Anonymous and Confidential Database Securing through Privacy Preserving Update

Poonam Joshi
Assistant Professor
Information Technology Department.
Atharva College of Engineering
Malad, Mumbai, India

Reena Somani
Assistant Professor
Information Technology Department
Atharva College of Engineering
Malad, Mumbai, India

Sejal D'mello
Assistant Professor
Information Technology Department.
Atharva College of Engineering
Malad, Mumbai, India

***Abstract*- Suppose Census Bureau receive private information of clients then the census bureau must publish anonymized version of data. So anonymization is done for security of clients. Today we are living in the Big Data world. In this, data privacy and confidentiality is the major concern or problem. Many algorithms are used for data privacy and confidentiality, which are not well-summarized because resulted dataset can be simply linked with public database so it reveals user identity. Suppose person-X having his own k-anonymous database and person-Y wants to insert a tuple.So, the problem is to check after inserting a tuple whether database retains its k-anonymity or not. We use two techniques, suppression and generalization.**

Keywords— Supression, Generalization,, K- Anonymity

I. INTRODUCTION

Information Security has become crucial issue since the information sharing have a become need of present technology. Privacy of information is a necessary to avoid fraud and theft for economic growth. New advances methods in data mining and knowledge discovery allow extraction of hidden knowledge in an enormous amount of data impose new threats on the seamless integration of information. Anonymization means identifying information is removed from original data to preserve private information. Data anonymization can be performed in different ways but in this k-anonymization approach is used. The data owners directly read the contents of the database simply it breaks the privacy of the user's data. If the users access the database content directly then the confidentiality of the data owners has been violated. So both privacy and confidentiality of the database are considered to be a major problem. In the existing system the privacy information gets lost in large amount and does not provide any security mechanism. It is impossible to consider every possible inference and also drastically reduces the quality of the data. The proposed system considers the privacy information is more valuable both in research and business areas. So the data sharing is common in order to provide the data remain k-anonymous even after the updates. In this project, we propose and implement two methods suppression and generalization to maintain privacy and confidentiality of database.

Database is important and critical thing for application so their security is very important .Data in the databases has its own relevant value. For example medical data collected by over the history of patients over years is an invaluable asset, which needs to be secured and can be used by people in various related areas of work. Nowadays, privacy accidents have become common problem in the information systems. For example, a hospital may have record of all the patients with various diseases critical and non-critical. If the hospital wishes to reveal the data to any pharmaceutical company or online market services, it should not be able to infer with particularity of patients with those diseases. It can give as a statistical view or just the superficial information such that privacy is not detained. There are huge numbers of databases that hold number of confidential information's such that people access those data correlating various information from various databases. Disclosure of confidential information to unauthorized persons may lead to data insecurity leading to dissatisfaction to users. For example, there was a company which sold health products online that also revealed the customer names phone numbers credit card numbers etc on the website. It leads to huge loss of information and breach of privacy. There was another issue when a researcher was enabled to retrieve health records from anonymous databases of insurance claims of employees. Privacy relates to what data can be safely disclosed without leaking information regarding the legitimate owner [1]. Thus, if one asks whether confidentiality is still required once data have been anonymized, the reply is yes if the anonymous data have 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. The term anonymized or anonymization means identifying information is removed from the original data to protect personal or private information. There are many ways to perform data anonymization. We only focus on the k-anonymization approach. To better understand the difference between confidentiality and anonymity, consider the case of a medical facility connected with a research institution. Suppose that all patients treated at the facility are asked before leaving the facility to donate their personal health care records and medical histories (under the condition that each patient's privacy is protected) to the research institution, which collects the records in a research database. To guarantee privacy to each patient, the medical facility only sends to the research database an anonymized version of the patient record. Once this anonymized record is stored in the research database, the nonanonymized version of the record is removed from the system of the medical facility. Thus, the

research database used by the researchers is anonymous. While k-anonymity protects against identity disclosure, it is insufficient to prevent attribute disclosure. Generalization and suppression such technique provides privacy by modifying data in such a way that it gives the same result for more than two tuples. So the problems of confidentiality and anonymization are different. The problem occurs when it comes to the updating of the database. When the tuple is to be inserted into the database, there are two problems: Is updated database still maintains privacy? And owner of the database really know data to be reply is yes if the anonymous data have 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. The term anonymized or anonymization means identifying information is removed from the original data to protect personal or private information. There are many ways to perform data anonymization. We only focus on the k-anonymization approach [2]. Thus, the problem is to check whether the database inserted with the tuple is still k-anonymous, without letting Data Provider and Database owner know the contents of the database and the tuple, respectively. We propose and implement two methods suppression and generalization to maintain privacy and confidentiality of database.

II. RELATED WORK

A number of techniques have been developed to provide privacy to the databases such as, randomization, secure multi-party computation and k- anonymity.

A. Randomization

The randomization method provides effective way of preventing the user from learning sensitive data which can be easily implemented because the noise added to the given record is independent from the other records. The amount of noise is large enough to smear original values, so individual record cannot be recovered. The randomization method is simple as compare to other methods because it does not require to knowledge of other records. Large randomization increases the uncertainty and the personal privacy of the user's .They claim that approaches may lose information as well as not provide privacy by introducing random noise to the data by using random matrix properties, [3]. It successfully separates the data from the random noise and subsequently discloses the original data.

B. Secure multi-party computing

Goal of secure party computation is to compute function when each party has some input. It generally deals with problems of function computation with distributed inputs. In this protocol, parties have security properties e.g., privacy and correctness regarding privacy a secure protocol must not reveal any information other than output of the function. *K-anonymity* and SMC are used in privacy-preserving data mining, but they are quite different in terms of efficiency, accuracy, security and privacy [4]. The goal of methods for secure multi-party computation is to enable parties to jointly compute a function over their inputs, while at the same time keeping these inputs

K-anonymous

A large number of privacy models were developed most of which are based on the *k*-anonymity property. The K-anonymity model was proposed to deal with 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 [2].

III. METHODOLOGY

Suppression Algorithm

The idea used in the Suppression algorithm is to mask some attributes by special value *, In suppression algorithm t stands for private tuple provided by Data provider, T stands for Anonymous database, QI stands for Quasi-Identifier which consist of set of attributes that can be used with certain external information to identify a specific individual. The suppression algorithm works as follows

Step1: User P sends User Q an encrypted version containing only the s non-suppressed attributes.

Step2: User Q encrypts the information received from User P and sends it to her, along with encrypted version of each value in his tuple t.

Steps 3-4: User P examines if the non suppressed QI attributes is equal to those of t. If true, t can be inserted to table T. Otherwise, when inserted to T, t breaks k-anonymity.

Generalization Based Algorithm

In generalization algorithm attributes are replaced with general value based on Value Hierarchy Graph (VGH). The protocol works as follows:

Step 1: User P randomly chooses a □ ε *Tw* (Witness Set).
Step 2: User P computes γ = GetSpec (□).
Step 3: User P and User Q collaboratively compute s= SSI (γ, τ).
Step 4: If s=u then t's generalized form can be safely inserted to T.
Step 5: Otherwise, User P repeats the above procedures until either s=u or witness set is empty. Let t is User Q's private tuple from

IV. ARCHITECTURE

The figure1 shows the flow of steps followed in the system. It starts with authentication of user. Each user is provided with username and password registered in system already. The authentication user has access to the database and system has particular access rights for each user. The anonymous database suppresses and generalizes the data according to data value. The database can be accessed by research centers forgathering statistical data regarding particular medicines, the percentage of curable medicines. The internal or private information of the patients are not revealed to the research centre computation. The research people can see the data's send by the database according to its access right. And allocate research peoples to each research data. And forward the data to research people. Here research people can't do any changes or modifications in patient database they only can use the database for reference purpose. Information of the patients is not revealed to the research centre computation.

The research people can see the data's send by the database according to its access right. And allocate research peoples to each research data. And forward the data to research people [5].

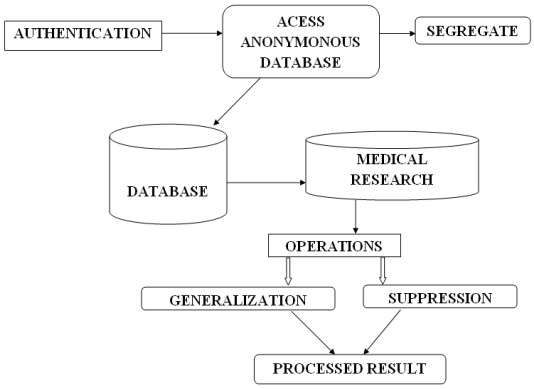

Fig. 1: System Block Diagram [5]

V.CONCLUSION AND FUTURE RESEARCH WORK

We have used two secure methods Suppression and Generalization are proposed that is used to check that if new tuple is being inserted to the database, Database updates has been carried out properly using proposed methods. Execution shows that once system verifies user 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. Thus the database is updated properly using the proposed methods. The data provider's privacy cannot be violated if user updates 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 sensitive information attribute used to maintain the k-anonymity in database. Thus such k-anonymity in table makes difficult for unauthorized user to identify record.

The important issues in future will be resolved:

- Implement database for invalid entries.
- Improving efficiency of protocol in terms of number of
- Messages exchanged between user and database.
- Implement real world database system
- Devising private update techniques to database systems

REFERENCES.

[1] Alberto Trombetta, Wei Jiang, Elisa Bertino and Lorenzo Bossi, "Privacy Preserving Updates to Anonymous and Confidential Distributed Databases IEEE Transactions on Dependable and Secure Computing. , vol. 8, no. 4, July/August 2011.

[2] M.K. Reiter and A. Rubin, "Crowds: Anonymity with Web Transactions," ACM Trans. Information and System Security (TISSEC),vol. 1, no. 1, pp. 66-92, 1998.

[3] K. Wang and B. Fung, "Anonymizing Sequential Releases," Proc. ACM Knowledge Discovery and Data Mining Conf. (KDD), 2006.

[4] J.W. Byun, T. Li, E. Bertino, N. Li, and Y. Sohn, "Privacy-Preserving Incremental Data Dissemination," J. Computer Security vol. 17, no. 1, pp. 43-68, 2009.

[5] Poonam Joshi and Prashant Jawade **"**Securing anonymous and confidential database through privacy preserving updates International Journal of Applied Information Systems (IJAIS) Proceedings on International Conference and workshop on Advanced Computing (ICWAC) July 2013, ISSN: 2249-0868, published by Foundation of Computer Science New York USA