# Security Preservation Methods to Confidential Databases

**Neha Gosai, S. H. Patil**

***Abstract- Anonymization means to remove personal identifier or converted into non readable form by human to protect private or personal information. Data anonymization can be performed in different ways but in this paper k-anonymization approach is used. Suppose one person A having his own k-anonymous database and needs to determine whether database is still k-anonymous if tuple inserted by another person B. For some applications (for example, Student's record), database needs to be confidential, So access to the database is strictly controlled. The confidentiality of the database managed by the owner is violated once others have access to the contents of the database. Thus, Problem is to check whether the database inserted with the tuple is still k-anonymous without letting the owner A and others (B) to know the content of the tuple and database respectively. In this paper, we propose a protocol solving this problem on suppression based k-anonymous and confidential database.***

***Keywords- Anonymization, Privacy ,Confidentiality, Anonymous***

## I. INTRODUCTION

For each application database is important or valuable thing, so their security is important. There are different security control methods are identified and each method have different criteria. For example, FERPA provides privacy protections for such records when held by federally funded educational institutions[1]. FERPA defines an education record as those records, files, documents, and other materials that contain information directly related to a student and are maintained by an educational agency or institution or by a person acting for such agency or institution. Students who are at least 18 years of age, or attending postsecondary institutions or otherwise their parents ,generally have a right to gain access to their education records within 45 days of a written request, seek to amend any information therein considered to be in error, control how information in such records is disclosed to other institutions ,in general, such disclosures must be authorized by the student or parent, with some exceptions and complain to the US Department of Education if these rights appear to have been violated. There is problem of providing security to statistical databases against disclosure of confidential information. There are various security control method classified into four groups: conceptual, query restriction, data perturbation, and output perturbation. Criteria for evaluating the performance of the various security-control methods are identified.

A detailed comparative analysis of the most promising methods for protecting dynamic-online statistical databases is also presented. To date no single security-control method prevents both exact and partial disclosures. There is big concern for privacy.

The problem of statistical disclosure control revealing accurate statistics about a population while preserving the privacy of individuals has a vulnerable history. Still, there is a difference between confidentiality and privacy-Confidentiality refers to limiting information access and disclosure to authorized users and preventing access by or disclosure to unauthorized users. Privacy refers to limiting access to individuals' personal information[5].Question is Confidentiality is still required if data have been anonymized-yes because anonymous data have business value for the party owning database or unauthorized disclosure of anonymous data may damage the party owning the data.There have been lots of techniques developed to protect privacy, but here we proposed k-anonymization[4]. K-Anonymity refers to attributes are suppressed or generalized until each row is identical with at least k-1 other rows. At this point the database is said to be k-anonymous. K-Anonymity prevents definite database linkages. The modification of the anonymous database DB can be naively performed as follows: the party who is managing the database or the server simply checks whether the updated database DB is still anonymous. Under this approach, the entire tuple t has to be revealed to the party managing the database server, thus violating the privacy of the patient. Another possibility would be to make available the entire database to the patient so that the individual can verify if the insertion of the data violates their own privacy. This approach however, requires making available the entire database to the patient thus violating data confidentiality. There is a protocol solving this problem on suppression-based k-anonymous and confidential databases. The protocol depends on well-known cryptographic assumptions. The huge numbers of databases recording a large variety of information about individuals makes it possible to find information about specific individuals by simply correlating all the available databases. Confidentiality is achieved by enforcing an access policy, or possibly by using some cryptographic tools. Privacy relates to data can be safely disclosed without leaking sensitive information regarding the legitimate owner. The confidentiality is still required once data have been anonymized, if the anonymous data have a business value for the party owning them or the unauthorized disclosure of such anonymous data may damage the party owning the data or other parties.So,

Manuscript received March 31, 2012

**Neha Gosai,** Department of Computer Engineering, Bharati Vidyapeeth Deemed University College of Engineering, Pune, Maharashtra, Mobile No:-9960687566,
(E mail :-Nehagosai@ymail.com).
**S. H. Patil,** Department of Computer Engineering , Bharati Vidyapeeth Deemed University College of Engineering, Pune, Maharashtra, Mobile No:-9960687566. (E mail :- shpatil@bvucoep.edu.in).

Problem is that can database owner assure privacy of database without knowing data to be inserted? It is important to assure that database maintains privacy of individual and also who maintain database.So,it needs to check that data entered in database do not violate privacy, and to perform such verification without seeing sensitive information of individual.

## II. RELATED WORK

There are some limitations of the protocol, if the database is not anonymous with respect to a tuple that has been inserted, the insertion cannot be performed. Therefore, one of the protocols is extremely inefficient. There are efficient protocols. The first research is based on algorithms for database anonymization. The database is protected by data reduction, data perturbation or generating synthetic data. However, the main concept of k-anonymity to maintain confidentiality of their contents. The problem is to protect privacy of data that has been divided into two groups depending on whether data are continuously released and anonymizaed or data released in different fashion and anonymized. The second research is related to Secure Multiparty Computation protocol which is subfield of cryptography. The third research is related to the private information retrieval, which can be seen as an application of the secure multiparty computation techniques to the area of data management. This allows a user to retrieve an data (or tuple) from database without revealing tuple one is retrieving. Here, main focus is to find efficient techniques to express queries over a database without letting the database know the actual queries [2]. Still, the problem of privately updation of database has not been resolved because these techniques deal with only data retrieval. These approaches that will not address the problem of k-anonymity since their goal is to encrypt the data hence external entities can obtain their data. Thus, the main goal is to protect the confidentiality of the data from the external entities that manages the data. Even though, the data are fully available to the clients that are not the case under our approach. In data anonymization, Insertion cannot be performed if database is not properly anonymized. The problem is private updates to k-anonymous databases The suppression based protocols deals with the problem of updating the databases.

Figure 1 shows Anonymous database system, we assume that information of single student stored in a tuple and database kept confidential at server. The users treated as database of educational record, only institution have right to access to database. Since database is anonymous, the data provider's privacy is protected from users. Since database have privacy sensitive data, so main aim to protect privacy of student' data. This can be achieved by anonymization.If database is anonymous, it is not possible to catch student's identity for database. Suppose new student has to be entered, this means database has to be updated in order to insert a tuple.The modification of anonymous database can be done as follows: the party who is managing the database checks whether the updated database is still anonymous after inserting a tuple. Under this approach, the entire tuple t has to be revealed to the party managing the database server, thus violating the privacy of the student. Another possibility would be to make available the entire database to the student so that the student can verify by himself/herself if the insertion of his/her data violates his/her own privacy. To get solution of these problem, several problem needs to be addressed: The first problem is: without revealing the contents of tuple t to be inserted and database DB, how to preserve data integrity by establishing the anonymity of DB U {t}.The second problem is: once such anonymity is established, how to perform this update?

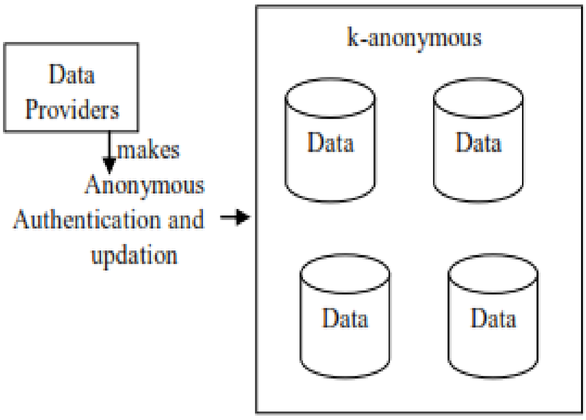

**Fig 1: Anonymous Database System**

The third problem is: what can be done if database anonymity is not preserved? Finally, The forth problem is: what is the initial content of the database, when no data about users has been inserted yet? In this paper, we propose a protocol solving first Problem, which is the central problem addressed by our paper.

## III. PROPOSED TECHNOLOGY

The protocol relies on the fact that anonymity of database does not affected ,if inserting tuple is already in database. then ,the problem of integrity while inserting tuple in database is equivalent to privately checking of inserting tuple with tuple already in database. The protocol is aimed at suppression based anonymous database and it allows the owner of database to properly anonymize the tuple t, without gaining any useful knowledge on its contents and without having to send to its owner newly generated data. To achieve such goal, the parties secure their messages by encrypting them. To assure higher level of anonymity to the party inserting a tuple ,we require that the communication between the party and database occurs through anonymous connection, as provided by crowd protocol[3]. Crowd protocol hides each user's communications by routing them randomly within a group of similar users. In order to perform the privacy-preserving verification of the database anonymity upon the insertion, the parties use a commutative and homomorphic encryption scheme.

### A Prototype Architecture

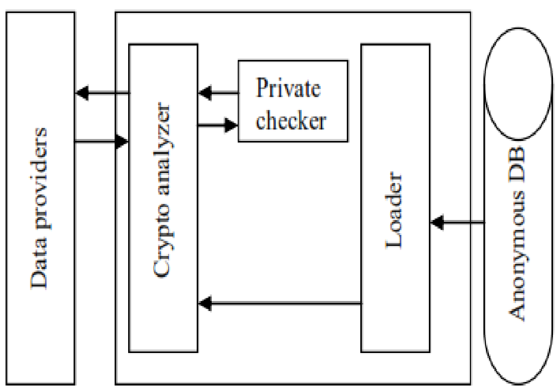

Fig 2 Prototype Architecture

In Fig 2, Data provider enters data is stored in crypto module which perform cryptography operation on all tuples exchanged between user and Private updater, using suppression based method. Loader module read anonymized tuples from k-anonymous database. And checker module checks whether the tuple from the user matches with the tuple in the database. If none of the tuple matches with the user tuple, then loader reads another tuple from k-anonymous database. The functionality provided by the Private Checker. Communication between user and database is carried out by anonymizer and that all the tuples are encrypted.

## IV. SUPRESSION BASED PROTOCOL

The suppression based protocol relies on well-known cryptographic techniques. We consider table T={t1,…t2} over the attribute set A. Generally in suppression based method we mask value of some special attributes with *, the value deployed by the user for anonymization. So the main idea behind this protocol is: To form subset of indistinguishable tuples by masking the value of some well chosen attributes.

TABLE 1 Original Dataset

| Birth date | Sex | Zip code |
|---|---|---|
| 21-01-79 | male | 53715 |
| 10-01-79 | female | 55410 |
| 21-02-83 | male | 2274 |
| 19-04-82 | male | 2237 |

TABLE 2 Suppressed Data with k=2

| Birth date | Sex | Zip code |
|---|---|---|
| */1/79 | person | 5**** |
| */1/79 | person | 5**** |
| */*/8* | male | 022** |
| */*/8* | male | 022** |

As shown in table 1 which contains original database (Table T) having three attributes Birth date, Sex, Zipcode.Table 2 shows a suppression based k-anonymization with k=2.As shown in table k=2 means at least k(=2) tuples should be indistinguishable by masking values. Suppression based attributes for every tuple of T is referred as anonymization problem, and finding the anonymization that minimizes the number of masked values.

### A Cryptography Primitive

The Diffie Hellmen 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. AES(Advanced Encryption Standard) algorithm is the advanced encryption standard form of algorithm which had been used as a symmetric form of encryption.

There are two encryption schemes, commutative and product homomorphic E to satisfy indistiguishability properly.. A commutative, product- homomorphic encryption scheme ensures that the order in which encryptions are preformed is irrelevant(commutativity) and it allows to consistently perform arithmetic operations over encrypted data (homomorphic property). Given a finite set K of keys and a finite domain D, a commutative, product homomorphic encryption scheme E is a polynominal time computable function E: K×D→D satisfying the following properties:

1. ***Commutativity*:**

In commutative, 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:***

In product homomorphic 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 distinguish an encryption from a randomly chosen value in the same domain and having the same length. The advantages are high privacy of data even after updation, and an approach that can be used is based on techniques for user anonymous authentication and credential verification.

### *B Algorithm*

Suppose Alice has control over database and Bob is data provider then protocol works as follows:

In step 1, Alice sends Bob encrypted version of tuple containing only non suppressed attributes. At step 2, bob encrypts the information received from Alice and sends it to her, along with encrypted version of each value in his tuple. In final step, Alice examines if the suppressed attributes of tuple is equal to the tuple sent by Bob. If yes then insert tuple in database.

## V. CONCLUSION

In this paper, we have proposed secure protocol for privately checking whether a k-anonymous database retains anonymity once a new tuple is being inserted to it. Since the proposed protocol ensures the updated database remains k-anonymous. Thus the database is updated properly using the proposed protocol. The data provider's privacy cannot be violated if user update a table. If updating any record in database violate the k- anonymity then such updating or insertion of record in table is restricted. If insertion of record satisfies the k-anonymity then such record is inserted in table and suppressed the sensitive information attribute by * to maintain the k-anonymity in database. Thus by making such k-anonymity in table that makes unauthorized user to difficult to identify the record.

The important issues for future work are as follows:

- Improve the efficiency of protocols, by the number of messages exchanged and sizes and algorithm used for encryption and decryption.
- The private update to database systems techniques supports notions of anonymity different than k-anonymity.
- In the case of malicious parties by the introduction of an untrusted third party, implementing a real-world anonymous database system.

## REFERENCES

1. U.S. Department of Education. General Family Educational Rights and privacy Act (FERPA).
2. B.C.M. Fung ,K. Wang, A.W.C. Fu and J. Pei, Anonymity for Continuous Data Publishing Proc. Extending database Technology Conference (EDBT),2008
3. M.K.Reiter, A. Rubin. Crowds: anonymity with Web transctions.ACM Transactions on Information amd System Security (TISSEC),1(1),1998;66-92
4. 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/Des.2001
5. University of Miami Leonard M. Miller School of Medicine, Information Technology.
6. Dr. Durgesh Kumar Mishra, Neha Koria, Nikhil Kapoor, Ravish Bahety ,A Secure Multi-Party Computation Protocol for Malicious Computation Prevention for preserving privacy during Data Mining, Vol. 3,No. 1,2009
7. C. Blake and C. Merz, ―UCI Repository of Machine Learning Databases, E. Bertino and R. Sandhu, ―Database Security—Concepts, Approaches and Challenges, IEEE Trans. Dependable and Secure Computing, vol. 2, no. 1, pp. 2-19, Jan.-Mar. 2005. www.wikipedia.com/wikifiles/.