ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

Secure Multiparty Protocol for Data Privacy in Distributed Environment

T.Renuka Devi[1], K.G.S. Venkatesan[*2]

Assistant Professor, Dept. of CSE, Jerusalem college of Engineering, Chennai, Tamil Nadu, India[1]

Assistant Professor, Dept. of CSE, Bharath University, Chennai, Tamil Nadu, India[2]

* Corresponding Author

ABSTRACT: Secure Multiparty Computation (SMC) protocols are one of the first techniques used in privacy preserving data mining in distributed environments. The protocol does not reveal anything other than the output of the function or anything that can be computed from it in polynomial time. Moreover, the protocol does not require a trusted third party. The main focus of this paper is to design the look ahead approach for SMC protocol with the help of distributed k-anonymity technique. These protocols prevent information disclosure other than the objective function. People are jointly conducting computation tasks based on the private inputs they each supplies. These computations could occur between mutually un-trusted parties, or even between Secure Multiparty Computations (SMC). The look-ahead operation is highly localized and its accuracy depends on the amount of information the parties are willing to share.

KEYWORDS: Privacy, secure multiparty computation, anonymity technique.

I. INTRODUCTION

Secure multiparty computation (SMC) protocols are one of the first techniques used in privacy preserving data mining in distributed environments. The idea behind these protocols is based on theoretical proof that two or more parties, both having their own private data, can collaborate to calculate any function on the union of their data [8]. While doing so, the protocol does not reveal anything other than the output of the function or anything that can be computed from it in polynomial time. More-over, the protocol does not require a trusted third party[1]. While these properties are Promising for privacy preserving applications, SMC may be prohibitively expensive. In fact, many SMC protocols for privacy preserving data mining suffer from high computation and communication costs. Furthermore, those that are closest to be practical are designed for the semi honest model, which assumes that parties will not deviate from the protocol. Theoretically, it is possible to convertprotocols in the semi honest model into protocols in the malicious model. However, the resulting protocols are even more costly[2]. To the best of our knowledge, this is the first work that looks ahead of an SMC protocol and gives an estimate for We state that an ideal look ahead satisfies the following:

1. The methodology is highly localized in computation, it is fast and requires little communication cost (at least asymptotically better than the SMC protocol).
2. The methodology relies on non sensitive data, or better, data that would be implied from the output of the objective function.

II. RELATED WORKS

In this section, we outline a number of characteristics we consider crucial to the design of a practical privacy criterion[10]. At the same time, we review the literature, indicating how previous work does not match our desired characteristics.

From our perspective, a practical privacy criterion should display the following characteristics:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

1. Intuitive: The data owner (usually not a computer scientist) should be able to understand the privacy criterion in order to set the appropriate parameters.
2. Efficiently checkable: Whether a release candidate satisfies the privacy criterion should be efficiently checkable.
3. Flexible: In data publishing, the data owner often considered tradeoff between disclosure risk and data utility. A practical privacy criterion should provide this flexibility.
4. External knowledge: The privacy criterion should guarantee safety in the presence of common types of external knowledge.
5. Value-centric: Often, different sensitive values have different degrees of sensitivity (e.g., AIDS is more sensitive than flu).

Thus, a practical privacy criterion should have the flexibility to provide different levels of protection for different sensitive values, not just uniform protection for all the values
in the sensitive attribute[4]. We call the latter attribute-centric. An attribute-centric criterion tends to over-protect the data. For example, to protect individuals having AIDS, the data owner must set the strongest level of protection, which is not necessary for individuals having flu. Instead, we take the more flexible value-centric approach[5]. 6. Set-valued sensitive attributes: In many real-world scenarios, an individual may have several sensitive values, e.g., diseases. No existing privacy criterion fully satisfies our desiderata. The most closely-related work is that of Martin et al.

While groundbreaking in the treatment of external knowledge, the approach has several important shortcomings:
- The knowledge quantification is not intuitive. It is hard to understand the practical meaning of k- implications.
- Martin et al. showed that their language can express any logic-based expression of external knowledge, when the number k of basic implications is unbounded[9]. However, their language cannot practically express some important types of knowledge, e.g., simply Flu ∈ Bob[S] (a very common kind of knowledge that the adversary may obtain from a similar dataset). Expressing such knowledge in their language requires (|S|−1) basic implications, where |S| is the number of sensitive values. However, with this number of basic implications, no release candidate can possibly be safe[7].

Thus, Flu ∈ Bob[S] will never be used in their criterion.
- The privacy criterion is attribute-centric, and there is no straightforward extension of the proposed algorithm to the more flexible value-centric case[11]. The reason is that the algorithm can only compute max {Pr(s ∈ t[S] | K, D*)} for the sensitive value that is most frequent in at least one QI-group. However, the sensitive values that need the most protection (e.g., AIDS) are usually infrequent ones.
- Each individual is assumed to have only one sensitive value. Our work builds upon and addresses the above issues. Note that our language can express some knowledge (e.g., Flu∈ Bob[S]) that cannot be practically expressed in their language, and vice versa.

III. METHODOLOGY

The earlier section demonstrated the viability of our approach using an example with eight potentially identifying attributes. In general, the size of the solution space depends on the number of such attributes and the granularity at which they need to be considered. Determining which attributes should be considered as potentially identifying is based on an assessment of possible links to other avail-able data[13]. This needs to be done with typical databases in each domain (e.g., retail). Clearly, as the number of potentially identifying attributes grows, identity disclosure risk in-creases[14]. The corresponding increase in the number of unique combinations of potentially identifying values will have an impact on the k-anonymity approach[15]. Also, the complex-it y of the optimization problem increases due to the larger solution space to be searched.

Further experiments are needed to investigate the applicability of this approach to wider data sets[17]. In each domain, in addition to the identifying attributes one needs to determine the sensitive attributes. It has been suggested that sensitive attributes be removed completely from data sets being publicly released [19]. Further work is needed to determine adequate ways of handling these at-tributes if they are included. Clearly, they cannot be targets of predictive modeling using our methods since that will result in their inferential disclosure. This is because the optimization we perform for predictive modeling would group together rows with similar values for the target attribute[19]. This optimization improves the model accuracy while satisfying the identity disclosure constraint, but it also increases the

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

inferential attribute disclosure for the sensitive attribute being targeted. While this is an explicit issue with the k-anonymity approach to anonymization, further investigation is needed on issues related to the inferential disclosure of sensitive attributes even for other approaches (e.g., additive noise and swapping). In many cases only a sample of the data is released. The privacy protection due to sampling has been considered in various works (e.g., [6, 16, 3]).
Applying the k-anonymity approach to the release of a sample opens up some new issues. One approach could be to require that the released sample satisfy the k-anonymity requirement.

Alternatively, the k-anonymity requirement could be rest applied to the entire population before a sample of the transformed table is released. The sizes of the groups in the released sample will depend on the form of sampling used (e.g., random, startled)[20]. Further work is needed to explore the k- anonymity approach in the context of sampling. For predictive modeling usage the metrics denned in consider predictability using only the potentially identifying attributes. This was done independent of the predictive capabilities of the other non-identifying attributes. Considering both identifying and non-identifying attributes during the data transformation process could lead to better solutions. Finding an effective way of doing this with potentially large numbers of non-identifying attributes needs further exploration.

IV. PROBABILISTIC MODEL

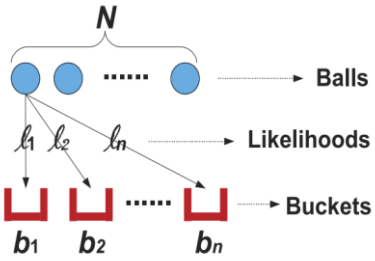

Fig:1 Model Diagram

A fast algorithm for distributed association rule mining is given in Cheung et. al. [2]. Their procedure for fast distributed mining of association rules (FDM) is summarized below[21].
1) Candidate Sets Generation: Generate candidate sets CGi(k) based on GLi(k−1) , item sets that are supported by the Si at the (k-1)th iteration, using the classic a-priori candidate generation algorithm. Each site generates candidates based on the intersection of globally large (k-1) item sets and locally large (k-1) item sets.

2) Local Pruning: For each X ∈ CGi(k) , scan the database DBi at Si to compute X.supi. If X is locally large Si , it is included in the LL i(k) set[22]. It is clear that if X is supported globally, it will be supported in one site.

3) Support Count Exchange: LL i(k) are broadcast, and each site computes the local support for the items in ∪ i LL i(k) 4) Broadcast Mining Results: Each site broadcasts the local support for item sets in ∪ i LL i(k) . From this, each site is able to compute L (k)

V. CONCLUSION

Most SMC protocols are expensive in both communication and computation. We introduced a look-ahead approach for SMC protocols that helps involved parties to decide whether the protocol will meet the expectations before initiating it. We presented a look-ahead protocol specifically for the distributed k-anonymity by approximating the probability that the output of the SMC will be more utilized than their local anonymizations. Experiments on real data showed the effectiveness of the approach. Designing look aheads for other SMC protocols stands as a future work. A wide variety

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

of SMC protocols have been proposed especially for privacy preserving data mining applications [12], [17], [25] each requiring a unique look-ahead approach.

As for the look-ahead process on distributed anonymization protocols, definitions of k-anonymity definitions can be revisited, more efficient techniques can be developed and experimentally evaluated.

REFERENCES

1. R.J. Bayardo and R. Agrawal, "**Data Privacy Through Optimal K-Anonymization**,"Proc. 21st Int'l Conf. Data Eng. (ICDE '05), pp. 217-228, 2005.
2. C. Blake and C.J. Merz, "**UCI Repository of Machine LearningDatabases**," http://www.ics.uci.edu/mlearn/MLRepository. html, Univ. of California, Irvine, Dept. of Information andComputer Sciences, 2012.
3. B.-C. Chen, K. LeFevre, and R. Ramakrishnan, "**Privacy Skyline: Privacy with Multidimensional Adversarial Knowledge**,"Proc. 33rd Int'l Conf. Very Large Data Bases (VLDB '07),pp. 770-781, 2007.
4. Udayakumar R., Khanaa V., Saravanan T., "Analysis of polarization mode dispersion in fibers and its mitigation using an optical compensation technique", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S6) (2013) pp. 4767-4771.
5. S.R. Ganta, S.P. Kasiviswanathan, and A. Smith, "**Composition Attacks and Auxiliary Information in Data Privacy**,"Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '08), pp. 265-273, http://doi.acm.org/10.1145/1401890. 1401926, 2008.
6. G. Ghinita, P. Karras, P. Kalnis, and N. Mamoulis, "**Fast Data Anonymization with Low Information Loss,**"Proc. 33rd Int'l Conf. Very Large Data Bases (VLDB '07),pp. 758-769, 2007.
7. Udayakumar R., Khanaa V., Saravanan T., "Chromatic dispersion compensation in optical fiber communication system and its simulation", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S6) (2013) pp. 4762-4766.
8. V.S. Iyengar, "**Transforming Data to Satisfy Privacy Constraints**," Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (KDD '02),pp. 279-288, 2002.
9. W. Jiang and C. Clifton, "**A Secure Distributed Framework for Achievingk-Anonymity**," VLDB J.,special issue on privacy-preserving data management, vol. 15, pp. 316-333, Sept. 2006.
10. Udayakumar R., Khanaa V., Kaliyamurthie K.P., "High data rate for coherent optical wired communication using DSP", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S6) (2013) 4772-4776.
11. D. Kifer and J. Gehrke, "**Injecting Utility into Anonymized Datasets,**"Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD '06), pp. 217-228, 2006.
12. S.N. Lahiri, A. Chatterjeea, and T. Maiti, "**Normal Approximation to the Hypergeometric Distribution in Nonstandard Cases and a Sub-Gaussian Berryesseen Theorem,**" J. Statistical Planning and Inference, vol. 137, no. 11, pp. 3570-3590, http://dx.doi.org/10. 1016/j.jspi.2007.03.033, Nov. 2007.
13. Udayakumar R., Khanaa V., Kaliyamurthie K.P., "Optical ring architecture performance evaluation using ordinary receiver", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S6) (2013) pp. 4742-4747.
14. Y. Lindell and B. Pinkas, "**Privacy Preserving Data Mining,**" J. Cryptology,vol. 15, pp. 36-54, 2000.
15. A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubrama-niam, "'-**Diversity: Privacy beyond k-Anonymity,**" Proc. IEEE 22nd Int'l Conf. Data Eng. (ICDE '06),Apr. 2006.
16. D.J. Martin, D. Kifer, A. Machanavajjhala, J. Gehrke, and J.Y. Halpern, "**Worst-Case Background Knowledge for Privacy-Pre-serving Data Publishing**," Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE '07), Apr. 2007.
17. M.E. Nergiz and C. Clifton, "**Thoughts on K-Anonymization,**" Data and Knowledge Eng.,vol. 63, no. 3, pp. 622-645, http:// dx.doi.org/10.1016/j.datak.2007.03.009, Dec. 2007.
18. Udayakumar R., Khanaa V., Kaliyamurthie K.P., "Performance analysis of resilient ftth architecture with protection mechanism", Indian Journal of Science and Technology, ISSN : 0974-6846, 6(S6) (2013) pp. 4737-4741
19. A. Øhrn and L. Ohno-Machado, "**Using Boolean Reasoning to Anonymize Databases**,"Artificial Intelligence in Medicine,vol. 15, no. 3, pp. 235-254, http://dx.doi.org/10.1016/S0933-3657(98)00056-6, Mar. 1999.
20. V. Poosala and Y.E. Ioannidis, "**Selectivity Estimation without the Attribute Value Independence Assumption,**"Proc. 23rd Int'l Conf. Very Large Data Bases (VLDB '97),pp. 486-495, 1997.
21. S.J. Schwager, "**Bonferroni Sometimes Loses,**"The Am. Statistician, vol. 38, no. 3, pp. 192-197, http://www.jstor.org/stable/2683651, 1984.
22. L. Sweeney, "**k-Anonymity: A Model for Protecting Privacy,**"Int'l J. Uncertainty, Fuzziness Knowledge-Based Systems, vol. 10, no. 5, pp. 557-570, 2002.
23. J. Vaidya, "**Privacy Preserving Data Mining Over Vertically Partitioned Data**," PhD dissertation, Dept. of Computer Sciences, Purdue Univ., West Lafayette, Indiana, http://www.cs.purdue. edu/homes/jsvaidya/thesis.pdf, 2004.
24. S. Zhong, Z. Yang, and R.N. Wright, "**Privacy-Enhancing K-Anonymization of Customer Data,**"Proc. 24th ACM SIGMOD-SIGACT-SIGART Symp. Principles of Database Systems (PODS '05), pp. 139-147, 2005.
25. Mehmet Ercan Nergiz, Abdullah Ercu¨ment C¸ic¸ ek, Thomas B. Pedersen, and Yu ¨cel Sayg□ n ,"**A Look-Ahead Approach to Secure Multiparty Protocols**"July2012.
26. Dr.K.P.Kaliyamurthie, D.Parameswari, Load Balancing in Structured Peer to Peer Systems, International Journal of Innovative Research in Computer and Communication Engineering, ISSN: 2249-2615,pp 22-26, Volume1 Issue 1 Number2-Aug 2011
27. Dr.R.Udayakumar, Addressing the Contract Issue,Standardisation for QOS, International Journal of Innovative Research in Computer and Communication Engineering, ISSN (Online): 2320 – 9801,pp 536-541, Vol. 1, Issue 3, May 2013
28. Dr.R.Udayakumar, Computational Modeling of the StrengthEvolution During Processing And Service Of9-12% Cr Steels, International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801,pp 3295-3302, Vol. 2, Issue 3, March 2014

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 4, April 2015

29. P.GAYATHRI, ASSORTED PERIODIC PATTERNS INTIME SERIES DATABASE USINGMINING, International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801, pp 5046- 5051, Vol. 2, Issue 7, July 2014.
30. Gayathri, Massive Querying For Optimizing Cost – CachingService in Cloud Data, International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801,pp 2041-2048, Vol. 1, Issue 9, November 2013