RESEARCH ARTICLE OPEN ACCESS

Private Data Transferring among 'N' Nodes in a group by Using Anonymous ID Assignment

M.Anisha Vergin, PG Scholar
Department of Information Technology,
Francis Xavier Engineering College,
Vannarpettai,Tirunelveli.
Email.id: anisha.vergin@gmail.com

T.Anto Theepak, Assistant Professor
Department of Information Technology,
Francis Xavier Engineering College,
Vannarpettai, Tirunelveli.
Email.id: theepak_a@yahoo.com

Abstract --
To overcome the problems in privacy preserving data mining, collision in communications, distributed database access in the wireless communications, the Secure Sum algorithm is used. This algorithm is also used to provide secure data transmission among 'N' parties in a group by assigning ID numbers ranging from 1 to N. This ID assignment is anonymous in that the identities received are unknown to the other members of the group. When private communication channels used, the resistance to collusion among other members is verified in an information theoretic sense. The required computations are distributed without using a trusted central authority. The algorithms for assigning the anonymous ID is examined by the trade-offs between the requirements of communication and computations. 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 the distributed solution of certain polynomials over finite field will enhances the scalability of the algorithms. The representation of the Markov Chain is used to find the statistics on the required number of iterations.

Keywords— Anonymization, distributed computing systems, secure data mining, privacy protection, security in cooperative communications.

I. INTRODUCTION

The popularity of internet as a communication medium for personal or business use depends in part of its support for anonymous communications. The businesses also have the legitimate reasons to engage in anonymous communication and avoid the problems in identity revelation. For example, to allow the dissemination of summary data without revealing the identity of entity in the underlying data is associated with or to protect the whistle-blower's right to be anonymous and free from political and economical retributions. Cloud-based website management tools will provide the capabilities for a server to anonymously capture the web actions of the visitor. The problem of sharing private data held the individuals who are the subjects of the data cannot be identified the research extensively. A secure computation is widely used in the literature is the secure sum that allows parties to compute the sum of their individual inputs without disclosing the inputs to one another. This function is familiar in data mining applications and also helps characterize the complexities of the secure multiparty computations.

This work deals with efficient algorithms for assigning identities (IDs) to the nodes of a network in the way that the IDs are anonymous using a distributed computation without a central authority. Given N number of nodes, this assignment is essentially a permutation of the integers {1…N} with each of the ID being known only to the node to which it is assigned. Our main algorithm is based on the method for anonymously sharing simple data and results in the methods for efficient sharing of complex data. Such IDs can be used as a part of the schemes for sharing or dividing communication bandwidths, data storage, and other resources anonymously and without any conflicts. This IDs are needed in the sensor networks for the security purposes or for the administrative tasks requiring reliability, such as the configuration and monitoring of individual nodes, and the download of binary codes of these nodes or data aggregation descriptions to these nodes.

To differentiate the anonymous ID assignment from anonymous communication, consider a situation of, where N parties wish to display their datas collectively, but anonymously, in N slots on a third party site. This IDs can be used to assign the N slots to the users, while anonymous communication can allow the parties to hide their identities from the third party.

The work described in this paper explores the connection between sharing secrets in an anonymous manner, distributed secure multiparty computations and the anonymous ID assignment. The use of the term "anonymous" here is differs from its meaning in the research dealing with symmetric breaking and the leader election in the unspecified or anonymous networks. Here, our network is not an anonymous but the participants are specialized in that they are known to others and can be addressed by the others.

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

II. TRANSMITTING A TROUBLE-FREE DATA

Suppose that our group of nodes wish to share the actual data values from their databases rather than relying on only statistical information. That is, each member of the group of N nodes has a data item which is to be communicated to all the other members of the group. Still, the data is to remain anonymous. We develop a method for collusion resistance for this task by using the secure sum as our core communication mechanism.

For Privacy Preserving we use an anonymous algorithm for sharing the key, since the key generation is a vast process and for it more memory allocation is required. In order to reuse the key we are using anonymous id as shown that it has N range of key which can be reused after the completion of the process. Thus the key is in redistributed manner.

The slot selection method was developed, in this variant of AIDA method, each node submits the Euclidean basis vector, but zero except for a single one in component, to a secure sum.

A node which has received an assignment in a previous round, however, submits the zero vectors. The sum of these vectors are computed over the abelian group GF$(1+N)^S$ using secure sum. The random numbers chosen and their multiplicities are simple to determine.

III. SHARING THE COMPLEX DATA WITH THE AIDA

Now consider the possibility of sharing of more complex data among the participating nodes. Each node has a data item of length -bits which wishes to make public but anonymously to the other participants of the group. Since the number of bits per data item and the number of nodes becomes larger, the method becomes infeasible. Instead of accomplishing this sharing, we will make use of an indexing of the nodes. These methods for finding such indexing are developed in subsequent sections. Let as assume that each node has a unique identification (ID) or serial number (1,2,…,N}. Further, if no node has knowledge of the ID number of any other node, and $s_1,\ldots,s_N$ are a random permutation of 1,…,N. This is termed as an Anonymous ID Assignment (AIDA). Such AIDA may be used to assign slots with respect to time and space for communications.

IV. ARCHITECTURE DIAGRAM

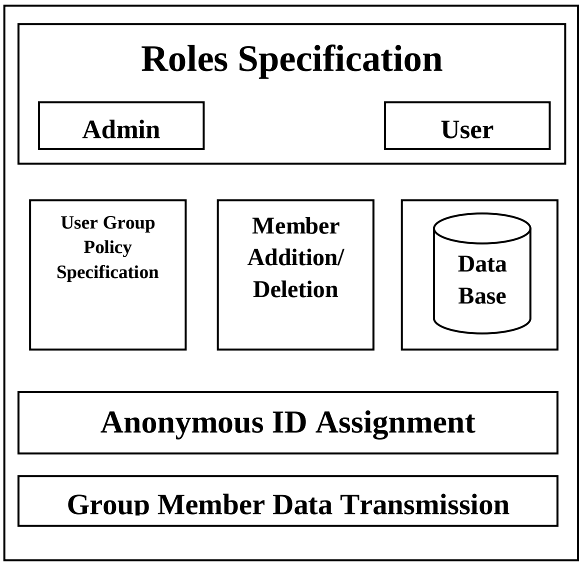
Roles Specification
Admin
User
User Group Policy Specification
Member Addition/ Deletion
Data Base
Anonymous ID Assignment
Group Member Data Transmission

Fig: 1 Architecture of the System

This work deals with efficient algorithms for assigning identities (IDs) to the nodes of a network in such a way that the IDs are anonymous using a distributed computation with no central authority. Given nodes, this assignment is essentially a permutation of the integers with each ID being known only to the node to which it has assigned. The main algorithm is based on the method for anonymously sharing the simple data and results in the methods for efficient sharing of complex datas. There are many applications which requires dynamic unique IDs for the 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. These IDs are needed in sensor networks for security or for the administrative tasks which required reliability, such as its configuration and monitoring of the individual nodes, and the download of binary code or the data aggregation descriptions to these nodes. An application where IDs need to be anonymous is grid computing where one may seek services without revealing the identity of the service requestor.

V. MODULES

1) Homomorphic Encryption Module

In this module, the first protocol is used which is aimed at the repression-based anonymous or unspecified databases, and it allows the owner of the database to anonymize the tuple of the node, without gaining any useful knowledge about its contents and without sending it to the particular tuple's owner that is the newly generated data. In order to achieve this goal, the parties secure their messages by encrypting the corresponding messages. To perform the privacy-preserving verification of the database, the parties use a commutative and a homomorphic encryption scheme.

2) Generalization Module

In this module, the second protocol is used for 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.

3) Cryptography Module

In this module, the process of converting ordinary information will be occur, which is the conversion of plaintext into incoherent rubbish or cipher text, which is called as encryption. Decryption is the conversion of the incoherent cipher text back to plaintext. A cipher is a pair of algorithms that used to create the encryption and the decryption. The specified operation of a cipher is controlled by the algorithm and by a key in each instance. This is a secret parameter (ideally known only to the communicants) for a specific message exchange context.

4) User and Admin Module

In this module, to arrange the database based on the patient's record and the doctor's details. The admin is used to encrypt the patient reports using encryption techniques by using the suppression and generalization protocols.

VI. COMPARISON OF AIDA AND ITS VARIANTS

In the previous section the algorithm to find an AIDA required that the random numbers be shared anonymously at step (3). We now look at three methods which are variants of that procedure. The parameter must be chosen in each case. The expected number of rounds depends only on the selection of S and not on the variant chosen.

A. Slot Selection AIDA

The slot selection method was developed where a more detailed explanation may be found. In this variant of the AIDA algorithm, each node n_i submits the euclidean basis vector e_{ri} € GF $(1+N)^S$, zero except for a single one in component r_i , to a secure sum algorithm. A node which has received an assignment in a previous round, however, submits the zero vector. The sum T of these vectors is computed over the abelian group $GF(1+N)^S$ using a secure sum algorithm. The random numbers chosen and their multiplicities are simple to determine as T_k=Card{i:$r_{i=}$k}.

This variant in the algorithm has its main drawback that is very long message lengths that are encountered when using the large S to keep the number of expected rounds as small.

B. Prime Modulus AIDA

A prime P>S is chosen. Generally,P will be chosen as small as possible subject to this restriction. The random numbers chosen are distributed at step (3) as in Section III using the field F=GF(P) to compute the required power sums, the Newton polynomial p(x) , and the polynomial roots. This variant will be seen to result in shorter message lengths for communication between nodes. Again, the computation required to find the roots of the Newton polynomial is addressed in the appendix. Though, this computation can be delayed and thus overlaps any additional required rounds. Additional rounds of the AIDA algorithm can proceed almost immediately as it is not necessary to solve p(x)=0 before proceeding to the next round. Each node n_i merely computes the derivative polynomial p′(x) and evaluates that polynomial at its chosen random value r_i. The value is a multiple root if and only if $p'(r_i)=0$. Thus, if $p'(r_i)=0$ then node n_i chooses a new random number r_i for use in the next round. If $p'(r_i)\neq 0$ then the n_i has an assignment and for subsequent rounds will use $r_i=0$.

C. Sturm's Theorem AIDA

It is possible to avoid solution of the Newton polynomial entirely. Sturm's theorem allows the determination of the number of roots of a real polynomial p(x) in an interval (a,b) based on the signs of the values of a sequence of polynomials derived from p(x) . The sequence of polynomials is obtained from a variant of the Euclidean Algorithm. As in the previous variant, the power sums are collected and the Newton Polynomial is formed. However, the field used for computation is the field of rational numbers **Q**. The test $p'(r_i)=0$ is again sufficient to determine whether or not n_i has received. There is a computational advantage which is arises in that nodes which do not need to solve the Newton polynomial p(x) to determine the (now implicitly) shared values. Assume that x=0 is not a root of p(x) as x^k has been factored out immediately if applicable. Each node n_i which has received an assignment must count separately multiple roots and also forms $g(x)=gcd(p(x),p'(x))$. A multiple roots version of Sturm's theorem [32] is then applied to calculate the number of roots for the polynomial p(x) in the range $(0,r_i)$. (Note that r_i itself is not a multiple root allowing application of the theorem). The polynomial $g(x)=gcd(p(x),p'(x))$ is a by-product of this computation. The same Sturm procedure is applied to g(x) thus obtaining a count of the multiple roots in the same range,$(0,r_i)$.

The collected power sums P_i are integers. To guarantee the privacy and the compute sums using a field GF(P) with P greater than any possible value of P_i. Our timings showed that using Sturm's theorem is not currently competitive with the various methods of polynomial solution using the "prime modulus" approach and runs twice as slow as best. Although, the construction is straight forward. The application of Sturm's theorem requires the use of an ordered field resulting in the large polynomial coefficients. Unfortunately, the analog of this result which is usable for a finite field of the corresponding polynomial coefficient. Still, some results in this direction are available.

VII. CONCLUSION

For private communication channels, the algorithm is secure in an information theoretic sense. In fact, this property is very delicate. The similar problem of mental poker was shown to have no solution with two players and three cards. The argument can be easily extended as, e.g., two sets each of N colluding players with a deck of 2N+1cards rather than our deck of 2N cards. In contrast to bounds on completion time developed in previously, the formula gives the completion time exactly. We speculate the asymptotic formula based on the computational experience as an original upper bound. The non-cryptographic algorithms have been broadly simulated, and we can say that this work does offer a source upon which the implementations will be constructed. The computational and communications requirements of the algorithms will depend upon the essential implementation of the chosen secure sum algorithm.

Since we are using an Anonymous Id assignment for preserving the private communication channels members, additionally here we are maintaining an indexing of the active group members in a private channel and assign an Anonymous id for the active user based on performance, so that the key assignment becomes more easier way and leads the private communication channel as a secure channel.

REFERENCES

[1] A.Friedman, R. Wolff, and A. Schuster, "Providing k-anonymity in data mining", VLDB Journal, vol. 17, no. 4, pp. 789–804, Jul. 2008.

[2] Junqiang Liu, Ke Wang, "Enforcing Vocabulary *k*-Anonymity by Semantic Similarity Based Clustering", 2010 IEEE International Conference on Data Mining.

[3] Maria E. Skarkala, Manolis Maragoudakis, Hannu Toivonen and Pirjo Moen, "Privacy Preservation by k-Anonymization of Weighted Social Networks", 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

[4] Hillol Kargupta and Souptik Datta, Qi Wang and Krishnamoorthy Sivakumar, "On the Privacy Preserving Properties of Random Data Perturbation Techniques", 2011 IEEE conference of privacy security.

[5] Q. Xie and U. Hengartner,"Privacy-preserving matchmaking for mobile social networking secure against malicious users", in Proc. 9th Ann. IEEE Conf. Privacy, Security and Trust, Jul. 2011, pp. 252–259.

[6] Jin Ma, Xiu-zhen Chen, Jian-hua Li," An Approach to Privacy-Preserving Alert Correlation and Analysis", 2010 IEEE Asia-Pacific Services Computing Conference.

[7] N. Eagle and A. Pentland,"Social Serendipity: Mobilizing Social Software",IEEE Pervasive Computing, 4(2):28–34, 2005.

[8] A. Yao," Protocols for secure computations",in Proc. 23rd Ann. IEEE Symp. Foundations of Computer Science, 1982, pp. 160–164, IEEE Computer Society.

[9] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu, "Tools for privacy preserving distributed data mining", ACM

SIGKDD Explorations Newsletter, vol. 4, no. 2, pp. 28–34, Dec. 2002.

[10] W. Du and M. J. Atallah., "Privacy-Preserving Cooperative Scientific Computations", In 14th IEEE Computer Security Foundations Workshop, pages 273{282, Nova Scotia, Canada, June 11-13 2001.

[11] J. Smith, "Distributing identity", IEEE Robot. Autom. Mag., vol. 6, no. 1, pp. 49–56, Mar. 1999.

[12] 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.

[13] D. M. Goldschlag,M. G. Reed, and P. F. Syverson, "Hiding routing information", in Proc. Information Hiding, 1996, pp. 137–150, Springer-Verlag.

[14] A.Karr, "Secure statistical analysis of distributed databases, emphasizing what we don't know", J. Privacy Confidentiality, vol. 1, no. 2,pp. 197–211, 2009.

[15] J. W. Yoon and H. Kim, "A new collision-free pseudonym scheme in mobile ad hoc networks", in Proc. 7th Int. Conf. Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT'09), Piscataway, NJ, 2009, pp. 376–380, IEEE Press J.W. Yoon and H. Kim, "A perfect collision-free pseudonym system", IEEE Commun. Lett., vol. 15, no. 6, pp. 686–688, Jun. 2011.

[16] J. Castella-Roca, V. Daza, J. Domingo-Ferrer , and F. Sebé, "Privacy homomorphisms for e-gambling and mental poker", in Proc. IEEE Int.Conf. Granular Computing, 2006, pp. 788–791.

[17] E. Karnin, J. Green, and M. Hellman, "On secret sharing systems", IEEE Trans. Information Theory, vol. IT-29, no. 1, pp. 35–41, 1983.

[18] Dennis Hofheinz, Dominique Unruh, "Simulatable Security and Polynomially Bounded Concurrent Composability", Proceedings of the 2006 IEEE Symposium on Security and Privacy (S&P'06) 1081-6011/06 $20.00 © 2006 IEEE.

[19] ,R. Crandall and C. B. Pomerance, "Network Coding: A Computational Perspective", 2nd ed. New York: Springer, 2005.