Privacy-Preserving Updates to Databases with Security

Pallavi Patil, Prof. K.B.Manwade, Prof. G.A.Patil

Abstract— **Today's databases speak to a critical stake for numerous requisitions. There is an expanded concern for security. The accessibility of tremendous amounts of databases recording an extensive assortment of data about people makes it conceivable by essentially corresponding all the accessible databases. Likewise, delicate data about the gathering embeddings the information may be spilled from the right to gain entrance control approaches received by the anonymous database framework. A methodology could be utilized is dependent upon methods for client anonymous authentication and qualification check. So the issue of anonymous updates to confidential databases is unpredictable and requires the combo of some systems.**

Some protocol and systems have been proposed in the literature, which manage the anonymity, confidentiality and privacy. Still a few issues are unaddressed like n-anonymity, malicious third party and efficiency improvement of protocol. In this, three protocols that is Formation 1 protocol, Formation 2 protocol and third party protocol are proposed to deal with above issues.

Index Terms — **Anonymous, Confidentiality, Privacy, Third party.**

I. INTRODUCTION

The accessibility of huge amounts of databases recording an expansive assortment of data about people makes it conceivable to identify data about particular people by essentially relating all the accessible databases. In spite of the fact that confidentiality and privacy are regularly utilized as equivalent words, they are diverse thoughts: information confidentiality is about the challenge by an unapproved client to take in anything about information saved in the database. Ordinarily, confidentiality is accomplished by authorizing a right to gain entrance arrangement, or potentially by utilizing some cryptographic tools. Privacy relates to what data can be safely disclosed without leaking sensitive information regarding the legitimate owner [1].

Anonymization means masking. That is identifying information is removed from the original data to protect personal or private information. Data Anonymization enables transferring information between two organizations, by converting text data in to non-human readable form using encryption method [2]. The K-anonymity is achieved by blocking all the dissimilar values (Suppression) or by replacing them with a less specific common consistent value (Generalization). Hence K-anonymity is trade-off between data utility and data confidentiality. It is important to note that the data loss happens for identifiers and quasi-identifiers only.

There have been lots of approaches developed. K-Anonymization is one of the approaches. In K-Anonymization approach, at least K-tuples should be indistinguishable by masking values [3]. So the probability of linking a given data value to a specific individual is very small, and the individuals cannot be uniquely identified by linking attacks.

So problem arises at this point where database needs to be updated. When tuple is to be inserted in the database problem occurs relating to privacy and confidentiality that is database owner decide that whether database preserve privacy without knowing what new tuple to be inserted. To carry out task of privacy, confidentiality to anonymous database, two approaches can be used. One is Suppression and the other is Generalization.

Manuscript received 23 January 2014
Pallavi S.Patil, Computer Department, JSPM, NTC ,Narhe, Pune, India, (e-mail: pal.patil21@gamil.com).
Prof.K.B.Manwade, Department of Computer Sci.Engg.**,** AMGI, Vathar, Kolhapur, India, (e-mail:mkarveer@gmail.com).
Prof.G.A.Patil,, Department of Computer Sci.Engg, DYCOET, Kolhapur, India, (e-mail: gapatil@gmail.com).

II. PAST WORK REVIEW

A. Private Updates to Anonymous Databases:

In this have some serious limitations of protocol, in that do not support generalization-based updates, which is the main strategy adopted for data anonymization. The availability of huge numbers of databases recording a large variety of information about individuals makes it possible to discover information about specific individuals by simply correlating all the available databases. Although confidentiality and privacy are often used as synonyms, they are different concepts: data confidentiality is about the difficulty by an unauthorized user to learn anything about data stored in the database.Therefore, if the database is not anonymous with respect to a tuple to be inserted, the insertion cannot be performed. In addition one of the protocols is extremely inefficient. [4]

B. Security-Control Methods for Statistical Databases:

This paper considers the problem of providing security to statistical databases against disclosure of confidential information. Security-control methods suggested in the literature are classified into four general approaches: conceptual, query restriction, data perturbation, and output

perturbation. Criteria for evaluating the performance of the various security-control methods are identified. Security-control methods that are based on each of the four approaches are discussed, together with their performance with respect to the identified evaluation criteria. A detailed comparative analysis of the most pr of the most promising methods for protecting dynamic-online statistical databases is also presented. The first research direction deals with algorithms for database anonymization. The idea of protecting databases through data suppression or data perturbation has been extensively investigated in the area of statistical databases. [5]

C. K-Anonymity: A Model for Protecting Privacy:

It presented the k-anonymity protection model, explored related attacks and provided ways in which these attacks can be thwarted. Initially proposed the notion of k-anonymity for databases in the context of medical data, which have developed complexity results concerning algorithms for k-anonymization. [4]

D. Privacy-Enhancing k-Anonymization of Customer Data:

In this paper, they focus how to create k-anonymous tables in a distributed scenario without the need for a central authority and while maintaining customer privacy. The problem of computing a k-anonymization of a set of tuples while maintaining the confidentiality of their content is addressed. [5]

E. Foundations of Cryptography:

The second research direction is related to Secure Multiparty Computation (SMC) techniques. SMC represents an important class of techniques widely investigated in the area of cryptography. General techniques for performing secure computations are today available. [6]

III. PROBLEM STATEMENT

A. Problem Definition

In the existing framework information are store in database straightforwardly. Anybody can effortlessly recover data like username, password and so on. The characterization of database is done from neighborhood framework just. Any unapproved individual can effortlessly access the database. Approved individual can see the other client's information as well. Information confidentiality is especially significant on account of the quality, regularly not just money related, The updating operation like by inserting a tuple holding data something like a given unique, presents two issues concerning both the anonymity and confidentiality of the information archived in the database and the security of the single person to whom the information to be embedded are identifies.

To devise a suitable result, some issues requirement to be tended to:

Issue 1: without uncovering the substance of tuple and database, how to save information respectability by masking the integrity of database?

Issue 2: what might be carried out if database anonymity is not saved?

Issue 3: what is the starting substance of the database, when no information about clients has been embedded yet?

Issue 4: How to enhancing the effectiveness of protocols, as far as number of messages exchanged

B. Proposed System

So, implementation of private updates techniques to database systems that support notions of anonymity different than k-anonymity [7].

In this paper, we propose two protocols [7] solving issue 1, which the central problem is addressed in paper as, suppression – based (Formation-1) and generalization - based (Formation-2) protocols. These protocols are relying on cryptographic techniques.

To solve issue 2 and 3 we proposed third protocol which dealing with the case of malicious parties by the introduction of an untrusted, noncolluding third party i.e. Third Party protocol [8].

To improve efficiency forth protocol is used i.e. Message Aggregation Protocol [9], which address in issue 4.

IV. METHODS & TECHNIQUES

Implementation of private update techniques to database systems that supports notions of anonymity different than k-anonymity [5]. For this two protocols will be used.

A. Suppression-based (Formation-1) protocol:

- Where sensitive information and all information that allows the inference of sensitive information are simply not released.
- To achieve such goal, the parties secure their messages by encrypting them.
- To perform the privacy-preserving verification, the parties use cryptographic schemes.

B. Generalization-based (Formation-2) protocol:

- It will rely on a secure set intersection protocol, to support privacy-preserving updates on a k-anonymous database.
- It may consist of following steps:
  - Random functions
  - GetSpec Function
  - SSI a secure protocol that computes the cardinality i.e.
    - Private Matching and Set Intersection which includes,
      1. Encryption
      2. Share public key
      3. Private Matching and Set Intersection
    - Intersection protocol which follows,
      1. Hash function
      2. Choose secrete key
      3. Lexicographical order

C. Third Party Protocol:

- It will communicate only with the specific participants identified for a computation. It will be identifies specific properties about each computation to every participant involved in that computation.
- To improve the efficiency of protocol number of messages exchanged and their sizes will be reduced.

V. PROPOSED ARCHITECTURE

Proposed architecture contains following modules:

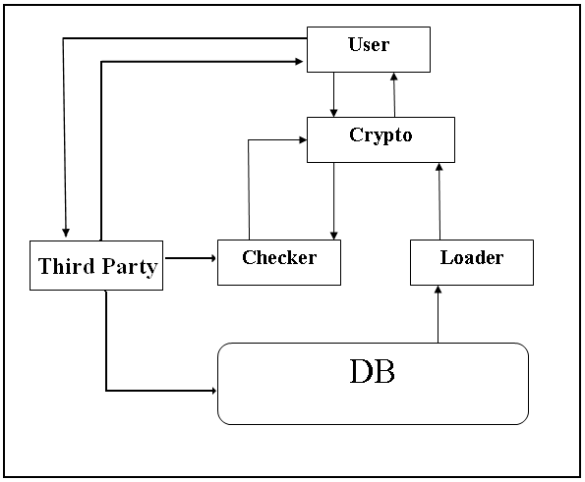

Fig. 1 Proposed Architecture

A. Module 1:

- **Private Checker:**
  - It will consist of the crypto module.
  - **Crypto Module:**
    - A crypto module that is in charge of encrypting all the tuples exchanged between a user and the Private Updater, using the Formation 1 protocol and Formation 2 protocol.

B. Module 2:

- ***Private Updater:***
  - It will consist of the checker module.
  - **Checker Module:**
    - A checker module that performs all the controls, as prescribed by Formation 1 protocol and Formation 2 protocol.

C. Module 3:

- **Loader Module:**
  - Loader module that reads chunks of anonymized tuples from the k-anonymous database.
  - The chunk size is fixed in order to minimize the network overload.

D. Module 4:

- **Third Party Module:**
  - All communication between a computation running on third protocol and the participants in the computation occurs through channels.
  - A channel is a bidirectional communication connection between a program on third protocol and a participant.

E. Module 5:

- **Message Aggregation Protocol:**
  - Several message aggregation techniques have been presented and applied to parallel simulations of a fine grain communication network model.
  - When comparing sender initiated and receiver-initiated message aggregation strategies, the former solution always performed better.
  - Receiver initiated succeeded to smooth the aggregate size distribution but failed to provide better performances.

VI. CONCLUSION

Achieving high security for large real time database is hard. Two protocols that privately checks whether a k-anonymous database retains anonymity for a new tuple. Since the proposed protocol ensures the updated database remains k-anonymous.

In case of suppression based k-anonymity approach, suppressed the sensitive information attribute by * to maintain the k-anonymity in database.

In case of generalization based k-anonymity approach, specific or original values are replaced by more general values so that attacker cannot identify correct values. This is particularly applicable in military application or health care system.

The important issues for future work are as follows:

- In the case of malicious parties by the introduction of an untrusted third party, implementing a real-world anonymous database system.
- Improve the efficiency of protocols, by reducing the number of messages exchanged and sizes.

VII. REFERENCES

1. 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.
2. D. Bonch The decision Diffic-Hellman problem in Proc. Of Int Algorithmic Number theory Symposium.
3. 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.
4. A. Trombetta and E. Bertino, "Private Updates to Anonymous Databases," Proc. Int'l Conf. Data Eng. (ICDE), 2006.
5. 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.
6. L. Sweeney, "k-Anonymity: A Model for Protecting Privacy," Int'l J. Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 5, pp. 557-570, 2002.

7. S. Chawla, C. Dwork, F. McSherry, A. Smith, and H. Wee, "Towards Privacy in Public Databases," Proc. Theory of Cryptography Conf. (TCC), 2005.
8. U. Feige, J. Kilian, and M. Naor, "A Minimal Model for Secure Computation," Proc. ACM Symp. Theory of Computing (STOC), 1994.
9. C. D. Pham RESAM, UniversitC Lyon 1, Claude Bernard B2t ISTIL"Comparison of Message Aggregation Strategies for Parallel Simulations on a High Performance Cluster"