IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 2, Issue 5, Oct-Nov, 2014
ISSN: 2320 – 8791 (Impact Factor: 1.479)
www.ijreat.org

Discretion Conserving Data Distribution with Anonymous ID

Chiranjeevi Mathe[1], Narasimha Rao[2], Venkataramaiah K[3]

[1]M.Tech Scholar, Department of Computer Science & Engineering, Chebrolu Engineering College, Chebrolu, Andhra Pradesh, India.

[2]Assoc. Professor, Department of Computer Science & Engineering, Chebrolu Engineering College, Chebrolu, Andhra Pradesh, India.

[3]Assoc. Professor, Department of Computer Science & Engineering, Chebrolu Engineering College, Chebrolu, Andhra Pradesh, India.

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:*Anonymization, cloud and distributed computing systems, privacy preserving data mining, privacy protection, security and trust in cooperative communications.*

1. 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 entity 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: 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. 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. 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. 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.

2. Literature Survey

Confidentiality, integrity and availability are requirementsfor just about any secure communication system. Nosystem can fulfil all three requirements in mostcircumstances; however the probability of the successfulattack that compromises some of the three needs to beminimal. The confidentiality requirement ensures that anemail's content must remain secret between its sender andits particular intended recipient. This content mustconsequently be protected in a way against unauthorizedreaders. Normally, this is done with a couple kind ofencryption, but ciphering alone is insufficient:

The communicating parties must also authenticate eachother to make certain what it's all about actually emanatesfrom the proper source and reaches the proper destination.Without an authentication mechanism it is also possiblefor a man in the centre to intercept what it's all about bypretending to be the recipient towards the sender andbecause the sender towards the recipient, rendering anyencryption useless. Maintaining availability is veryimportant. If the adversary can disable a secure way ofcommunication, it might be capable of force thecommunication to some less secure channel. If you wantto maintain availability, a communication system shouldnever have any single point of failure.

The purpose of failure can be technical, for instancereliance upon an individual server. In this case amalfunction within the server or possibly a denial ofservice attack against it might disable the system. It isalso organizational, in which particular case an individualentity has the strength to disable the system. The priorparagraphs have listed many requirements to get a secureand anonymous communication system. Theserequirements are summarized from the following list:

- The system must use cryptographically strongencryption to defend the information ofmessages.
- The communication parties will need to have theopportunity to authenticate one another by usinga cryptographically strong authenticationscheme.
- The system should have a robust and secureprocedure to manage and exchange encryptionkeys and codes.
- There should be no single point of failure withinthe system, whether it is organizational ortechnical.
- The system need to be resilient to attacks againstits infrastructure that make an effort tocompromise availability.
- The system will need to have the opportunity tohide the identities of communication participantsfrom third parties.

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.

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. 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.

- First of all, data confidentiality should be guaranteed.
- Secondly, personal information (defined by a user's attributes) is at risk because one's identity is authenticated according to his information.
- Last but not least, the cloud computing system should be resilient in the case of security breach in which some part of the system is compromised by attackers.

The privacy of the data must be preserved while disclosing it to third party or while placing it in long time storage.

According to Data Protection Act,1998 (DPA) 'Personal Data' can be defined as data related to living individual who can be identified from that data or from that data and other information which includes expression of opinion about the individual.Privacy preserving methods generally there are wide andvaried methods in preserving privacy of data in cloud.

- Anonymity-based Method
- A privacy-preserving Architecture
- Privacy-Preserved Access Control
- A Privacy Preserving Authorization System
- A Privacy Preserving Data Outsourcing.
- PccP Model for Cloud
- Dynamic Metadata Reconstruction

3. Proposed System Analysis and Design

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.

In this paper we have discussed about *anonymity-based method.* Our 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.There are many applications which requires unique dynamic IDs for network nodes. An application where IDs need to be anonymous is grid computing where one may take service without disclosing the identity of service requestor.

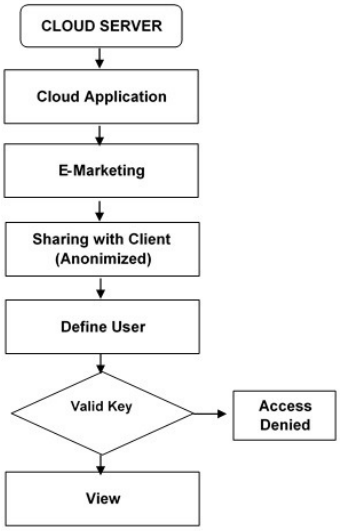

Figure 1 System Flow

An algorithm for anonymous sharing of private data among parties is developed. This technique is used iteratively to assign these nodes ID 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.The assignment of serial numbers allows more complex data to be shared. The required computations are distributed without using a trusted central authority.

3.1 Algorithms

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. There are some algorithms are as follows:

3.1.1 Secure SUM

Given nodes n_1, n_2,……n_n each holding an data item $d_i$from a finitely representable abelian group, share the values T= $\Sigma\, d_i$ among the nodes without revealing the values d_i .

1) Each node n_i, where i = 1,….N chooses random values $r_{i,1}$,….$r_{i,N}$ such that

$$r_{i,1},\ldots.r_{i,N} = d_i$$

2) Each "random" value $r_{i,j}$is transmitted from node $n_i$to node n_j. The sum of all these random numbers $r_{i,j}$is, of course, the desired total T.

3) Each node n_j totals all the random values received as:

$$s_i = r_{i,1}+\ldots.+r_{i,N}$$

4) Now each node n_i simply broadcasts s_j to all other nodes so that each node can compute:

$$T = S_1,\ldots.S_n$$

3.1.2 Anonymous Data Sharing With Power Sums

Given nodes n_1, n_2,……n_n each holding an data item $d_i$from a finitely representable field F make their data items public to all nodes without revealing their sources.

1) Each node n_i computes d_i^n over the field F for$n = 1,\ldots N$. The nodes then use secure sum to share knowledge of the power sums:

$$P_1 = \sum_{i=1}^{N} d_i^1 \qquad P_1 = \sum_{i=1}^{N} d_i^2 \qquad \ldots. \qquad P_1 = \sum_{i=1}^{N} d_i^N$$

2) The power sums $P_1, \ldots\ldots P_N$ are used to generate a polynomial which has $D_1, \ldots\ldots D_N$ as its roots using Newton's Identities as developed in. Representing the Newton polynomial as:

$$p(x) = C_{Nx^n} + \cdots..+C_{1x} + C_0$$

the values $SC_0, \ldots. C_N$are obtained from the equations:

$$C_N = -1$$

$$C_{N-1} = -\frac{1}{1}(C_N P_1)$$

$$C_{N-2} = -\frac{1}{2}(C_{N-1}P_1 + C_N P_2)$$

...

...

$$C_{N-m} = -\frac{1}{m}\sum_{k=1}^{m} C_{N-m+k}\, P_k$$

3) The polynomial *p(x)* is solved by each node, or by a computation distributed among the nodes, to determine the roots d_1,.... d_n.

3.1.3 Sharing complex data with AIDA

Given nodes $n_1, \ldots . n_N$ use distributed computation (without central authority) to find an anonymous indexing permutation $s: \{1, \ldots, N\} \rightarrow \{1, \ldots, N\}$

1) Set the number of assigned nodes$A = 0$.

2) Each unassigned node n_i chooses a random number in the range 1 to S. A node assigned in a previous round chooses$r_i = 0$.

3) The random numbers are shared anonymously. One method for doing this was given in Section III. Denote the shared values by $q_1 \ldots \ldots q_N$

4) Let $q_1 \ldots \ldots q_k$ denote a revised list of shared values with duplicated and zero values entirely removed where k is the number of unique random values. The nodes n_i which drew unique random numbers then determine their index s_i from the position of their random number in the revised list as it would appear after being sorted:

$$s_i = A + Card\{q_j : q_j \leq r_i\}$$

5) Update the number of nodes assigned: A = A + k

6) If A < N then return to step (2).

4. Conclusions

The actual suggested system should be to secure privacyof shared data by Anonymous ID Assignment, byimplementing discussed algorithms. This systemeffectively preserves both information utility andindividual's privacy. Privacy preserving keepsgrowing field of research. It can be clear that we nowhave much privacy preserving techniques availablehowever they have got shortcomings. Anonymitytechnique gives privacy protection and usability ofknowledge. This system will secure anonymous sharing ofprivate data by anonymous ID assignment.

References

[1] Information Commissioner's office, "Anonymization: managing data protection risk, code of practice", 2012

[2] Tomas Isdal, Michael Piatek, Arvind Krishnamurthy, Thomas Anderson: "Privacy-preserving P2P data sharing with OneSwarm".

[3] A. Shamir, "How to share a secret," Commun. ACM, vol. 22, no. 11, pp. 612–613, 1979.

[4] Denis Reilly, Chris Wren, Tom Berry, "Cloud Computing_Pros and Cons for Computer Forensic Investigations" International Journal Multimedia and Image Processing (IJMIP), Volume 1, Issue 1, March 2011

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

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

[7] 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.

[8] 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.

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

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

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

[12] 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.

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

[14] T. Jothi Neela1* and N. Saravanan2A. Dunning "Privacy Preserving Approaches in cloud – A Survey" International journal Of Science and Technology.