ASSIGNMENT OF ID USING AIDA FOR SECURE DATA SHARING IN DISTRIBUTED DATA MINING

[1]R. Annamalai Saravanan and [2]Dr. M. Rajesh Babu

[1]Research and Development Center, Bharathiar University, Coimbatore-46

[2]Professor - Department of Computer Science & Engg, Karpagam College of Engineering (Autonomous), Coimbatore-32

[1]annamalai22phd@gmail.com , [2]drmrajeshbabu@gmail.com

Abstract: A distributed system consists of enormous independent devices that appears as a single coherent system to its users. In order to complete the tasks allocated data is shared among several computers in distributed computed systems. Here, security is a main concern in data sharing and so anonymous communication is needed in which, the private data is to be shared anonymously without revealing their identity. The nodes are assigned IDs in such a way that the assigned identities are unknown to others nodes. Using AIDA (Anonymous ID Assignment) algorithm, the IDs are assigned to the N nodes ranging from 1 to N in iterative manner. Based on the choice of sharing the data, the variants of AIDA algorithm are chosen. The sharing of more critical data is possible using AIDA in distributed data access. The required computations are performed without the use of central authority. Algorithms for assigning anonymous IDs are examined with respect to computational needs.

1. INTRODUCTION

A distributed system contains large number of independent computers in which the tasks are broken up and sent out to the nodes where the application processes the data and send back the computed solution data. An important class of distributed system is distributed computing system which is used for high-performance computing tasks.

Main benefits of distributed systems are performance, high availability, scalability, fault tolerance and reliability. But security is a main issue while sharing the secret data among nodes of distributed systems. The identity of a node that needs to send data should not be known to other nodes. Anonymous communications is needed in some areas in order to share data without leaking the identity of the entity. In secure multiparty computation, anonymity is needed to allow parties to carry a computation using their individual data while the data held by each party remains unknown to other parties.

The main to the nodes and the assigned IDs should be known only to the node to which the IDs are assigned.

AIDA algorithm is mainly used aim is to provide an efficient algorithm for assigning unknown identifiers (IDs) to the network nodes and the IDs should be anonymous with no central authority using distributed computation. For N nodes, a permutation of the integers are involved in assigning IDs for sharing simple data anonymously which results in sharing of complex data. Dynamic unique IDs are required in sharing/dividing communication bandwidth, resources and data sources anonymously and without conflict. To provide security for sensor networks and monitoring of individual nodes, the IDs are needed. In grid computing, the IDs need to be anonymous where the services are provided without revealing the identity of the service requestor.

Using secure sum is possible to allow one to opt-out of a computation beforehand based on certain rules in statistical disclosure limitation or during a computation and even to do so anonymously. This algorithm explores the connection between sharing secret data anonymously, anonymous ID assignment and distributed secure multiparty computation.

2. PROBLEM DEFINITION

Though the distributed systems provide many benefits such as scalability, high availability, fault tolerance and reliability, security is yet a main issue which is to be solved. The nodes in the distributed system are semi

honest and so they can trace the protocol which is used to share the in order to retrieve the secure data.

A new algorithm is to be built in which the sharing of data is performed in anonymous manner and to avoid collision resistance. The proposed algorithm should not allow the nodes of a distributed system to retrieve the data by tracking the computational protocol. Assign ID that should be anonymous and also the computational overhead is to be reduced.

3. RELATED WORK

Distributed data mining (DDM), that mines data from multiple sites, offers the miner a larger data set with the possibility of stronger and, perhaps, novel association rule findings. Privacy preserving data mining (PPDM) [1] algorithms attempt to mine distributed data while addressing privacy concerns. Useful algorithmic primitives for PPDM, including secure set union, secure size of intersection, and secure sum (SS). The goal of SS is simple: let the value of each site's individual input be masked while the global sum of all inputs is universally known. A drawback is that for any arbitrary *Ni*, *Vi* can be found if nodes *Ni*−1 and *Ni*+1 collude. Shepard et. al., (2009) present a cycle-partitioned secure sum (CPSS) algorithm, and provide a mathematical proof for its collusion resistance based on edge-disjoint Hamiltonian cycles. There is a chance that a non-viable global sum may be accepted as a viable sum.

The huge growth of the Internet and its easy access by common man created opportunities for joint computations by multiple parties. For the sake of their mutual benefit all the participating parties want to compute the common function of their inputs but at the same time they are afraid about the privacy of their data. This subject of the information security is called Secure Multi-Party Computation (SMC) [2].

The secure sum protocol proposed by Clifton *et al*. [2] used random numbers for privacy of individual data inputs. In this protocol any two parties *Pi-1* and *Pi+1* might collude each other to find the secret data of *Pi* by performing only one computation. We proposed k-Secure Sum Protocol and extended k-Secure Sum Protocol where the probability of data leakage is minimized considerably by splitting the data block of individual party into number of segments. As the number of segments in a data block is increased, the probability decreases. In another paper we have proposed a zero probability protocol ck-Secure Sum.

The rising familiarity of the Internet and its accessibility to high-speed networks with large powerful personal computers, servers and workstations increased the computing usage by leaps and bounds. With the advent of distributed computing with distributed data, sophisticated load balancing and concurrent computing power using clustered servers, the need of collaborative computing is felt. The grid computing has emerged to cater the need of computing-on-demand. In grid computing, [3] geographically dispersed heterogeneous computing stations belonging to diverse administrative domains can connect to the grid and offer or request services in a loosely coupled environment with services provided on-demand. Grid computing takes advantage of idle or unused processing cycles of all capable computers in a network to solve problems, otherwise which may be excessively intensive for single computing station. Similar to electric power grid, the computational grid provides a collaborative infrastructure with easy, consistent and inexpensive access to diverse computational sources belonging to heterogeneous administrative domains through a unified view of single virtual resource. When messages are to be exchanged among grid nodes belonging to public realm, the service providers and service consumers may need to conceal the identity.

The enormous growth of online social networking and mobile phone services has paid increased attention to mobile social networking. Matchmaking is the primary component of mobile social networking [4]. It notifies nearby users who fulfill certain conditions, such as having same interests, and who are therefore considered as good candidates for being added to a user's social network. This approach reveals users more personal information than necessary information. The other approach needs a trusted server that involves in each matchmaking operation. Namely, the server knows the current location and interests of each user and performs matchmaking based on this information. This technique allows the server to track user's location and .information. This paper introduces a privacy-preserving matchmaking for mobile social networking that allows a potentially malicious user know only the interests that he has in common with a neighbor user, but no other information. Additionally, this protocol is distributed and hence it does not require a trusted server that can track users or that needs to be involved in each

matchmaking operation. The complexity of matchmaking increases as the no. of interests increases. Each mobile node possesses various types of interests. And it is not practical on current smartphones.

4. PROPOSED SYSTEM

The proposed work provides an efficient algorithm for assigning anonymous identifiers (IDs) with no central authority using distributed computation. Our main algorithm is based on a method for anonymously sharing simple information and results in methods for efficient sharing of complex information.

There are many applications requiring dynamic anonymous IDs for network nodes. Such IDs can be used as part of schemes for sharing/dividing communications bandwidth, data storage, and other resources anonymously and without conflict. The IDs are needed in sensor networks for security and for administrative tasks requiring reliability, such as configuration and monitoring of individual nodes, and download of binary code or data aggregation descriptions to these nodes.

In such application where IDs need to be anonymous like grid computing where one may seek services without revealing the identity of the service requestor. To differentiate anonymous communication from anonymous ID assignment, consider a situation where parties want to collectively display their data, but anonymously in slots on a trusted third party site.

The IDs are used for assigning the slots to users and anonymous communication can allow the parties to hide their identities from the trusted third party. Use of secure sum algorithm is possible to allow one to opt-out of a computation on the basis of certain rules in statistical disclosure or during a computation to do in an anonymous manner.

A simple algorithm is presented for finding an AIDA which has several variants depending on the choice of the data sharing method at step (3) below. At one ste random integers or "slots" between 1 and are chosen by each node, a node's position will be determined by its position among the chosen slots, but provisions must be made for collisions. The parameter should be chosen so that round chooses. The random numbers are shared anonymously. One method for doing this was given in Section III.

Let denote a revised list of shared values with duplicated and zero values entirely removed where the number of unique random values is the nodes which drew unique random numbers then determine their index from the position of their random number in the revised list as it would appear after being sorted:

- **Algorithm (Find AIDA):** Given nodes, use distributed computation (without central authority) to find an anonymous indexing permutation.

1) Set the number of assigned nodes.
2) Each unassigned node chooses a random number

Update the number of nodes assigned: If then return to step (2).

4.1 Advantages

- More complex data can be shared by using the assignment of serial numbers and provides applications to problems such as collision avoidance in communication, distributed database access and privacy preserving data mining.

- The resulting solution of a polynomial equation can be avoided by using Sturm's theorem.

4.2 Disadvantages

- Here the nodes of the network are semi-in the range 1. A node assigned in a previous honest and they can follow the protocol honestly and try to infer additional information.

- If the nodes of a network are semi-honest, and trusted third party is involved, a permutation can also be included by using an anonymous routing protocol

4.3 Architecture diagram

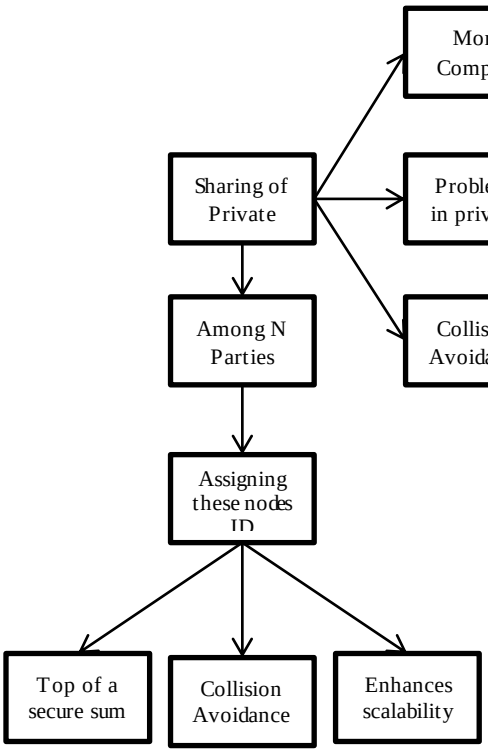

Figure 1: Architecture diagram

5. CONCLUSION AND FUTURE WORK

It provides an efficient assigning of ID to securely data sharing in communication. Each algorithm provides an effective processing of data accessing. It has a several advantages. Communication overhead can be decreased by using Newton's identities. This can allow the use of a more number of "slots" with a consequent reduction in the number of rounds required. The result of a polynomial equation also can be avoided by using Sturm's theorem. The building of a result similar to the Sturm's method over a definite field can be done. This assignment is anonymous in that the identities received are unknown to the other members of the group. Verifying the resistance to collusion among other group is verified in an information theoretic sense when individual communication channels are used.

This assignment of individual anonymous numbers allows complex data to be shared and provides applications to other fields such as privacy preserving data mining, accessing distributed database and collision avoidance in communications. Without using the trusted central authority the distribution of computations are performed. The new algorithms are provided based on the principles of Sturm's theorem and Newton's theorem. An algorithm for solution of certain polynomial is built to enhance the scalability of the algorithms. The communication requirements of the algorithms depend heavily on the underlying implementation of the chosen secure sum algorithm. In some situations, computation overhead can be reduced by merging the two layers. And the number of rounds in assigning Id can be decreased by increasing the parameters.

REFERENCES

[1] Shepard SS, Dong R, Kresman R, Dunning L. Anonymous id assignment and opt-out. InElectronic Engineering and Computing Technology 2010 (pp. 419-431). Springer Netherlands.

[2] Sheikh R, Kumar B, Mishra DK. A Modified ck-Secure Sum Protocol for Multi-Party Computation. arXiv preprint arXiv:1002.4000. 2010 Feb 21.

[3] D. Jana, A. Chaudhuri and B.B. Bhaumik, "Privacy and anonymity protection in computational grid services," Int. J. Comput. Sci. Applicat., vol. 6, no. 1, pp. 98–107, Jan. 2009.

[4] Xie Q, Hengartner U. Privacy-preserving matchmaking for mobile social networking secure against malicious users. InPrivacy, Security and Trust (PST), 2011 Ninth Annual International Conference on 2011 Jul 19 (pp. 252-259). IEEE.

[5] Urabe S, Wang J, Takata T. A collusion-resistant approach to distributed privacy-preserving data mining. InParallel and distributed computing and systems 2004 Nov (Vol. 436, No. 088, pp. 626-631). MIT Cambridge: ACTA Press.