ISSN: 2348 - 2117

International Journal of Engineering Technology and Computer Research (IJETCR)
Available Online at www.ijetcr.org
Index Copernicus Value 2016: 69.88
Volume 6; Issue 6; November – December ; 2018; Page No. 01-04

M-PRIVACY FOR COLLABORATIVE DATA PUBLISHING: COLLABORATIVE DATA PUBLISHING PROBLEM

Akanksha Jain, Dr. Dinesh Shrimali

JRN Rajasthan Vidhyapeeth University

Abstract

The collaborative data publishing problem for anonymizing horizontally partitioned data at multiple data providers is considered. A new type of "insider attack" by colluding data providers who may use their own data records (a subset of the overall data) in addition to the external background knowledge to infer the data records contributed by other data providers. This new threat and makes several contributions. The notion of m-privacy, which guarantees that the anonymized data satisfies a given privacy constraint against any group of up to m colluding data providers. A heuristic algorithms exploiting the equivalence group monotonicity of privacy constraints and adaptive ordering techniques for efficiently checking m-privacy given a set of records is presented. A data provider-aware anonymization algorithm is presented with adaptive m- privacy checking strategies to ensure high utility and m-privacy of anonymized data with efficiency. Experiments on real-life datasets suggest that this approach achieves better or comparable utility and efficiency than existing and baseline algorithms while providing m-privacy guarantee.

The goal is to publish an anonymized view of the integrated data such that a data recipient including the data providers will not be able to compromise the privacy of the individual records provided by other parties.

Key words: Anonymization, Adversary, Algorithm.

INTRODUCTION

There is an increasing need for sharing data that contain personal information from distributed databases. For example, in the healthcare domain, a national agenda is to develop the Nationwide Health Information Network (NHIN) to share information among hospitals and other providers, and support appropriate use of health information beyond direct patient care with privacy protection. Privacy preserving data analysis and data publishing have received considerable attention in recent years as promising approaches for sharing data while preserving individual privacy. When the data are distributed among multiple data providers or data owners, two main settings are used for anonymization. One approach is for each provider to anonymize the data independently, which results in potential loss of integrated data utility. A more desirable approach is collaborative data publishing, which anonymizes data from all providers as if they would come from one source, using either a trusted third-party (TTP) or Secure Multi-party Computation (SMC) protocols to do computations.

Problem Definition

An m-adversary as a coalition of m colluding data providers or data owners, who have access to their own data records as well as publicly available background knowledge BK and attempts to infer data records contributed by other data providers.

Data Publishing and Data Privacy

Corresponding author: Akanksha Jain

Society is experiencing exponential growth in the number and variety of data collections containing person-specific information. This collected information is valuable both in research and business. Data sharing is common. Publishing the data may put the respondent's privacy in risk.

Objective:

Maximize data utility while limiting disclosure risk to an acceptable level

K-Anonymity

If the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release. Ex. If you try to identify a man from a release, but the only information you have is his birth date and gender. There are k people meet the requirement. This is k-Anonymity.

Proposed System

Attacks by Data Providers Using Anonymized Data and Their Own Data

- Collaborative data publishing setting with horizontally partitioned data across multiple data providers, each contributing a subset of records is considered.
- A data provider could be the data owner itself who is contributing its own records.
- Each provider has additional data knowledge of their own records, which can help with the attack.

This issue can be further worsened when multiple data providers collude with each other.

Advantages

- "Insider attack" by data providers is considered.
- High privacy for published data

Methodology

The project is developed in the following stages

* Analysis – analysis of the customer need
* Design – design of the desired solution
* Development – (technical) development of the solution
* Implementation – deployment of the developed solution in the organization
* Evaluation – evaluation the implemented solution

System Architecture

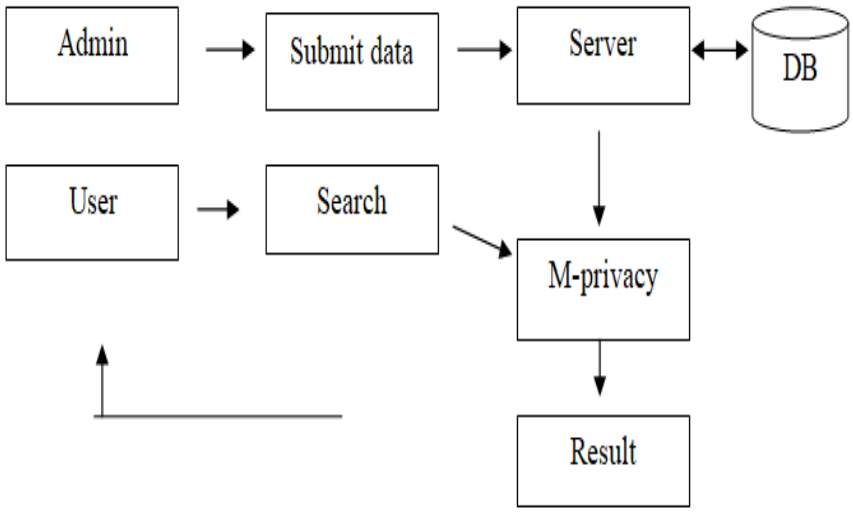

Figure 1: System architecture

The above figure represents the system architecture of m-privacy, collaborative data publishing model. The above figure represents each model designed in our study.

Project Flow Chart

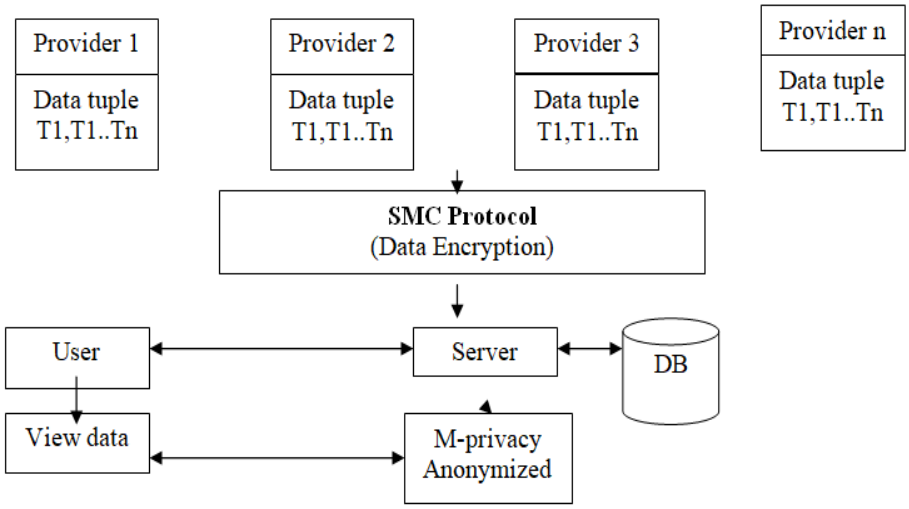

Figure 2: Project Flow Chart

Activity Diagram

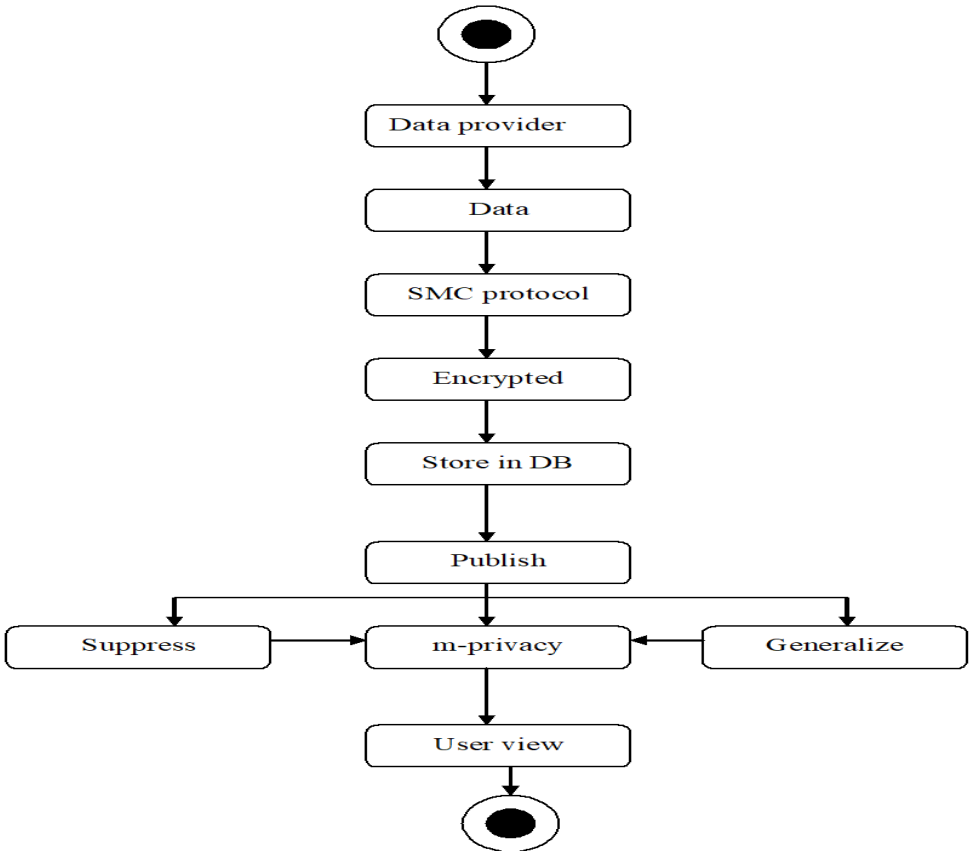

Figure 3: Activity Diagram

The above figure shows various activities handled by the system such as encryption, suppression, generalization and m-privacy.

Conclusion

A new type of potential attackers in collaborative data publishing – a coalition of data providers, called m-adversary is considered. To prevent privacy disclosure by any m-adversary we showed that guaranteeing m-privacy is enough. Heuristic algorithms are presented exploiting equivalence group monotonicity of privacy constraints and adaptive ordering techniques for efficiently checking m-privacy. We introduced also a provider-aware anonymization algorithm with adaptive m-privacy checking strategies to ensure high utility and m-privacy of anonymized data.
There are many remaining research questions. Defining a proper privacy fitness score for different privacy constraints is one of them. It also remains a question to address and model the data knowledge of data providers when data are distributed in a vertical or ad-hoc fashion. It would be also interesting to verify if our methods can be adapted to other kinds of data such as set-valued data.

References

1. C. Dwork, “Differential privacy: a survey of results,” in Proc. of the 5th Intl. Conf. on Theory and Applications of Models of Computation, 2008, pp. 1–19.
2. B. C. M. Fung, K. Wang, R. Chen, and P. S. Yu, “Privacy-preserving data publishing: A survey of recent developments,” ACM Comput. Surv.,vol. 42, pp. 14:1–14:53, June 2010.
3. C. Dwork, “A firm foundation for private data analysis,” Commun. ACM vol. 54, pp. 86–95, January 2011.
4. N. Mohammed, B. C. M. Fung, P. C. K. Hung, and C. Lee, “Centralized and distributed anonymization for high-dimensional healthcare data,” ACM Transactions on Knowledge Discovery from Data (TKDD), vol. 4, no. 4, pp. 18:1–18:33, October 2010.
5. W. Jiang and C. Clifton, “Privacy-preserving distributed k-anonymity,” in Data and Applications Security XIX, ser. Lecture Notes in Computer Science, 2005, vol. 3654, pp. 924–924.
6. W. Jiang and C. Clifton, “A secure distributed framework for achieving k-anonymity,” VLDB J., vol. 15, no. 4, pp. 316–333, 2006.
7. O. Goldreich, Foundations of Cryptography: Volume 2, Basic Applications. Cambridge University Press, 2004.
8. Y. Lindell and B. Pinkas, “Secure multiparty computation for privacypreserving data mining,” The Journal of Privacy and Confidentiality, vol. 1, no. 1, pp. 59–98, 2009.
9. A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, “l-diversity: Privacy beyond k-anonymity,” in ICDE, 2006, p. 24.
10. P. Samarati, “Protecting respondents’ identities in microdata release,” IEEE T. Knowl. Data En., vol. 13, no. 6, pp. 1010–1027, 2001.