International Journal of Communication Networks and Security

Volume 2 | Issue 1 Article 16

January 2013

SMC PROTOCOL FOR DISTRIBUTED K- ANONYMITY

V. SIREESHA
Audisankara Institute Of Technology , Anantapur, A.P, V.SIREESHA@gmail.com

B. OBULESU
Audisankara Institute Of Technology , Anantapur, A.P, B.OBULESU@gmail.com

Recommended Citation

SIREESHA, V. and OBULESU, B. (2013) "SMC PROTOCOL FOR DISTRIBUTED K- ANONYMITY," *International Journal of Communication Networks and Security*: Vol. 2 : Iss. 1 , Article 16.
DOI: 10.47893/IJCNS.2013.1073
Available at: https://www.interscience.in/ijcns/vol2/iss1/16

SMC PROTOCOL FOR DISTRIBUTED K- ANONYMITY

[1]V.SIREESHA, [2]B.OBULESU

[1]Associate Professor, [2]M.Tech (CS) Student, Audisankara Institute Of Technology , Anantapur, A.P

Abstract- Secure multiparty protocols have been proposed to enable non colluding parties to cooperate without a trusted server. Even though such protocols put off information exposé other than the objective function, they are quite costly in computation and communication. The high overhead motivates parties to estimate the utility that can be achieved as a result of the protocol beforehand. To avoid this issue we propose a look-ahead approach, specifically for secure multiparty protocols to achieve distributed k-anonymity, which helps parties to decide if the utility benefit from the protocol is within an acceptable range before initiating the protocol. The look-aheadoperation is highly localized and its accuracy depends on the amount of information the parties are willing toshare. Experimental results show the effectiveness of the proposed methods.

I. INTRODUCTION

DATA mining is widely used by researchers for science and business purposes. Data collected from individuals are important for decision making or pattern recognition. Therefore, privacy-preserving processes have been developed to sanitize private information from the samples while keeping their utility.

A large body of research has been devoted to the protection of sensitive information when samples are given to third parties for processing or computing. It is in the interest of research to disseminate samples to wide audience of researchers, without making strong assumptions about their trustworthiness. Even if information collectors ensure that data are released only to third parties with nonmalicious intent (or if a privacy preserving approach can be applied before the data are released, see Fig. 1a), there is always the possibility that the information collectors may inadvertently disclose samples to malicious parties or that the samples are actively stolen from the collectors (see Fig. 1b). Samples may be leaked or stolen anytime during the storing process or while residing in storage. This paper focuses on preventing such attacks on third parties for the whole lifetime of the samples.

Contemporary research in privacy preserving data mining mainly falls into one of two categories: 1) perturbation and randomization-based approaches, and 2) secure multiparty Computation (SMC)-based approaches. SMC approaches employ cryptographic tools for collaborative data mining computation by multiple parties. Samples are distributed among different parties and they take part in the information computation and communication process. SMC research focuses on protocol development for protecting privacy among the involved parties or computation efficiency; however, centralized processing of samples and storage privacy is out of the scope of SMC.

We introduce a new perturbation and randomization based approach that protects centralized sample data sets utilized for decision tree data mining. Privacy preservation is applied to sanitize the samples prior to their release to third parties in order to mitigate the threat of their inadvertent disclosure or theft. In contrast to other sanitization methods, our approach does not affect the accuracy of data mining results. The decision tree can be built directly from the sanitized data sets, such that the originals do not need to be reconstructed. Moreover, this approach can be applied at any time during the data collection process so that privacy protection can be in effect even while samples are still being collected.

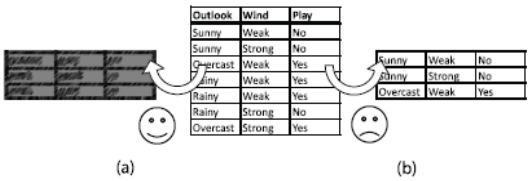


Fig. 1. Two forms of information release to a third party: (a) the data collector sends the preprocessed information (which was sanitized through extra techniques, such as cryptographic approaches or statistical database) at will or (b) hackers steal the original samples instorage without notifying the data collector.

Existing system:

Many SMC protocols for privacy conserve data mining suffer from high computation and communication costs. The high overhead of SMC Protocol raise the question of whether the information gain after the protocol execution is more than the cost. This is the valid question for protocols working on horizontally or vertically portioned data. While publishing person specific sensitive data, simply removing uniquely identifying information (SSN, name) from data is not sufficient to prevent identification because partially identifying information, quasi identifiers, (age, sex, nation, ...) can still be mapped to individuals (and possibly to their sensitive information such as salary) by using an external knowledge. The goal of privacy protection based on k-anonymity is to bind the linking of a

record from a set of released records to a specific individual even when adversaries can link individuals via QI. Previous work does not accurately achieve the goal of privacy data preserving method.

Proposed System:

Our Proposed system focuses on the SMC protocol for distributed k- anonymity. k-Anonymity is a well-known privacy preservation technique proposed in to prevent linking attacks on shared databases One way to enable effective data mining while preserving privacy is to anonymize the data set that includes private information about subjects before being released for data mining. One way to anonymize data set is to manipulate its content so that the records adhere to k-anonymity. Two common manipulation techniques used to achieve k-anonymity of a data set are generalization and suppression. Generalization refers to replacing a value with a less specific but semantically consistent value, while suppression refers to not releasing a value at all.

MODULES

1. SECURE MULTIPARTY COMPUTATION DESIGN:

Secure multiparty computation (SMC) protocols are one of the first techniques used in privacy preserving datamining in distributed environments. The idea behind these protocols is based on the 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. 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. Moreover, the protocol does not require a trusted third party. Theseproperties are promising for privacy preserving applications.

2. K-ANONIMITY FORMULATION:

K-Anonymity is a well-known privacy preservation technique proposed in to prevent linking attacks on shared databases. Linking attacks are performed by adversaries who know some attributes (quasi-identifier attributes) of an individual to identify him/her in the data set. A database is said to be k-anonymous if every tuple projected over the quasi-identifier attributes appears at least k times in the data base. k-Anonymization is the process of enforcing the k-anonymity property on a given database by using generalization and suppression of values.

3. GENERALIZATION AND SUPPRESSION ON DATAS:

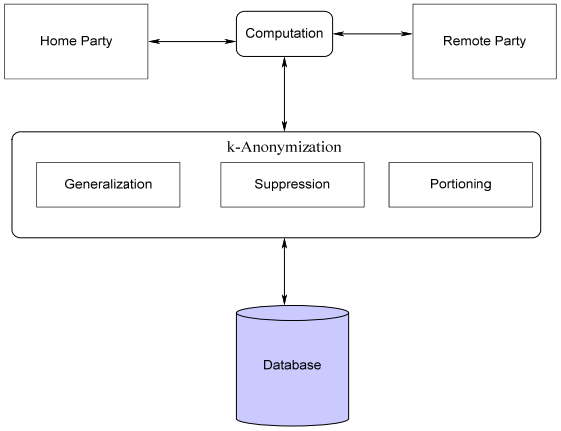
Merging the similar data types of a given selected mining algorithm into a generalized data type seems to be a good approach to reduce the transformation complexity in SMC Protocols. The Generalization process, including merging and transforming phases is proposed. In the merging phase, the original data types of the data sources to be mined are first merged into the generalized one. The transforming phase is then used to convert the generalized data types into the target ones for the selected mining algorithm. Some of the data are not visible to the partiesto improve the performance of k-anonimity using suppression methodology.

4. TABLE PORTIONING:

Vertical partitioning involves creating tables with fewer columns and using additional tables to store the remaining columns. A common form of vertical partitioning is to split dynamic data (slow to find) from static data (fast to find) in a table where the dynamic data is not used as often as the static. The database security is achieved using the vertical portioning by this module. If the adversary wants to hack the data, they have to compromise all the databases which are not easier process to perform the database injection attacks.

Design:

System Flow Diagram:



Use Case Diagram:

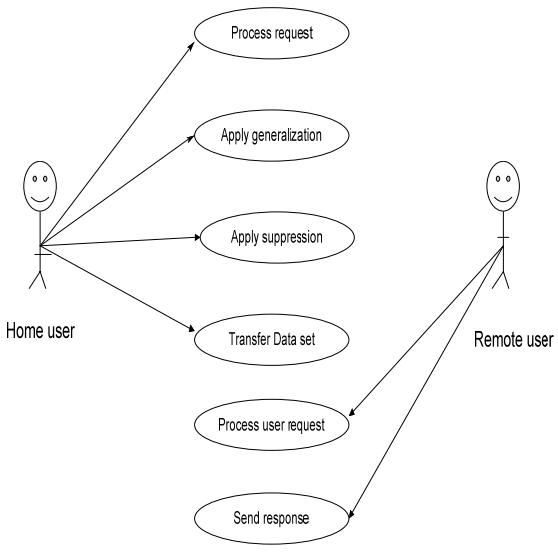
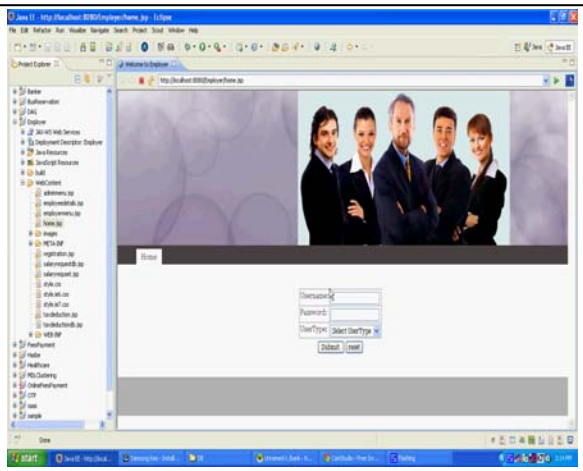
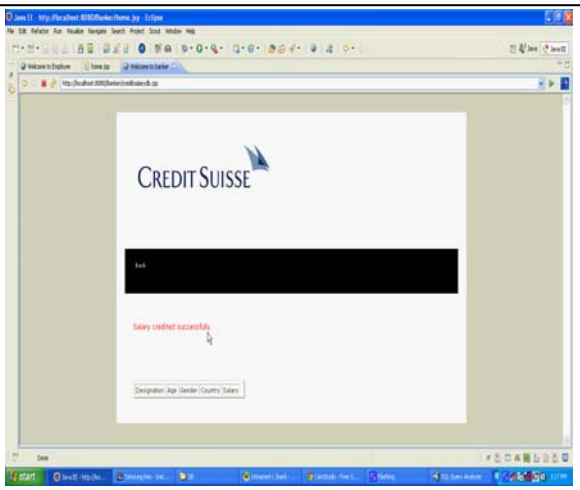
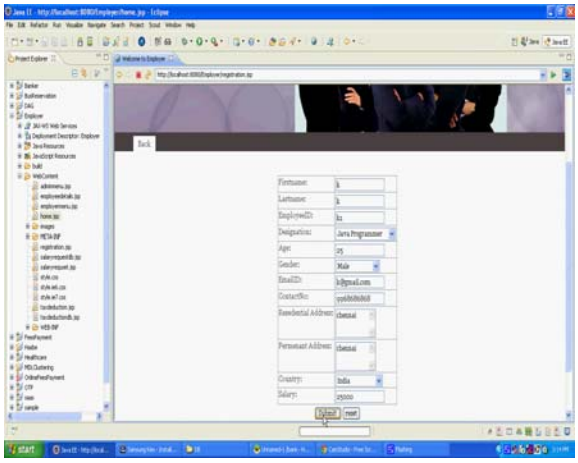
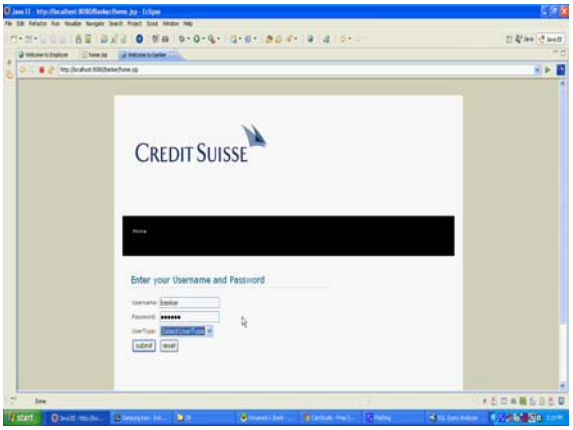


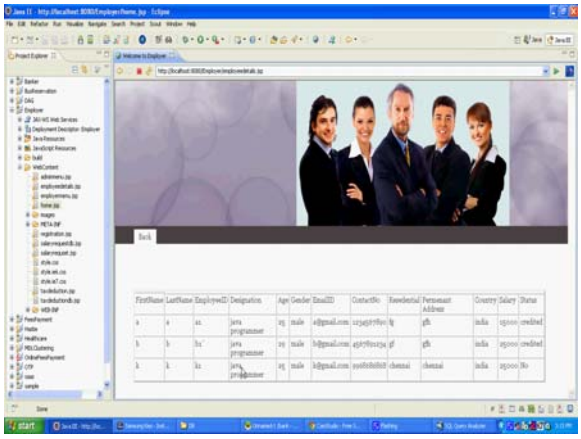
Figure 1.2

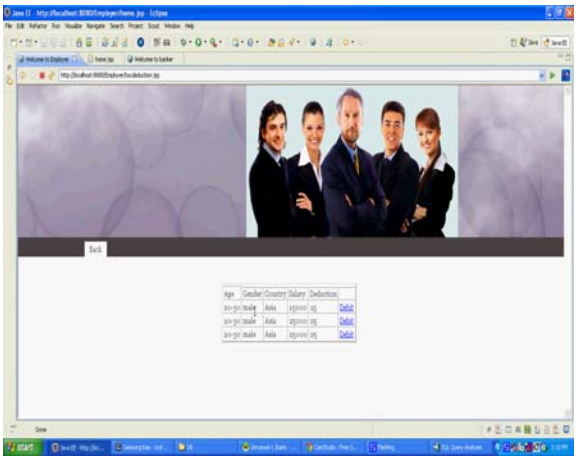


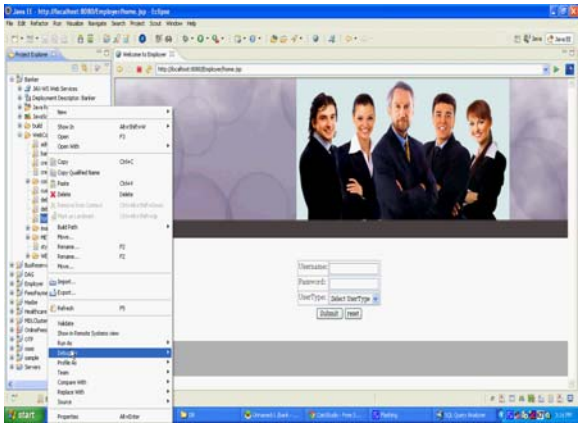

CREDIT SUISSE

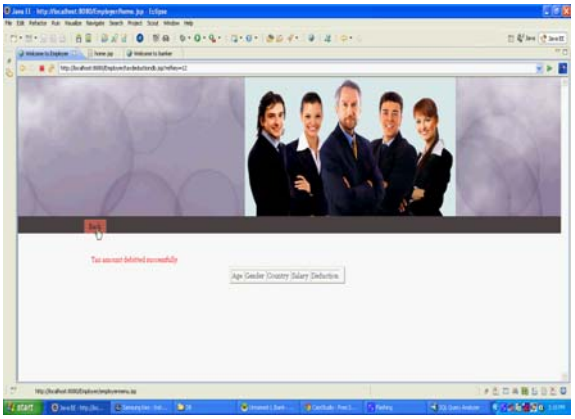



CREDIT SUISSE
Enter your Username and Password









CONCLUSION& FUTURE ENHANCEMENT

Designing look aheads for other SMC protocols stands as a future work. A wide variety of SMC protocols have been proposed especially for privacy preserving data mining applications 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. 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.

REFERENCES

[1] R.J. Bayardo and R. Agrawal, "Data Privacy through Optimal KAnonymization," Proc. 21st Int'l Conf. Data Eng. (ICDE '05), pp. 217-228, 2005.

[2] C. Blake and C.J. Merz, "UCI Repository of Machine Learning Databases," h ttp://www .ics.uci .edu /mlearn/ MLRepo si tory. html , Univ. of California, Irvine, Dept of Information and Computer 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] J. Doming o-Ferrer and V.Torra," Ordinal, Continuous and Heterogeneous K-Anonymity through Microaggregation," Data Mining and Knowledge Discovery, vol. 11, no. 2, pp. 195-212, 2005.

[5] W. Feller,An Introduction to Probability Theory and Its Applications, vol. 1, Wiley, 1968.[6] 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.

[7] 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.

[8] O. Goldreich, The Foundations o f Cryptography, vol. 2, Cambridge Univ. P ress, htt p : //www.w i sdom. weizmann.ac.il/oded/PSBookFrag/enc.ps, 2004.

[9] 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.

[10] W. Jiang and C.Clifto n, "Privacy - Preserving Distributed k Anonymity," Proc. 19th Ann. IFIP WG 11.3 Working Conf. Database and Applications Security, Aug. 2005.