Survey on Privacy Preserving Updates for Anonymous and Confidential Databases

Gaurang Patel[1] , Sanket Patel[2]
[1]*M.E. Student, Department of Computer Engineering,*
[2]*Assistant Professor, Department Of Computer Engineering,*
Kalol Institute of Technology & Research Center

Abstract- **In today life database is widely used in industries in order to maintain data. It becomes crucial part of our life so to maintain information in database is very important. Some databases contain private information about individuals so it is necessary to keep this information private in database so for this reason we provide security to database so data remains secure. In the existing system the private information gets lost in big amount and doesn`t provide any security mechanism. Today many applications employ data mining techniques on databases which include private and sensitive information. In order to protect database from unauthorized access one should keep database anonymized or confidential. Anonymization means hide the identifying information from original data to protect private data. Confidentiality is a set of rules that limits access on certain types of information. There are different ways by which we can perform anonymization. In this paper we use k-anonymization. Suppose one person A owns a k-anonymous database and needs to determine whether his database is still k-anonymous if tuple inserted by another person B. For some applications database needs to be confidential, so we must keep access to the database strictly controlled. The confidentiality of the database is violated once other persons 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 B to know the content of the tuple and database respectively.**

Index Terms- **Privacy, preserving, Confidentiality, K-anonymity**

I. INTRODUCTION

Today everyone is focusing on data collection, containing personal details or some other confidential information. Database is known as collection of data that can be accessed, updated and retrieved by user. To provide security or maintain security of confidential databases like medical data is a big issue. Maintaining privacy of database is very important. Privacy or security of database is very crucial because databases represent an important asset of many users. For example in medical, hospital database contains all details of patients including patient`s private details so it is important to keep this data secure or private so unauthorized users can not use it. Confidential database must be accessed by authorized users. Today huge numbers of databases available with large variety of information about individuals so it makes it possible to find information about specific person. Confidentiality is a set of rules that limits access on certain types of information or it is achieved by some cryptographic tools. Anonymity means masking the identifying data. One well known technique is k-anonymization this technique protects privacy by modifying the data so the possibility of linking sensitive data is very small. In anonymity identifying information is removed from original data to protect private information. In data anonymization allows transferring information between two parties by converting text data into non-human readable form using encryption method. It is also important to maintain database integrity while data is transferring from source to destination. K-anonymization approach is good it protects privacy of original data. So problem arises when database needs to be updated. When tuple is to be inserted in the database problems occurs relating to privacy and confidentiality that is to check whether the database is still k-anonymous, without owner knows what tuple to be inserted. Suppose Person A has a K-anonymous database and Person B inserts a tuple in database. Clearly allowing person A to directly read the contents of tuple breaks the privacy of person B and confidentiality managed by person A is broken when person B had access the contents of database so problem is to check whether the database inserted with the tuple is still k-anonymous.

II. PROBLEM STATEMENT

There are so many methods available for k-anonymity of database. To achieve the goal to check whether the database inserted with the tuple is still k-anonymous, without letting admin & user know the contents of tuple & database. Also to ensure privacy preserving and confidentiality during updating of database. The operation of updating such a database e.g., by inserting a tuple containing information about a given individual, introduces two problems.

1) Is the updated database still privacy preserving?
2) Database owner need to know the data to be inserted?

In some existing systems data are stored directly in database. So anyone can easily access information like password, address etc. Any unauthorized person can easily access the database. Data confidentiality is needed because the availability of large numbers of databases. So many methods for anonymity of database have been developed but these methods have some problem or drawbacks. Existing methods do not implement on real world database system. Also these methods do not implement database for invalid entries. Another problem is improving efficiency of protocol in terms of number of messages exchanged between user and database and solve problem of anonymity when initially table is empty. In existing methods there is a possibility to lost original data when we are trying to perform suppression and generalization on data. There is no facility for protecting original data.

III. LITERATURE REVIEW

Today everyone uses databases for storing their data, these databases contain private or personal information about users so these databases must be kept secure from unauthorized or other users and also privacy must be maintained. To maintain privacy is big issue. Databases like hospital, bank need to be protected. By anonymization we can remove personal identifier to protect private information. k-anonymization approach[1] is also used for anonymization. Suppression and generalization are the properties of k-anonymization. The idea used in the Suppression algorithm is to mask some attributes by special value *, in generalization algorithm attributes are replaced with general value based on Value Hierarchy Graph. In Suppression approach to provide privacy preserving updates to confidential database is designed. Data entered by the user directly replaced by special value "*" and these values being inserted into table. In generalization If the data matches with the general value then this record will replaced by the general value and these general values being inserted into table.

If the data owners can directly read the contents of the database simply it breaks the privacy of the user's data. If the users can access the database content directly then the confidentiality of the data owners has been violated. In the existing systems the privacy information gets lost in huge amount and does not provide any security mechanism. The Commutative Homomorphic Encryption Scheme[3] is used to make the data privacy and confidentiality of the database. It provides the security of data using RSA algorithm. It has several important properties which overcomes the drawbacks of the existing system. The properties are anonymous authentication, privacy preservation, confidential access, anonymity maintenance, and also access control. The mechanism of trusted third party has been introduced here. In this system, users and surveyor registers with the trusted third party. The trusted third party registers with the admin database. The admin maintains the user records and also categorizes the attributes into sensitive and non sensitive. The user and surveyor communicate with the trusted third party using Commutative Homomorphic Encryption Scheme (CHES). Similarly the trusted third party communicates with the admin using RSA technique. This system ensures the users privacy and confidentiality and also providing security to confidential data.

Another approach for achieving k-anonymity named K-anonymity of Classification Trees Using Suppression (kACTUS)[4]. In kACTUS, efficient multidimensional suppression is performed, i.e., values are suppressed only on certain records depending on other attribute values. kACTUS, wraps a decision tree inducer which is used to induce a classification tree from the original data set. kACTUS was specifically designed to support classification. It introduces Quasi-identifier, Quasi-identifier is a set of features whose associated values may be useful for linking with another data set to reidentify the entity that is the subject of the data. The kACTUS algorithm contains of two main phases: In the first phase, a classification tree is induced from the original data set, in the second phase, the classification tree is used by a new algorithm developed in this study to k-anonymize the data set. kACTUS is capable of applying k-anonymity on a given table with no significant effect on classification accuracy. kACTUS works well with large data sets.

There are different techniques to provide confidentiality and privacy to anonymous database like Data Reduction, Data perturbation and Secure Multiparty Computation etc.In suppression based anonymous database[5] secure protocol presented for checking that if new tuple is being inserted to the database, it does not affect anonymity of database. It has been said that first find Quasi Identifiers because Attack is mainly using Quasi-Identifier. To prevent the attack, masks Quasi-Identifier`s values using either suppression based or Generalization based Anonymization methods. In Suppression based anonymization method, mask the Quasi-Identifiers value using a special symbol like * by making such kanonymity in table that makes unauthorized user too difficult to identify the record.

The privacy is so important for databases the paper[6] also focuses on survey on privacy preserving on anonymous database and on devising private update techniques to database systems that supports notions of anonymity diverse than k-anonymity. The existing methods provide the same amount of privacy for all people, and may be offering insufficient protection to a subset of people, while applying more privacy control to another subset. Motivated by these the concept of personalized anonymity[6] is used which performs the least generalization for satisfying everybody's requirements, and thus, retains the largest amount of information from the data. So the proposed work based on the concept of personalized anonymity, and updates will be performed on the personalized anonymous databases by using SA- generalization algorithm. So whenever a new tuple is inserted the individual will decide the level of privacy from taxonomy tree for sensitive attributes.

IV. CONCLUSION

This paper proposed different algorithms and techniques for privacy preserving updates for anonymous and private databases. There are many techniques available for privacy preserving every technique has different solution, advantages and disadvantages. Anonymization is one way to protect data. The goal is to make the process of anonymity more effective and we will get good result to achieve that we will propose new approach in which we will get effective results and no loss of data.

REFERENCES

[1] Prashant Jawade Assistant Professor Thakur College of Engineering &Technology, Mumbai University, Kandivali (E) Mum-101 and Poonam Joshi ME student Thakur College of Engineering &Technology, Mumbai University, Kandivali (E) Mum-101 "Securing Anonymous and Confidential Database through Privacy Preserving Updates"

[2] Dr.K.P.Thooyamani, Dr.V.khanaa, Er.M.R.Arun Venkatesh "Privacy-Preserving Updates to Anonymous and Confidential Database"

[3] Kavitha Department of Information Technology Rajalakshmi Engineering College, Chennai "Suppression and Generalization – Based Privacy Preserving Updates to Confidential Databases"

[4] Mrs.M.Aruna Safali, Mr.T.Bala Murali Krishna, Mr. G Sai Chaitanya Kumar "Suppression of Multidimensional Data Using K-Anonymity"

[5] SIVASUBRAMANIAN.R, K.P. KALIYAMURTHIE Department of Information Technology, Bharath University, India "Privacy-Preserving Updates to Anonymous Databases"

[6] E. Gokulakannan, Dr.K.Venkatachalapathy [1]Department of CSE, MRK Institute of Technology, Kattumannarkoil, [2]Department of Computer Science and Engineering Annamalai University, Annamalai Nagar "SURVEY ON PRIVACY PRESERVING UPDATES ON UNIDENTIFIED DATABASE"

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

[8] S. Zhong, Z. Yang, and R.N. Wright, "Privacy-Enhancing k-Anonymization of Customer Data," Proc. ACM Symp.Principles of Database Systems (PODS), 2005.

[9] V.S. Verykios, E. Bertino, I.N. Fovino, L.P. Provenza, Y. Saygin, and Y. Theodoridis, "State-of-the-Art in Privacy Preserving Data Mining," ACM SIGMOD Record, vol. 33, no. 1, pp. 50-57, 2004.

[10] Trombetta and E. Bertino, "Private Updates to Anonymous Databases," Proc. Int'l Conf. Data Eng. (ICDE), 2006.