Neha Gosai, S.H.Patil / International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 www.ijera.com
Vol. 2, Issue 4, July-August 2012, pp.856-861

Privacy Conserving Approach to Confidential Database

Neha Gosai*,S.H.Patil**
*Department of Computer engineering, Bharati Vidyapeeth University,Pune, Maharashtra-411043
**Department of Computer engineering, Bharati Vidyapeeth University,Pune, Maharashtra-411043

ABSTRACT

Suppose Bank has private data of each account holder which contains account number, details of account hoder.Now bank wants to share data with researchers for some specific purpose in such a way that privacy of individual account holder not violated by disclosing his data to researchers. If allow account holder to directly add data into the database of researcher it will violate confidentiality of research database and if allow researchers to directly read the contents, violates privacy of account holder. To preserve privacy and confidentiality we have proposed approach on suppression based k-anonymous method to protect privacy of individual.

Keywords- **Privacy, Anonymous, Suppression, Confidentiality**

1. INTRODUCTION

Today society is growing in terms of data collection, containing personal details or some other confidential information. Because, use of computers increasing, its privacy or security becomes crucial. Database modelled as collection of data that can be accessed, updated and it enables user to retrieve data. To provide security to these databases is big issue. For example, Medical data of each patient should be protected for years. There are different approaches to protect database. The emphasis in database privacy should fall on a balance between confidentiality, integrity and availability of personal data, rather than on confidentiality alone. Database privacy concerns the protection of information about individuals that is stored in a database. Sometimes, without your knowledge, health records used by insurance companies, drug manufacturing companies. Because medical records may contain some sensitive information like effect of drug on patient it is important to keep this information private. Confidential data is personal information relating to a person, it could be reference to an identification number or other factors like social identity. To maintain confidentiality, unauthorized third parties must be prevented from accessing and viewing medical data. It is also essential to maintain database integrity while data is transferring from source to destination. Confidentiality is achieved by Cryptography methods or tools. Though data is anonymized, Confidentiality is necessary.Anonymized or anonymization means remove personal identifier to protect private information. There are many ways of anonymization but we will focus on k-anonymization approach only.

Data anonymization enables transferring information between two organizations, by converting text data into non human readable form using encryption method[1].There have been lots of techniques developed.K-anonymization is one of the approach[2].his technique protects privacy of original data by modification. So problem arises at this point where database needs to be updated. So when tuple is to be inserted in the database problems occurs relating to privacy and confidentiality that Is database owner decide that whether database preserve privacy without knowing what new tuple to be inserted. In this paper, we propose a protocol called Suppression based approach to solve above problem.

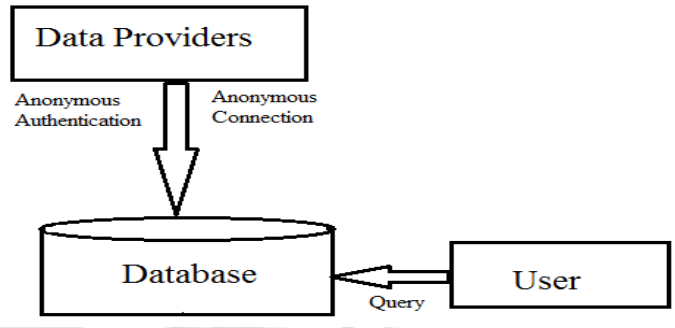

Fig 1 Anonymous Database

Fig 1 shows approach to update anonymous database. Suppose Alice is user who owns the database and data provider that is Bob who wants to insert his own tuple.So it is necessary to check whether it is possible to update database or inserting tuple without knowledge of Bob so that privacy of Bob cannot be violated and confidentiality of database does not violated if Bob have access to the contents of database.

2. PROBLEMS WITH CURRENT APPROACH

There are various techniques to provide confidentiality and privacy to anonymous database like Data Reduction, Data perturbation and Secure Multiparty Computation etc.

The first approach is Data perturbation technique derived from secure database techniques to overcome privacy preserving problem. It is effective application to protect health care system.

Basically there are two types of data perturbation. First type Probability distribution approach and the second type are called the value distortion approach. In the probability distribution, Original database is replaced by sample from distribution or by distribution itself [4] and the value distortion approach perturbs data elements or attributes directly by either additive noise, multiplicative noise, or some other randomization procedures [5]. Agrawal et al. [6] proposed a value distortion technique to protect the privacy by adding random noise from a Gaussian distribution to the actual data. They showed that this technique appears to mask the data while allowing extraction of certain patterns like the original data distribution and decision tree models with good accuracy.

Second research approach is Secure Multiparty Computation method consider problem of evaluating function of two or more parties' secret input in such a way that each party does not get anything else except specified output. Secure computation was formally introduced in 1982 by A. Yao This concept is important in field of cryptography. For example two parties having some secret information-a and b respectively. They compute joint function $f(a,b)$ without disclosing information about a and b .Main aim of using SCM is to allow maximum use of information without compromising user privacy .However, computational complexity makes this approach infeasible for large dataset. There are different tool available for large dataset [8].

The third approach is private data retrieval related to SMC. It focuses on Queries for retrieving data from database. But still it doesn't concern with privately updating database. Other techniques have been developed that is data anonymization, which protects data through suppression or perturbation in stastical database. Sweeny who proposed concept of k-anonymity[9].But after researching on algorithm result comes out that it still not resolves problem of privately update of database.

The fourth approach is Query Processing Techniques for encrypted data [10].These approach provide whole data to client, though it encrypts all data. But this is not concern to our approach. Most of privacy models developed are based on k-anonymity property-anonymity property deals the possibility of indirect identification of records form public databases-anonymity means each released record has at least (k-1) other records in the release whose values are indistinct[15]. K-anonymity and SMC are used in privacy-preserving data mining, but they are quite different in terms of efficiency, accuracy.

3. PROPOSED PROTOCOL

The protocol Suppression based method for anonymous database allows the owner of database DB to anonymize tuple t, without knowledge of content of tuple and without sending tuple to owner [11].To achieve this goal two party exchange message by encrypting. Or say by anonymous connection using protocol like Crowds [12].Crowd protocol is anonymity protocol which hides each user's communication by routing them randomly within group of similar users. This is necessary because attacker can easily reveal IP address and get sensitive information. It can be leaked from access control policies, so it is necessary to provide authorization so that only authorized party can access data. This is based on user anonymous authentication [13].Our problem is to privately updates of anonymous and confidential databases.

3.1 Prototype Architecture

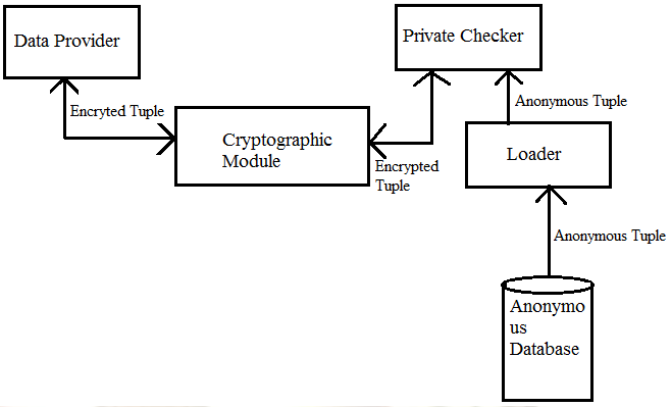

Fig 2 Proposed System

In fig 2 There are different modules shown:Cryptographic Module,Private Checker,Loader Module.Job of cryptographic module is to encrypt all tuples between Data Provider and Private checker.Loader module read and transfer tuple from Anonymous database. Private checker module performs all the controls that is it checks whether the inserted data is matched with the data's in the anonymous database using suppression method. The main concept behind private checker is to check whether insertion is possible into the k anonymous database.

3.2 Suppression Based Method to Update Confidential Databases

Consider Table T ={t1,…,tn} over the attribute set A. The idea of this algorithm is mask some attributes by special value *,the value employed by Alice for the anonymization.

Table 1 Original Dataset

AREA	POSITION	SALARY
Networking	Professor	12,000
Networking	Professor	20,000
Programming Language	Professor	25,000
Programming Language	Professor	20,000
Database System	Assistant	50,000
Database System	Assistant	40,000

Table 2 Suppressed Data with k =2

AREA	POSITION	SALARY
*	*	*
*	*	*
Programming Language	Professor	*
Programming Language	Professor	*
*	*	*
*	*	*

In suppression based method, Every attribute is suppressed by *.So third party cannot differentiate between any tuples. Table 2 shows Suppressed attributes or data with k= 2.Here,k=2 indicates k-anonymity that is each row in the table cannot be distinguished from at least other k-1 rows by only looking a set of attributes[9].

3.3 Cryptography Assumption

Suppression based k-anonymity protocol uses encryption scheme of commutative and product homomorphic. This encryption scheme allows performing mathematical operation over encrypted data. We provide definition of Commutativity, product homomorphic and indistiguishability [14].

Given a finite set K of keys and a finite domain, a commutative, product homomorphic encryption scheme E is a polynomial time computable function $E: K \times D \rightarrow D$ satisfying the following properties:

1) Commutativity: For all key pairs $k1, k2 \in K$ and value $d \in D$,the following equality holds:

$$Ek1\big(Ek2(d)\big) = Ek2(Ek1(d))$$

2) Product Homomorphism: For every $k \in K$ and every value pairs $d1, d2 \in D$,the following equality holds:

$$Ek(d1) \cdot Ek(d2) = Ek(d1 \cdot d2)$$

3) *Indistinguishability:* It is infeasible to obtain data of plaintext from cipher text. The advantages are high privacy of data even after updating, and an approach that can be used is based on techniques for user anonymous authentication and credential verification.

The Diffie Hellman key exchange algorithm allows two users to establish shared secret key over insecure communication without having any prior knowledge. Here, Diffie Hellmen is used to agree on shared secret key to exchange data between two parties. Here, we have assumed that database is k-anonymous. So it needs to check that after inserting properly anonymized tuple, by Bob, whether database (Alice) maintains its k-anonymity. If this is the case, tuple can be inserted otherwise tuple will be rejected.AES (Advance Encryption Scheme) is symmetric key algorithm, means same key used for encryption and decryption, for encryption of data.

4. ALGORITHM

In suppression based method of anonymous database, our main aim to compute anonymized version of tuple t without letting Alice and Bob know about the contents of tuple t and what are the suppressed attributes in tuple t. Suppose, Alice and Bob agree on the commutative and product homomorphic encryption scheme. Steps are described as below:

1) Alice encrypts tuples with her private key and sends it to Bob.
2) Upon receiving encrypted version of tuple Bob it again encrypts tuple with his key send it to Alice.
3) Alice decrypts values to get Bob's key.
4) Alice examines if the suppressed attributes of tuple is equal to the tuple sent by Bob. If yes then insert tuple in database.

4.1 Implementation

Suppression based k-anonymity approach to provide privacy updates to confidential database is designed by using Java. The implementation setup Considered attributes AREA, POSITION and SALARY. We have considered POSITION as non suppressed attribute. Figure 2 shows home page of Suppression based approach. Data entered by the user directly replaced by special value '*' and these values being inserted into table. To carry out this task, we have made separate table for original values. When user enters data it checks value in original table if it is valid then it replaces original value with suppressed values. Based on this outcome data will get inserted or rejected as shown in snapshot below.

Fig 3 Attributes with wrong data value

Fig 4 Output screen of wrong data values

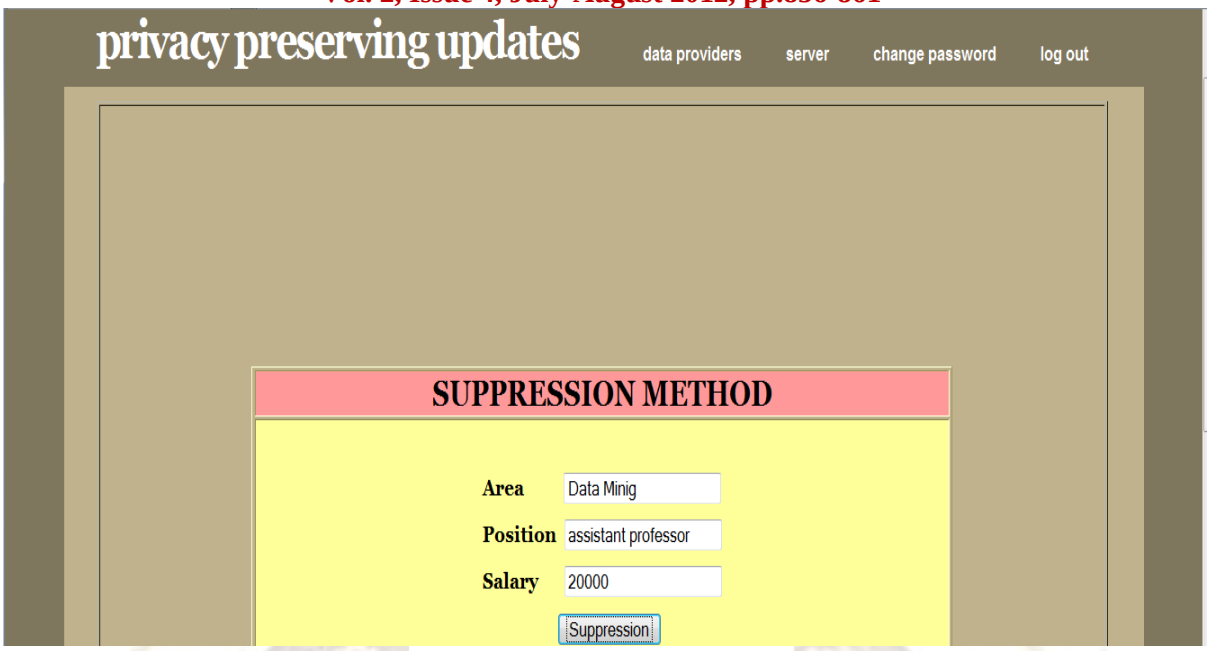

Fig 5 Attributes with correct Values

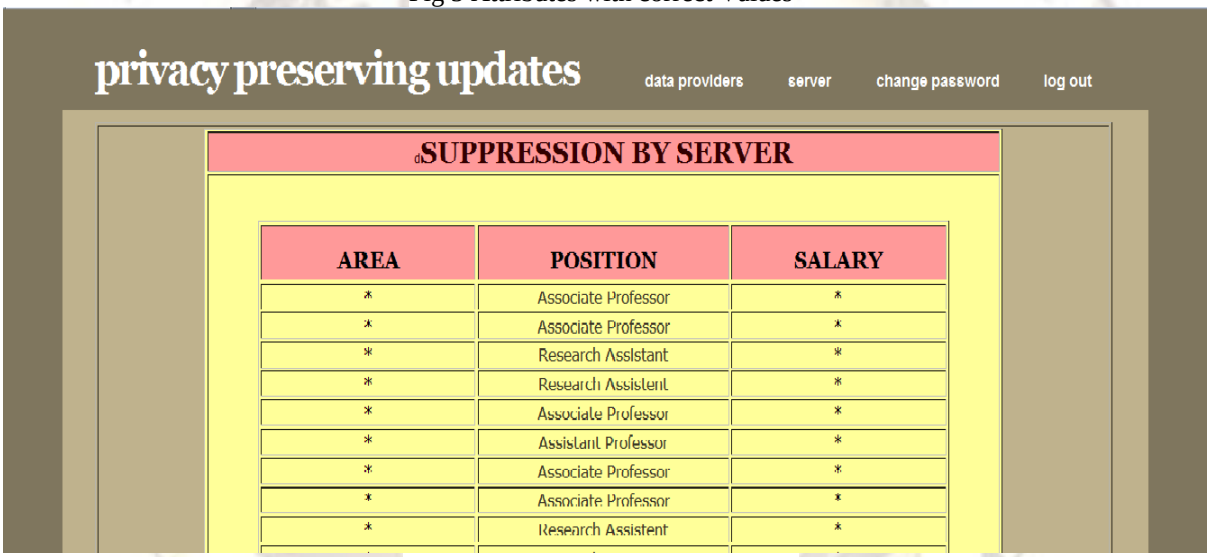

Fig 6 Output screen of correct values

4.2 Result

From the implementation, we can say that the complexity of protocol depends on number of message exchanged and their size. The complexity of protocol depends on the size of T. We have used java as front end and My SQL for database. Experiment executed on Pentium 1GHz with 1 GB physical memory. We can make result that if all values entered by data provider is correct then database will be updated successfully otherwise tuple will not be inserted to the database. Thus we can say that database successfully updated while preserving privacy and k-anonymity.

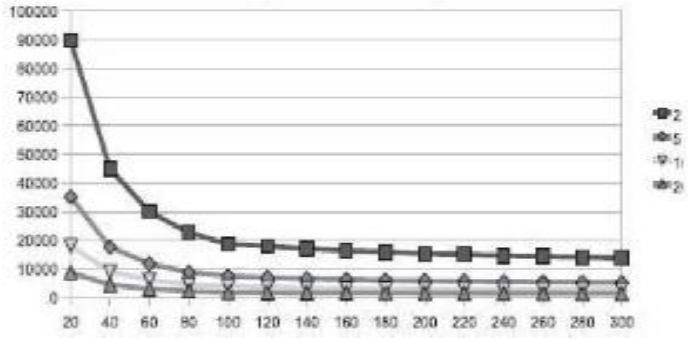

Fig 7 Execution time of Suppression based protocol

Above figure shows average execution time of the protocol. From the experiment we can say that as k increase tuple is safely inserted into the database. Execution time depends on the size of

k.Execution time increases according to the equation dataset size/*k*.

5. ACKNOWLEGDEMENT

I must thank, first and foremost, my guide and Head of the Department, Dr. prof. S. H., Department of Computer Science and Engineering, who gave me opportunity to write this paper without his guidance and patience, this dissertation would not be possible. Finally, I thank him to review my paper and his invaluable suggestion that make to improve quality of my paper.

6. CONCLUSION AND FUTUREWORK

In this paper, we have proposed secure protocol to check that if new tuple is being inserted to the database, it does not affect anonymity of database. It means when new tuple get introduced, k-anonymous database retains its anonymity. Database updates has been carried out properly using proposed protocol. Execution shows that once system verifies user's tuple, it can be safely inserted to the database without violating k-anonymity. Only user required to send non suppressed attributes to the k-anonymous database. This is useful in medical application. Suppression based method is not fully sufficient as if a tuple fails to check, it does not insert to the database and wait until k-1.because of longer process and waiting time.

The important issues in future will be resolved:

1) Implement database for invalid entries.
2) Solve problem of anonymity when initially table is empty.
3) When system fails to check tuple, it checks these tuple in wait state called hanging tuples.try to resolve this problem.
4) Improving efficiency of protocol in terms of number of messages exchanged between user and database.
5) Implement real world database system.

7. REFERENCES

[1] U.S. Department of Justice, Privacy Technology Focus Group Final Report, IJIS Institute.

[2] P. Samarati. Protecting respondent's privacy in micro data release. IEEE Transactions on Knowledge and Data Engineering, vol. 13, no.6, pp. 1010–1027, Nov/Dec. 2001.

[3] Agrawal and Srikant, 2000; Rizvi and Haritsa,2002; Evfimievski *et al*., 2004

[4] C.K. Liew, U.J. Choi, and C.J. Liew, "A Data Distortion by Probability Distribution," ACM Trans. Database Systems (TODS), vol. 10, no. 3, pp. 395-411, 1985.

[5] N.R. Adam and J.C. Worthmann, "Security-Control Methods for Statistical Databases: A Comparative Study," ACM Computing Surveys (CSUR), vol. 21, no. 4, pp. 515- 556, 1989.

[6] R. Agrawal and R. Srikant, "Privacy preserving data mining," in Proceedings of the ACM SIGMOD Conference on Management of Data, Dallas, TX, May 2000, pp. 439–450.

[7] Andrew C. Yao, Protocols for secure computations, University of California Berkeley, California 94720, 1982.

[8] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Zhu, "Tools for Privacy Preserving Distributed Data Mining," ACM SIGKDD Explorations, vol. 4,no.2, 2003.

[9] L. Sweeney. K-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10(5), 557–570, 2002.

[10] D. Boneh, G. di Crescenzo, R. Ostrowsky, G. Persiano. Public key encryption with keyword search. In Proc. of Euro crypt Conf.,Interlaken, Switzerland, 2004

[11] Privacy-Preserving Updates to Anonymous and Confidential Databases ,Alberto Trombetta, Wei Jiang, Elisa Bertino and Lorenzo Bossi, Department of Computer Science and Communication, University of Insubria, Italy.

[12] M. K. Reiter, A. Rubin. Crowds: anonymity with Web transactions.ACM Transactions on Information and System Security (TISSEC), 1(1), 1998; 66–92.

[13] J. Li, N. Li, W. Wins borough. Policy-hiding access control in open environment. In Proc of ACM Conf. on Computer and Communications Security (CCS), Alexandria, Virginia, 2005.

[14] S. Brands, Untraceable off-line cash in wallets with observers. In Proc. Of CRYPTO onf. Lecture Notes in Computer Science, 773, 1994; 302-318.

[15] Generalization Based Approach to Confidential Database Updates , Neha Gosai, S H Patil, Department of ComputerScience,pune,Maharashtra,2012

[16] www.wikipedia.com/wikifiles/.