ISSN 23499842(Online), Volume 1, Special Issue 2(ICITET 15), March 2015
International Journal of Innovative Trends and Emerging Technologies

SECURITY AND PRIVACY PRESERVING DATA SHARING USING ANONYMOUS IDENTITY ASSIGNMENT (AIDA)

B TULASI[1], N.PAVITHRA SAHARI[2], M GEETHA[3], O. PANDITHURAI[4]
[1]UG [Scholar], [1, 2, 3 ,45]Department of Computer Science and Engineering,
[1, 2, 3 ,4]Rajalakshmi Institute of Technology, Chennai, India
[1]tulasirit@gmail.com , [2]pavithrasahari@gmail.com, [3]geethapandian@yahoo.com, [4]pandics@gmail.com

***Abstract*-It is very important that the people be able to exercise control over how their personal data is distributed, shared and sold. The control of information remains in the hands of the owner of that information. Hence, there is a necessity to maintain the privacy and security of shared data communication across many entities in the distributed system. Anonymity which is used in the multiparty security, allows multiple parties on a network in collaboration carry out a global computation. The computations depend on data from each party while the data held by each party remains unknown to other parties. Existing and new algorithms are examined for assigning anonymous IDs to the nodes of the network where the IDs are anonymous using a distributed computation. These IDs which are created when the user is registered are considered as dummy IDs.**

Keywords -Anonymity,multiparty security,distributed computation.

I. INTRODUCTION

Encryption, or the art of concealing information, dates back thousands of years. Encryption only hides what is being said. It does not hide who is talking to whom. This information by itself can be used to one's advantage or disadvantage. Anonymity as used in the multiparty security, allows multiple parties on a network to jointly carry out a global computation. The computations depend on data from each party while the data held by each party remains unknown to other parties. The application can use either private or public anonymous proxy servers, encryption or even spoofing to ensure an anonymous and/or difficultly traceable access to a resource. Most commentary on the Internet is essentially done anonymously, using unidentifiable false user names. While these usernames can take on an identity of their own, they are anonymous from the actual author. Their distributed technology approach may grant a higher degree of security than centralized anonym zing services where a central point exists that could disclose one's identity. It was shown that it is possible to compromise the anonymity of many Internet users by a group of collaborating eavesdroppers even when the most central routers are protected from eavesdropping. The relevance of anonymity in various application domains: whistle blowers, human rights work, health care, elections, e-cash, political speech. Another form of anonymity, Alice wants to talk to Bob without anyone, including Bob, knowing her identity (sender anonymity). She wants Bob to reply without anyone knowing her identity (receiver anonymity). A network is established across multiple heterogeneous users. Not only the critical data but also their identities should not be revealed to other members of the network. The critical data is to be shared with maximum privacy among trusted members when demanded.

The work reported in this paper further explores the connection between sharing secrets in an anonymous manner using distributed secure multiparty computation and anonymous ID assignment. The use of the term "anonymous" here differs from its meaning in various research dealing with applications such as leader election in anonymous networks. Various networks are not anonymous and the participants are identifiable in that they are known to and can be addressed by the others. These IDs can be used for sharing/dividing communication bandwidth, data storage, and other resources without conflict. The IDs are commonly needed in sensor networks for security or for administrative tasks, 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.

II. RELAED WORK

Larry A. Dunning, et al [1] proposes a technique which is used to assign the nodes ID numbers iteratively. This assignment is anonymous where the identities received are unknown to the other members of the group. The limitation of this technique is that the heterogeneity is not achieved.

Q. Xie, et al [2]proposes that matchmaking is a key component of mobile social networking where it notifies users having shared interest and adds them to theusers social network. The limitation is that this approach reveals more personal information than necessary.

A. Yao, et al [18] deals with secret voting such that suppose a committee of m members wish to decide on an yes-no action, each member is to write an opinion. The results are obtained without knowing

the opinion of any other members. This paper proposes private querying of database supposing Alice wishes to compute a function f1 and is asking a query from the database query system, then the query is answered without knowing anything else about the data in it. The limitation in this paper is that the computation requires a larger number of roundsinorder to serve anonymity.

J. Smith, et al [21]proposes that the distributed protocol makes use of hardware random generation to create unique ids. Random IDs of sufficient length are independently chosen and does not require any communication. The limitation is that computation time is high.

J. Castellà-roca, et al [4] that e-gambling requires standards of security similar to those in physical gambling. Cryptographic tool have been commonly used so far to provide security to e-gambling. They offer the possibility of manipulating cards in encrypted form.

III. MOTIVATING APPLICATIONS

We put forward two applications to make the above paradigm concrete.

Application 1: Document Sharing

Enterprise R is shopping for technology and wishes to find out if enterprise S has some intellectual property it might want to license. However, R would not like to reveal its complete data, nor would S like to reveal all its unpublished data of intellectual property. Rather, they would like to first, find the specific technologies for which there is a match, and then reveal information only about those technologies. This problem can be abstracted as follows. We have two databases DR and DS, where each database contains a set of documents. The document shave been preprocessed to only include the most significant words, using some measure such as term frequency times inverse document frequency. We wish to find all pairs of similar documents dR 2 DR and dS 2 DS, without revealing the other documents. In database terminology, we want to compute the join of DR and DS using the join predi-catef(jdR \ dSj;jdRj;jdSj) >, for some similarity function f and threshold. The function f could be jdR\dSj=(jdRj+jdSj), for instance. Many applications map to this abstraction. For example, two government agencies may want to share documents, but only on a need-to-know basis. They would like to find similar documents contained in their repositories in order to initiate their exchange.

Application 2: Medical Research

Imagine a future where many people have their DNA sequenced. A medical researcher wants to validate a supposition connecting a DNA sequence D with a reaction to drugG. People who have taken the drug are separated into four groups, based on whether they had an adverse reaction and whether their DNA contained the specific sequence; the researcher needs the number of people in each group. DNA sequences and medical records are stored in databases in autonomous enterprises. Due to privacy concerns, the enterprises do not wish to provide any information about an individual's DNA sequence or their medical records, but still wish to help with the research.Assume that the table TR(person id, pattern) stores whether a person's DNA contain patternD and TS(person id, drug, reaction) checks whether a person took drug G and whether the person had an adverse reaction. TR and TS belong to two different enterprises. The researcher should get to know the counts and nothing else, and the enterprises should not learn any new information about any individual.

IV. EXISTING SYSTEM

Here we discuss some existing techniques that is used for building the above applications, and why they are insufficient.

Trusted Third Party: The main parties give the data to a "trusted" third party and have the third party do the computation. However, the third party has to be completely trusted with respect to competence against security violation. The level of trust is expected is to be too high for this solution to be agreed.

Secure Multi-Party Computation: Given two parties with inputs x and y respectively, the goal of secure multi-party computation is to compute a function f(x,y) such that the two parties learn only f(x,y), and not x, y.

LIMITATIONS:

Schema discovery and Heterogeneity: We do not address the question of how to find which database contains which tables and what the attribute names are. We assume that the database schemas are known and we also do not address the issues of heterogeneity.

V. PROBLEM STATEMENT:

The problem we study in this paper is now formally stated.

Problem Statement (Ideal): Let there be two parties R (receiver) and S (sender) with databases DR and DS respectively. Given a database query Q spanning the tables in DR and DS, compute the answer to Q and return it to R without revealing any additional information to either party.

Problem Statement (Minimal Sharing): Let there be two parties R and S with databases DR and DS respectively. Given a database query Q spanning the tables in DR and DS, and some categories of information I, compute the answer to Q and return it to R with- out revealing any additional information to either party except for information contained in I. 2
For example, if the query Q is a join TR 1 TS over two tables TR and TS, the additional information I might be the number of records in each table: jTRj and jTSj. Note that whatever R can infer from knowing the answer to the query Q and the addi- tional information I is fair game. For instance, if the query Q is an intersection VS \ VR between two sets VS and VR, then for all v 2 (VR (VS \VR)), R knows that these values were not in VS. We assume that the query Q is revealed to both parties. One can think of other applications where the format of Q is revealed, but not the parameters of Q}

MINIMAL INFORMATION SHARING USING SECURE SUM:
Efficient algorithms are developed for secure sum data mining operation. It is a secure computation that allows parties to compute the sum of their individual inputs without disclosing the inputs to one another and also helps characterize the complexities of the secure multiparty computation. Let us consider a group of hospitals with individual databases. The hospitals compute and share only the average of a data item, such as the number of hospital acquired infections. The value of the data item is not revealed to any member of the group. Thus N nodes n1, n2,...nN have data items d1, d2,.... dN, and wish to compute and share only the total value T = d1 + d2 + ... dN.

SHARING COMPLEX DATA WITH AIDA:
The focus is to develop efficient algorithms on top of a secure sum data mining operation. The identities received are unknown to other members of the network. Our algorithm is based on a method for anonymously sharing simple data and results in methods for efficient sharing of complex data.
Anonymous Identity Assignment (AIDA) is used for assigning Identities to the nodes of a network. The IDs are anonymous and required computations are distributed using a central anonymous server. This assignment is essentially a permutation of the integers with each ID being known only to the node to which it is assigned. Consider a situation where parties wish to display their data collectively, but anonymously on a third party site. The anonymous IDs are used to assign the slots to users, while the anonymous communication allows the parties to conceal their identities from the third party.

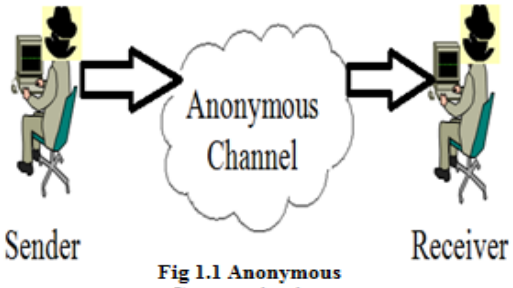


Fig 1.1 Anonymous Communication

Consider the possibility that more complex data is to be shared amongst the participating nodes. Each node has a data item of length b-bits which it wishes to make public anonymously to the other participants. As the number of bits per data item and the number of nodes becomes larger, the methods performed in the previous section becomes infeasible. Instead, to establish this sharing, we will utilize an indexing of the nodes. Methods for finding such an indexing are developed in subsequent sections. Assume that each node has a unique identification (ID) and no node has knowledge of the ID number of any other node.

Find AIDA:
Given N nodes, use distributed computation to find an anonymous indexing permutation.
1) Set the A = 0 [Initial number of assigned nodes].
2) Each unassigned node chooses a random number in the range 1 to S. A node assigned in a previous round chooses r = 0.
3) The shared values are denoted by q.
4) Consider k is the number of unique random values. The index of the nodes is derived from the position of their random number in the revised list after being sorted:
s = A + Card {q <= k}
5) Now, the number of nodes assigned A = A + k
6) If A < k then return to step (2).

Protocol for Secure Computations:
For definiteness, suppose Alice requests for the current time to Bob considering the central server anonymous. Now Alice is sending a query to the database requesting for the current time. The request is made anonymous using the anonymous server and is sent Bob. Bob has no knowledge about Alice while he can view only the query from the server (sender anonymity). Using secure computations, Bob responds to the server and the server then sends the current time to Alice without revealing Bob's identity (receiver anonymity).

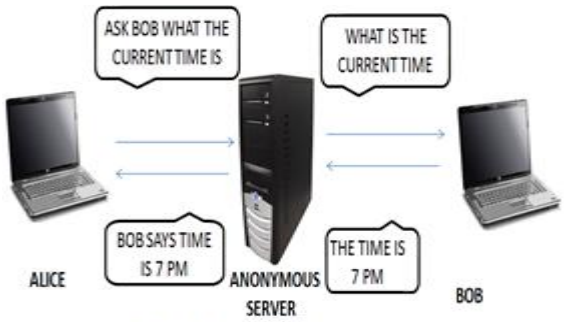
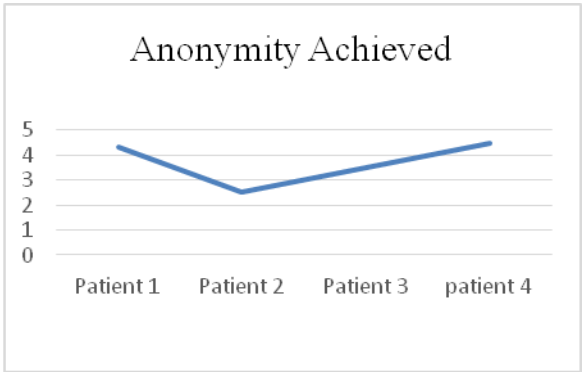


Fig 1.2 Secure Computation

VI. EXPERIMENT AND RESULTS

The proposed application is made available to a variety of hospitals. In this system the patient's medical records are saved and the system keeps track if these records. The information related to the patients are stored anonymously in the database. To achieve this, the health care management system's information is stored and retrieved by preserving its anonymity. Whenever the patient's information is entered, it is stored into the anonymous server and an anonymous identity is created in the database. This ID is unique and concealed. The doctor's information is also fed into the system in order to identify who treated the patient in case of any emergency. During the time of retrieval, only the patient's information can be viewed, whereas the anonymous identities remain unknown.



VII. CONCLUSION

Each algorithm compared in the above sections can be reasonably implemented and each has its own advantages. With private communication channels, our algorithms are secure in an information hypothetic sense. The problem of mental poker is also similar to our algorithm and has 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 non-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 secure sum algorithm.

REFERENCES

[1]Larry A. Dunning, Member, IEEE, and Ray Kresman "Privacy Preserving Data Sharing With Anonymous ID Assignment", *IEEE transactions on information forensics and security*, vol. 8, no. 2, February 2013.

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

[3]S. S. Shepard, R. Dong, R. Kresman, and L. Dunning, "Anonymous id assignment and opt-out," in *Lecture Notes in Electrical Engineering*, S. Ao and L. Gleman, Eds. New York: Springer, 2010, pp. 420–431.

[4] J. Castellà-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.

[5]R. Agrawal and R. Srikant, "Privacy-preserving data mining," in *Proceedingsof ACM SIGMOD Conference*, 2000.

[6]K. Chen and L. Liu, "A random rotation perturbation approach to privacy preserving data classification," in *Proceedings of International Conference on Data Mining (ICDM)*, 2005.

[7]A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke, "Privacy preserving mining of association rules," in *Proceedings of ACM SIGKDDConference*, 2002.

[8] A Survey: Electronic Voting Development and Trends KomministWeldemariam and Adolfo VillafioritaFondazione Bruno Kessler, Center for Scientific and Technological Research (FBK-IRST) viaSommarive 18 I-38050 Trento, Italy.

[9]Gmatch: Secure and Privacy-Preserving Group Matching in Social Networks Boyang Wang , Baochun Li and HuiLiState Key Laboratory of Integrated Services Networks, Xidian University, Xi'an, China Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada.

[10] S. Urabe, J. Wang, and T. Takata, "A collusion-resistant approachto distributed privacy-preserving data mining," in *Parallel and DistributedComputing*

and Systems, T. Gonzalez, Ed. MIT Cambridge:ACTA Press, Nov. 2004, vol. 436, no. 088, pp. 626–631.

[11] 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, Spinger-Verlag.

[12] R. Canetti, "Security and composition of multi-party cryptographic protocols," *J. Cryptol.*, vol. 13, no. 1, pp. 143–202, 2000.

[13] S. Ajmani, R. Morris, and B. Liskov.A trusted third-party computation service. Technical Report MIT-LCS-TR-847, MIT, May 2001.

[14] B. Chor and N. Gilboa.Computationally private information retrieval. In *Proc. of 29th ACM Symposiumon Theory of Computing*, pages 304–313, 1997.

[15] B. Chor, O. Goldreich, E. Kushilevitz, and M. Sudan.Private information retrieval. In *IEEE Symposium on Foundations ofComputer Science*, pages 41–50, 1995.

[16] D.Dobkin,A. Jones, andR. Lipton. Secure databases: Protection against user influence. *ACM Transactions on DatabaseSystems*, 4(1):97–106,March 1979.

[17] A. Evfimievski, R. Srikant, R. Agrawal, and J. Gehrke.Privacy preserving mining of association rules.In *Proc. of the8th ACMSIGKDD Int'l Conference on Knowledge Discoveryand Data Mining*, Edmonton, Canada, July 2002.

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

[19] 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.

[20] O. Goldreich, S. Micali, and A. Wigderson, "How to play any mental game," in *Proc. 19th Ann. ACM Conf. Theory of Computing*, Jan. 1987, pp. 218–229, ACM Press.

[21]J. Smith, "Distributing identity [symmetry breaking distributed access protocols]," *IEEE robot. autom.* Mar. 1999