RESEARCH ARTICLE OPEN ACCESS

Secure Multiparty Computation and Privacy Preserving Data Sharing with Anonymous ID Assignment

Shiny. I.S , S. Gayathri2

1Shiny.I.S. Author is currently pursuing M.E (Computer Science and Engineering) in Vins Christian College of Engineering. e-mail:shinyissacnov7@g mail.com,
2S. Gayathri, currently working as Asst. Professor Department of Computer Science and Engineering, in Vins Christian College of Engineering.

Abstract:
Sharing of private data among N parties was developed by using anonymous sharing. Each member in the group has specific anonymous id. Id received is unknown to the other members of the group. Anonymous id assignment algorithm (AIDA) is utilized for this approach. Serial number allows more complex data to be shared and has applications to other problems in privacy preserving data mining, collision avoidance in communication and distributed data base access. Required computations are distributed with using a trusted administrator. Algorithm for assigning Anonymous id is examined between communication and computational requirement. This paper builds an algorithm for sharing simple integer data on top of secure sum. A secure sum algorithm allows the sum to be collected with some guarantees of anonymity. Secure computation function is popular in data mining applications and also helps characterize the complexities of the secure multiparty computation.Data encryption is an anonymization technique that replaces sensitive data with encrypted data. The process provides effective data confidentiality, but also transforms data into an unreadable format. The Anonymous IDs are needed in sensor networks for security or 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.

Keywords— secure multiparty computation, Anonymization, Deanonymization, privacy preserving data mining, privacy protection.

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 [2]. 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 entity the underlying data is associated with, or to protect whistle-blower’s right to be anonymous and free from political or economic retributions [1]. Cloud-based website management tools provide capabilities for a server to anonymously capture the visitor’s web actions [17]. 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: patient medical records [10], electronic voting [6], e-mail [2], social networking [15], etc. 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 [5]. This function is popular in data mining applications and also helps characterize the complexities of the secure multiparty computation. This work deals with efficient algorithms for assigning identifiers (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 N nodes, this assignment is essentially a permutation of the integers {1…N} with each ID being known only to the node to which it is assigned. The main algorithm is based on a method for anonymously sharing simple data and results in methods for efficient sharing of complex data. There are many applications that require dynamic unique IDs for network nodes[17]. 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 or 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.An application where IDs need to be anonymous is grid computing where one may seek services without divulging the identity of the service requestor[16]. In another application, it is possible to use secure sum to allow one to opt-out of a computation based on the results of the analysis.

II. EXISTING SYSTEM

It allows parties to compute the sum of their individual inputs without disclosing the inputs to one another.In this system database does not maintained anonymously. Anyone can easily access thedatabase. This is the drawback of existing system. Game theoretic approach (LFA) to active distributed data mining approach makes the data mining process as scalable. This scheme is not possible for large data bases. Solution concept can be generated by Nash equilibrium method. Secure sum calculates the sum of values from the individual sites.

Secure multiparty computation allows the parties to jointly compute the sum of their individual input without disclosing the input to another [3], [13]. Secure set union avoids the duplicates during the data mining. In this scheme information are not guarantee that are properly correct, attacker can add additional information to data records. Secure size of set intersection gets the common details during the data mining. Association rule is new information discovered at the result of data mining. In EM clustering items can be partitioned into set of similar elements. Routing information is a part of each packet. By watching the routing information sender and receiver of the data can be easily identified. Onion Routing is a method which limits the network vulnerability. It provides the anonymous socket connection over the computer network [12].

This approach secure only in web service. Homomorphic encryption is used to provide security to E-Gambling. In E-Gambling set of players remotely play a game, for earning money so that security will be needed. Mental Poker protocol guarantees the fairness of the game. Homomorphic properties of cryptosystem can be used in Mental Poker protocols [10]. Homomorphic encryption allows the player to manage the cards co-operatively. Mental Poker protocols use zero-knowledge proof to ensure the honesty of the game. In Entity generated pseudonym scheme entity can generate own pseudonyms. In centralized pseudonym assignment admin collects the set of unique pseudonym to avoid repetition of same pseudonyms. In Hybrid Pseudonym (HP) scheme pseudonyms are locally generated and centrally controlled to prevent collisions [8].

III. PROPOSED SYSTEM

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: patient medical records, electronic voting, e-mail, social networking, etc.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.

A.System Design

In system design Advanced Encryption Standard is used for both encryption and decryption (AES).

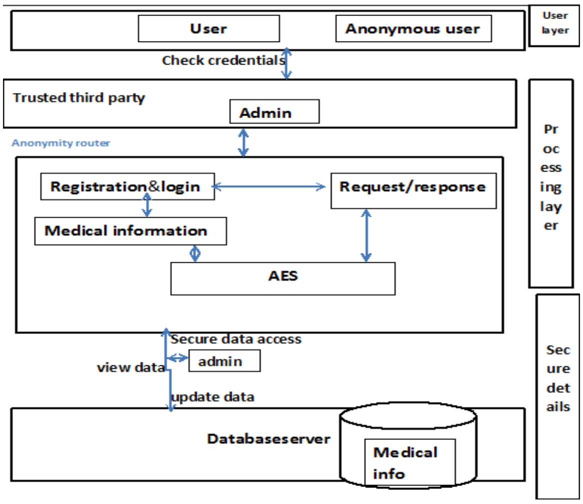

Fig1: System Architecture

B. Review of Secure Sum

Suppose that a group of hospitals with individual databaseswish to compute and share only the average of a data item, such as the number of hospital acquired infections, without revealing the value of this data item for any member of the group. Thus, nodes have data items, andwishto compute and share only the total value. A secure sum algorithm allows the sum to be collectedwith some guarantees of anonymity. Again, assumethe semi-honest model of privacy preserving data mining.Under this model, each node will follow the rules of the protocol, but may use any information it sees during the executionof the protocol to compromise security.Should all pairs of nodes have a secure communicationchannel available; a simple, but resource intensive, secure sumalgorithm can be constructed. In the following algorithm, it is useful to interpret the values as being integer to read first.

C.Data Anonymization Techniques

There are a number of data anonymization techniques that can be used, including data encryption, substitution, shuffling, number and date variance, and nulling out specific fields or data sets. Data encryption is an anonymization technique that replacesthe sensitive data with encrypted data. The process provides effective data confidentiality, but also transforms the data into an unreadable format. For example, once data encryption is applied to the fields containing usernames, "John Doe" may become "@Gek1ds%#$". Data encryption is suitable from an anonymization perspective, but it’s often not as suitable for practical use. Other business requirements such as data input validation or application testing may require a specific data type such as numbers, cost, dates or salary and when the encrypted data is put to use, it may appear to be the wrong data type to the system trying to use it.

D. Transmitting Simple Data with Power Sums

Suppose that our group of nodes wishes to share actual datavalues from their databases rather than relying on only statisticalinformation as shown in the previous section. That is, eachmember of the group of nodes has a data itemwhich is to be communicated to all the other members of thegroup. However, the data is to remain anonymous. A collusion resistant method was developed for this task, using secure sum as underlying communication mechanism.

E.Sharing Complex Data with AIDA1

Now consider the possibility that more complex data is tobe shared among the participating nodes. Each node has a data item of length bits which it wishes to make publicanonymously to the other participants.As the number of bits per data item and the number ofnodes becomes larger, the method of the previous section becomes infeasible. Instead, to accomplish this sharing, will utilize an indexing of the nodes.

F. Comparison of AIDA Variants

The algorithm to find an AIDA required that the random numbers be shared anonymously at step. 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 and not on the variant chosen.

G.Slot Selection AIDA

The slot selection method was developed where a more detailed explanation may be found. In this variant of theAIDA algorithm, each node $n_i$submits the Euclidean basisvector $er_i \in$ G F (N+1)s

Zero except for a single one in component, to a secure sum algorithm. A node which hasreceived an assignment in a previous round, however, submits the zero vectors. The sum T of these vectors is computed overthe Aeolian group using a secure sum algorithm. The random numbers chosen and their multiplicities are simpleton determine as.$T_k = Card\{i: r_i = k\}$ (1)

H.Anonymous ID Assignment

It builds an algorithm for sharing simple integer dataon top of secure sum. The sharing algorithmwill be used at each iteration of the algorithm for anonymous ID assignment (AIDA). This AIDA algorithm, and the variants that already discussed, can require a variable and unbounded number of iterations. 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.

I.Anonymous ID Creation Algorithm

Create an anonymous ID based on a personal ID number and a combination of the Initials and the year of birth (Initials+YoB). An exhaustive search can find the original values matching the anonymous ID unless a strong Key is added for security. Therefore, make sure the personal ID number cannot easily be used to identify a list of Initials+YoB/Personal ID pairs on separate lines in the text field. Initials+YoB and ID fields must be separated by spaces, tabs, commas, or semicolons. Only the first two fields are used, unless the all field box was checked. With the All fields box checked, all text on a row is used to generate an ID. A "-" dash is inserted as a separator. A tab-delimited list of Initials+YoB, Personal ID, and Anonymous ID will be returned. Formats for Initials+YoB and Personal ID fields are not fixed. You can use any number and type of printable characters, (comma nor semicolon) nor white-space characters (space, tab, newline). Note that the fields are case sensitive "june", "June", and "JUNE" all give different Anonymous IDs.The minimum length chosen for the anonymous ID should be at least 1 1/3 times the number of digits of the maximal number of person IDs ever needed.

J. Basic formula

P (collision) =1-e^(-N^2/(2*H))=alpha
N=sqrt (2*ln (1/1-alpha))*sqrt (H)
N: number of evenly distributed hashes to compare
H: the Size of the element count of all possible hashes
H=2^ (IDlength *ln (36)/ln (2))
Hashes calculated with SHAI (IDlength<=30).

K.Communications Requirements of AIDA Methods

Consider the required number of data bits for each of the three variant methods just described. This is the number of data bits that would be transmitted in each packet by the secure sum algorithm introduced earlier. The requirednumbers of data bits B are slightly overestimated by the formulae:

B_{prime} = N. [log_2(P+1)]
B_{prime} = N. [log_2(N)] /2.[log_2(S)]
B_{prime} = S. [log_2(N+1)] (2)

The computational requirements of the “slot selection” appear, at first, to be trivial. However, for every root that the “prime modulus” method must check, “slot selection”.

L.The Completion Rate after R Rounds

Two nodes might make identical choices of random numbers, or slots as they will be termed in this section. One can only guarantee that a complete assignment of nodes using possibilities for slots or random number choices and rounds will occur with at least a desired probability. The formulae are derived by assuming that N -1 node have chosen slots and looking at the next choice. The N_{th} node into choose a slot resulting in assignments and conflicts. The slot it chooses could be unassigned, already in conflict withmultipleoccupants, or already assigned with exactly one
occupant.

IV. EXPERIMENTAL RESULTS

Sl.no	N	H	P
1.	189	45	396
2.	210	55	400
3.	290	95	442
4.	549	225	669
5.	999	450	1108

Table1: Anonymous ID creation Algorithm

Results are given in terms of identities. N represents the number of evenly distributed hashes to compare. H indicates the size of the element count of all possible hashes. P Indicates the value of anonymous ID. This is unique to every members of the group. Anonymous ID received is unknown to other members of the group.Anonymous id is examined between communication and computational requirement.

R Step	A	r1 r2 r3 r4	q1 q2 q3q4	s1 s2 s4
1 2 1 3 1 4 1 5 2 2 2 3	0 0 0 2 2 2	6 10 6 2 6 10 6 2 6 10 6 2 5 0 6 0 5 0 6 0	2 6 6 10 2 10 0 0 5 6	2 2 2 1 32 4 1

Table 2:Random Serial number Generation

Where,
Ri= rounds
r1, r2……r n=Random value within Range S
q1, q2….q n= given serial number

V. CONCLUSION

Proposed paper greatly decreases communication overhead. By using private communication channel that is anonymity router to transmit the data more securely. To overcome the problem of identifying details and changing information anonymous id was utilized. Random serial number is used to identify whether the data requesting person is a correct authorized person or hackers. The use of the Newton identities greatly decreases communication overhead. This can enable the use of a larger number of “slots” with a consequent reduction in the number of rounds required. The solution of a polynomial can be avoided at some expense by using Sturm’s theorem. The development of a result similar tothe Sturm’s method over a finite field is an enticing possibility.

With private communication channels, the algorithms are secure in an information theoretic sense. Apparently, this property is very fragile. The very similar problem of mental pokerwas shown to have no such solution with two players andthree cards. The argument can easily be extended to, e.g., two sets each of N colluding players with a deck of 2N+1 cardsrather than our deck of 2N cards. In contrast to bounds on completion time developed in previous works, our formulae give the expected completion time exactly. All of the no cryptographic algorithms have been extensively simulated, and 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.

REFERENCES

[1] Sarbanes–Oxley Act of 2002, Title 29, Code of Federal Regulations,Part 1980, 2003.

[2] Y.Zhang,W.Liu,andW.Lou.Anonymous communications in mobilead hoc networks. In INFOCOM 2005, 24th annual joint conference the IEEE Computer Societies.Proceeding IEEE, volume 3, pages 1940– 1951, March 2005.

[3] D. Chaum, “Untraceable electronic mail, return address and digital pseudonyms,” *Commun. ACM*, vol. 24, no. 2, pp. 84–88, Feb.1981.

[4] C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu, “Tools for privacy preserving distributed data mining,” ACM SIGKDD ExplorationsNewsletter, vol. 4, no. 2, pp. 28–34, Dec. 2002.

[5] J. Wang, T. Fukasama, S. Urabe, and T.Takata, “A collusion-resistantapproach to privacy-preserving distributed data mining,” IEICE Trans.Inf.Syst. (Inst. Electron. Inf. Commun. Eng.), vol. E89-D, no. 11, pp.2739–2747, 2006.

[6] F. Baiardi, A. Falleni, R. Granchi, F. Martinelli, M. Petrocchi, and A.Vaccarelli, “Seas, a secure e-voting protocol: Design and implementation,”Comput. Security, vol. 24, no. 8, pp. 642–652, Nov. 2005.

[7] Sanil, A. P., Karr, A. F., Lin, X., and Reiter, J. P. (2007). Privacy preserving analysisof vertically partitioned data using secure matrix products.

[8] J. Kong and X. Hong. Anodr: anonymous on demand routing withuntraceable routes for mobile ad-hoc networks. InMobihoc’03:Proceedings of the 4th ACM international symposium on Mobile ad hoc networking ages 291–302, New York, NY, USA,2003.

[9] J. Yoon and H. Kim, “A new collision-free pseudonym scheme in mobile ad hoc networks,”5th workshop on Resource allocation, Cooperation and competition in wireless networks. June 2009.

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

[11] J. Domingo-Ferrer, “A new privacy homomorphism and applications,”Information Processing Letters, vol. 60, no. 5, pp. 277–282, December1996.[Online]. Available: citeseer.nj.nec.com/290190.html

[12] J. Domingo-Ferrer, “A provably secure additive and multiplicative privacyhomomorphishm,” in Information Security, ser. Lecture Notes in Computer Science, A. Chan and V. Gligor, Eds., vol. 2433. SpringerVerlag, 2002, pp. pp. 471–483.

[13] D. M. Goldschlag,M. G. Reed, and P. F. Syverson, “Hiding routing information,”inProc. Information Hiding, 1996, pp. 137150,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] Q. Xie and U. Hengartner, “Privacy-preserving matchmaking for mobilesocial networking secure against malicious users,” in Proc. 9thAnn. IEEE Conf. Privacy, Security and Trust, Jul. 2011, pp. 252–259.

[16] D. Jana, A. Chaudhuri, and B. B. Bhaumik, “Privacy andanonymityprotection in computational grid services,” Int. J. Comput. Sci. Applicat.,vol. 6, no. 1, pp. 98–107, Jan. 2009.

[17] J. Smith, “Distributing identity [symmetry breaking distributed access protocols],” IEEE Robot. Autom.Mag., vol. 6, no. 1, pp. 49–56, Mar.1999.

[18] U. Maurer, “Secure multi-party computation made simple,” in Proc.3rd Int. Conf. Security in Communication Networks (SCN’02), Berlin,Heidelberg, 2003, pp. 14–28, Springer-Verlag.

[19] R. Canetti, C. Dwork, M. Naor, and R. Ostrovsky, “Deniable encryption,”in*CRYPTO*, ser. Lecture Notes in Computer Science, B. S. K.Jr., Ed. New York: Springer, 1997, vol. 1294, pp. 90–104.

[20] J. A. Eidswick, “A proof of Newton’s power sum formulas,” Amer.Math. Monthly, vol. 75, no. 4, pp. 396–396, Apr. 1968.

[21] S. S. Shepard, R. Dong, R. Kresman, and L. Dunning, “Anonymous idassignment and opt-out,” in Lecture Notes inElectrical Engineering,S. Ao and L. Gleman, Eds. New York: Springer, 2010, pp. 420–431.