows(State)
3. AddRoundKey(State, KeyNr)

Y = State

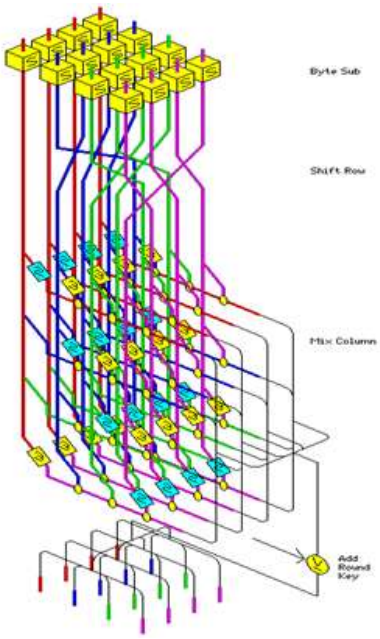

Figure 3. Basic RiJndael Algorithm Diagram

IV. EXPERIMENTAL RESULT

The proposed algorithm implemented in Microsoft Visual Studio 2012 as stated earlier. At first the server is to be configured for stated environment. A database consisting of DNA entries where used to evaluate the effectiveness of the proposed method. The test set results are comprised of entries taken from patients registration as well as DNA entries of patients to demonstrate the robust matching of the proposed description.

The procedure and screen shot of the project is as shown below. The complete result with output is demonstrated below. Each step has its own importance. In this chapter every step is explained with the screen shot. According to that the procedure goes on and according to input the result will show.

There will be a three important person in this project on health care data mining in outsources database.

1. Client
2. Owner
3. Server

- **Server configuration**

The Server Configuration is the main process of this project. The system have to connect with server. It requires some fields like server type, server name, Authentication, User name, Password. In this project the server type is of data engine. Server Name is System default sql server name and it contains windows authentication. The User name and password is required for connection of server.

- **Web config**

Web config is used to show the connection of database.Configuration file is used to manage various settings that define a website.Generally a website contains a single Web config file stored inside the application root directory. There are number of important settings that can be stored in the configuration file. Some of the most frequently used configurations, stored conveniently inside Web config file are:

- Database connections
- Caching settings
    - Session States
    - Error Handling
    - Security

- **Owner Login**

Owner is one of the important people in this project.Data Owner has the authority to change the clients DNA data can view client details,can do client entry.as well as the owner can search data.When client registered his data and submit it the data owner get all the client data as we can see in the screen

- **Client Login**

The client has to login first to view his entire details for that they enter their client Id and password. Then Client home page is appeared on screen. It contains client details like DNA Id, Client Id, and Testof Month and can view full details .DNA entries shows here in encrypted form.

- **Server Login**

The server is one of the most important people in this project. Server has to enter their login id and password for login. The server home page contains the client DNA encrypted result. For viewing the decrypted result server has to verify the DNA key. After verifying the DNA key server can check the DNA data

RESULT DETAILS

And encrypted data of each client DNA key is get by ID to verify each client data for security purpose. The entire association rule mining task is done by server side.And there will be three type of computation result will show for data owner, DNA entries, data verify in graphical format at the end.

Computation Result for data owner

The following graph shows the computational and communication costs result on the basis client's details.

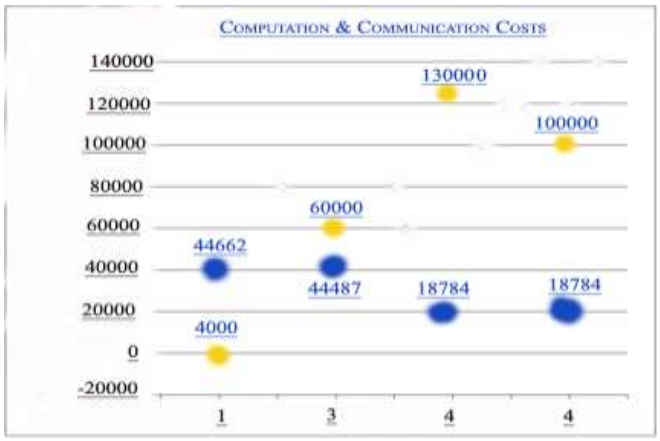

Figure 5. Computation Result for data owner

Computation Result for data Verification

The following graph shows computational result for data verification on the server side on the basis of considering client's details.

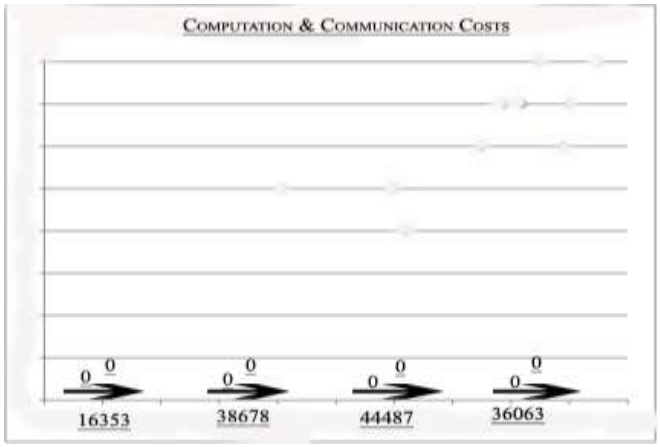

Figure 6. Computational Result For data Verification

Computation Result for DNA entries

The following graph shows the computational and communication cost entries for DNA entries base on DNA entries.

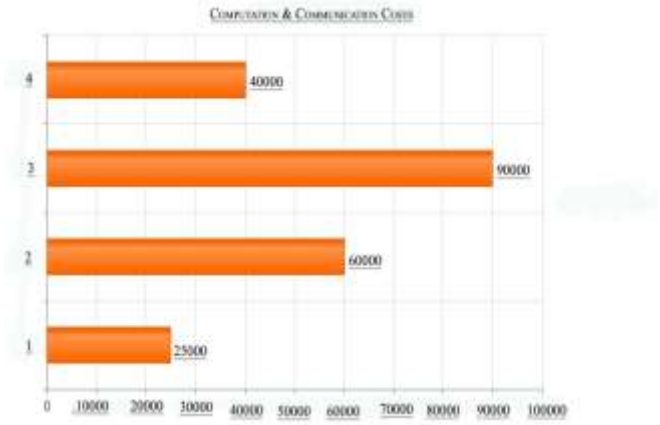

Figure 7. Computation Result for data entries

DNA Data Object Graph

The following graph shows DNA Data Object Graph. On X axis DNA Data entries and on Y axis no of objects shows.

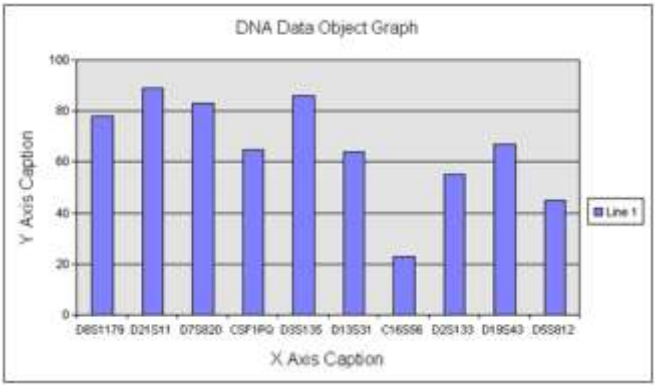

Figure 8.DNA Data Object Graph

V. CONCLUSIONS

The main goal is to make a system which provides a security and privacypreservingtaskin outsource database.A making query to database server and only that much information is provided to user.The proposed framework will ensures the security of data by using semantically-secure encryption schemes to keep data encrypted in outsourced databases of the health care database. This is more sensitive data for organization. The framework will also provide a differentially-private query interface that can support a number of SQL queries and complex data mining tasks. Multiparty Protocol will also guarantee the private access of outsourced data.

REFERENCES

[1] F.McSherry, “Privacy integrated queries,” in Proceedings of the 35th ACM International Conference on Management of Data (SIGMOD), 2009.

[2] Y. Lindell and B. Pinkas, “Privacy preserving data mining,” Journal of Cryptology, vol. 15, no. 3, pp.177–206, 2002

[3] S. Barouche, D. Alhadidi, and M. Debbabi, “Symmetrically private database search in cloud computing,” in Cloud Computing Technology and Science (CloudCom), International Conference on, vol. 1. IEEE, 2013, pp. 671–678.

[4] N. Cao, Z. Yang, C. Wang, K. Ren and W. Lou, "Privacy-Preserving Query over Encrypted Graph- Structured Data in Cloud Computing," 2011 31st International Conference on Distributed Computing Systems, Minneapolis, MN, 2011, pp. 393-402.

[5] F. Chen and A. X. Liu, "SafeQ: Secure and EfficientQuery Processing in Sensor Networks," 2010Proceedings IEEE INFOCOM, San Diego, CA, 2010, pp. 1-9.

[6] M. A. ALzain and E. Pardede, "Using Multi Shares for Ensuring Privacy in Database-as-a-Service," 201144th Hawaii International Conference on SystemSciences, Kauai, HI, 2011, pp. 1-9.

[7] R. Mishra, D. P. Mishra, A. Tripathy and S. K. Dash, "A privacy preserving repository for securing data across the cloud," 2011 3rd International Conference on Electronics Computer Technology, Kanyakumari,2011, pp. 6-10.

[8] B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, “Privacy preserving data publishing: A survey of recent developments,” ACM Computing Surveys, vol. 42, no. 4, pp. 1–53, June 2014.

[9] C. Dwork, “Differential privacy,” in Proceedings of the International conference on Automata, Languages and Programming (ICALP), 2012.

[10] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” in Proceedings of the 3rd conference on Theory of Cryptography (TCC), 2008.

[11] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed. Morgan Kaufmann Publishers Inc., 2011.

[12] M. D. Berg, O. Cheong, M. V. Krefeld, and M.Overmars, Computational Geometry: Algorithms and Applications, 3rd ed. Springer-Verlag TELOS, 2008.

[13] A.Frankand A. Asuncion, “UCI machine learning repository,” 2010. [Online]. Available: http://archive.ics.uci. edu/ml

[14] Z. Zhu and W. Du, “Understanding privacy risk of publishing decision trees,” in Proceedings of the 24th Annual IFIP WG

[15] Working Conference on Data and Applications Security and Privacy (DBSec), 2010.

[16] G. Jagannathan, K. Pillaipakkamnatt, and R. N. Wright, “A practical differentially privaterandon decisiontreeclassifier,” Trans.DataPrivacy,vol.5,no.1,pp.27