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. 2, Issue 11, November 2014

Enhanced M-Privacy for Collaborative Data Publishing

Aseema Jana, Shubham Joshi

Research Scholar, Dept. of C.S.E., D.P.C.O.E, Savitri bai Phule Pune University, Pune, Maharashtra, India.

Research Supervisor, Dept. of C.S.E., D.P.C.O.E, Savitri bai Phule Pune University, Pune, Maharashtra, India.

ABSTRACT: In recent years, privacy takes an important role to secure the data from various probable attackers. When for public advantage data need to be shared as required for Health care and researches, individual privacy is major concern regarding sensitive information. So while publishing such data, privacy should be conserved .While publishing collaborative data to multiple data provider's two types of problem occurs, first is outsider attack and second is insider attack. Outsider attack is by the people who are not data providers and insider attack is by colluding data provider who may use their own data records to understand the data records shared by other data providers. The paper focuses on insider attack, and makes some contributions. This problem can be overcome by combining slicing techniques with m-privacy techniques and addition of protocols as secure multiparty computation and trusted third party will increase the privacy of system effectively.

KEYWORDS: Privacy, security, integrity, collaborative publishing, slicing, distributed databases.

I. INTRODUCTION

In current years, for public advantage data need to be shared. Generally data is collected from distributed databases for e.g. in case of Health care and researches, data is collected from different providers and gathered in central network. In health care all information related to patient is present in central network which includes disease details, corresponding treatment and test details.

By using anonymization technique the data is modified and then released to the public. This process is known as the privacy preservation data publishing. The attributes are classified by three types which are Key attribute, quasi identifier and sensitive attribute. Key attribute represents unique identification such as names, SSN and it is always removed before publishing.

Quasi-identifiers are segments of information that are not unique identifiers but well correlated with an entity; they can be combined with other quasi-identifier to create a unique identifier. Example birth date, gender, which can be used link unionized dataset with other data. Last one is sensitive attributes example diseases, policy detail, and salary. As a special case, a data provider could be the data owner itself who is contributing its own records. A data recipient may have access to some background knowledge which represents any publicly available information about released data, e.g., Census datasets. Consider the set of records in d1, d2…. dn, which are provided by the provider.

The record is a collection of some data. Before publishing the records to the public the anonymization technique is applied to the data, then it generate the subset of records in d1, d2….. dn. Goal is to secure the original data or individual information from the different malicious user by using the anonymization using either a trusted third-party or Secure Multi-party Computation protocols, when the data is published to the public.

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. 2, Issue 11, November 2014

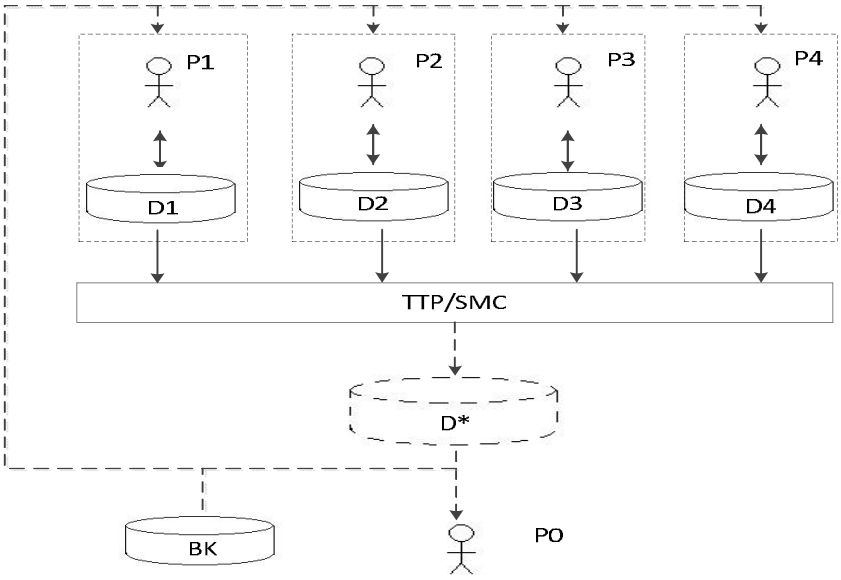

Figure1. Collaborative data publishing

II. RELATED WORK

Privacy preserving data analysis and collaborative data publishing has received considerable attention in current years as promising approaches for sharing data while preserving individual privacy.

Privacy hazard occurs when attacker is able to establish any type of link as record linkage, attribute linkage, table linkage and probabilistic attack, which he can violate individual privacy. B.C.M. Fung et al. [1] proposed the concept of privacy preserving data publishing (PPDP). PPDP provides methods and tools for publishing useful information while preserving data privacy. These methods include K- anonymity, L-diversity and δ-Presence which encounter the attack of record linkage, attribute linkage and table linkage respectively.

N. Mohammed et al. [2], proposed LKC privacy model for high dimensional relational data for healthcare system. This LKC model gives better result than traditional k anonymization model. But LKC model consider only relational data and healthcare data is complex, may be a combination of relational data, transaction data and textual data.

According to Yehuda Lindell et al. [7], the major problem related to privacy preserving is, finding the computation function where individual privacy is preserved. For example, computation on confidential medical or criminal data in such a way that information is not revealed. This is called secure multiparty computation where number of parties wants to mutually compute some functions on their confidential inputs and through the result of this computation, parties only study the correct output and nothing else, even if some of the parties nastily plan to obtain more information. Secure multiparty computation (SMC) protocol is useful in handling above discussed scenario.

D.K. Mishra et al. [8] have proposed Distributed K-secure sum protocol for secure multiparty computation. Secure sum computation of private data inputs is an example of SMC which can give a secure protocol with lower probability of data leakage. Here the idea of secure sum protocol has been enhanced which is proposed by C. Clifton et al. [9]. Distributed K-secure sum protocol compute the sum of individual data inputs with zero probability of data leakage when two neighbor parties plan to know the data of a middle party. Each data block is broken into k segments where k is equal to the number of parties. Then the segments are distributed to other parties before computation. This protocol we call as dk-Secure Sum Protocol.

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. 2, Issue 11, November 2014

The concept of 'insider attack' is given by B. C. M. Fung et al. [10] and according to him insider attack is caused by those data provider who may use their own data records to understand the data records shared by other data providers. Most of the discussed work was focused on outsider attack, Insider attack is a new threat and author has used m-privacy techniques for encounter the same.

III. SCOPE OF WORK

Proposed concept can be used in many applications like hospital management system, many industrial areas where we like to protect a sensitive data like salary of employee. Pharmaceutical company where sensitive data may be a combination of ingredients of medicines, in banking sector where sensitive data is account number of customer, this system can be used. It can be used in military area where data is gathered from different sources and need to secure that data from each other to maintain privacy.

IV. PROBLEM DEFINITION

Major problem while publishing collaborative data is attacks. Attacks are executed by insider or external attackers, which may be a single or a group of internal and external bodies that wants to violate privacy of collaborative data using background information/knowledge, also anonymized data. Privacy is violated if one knows anything about data.

Main goal is to publish an anonymized view of incorporated data, D* which will be resistant to internal or external attacks. This improves the security and privacy with combination of, m-privacy techniques and slicing technique which accomplish privacy verification with better performance than encryption algorithm and provider aware (base algorithm).

V. PERSPECTIVE SOLUTION

In Privacy for collaborative data publishing, main focus is on insider attacks. This problem can be solved by using various approaches as m-privacy, Heuristic algorithms, Data provider aware anonymization Algorithm and SMC/TTP protocols. M-Privacy helps in protecting anonymized data against m-adversary with respect to privacy constraint as K-anonymity and L-diversity. M-Privacy can also be sure when there are duplicate records; it also contains syntactic privacy constraint, monotonicity of privacy constraints and differential privacy constraint. Verification of m-privacy can be done by Binary m-Privacy verification algorithm, Top-Down and Bottom-Up algorithms are also used for this. This process primarily analyze the problem by reproducing adversary space and using heuristic algorithms with efficient pruning strategy and adaptive ordering techniques for efficiently checking m-privacy with respect to equivalence group monotonicity constraints.

VI. PROPOSED WORK

The proposed model provides a competent approach to achieve enhanced privacy for collaborative data publishing. This model combines slicing techniques with m-privacy techniques. Slicing overcomes the limitations of generalization and Bucketization and preserves better utility while protecting against privacy threats. M-privacy techniques assure that the anonymized data fulfils a given privacy constraint against any range of m-colluding data providers (where m can be varied between certain ranges 1 to m), additionally it's using monotonicity constraints for efficiently checking m-privacy. Model uses Slicing for partitioning the data records and then follows m-privacy techniques and its related algorithms.

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. 2, Issue 11, November 2014

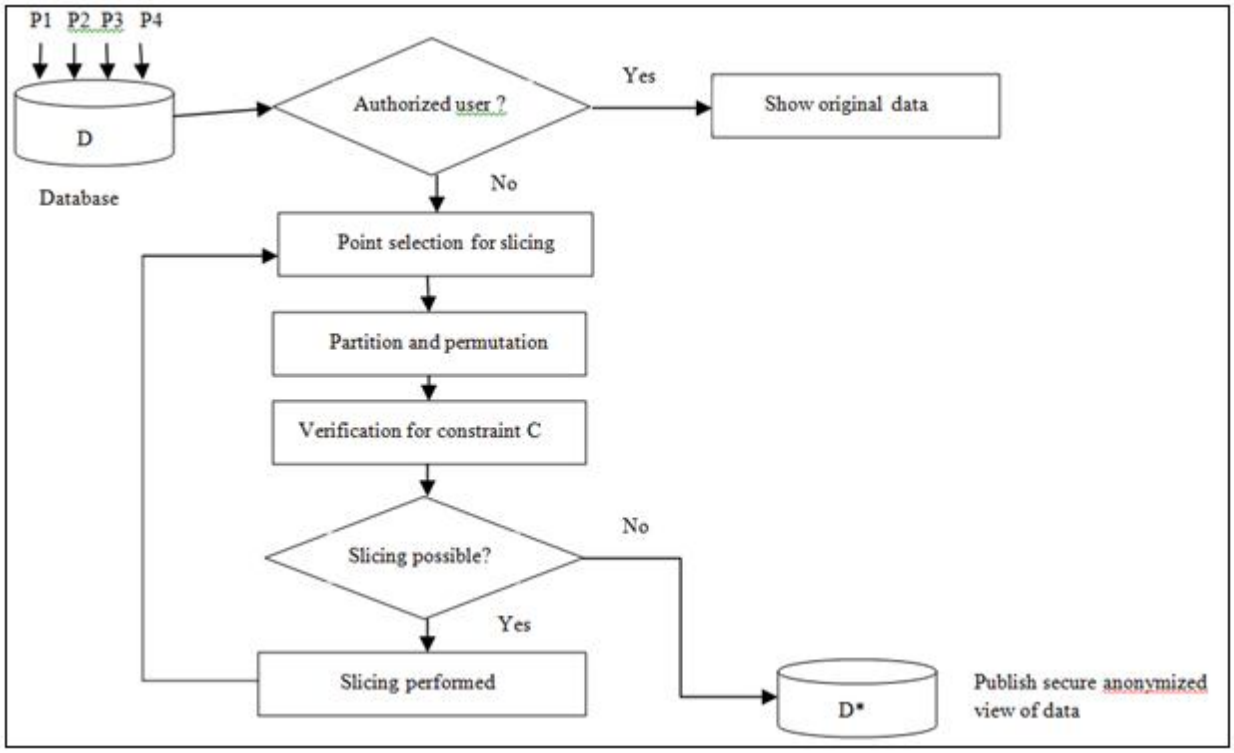

Figure2. System Flow Diagram

VII. DICUSSION AND FUTURE WORK

Above discussed approaches help to enhance the data privacy and security when data is collected from various resources and output should be in collaborative style. In future, this system can consider for data, which are distributed in ad hoc grid computing. Also the system can be considered for set valued data. The consumption of various protocols can address various data publishing paradigms. The consumption of these protocols can make collaborative data publishing more effective and enhanced using m-privacy.

VIII. ACKNOWLEDGMENTS

Sincere thank to the reviewers for reviewing this manuscript and providing inputs for greatly improving the quality of this paper.

REFERENCES

1. 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.*
2. N. Mohammed, B. C. M. Fung, P. C. K. Hung, and C. Lee,"Centralized and distributed anonymization for high-dimensional healthcare data," *ACM Trans. on Knowl. Discovery from Data, vol. 4,no. 4, pp. 18:1–18:33, October 2010.*
3. W. Jiang and C. Clifton, "A secure distributed framework for achieving k-anonymity", *VLDB J., vol. 15, no. 4, pp. 316–333, 2006.*
4. Machanavajjhala, A.Gehrke J., Kifer D. and Venkitasubramaniam M. "l-diversity: Privacy beyond k-anonymity" *In Proceedings of the 22nd IEEE International Conference on Data Engineering (ICDE).*
5. CHAWLA, S., DWORK, C.MCSHERRY, F., SMITH, A., AND WEE, H. "Toward privacy in public databases". *In Proceedings of the Theory of Cryptography Conference (TCC),2005.*
6. NERGIZ, M. E., CLIFTON, C., AND NERGIZ, A. E. Multirelational k-anonymity. *In proceedings of the 23rd International Conference on Data Engineering (ICDE), 2007.*
7. Y. Lindell and B. Pinkas, "Secure multiparty computation for privacy-preserving data mining," *The Journal of Privacy and Confidentiality,vol. 1, no. 1, pp. 59–98, 2009.*

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. 2, Issue 11, November 2014

8. C. Clifton, M. Kantarcioglu, J. Vaidya, X. Lin, and M. Y. Zhu,"Tools for privacy preserving distributed data mining," *SIGKDD Explor. Newsl., vol. 4, pp. 28–34, December 2002.*
9. R. Sheikh, B. Kumar, and D. K. Mishra, "A distributed k-secure sum protocol for secure multi-party computations," *J. of Computing,vol. 2, pp. 68–72, March 2010.*
10. S. Goryczka, L. Xiong, and B. C. M. Fung, "m-Privacy for collaborative data publishing", *In Proc. of the 7th Intl. Conf. on Collaborative Computing: Networking, Applications and Work sharing, 2011.*
11. S. Goryczka, L. Xiong, and B. C. M. Fung, "m-Privacy for collaborative data publishing," *In Proc. of the 7th Intl. Conf. on Collaborative Computing: Networking, Applications and Work sharing, 2013.*
12. Tiancheng Li, N inghui Li, Senior Member, IEEE, Jian Zhang, Member, IEEE, and IanMolloy "Slicing: A New Approach for Privacy Preserving Data Publishing" *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. MARCH 2012.*
13. S.Kiruthika and Dr. M.Mohamed Raseen "Enhanced Slicing Models For Preserving Privacy In Data Publication", *ICCTET, 2013.*