ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

Preserving Privacy using Secret Sharing in Distributed Environment on Perturbed Data

Sagar S. Phake [1], Vikrant S. Moon [2], Avinash A. Waghmare [3], Gaurav B. Ijare [4]

Graduate Student, Department of Computer Engineering, Pimpri Chichwad College of Engineering, Pune, India[1]

Graduate Student, Department of Computer Engineering, Pimpri Chichwad College of Engineering, Pune, India[2]

Graduate Student, Department of Computer Engineering, Pimpri Chichwad College of Engineering, Pune, India[3]

Graduate Student, Department of Computer Engineering, Pimpri Chichwad College of Engineering, Pune, India[4]

ABSTRACT: With the rapid growth in networking technology, hardware and software there is large amount of data collection. Organizations take large volumes of data from heterogeneous databases. This data is huge and contains sensitive data. Organizations data contains private information. The data mining takes patterns from sensitive data used for decision making. There may be chance of misusing information without knowledge of actual data owner.
Privacy preserving with secret sharing (PPSS) technique gives new direction to solve this problem. PPSS maintains privacy for data. Secret sharing takes keys for providing privacy to individual data. After the modification data is done in such a way that original data remains private even after secret sharing process.

In this paper we have proposed a framework that allows simple transformation of original data using encryption and secret sharing technique and providing security for keys used for encryption by secret sharing while transformation of data. Proposed technique increased the security and maintaining privacy for data as well as key. Using this approach we can achieve more confidentiality at client as well as data owner sites in distributed environment.

KEYWORDS: Data Mining, Cryptography, Secret Sharing, Data perturbation, Privacy.

I. INTRODUCTION

Data mining is a[15] powerful tool that can find and collect previously unknown patterns from large amounts of data. Data mining process collects large amount of data. Recently the application of data mining is increased in various domains like business, academic, communication, bioinformatics and medicine. Organizations are extremely dependent on data mining in results to provide better service, achieving good and better profit, and better decision-making. Privacy can for instance be threatened when data mining techniques uses the identifiers which themselves are not very sensitive, but are used to connect personal identifiers such as addresses, names etc., with other more sensitive personal information. The simplest solution to this problem is to completely hide the sensitive data. But this solution is not correct in many applications, like medicine research, DNA research etc. Privacy preserving using secret sharing is a special privacy maintaining technique which has emerged to protect the privacy of sensitive data and also give valid data mining results. In this project the total process is divided into three components the customer, mediator and a group of service data providers. At any time the role of service data provider and customer can be interchanged. The role of mediator is purely passive; it keeps the record of no. of data providers and transfers the query from customer to data providers. Initially there is no communication between customer and data provider. When the client sends a query, the mediator send the information to all the data holders and through exchange of acknowledgements, the mediator establishes the connection with data providers.

The rest of the paper is organized as follows. In Section 2 we provide some background information in Literature Survey and also provide some techniques which provide the more security. Section 3 the project framework, RSA algorithm and our PPSS algorithm is mentioned. The analysis is presented in Section 4. The paper concludes with a discussion of future work.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

1.1 Previous work

The data mining results not only gives the valuable information hidden in these databases, but sometimes also reveals private information about individuals. This data mining results have to maintain privacy for information. The true problem is not data mining, but the way data mining is done. PPSS is technique in data mining where privacy and data mining can coexist. It gives the results without any loss of privacy. In previous PPDM approach Data transformation and Cryptographic method approaches are used. The set of protocols used in cryptographic technique is called as security multiparty communication (SMC).In SMC the participating parties know only the final result of the computation and no additional information is given during the computation. Perfect privacy in the SMC is achieved because any important information is not released to any third party.

1.2 Our Contribution

In this framework the total process is divided into three components the client, mediator and a group of service data providers. At any time the role of service data provider and client can be interchanged. The role of mediator is purely passive. A mediator keeps the record of number of data providers and sends the query from client to data providers.

Initially there is no communication between customer and data provider. When the client sends a query, the mediator send the information to all the data holders and through exchange of acknowledgements and maintain data privacy, the mediator establishes the connection with data providers. [1]

II. LITERATURE SURVEY

Recently the application of data mining is increased in various domains like business, communication, and medicine. The data mining results not only gives the valuable information hidden in these databases, but sometimes also reveals private information about individuals. PPDM is an emerging technique in data mining where secret sharing and data mining can coexist. In general there are two main approaches in PPDM:

i) Data transformation based

ii) Cryptographic-based methods

The data transformation based approach modifies sensitive data in such a way that it loses its sensitive meaning.

2.1 Randomization Perturbation Technique

In randomization perturbation approach the privacy of the data can be protected by perturbing sensitive data with randomization. In this method privacy of confidential data can be obtained by adding small noise component which is obtained from the probability distribution.

In a set of data records denoted by X = {x1 . . . x *N}*. We add a noise component which is drawn from the probability distribution. These noise components are drawn independently, and are denoted y1 . . . yN. Thus, the new set of distorted records are denoted by x1 + y1 . .x*N* + y *N*. We denote this new set of records by z1 .. z*N*.

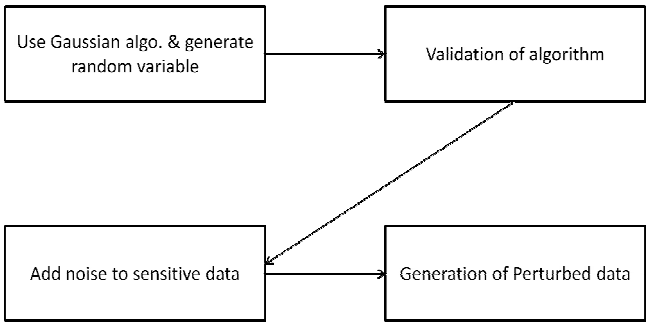

Fig. 2.1 Generating perturbation technique

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

2.2 Secret Sharing Technique

Cryptographic technique performs traditional tasks of encryption and authentication to protocols for securely distributing computations among a group of mutually distrusting parties. In this situation there is one additional party called "trusted party" who does not deviate the activities performed by him. All other parties sends their inputs to trusted party, then trusted party computes the function and send back the appropriate results to parties. It does not leak any information in this process.

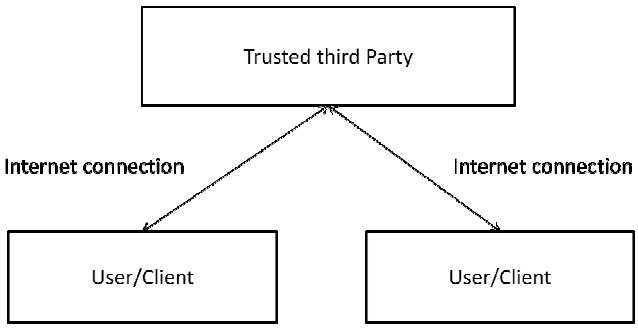

Fig. 2.2 Implementation of trusted third party

In secure multiparty computation a set of N parties with their private inputs x1,…,xn on a network can compute a joint function of their inputs. This joint computation should have the property that the parties learn the correct output y = f(x1,…,xn)

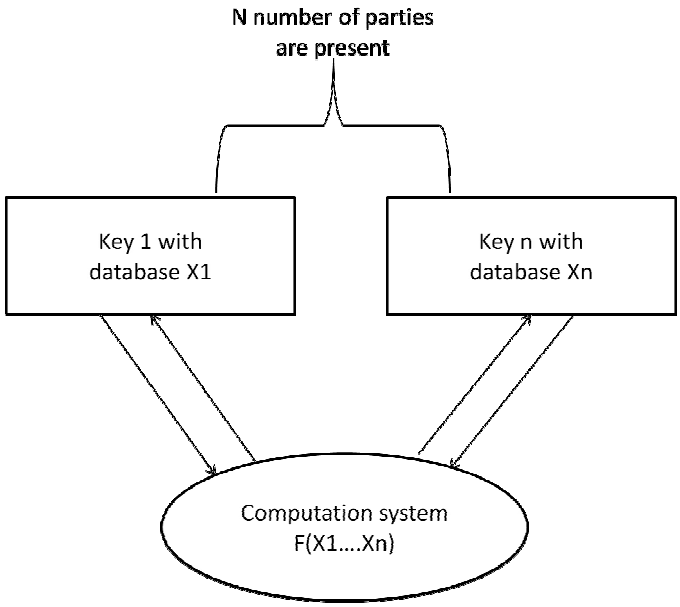

Fig. 2.3 Secure multiparty computation system

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

III. PROPOSED ALGORITHM

Step 1: To obtained available information Client send query to Server.
Step 2: Server will check the requested data is available on which Client in the database.
Step 3: If data is available then give reply to the requested client (i.e. Data is available).
Step 4:Then the another client on which data is available performing encryption on data and generate 3 shares of key. This encrypted data with one share of key will give back to requested client through server.
Step 5: To decrypt the data requested client will request for second share to server. This request send to another client on which data available.
Step 6: Client will check that the request client is authenticated by server if yes then second share is directly send to requested client.
Step 7: After receiving second share client will reconstruct key.
Step 8: Client decrypts the data using its received key and obtains the required information in a consolidated form.

IV. FRAMEWORK

4.1 Client

Client is one who sends a query to mediator to obtain the number of actual data providers. Client wants the available information of data providers which gives the required information.

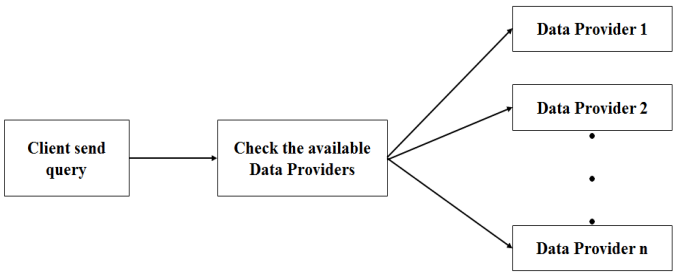

Fig. 4.1 Checking Availability of Data Providers

4.2 Mediator

Mediator is the trusted party which keeps the information about the available data providers. It works as the mediator between client and the data provider. It also generates the secret key for maintaining the security.

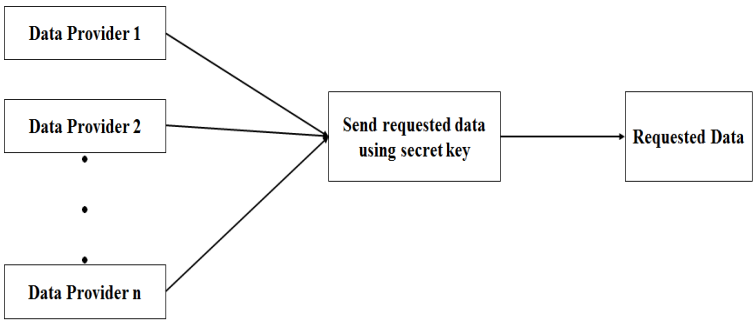

Fig. 4.2 Send requested data

4.3 Data Provider

Data Provider is one who stores the sensitive information about data. It retrieves the key from mediator and sends requested data to the mediator. Mediator will send back the data to client.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

During this process privacy is maintained at both the client and the service provider's location. The privacy is achieved because the customer who is query generator doesn't know the exact details of data providers who actually contributed the results except the value N i.e. total no. of data owners. Similarly the data owner don't know which key out of given set of keys have been selected by the customer for encryption purpose. The result obtained by the client is also in the perturbed form in which small amount of noise is added to sensitive data such that the properties and the meaning of the original data is not changed but privacy is maintained.

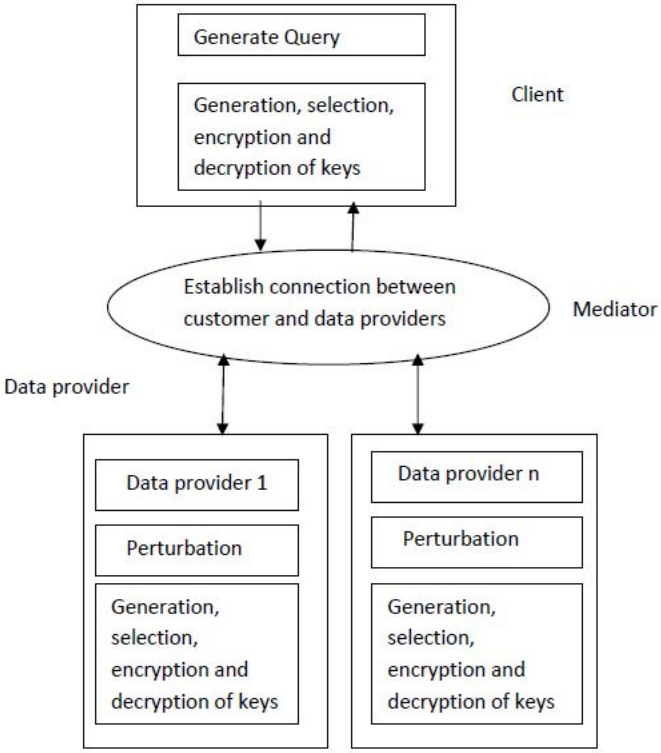

Fig. 4.3 Framework for sharing the data using secret sharing
And Perturbation technique [1].

V. TECHNICAL SPECIFICATION

5.1 Advantages

5.1.1 To discover and distribute the databases, without compromising the privacy of the individual's data.
5.1.2 The confidentiality is guaranteed among the distributed sites because the details of data owner are hidden from
5.1.3 client, and the data owner is completely unaware of the selection of secrets by client during process.
5.1.4 The unauthorized groups gain no information about the secret are referred to as a perfect have proved using the concept of entropy.
5.1.5 Providing secrets makes more privacy to confidential data.
5.1.6 Preventing loss of secret information and attacks.

5.1.6 The confidentiality is guaranteed among the distributed sites because the details of data owner are hidden from client, and the data owner is completely unaware of the selection of secrets by client during process.
5.1.7 To discover and distribute the databases, without compromising the privacy of the individual's data.

Secret sharing provides reliability for reconstruction of key.
Preventing loss of secret information and attacks. Providing secrets makes more privacy to confidential data.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 3, March 2015

5.2 Disadvantages

5.2.1 It is mandatory to have computer system with all users of this product along with internet connectivity.

5.2.2 If intermediate nodes will enter the wrong data in the database, then the decision making will affect.

5.3 Applications

5.3.1 This system connects different major areas like databases, artificial intelligence and banking application.

5.3.2 This type of system is use in distributed environment for maintaining security on sensitive data.

VI. CONCLUSION

Data mining is a powerful tool that can find and collect previously unknown patterns from large amounts of data. Data mining process collects large amount of data. Organizations are extremely dependent on data mining in results to provide better service, achieving good and better profit, and better decision-making. Privacy preserving using secret sharing is a special privacy maintaining technique which has emerged to protect the privacy of sensitive data and also give valid data mining results. The ever increasing ability to identify and collect large amounts of data, analysing the data using data mining process and decision on the results gives prospective benefits to organizations. Where each data owner has its own data and also wants to share the data with other data owners, but at the same time want to preserve the privacy of sensitive data in the data records. The confidentiality is guaranteed among the distributed sites. The details of data owner are hidden from client and data owner is completely unaware of the selection of keys by client during cryptographic process.

REFERENCES.

1. Agrawal R. Srikant R, "Privacy Preserving Data Mining", In the Proceedings of the ACM SIGMOD Conference. 2000.
2. K.Muralidhar., R.Sarathi, "A General additive data perturbation method for data base security" , journal of Management Science.
3. Agrawal D. Aggarwal C.C. "On the Design and Quantification of Privacy Preserving Data mining algorithms." ACM PODS Conference, 2002.
4. R.Agarwal and R.Srikant, "Privacy preserving data mining", In Proceedings of the 19th ACM SIGMOD conference on Management of Data, Dallas, Texas, USA, May2000.
5. Y. Lindell and B. Pinkas "Privacy preserving data mining". In Advances in Cryptology - crypto2000, Lecture Notes in Computer Science, volume 1880, 2000.
6. J. Vaidya and C. Clifton, "Privacy preserving association rule mining in vertically partitioned data". In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, July 23-26 2002.
7. Kargupta. H., Datta, S.Wang, and Siva Kumar. K. "On the privacy preserving properties of random data perturbation techniques", Proc. of Intl. Conf. on Data Mining (ICDM) , 2003.