An Anonymous Key Assignment For Privacy Preserving Data Sharing In Data Mining

Vaheeda Kuari, M.Tech Student, Dept.of CSE, DJR College Of Engineering & Technology, Andhra pradesh, India
Sk.N.Rehmathunnisa, Assoc Professor, Dept.of CSE, DJR College Of Engineering & Technology, Andhra pradesh, India

vaheeda.mathineni@gmail.com, shaikrehmathunnisa@gmail.com

ABSTRACT

This assignment of serial numbers allows more complex data to be shared and has applications to other problems in privacy preserving data mining, collision avoidance in communications and distributed database access. The required computations are distributed without using a trusted central authority. An algorithm for anonymous sharing of private data among parties is developed. This technique is used iteratively to assign these nodes key numbers ranging from 1 to N. This assignment is anonymous in that the identities received are unknown to the other members of the group. Resistance to collusion among other members is verified in an information theoretic sense when private communication channels are used. Existing and new algorithms for assigning anonymous keys are examined with respect to trade-offs between communication and computational requirements. The new algorithms are built on top of a secure sum data mining operation using Newton’s identities and Sturm’s theorem. An algorithm for distributed solution of certain polynomials over finite fields enhances the scalability of the algorithms.

Index Terms: Anonymization and deanonymization, cloud and distributed computing systems, multiparty computation, privacy preserving data mining, privacy protection, security and trust in cooperative communications.

I. Introduction:

The popularity of internet as a communication medium whether for personal or business use depends in part on its support for anonymous communication. Businesses also have legitimate reasons to engage in anonymous communication and avoid the consequences of identity revelation. For example, to allow dissemination of summary data without revealing the identity of the underlying data is associated with, or to protect whistle-blower’s right to be anonymous and free from political or economic retributions. Cloud-based website management tools provide capabilities for a server to anonymously capture the visitor’s web actions. The problem of sharing privately held data so that the individuals who are the subjects of the data cannot be identified has been researched extensively. Researchers have also investigated the relevance of anonymity and/or privacy in various application domains. Another form of anonymity, as used in secure multiparty computation, allows multiple parties on a network to jointly carry out a global computation that depends on data from each party while the data held by each party remains unknown to the other parties.

A secure computation function widely used in the literature is secure sum that allows parties to compute the sum of their individual inputs without disclosing the inputs to one another. This function is popular in data mining applications and also helps characterize the complexities of the secure multiparty computation. To differentiate anonymous ID assignment from anonymous communication, consider a situation where parties wish to display their data collectively, but anonymously, in slots on a third party site. The IDs can be used to assign the slots to users, while anonymous communication [6],[14] can allow the parties to conceal their identities from the third party.

In another application, it is possible to use secure sum to allow one to opt-out of a computation beforehand on the basis of certain rules in statistical disclosure limitation [15] or during a computation and even to do so in an anonymous manner. However, very little is known with respect to methods allowing agencies to opt-out of a secure computation based on the results of the analysis, should they feel that those results are too informative about their data.

II. EXISTING SYSTEM:

Existing and new algorithms for assigning anonymous keys are examined with respect to trade-offs between communication and computational requirements. Also, suppose that access to the database is strictly controlled, because data are used for certain experiments that need to be maintained confidential. Clearly, allowing Alice to directly read the contents of
the tuple breaks the privacy of Bob; on the other hand, the confidentiality of the database managed by Alice is violated once Bob has access to the contents of the database. Thus, the problem is to check whether the database inserted with the tuple is still k-anonymous, without letting Alice and Bob know the contents of the tuple and the database respectively.

Disadvantage:
1. The database with the tuple data does not be maintained confidentially
2. The existing systems another person to easily access database.

III. PROPOSED SYSTEM:

An algorithm for anonymous sharing of private data among parties is developed. This technique is used iteratively to assign these nodes key numbers ranging from 1 to N. This assignment is anonymous in that the identities received are unknown to the other members of the group. Resistance to collusion among other members is verified in an information theoretic sense when private communication channels are used. This assignment of serial numbers allows more complex data to be shared and has applications to other problems in privacy preserving data mining, collision avoidance in communications and distributed database access. The required computations are distributed without using a trusted central authority.

Advantage:
1. The anonymity of DB is not affected by inserting the records.
2. We provide security proofs and experimental results for both protocols.

MODULES
1. Homomorphic encryption Module.
2. Generalization Module.
3. Cryptography Module.

Homomorphic encryption Module:
This module is used to encode the first protocol is aimed at suppression-based anonymous databases, and it allows the owner of DB to properly anonymize the tuple t, without gaining any useful knowledge on its contents and without having to send to t’s owner newly generated data. To achieve such goal, the parties secure their messages by encrypting them. In order to perform the privacy-preserving verification of the database anonymity upon the insertion, the parties use a commutative and homomorphic encryption scheme.

Generalization Module:
In this module, the second protocol is aimed at generalization-based anonymous databases, and it relies on a secure set intersection protocol, such as the one found in, to support privacy-preserving updates on a generalization based k-anonymous DB.

Cryptography Module:
In this module, the process of converting ordinary information called plaintext into unintelligible gibberish called cipher text. Decryption is the reverse, in other words, moving from the unintelligible cipher text back to plaintext. A cipher (or) crypt is a pair of algorithms that create the encryption and the reversing decryption. The detailed operation of a cipher is controlled both by the algorithm and in each instance by a key. This is a secret parameter (ideally known only to the communicants) for a specific message exchange context.

User and Admin Module:
In this module, to arrange the database based on the patient and doctor details and records. The admin to encrypt the patient reports using encryption techniques using suppression and generalization protocols.

IV. Related work:

This paper builds an algorithm for sharing simple integer data on top of secure sum. The sharing algorithm will be used at each iteration of the algorithm for anonymous ID assignment (AIDA). This AIDA algorithm, and the variants that we discuss, can require a variable and unbounded number of iterations. Finitely-bounded algorithms for AIDA are discussed in Section IX. Increasing a parameter in the algorithm will reduce the number of expected rounds. However, our central algorithm requires solving a polynomial with coefficients taken from a finite field of integers modulo a prime. That task restricts the level to which can be practically raised. We show in detail how to obtain the average number of required rounds, and in the Appendix detail a method for solving the polynomial, which can be distributed among the participants.

V. Conclusion:

With private communication channels, our algorithms are secure in an information theoretic sense. Apparently, this property is very fragile. The very similar problem of mental poker was shown to have no such solution with two players and three cards. The argument of can easily be extended to, e.g., two sets each of colluding players with a deck of cards rather than our deck of cards.

In contrast to bounds on completion time developed in previous works, our formulae give the expected completion time exactly. We conjecture the asymptotic formula of Corollary 9, based on computational experience, to be a true upper bound.

All of the no cryptographic algorithms have been extensively simulated, and we can say that the present work does offer a basis upon which implementations can be constructed. The
communications requirements of the algorithms depend heavily on the underlying implementation of the chosen secure sum algorithm. In some cases, merging the two layers could result in reduced overhead.

VI. REFERENCES:

[26] T. M. Cover and J. A. Thomas, Elements of


Authors:

First Author: Vaheeda Kumari

Kumari M received her B.Tech degree in computer science and engineering in the year 2013 and pursuing M.Tech degree in computer science and engineering from, DJR College of Engineering & Technology.

Second Author:

Sk. N. Rehmatunnisa

M.Tech received her M.Tech degree and B.Tech degree in computer science and engineering. She is currently working as an Assoc Professor in, DJR College of Engineering & Technology.