



# Review of Literature on Some Privacy Preserving Techniques, Algorithms and Methods for Data Privacy and Preserving in Data Mining

Mr.Paresh T.Patel<sup>1st</sup>

Assi. Prof In M.C.A at SCMT-Limbdi  
Ph.D Research Scholar,  
C.U.Shah University,  
Surendranagar Gujarat (India)

Dr. Keyur Shah<sup>2nd</sup>

H.O.D in M.C.A Dept at LDRP  
Ph.D Research Guide,  
C.U.Shah University,  
Surendranagar Gujarat (India)

**Abstract:** *Now in a days in business world and in society, data exchanging process is need to manage today's world of universal data exchange, there is a need to manage the risk of unintended information disclosure. Publishing the data about the individuals, without revealing sensitive information about them in an important problem. So Privacy is one of the most important properties of the data sharing and data mining field. To protect client privacy in the data mining process, a hundred of techniques and algorithms are available. Here some reviews of papers are available which give us some algorithms and techniques to preserve and maintain privacy of the important data from universe.*

**Keywords:** *Privacy, Preserving data, Privacy management.*

## I. INTRODUCTION

One of the most study problems in data mining is privacy of the data, dataset, frequently item set, share data etc. privacy preserving of data in data mining field is the most important and now a days it required part of the business world or universe. Everyone wants to save and secure their private data likes banks, govt. sectors, business world etc. The goal of data mining is discovering new and useful knowledge from data. Sometimes the data contains sensitive information and it needs to be privacy preserving before it is given to data mining researcher and the public <sup>[12]</sup>.

Here we can review the different types of techniques, methods and ideas to solve the breaches for privacy preserving the data and try to secure and safe data from others. Here we explain different research papers on privacy preserving the data with different techniques, methods and algorithms.

Here we explain framework for privacy preserving data sharing, using graph data to preserve sensitive relationship data, cryptography technique to preserve the data, worked on frequent item set in data mining, used k- means clustering over vertically partitioned data, used logistic regression technique, work on k-anonymization, feature selection in data mining using centralized network, used association rule for centralized database.

## II. LITERATURE REVIEW

### A. Apache Hadoop yarn parameter configuration challenges and optimization.

Bhavin J. Mathiya et al <sup>[1]</sup> Apache Hadoop yarn parameter configuration challenges and optimization. They proved that default Hadoop parameter configuration setting is not suitable for all kind of application. Future work for that to develop customize framework for performance tuning in Hadoop <sup>[1]</sup>.

### B. Privacy Preserving Mining of Association Rules.

Alexander Evfimievski et al <sup>[2]</sup> worked on a framework for mining association rule for categorical randomized data to preserve privacy of individual transactions. They derive formulae for an unbiased support estimator and its variance, which allow us to receive item set from randomized dataset and show how to incorporate these formulae into mining algorithm <sup>[2]</sup>.

According to them future research on privacy preserving for all kind of breaches, what are the theoretical limits on discoverability for a given level of privacy? <sup>[2]</sup>.

### C. An Improved Privacy Preserving Algorithm using Association Rule Mining in Centralized databases.

Archana Tomar et al <sup>[3]</sup> worked on perturbation technique for privacy preserved for centralized data. In future instead of perturbation technique they can use SHA Algorithm (Secure Hash Algorithm) for data encryption and get better result to hides logical instances from others <sup>[3]</sup>.

**D. Privacy Preserving for Future Selection in Data Mining using Centralized Network.**

Hemanta Kumar Bhuyan et al<sup>[4]</sup> worked on centralized data for future selection in data mining. They used perturbation technique and apply Gain Ratio technique for future selection. They used the Chi-Square testing for the classification of data by centralized data mining<sup>[4]</sup>.

**E. On Private Scalar Product Computation for Privacy Preserving Data Mining.**

Bart Goethals et al<sup>[5]</sup> worked on private scalar product protocols. Those are based on homomorphic encryption and improve its efficiency and proved that it is secure. So that can also be used on massive datasets<sup>[5]</sup>.

**F. Privacy Preserving Logistic Regression.**

Kamalika Chaudhari et al<sup>[6]</sup> worked on privacy preserving regularized logistic regression algorithm, which is solving a perturbed optimization problems. This algorithm was not depend on the sensitivity of the function and extends easily to a class of convex less function<sup>[6]</sup>.

**G. Privacy Preserving K-means clustering over vertically partitioned data.**

Jaideep Vaidya et al<sup>[7]</sup> they prove that on k-means clustering method when different sites contain different attributes for a common set of entities. Each site learns the cluster of each entity, but learns nothing about the attributes at other sites<sup>[7]</sup>.

**H. Privacy Preserving Frequent Item set Mining.**

Stanly R.M.Oliveira et al<sup>[8]</sup>. They worked on hiding restrictive rules on inverted files, (1) For indexing the transactions per item. (2) For indexing the sensitive transaction per restrictive pattern and Boolean queries for frequent item set mining. By using a set of algorithm which is combine in a single framework. They used (1) The Sanitization Algorithm. (2) The Naïve Algorithm. (3) The Minimum frequency item Algorithm. (4) The Item Grouping Algorithm<sup>[8]</sup>.

**I. A Framework for High-Accuracy Privacy Preserving Mining.**

Shipra Agrawal et al<sup>[9]</sup> they developed FRAPP ( Framework for Accuracy in Privacy Preserving Mining) a generalized model for random perturbation based method, operating on categorical data under strict privacy constraints. They presented an implementation technique for gamma diagonal based perturbation whose complexity is proportional to the sum of the domain cardinalities of the attributes in the database. They also worked on the novel strategy of having the perturbation matrix composed of not value, but random variables intend at marginal cost in accuracy<sup>[9]</sup>.

**J. Limiting Privacy Breaches: in Privacy Preserving Data Mining.**

Alexander Evfimievski et al<sup>[10]</sup> worked on an amplification which makes possible to guarantee limits on privacy breaches without any knowledge of the distribution of the original data they gave a method for compressing long randomized transaction by using pseudo random generators and showed that this cloud reduce their sized by orders of magnitude, for future research

1. How do we extend amplification to continue distributions?
2. What is the relationship between the specific randomization operation and the tradeoff between privacy and accuracy?
3. Are there ways t combine the randomization and the secure multi-party computation approach?<sup>[10]</sup>.

**K. Preserving the Privacy of Sensitive Relationship in Graph data.**

Elena Zheleva et al<sup>[11]</sup> worked on link re- identification for sensitive relationship in graph data. They focused on the five different link re-identification strategies which vary in terms of the amount of data removed and the amount of privacy preserved. Strategies comes are

1. Link re-identification using observation.
2. Amount of information disclosed.
3. Utility.
4. Link re-identification in cluster-edge anonymization.
5. Link re-identification in cluster-edge anonymization with constraints<sup>[11]</sup>.

**L. A New Framework of Privacy Preserving Data Sharing.**

Xia Chen et al<sup>[12]</sup> they used a constraint based inverse item set lattice mining technique to automatically generate a sample dataset to be related for sharing. They also balance between privacy and preserving and data sharing<sup>[12]</sup>.

**M. Implementation of Cryptography for Privacy Preserving Data Mining.**

Anand Sharma et al<sup>[13]</sup> they worked on Cryptographic protocols for secure and store sensitive data and providing access to the stored data based on an individual's role. They described the generic construction for the two party and multi party scenario. The main parameter of that research is the size of the best combinatorial circuit that computes the function that is evaluated<sup>[13]</sup>.

**N. Privacy and utility Preserving Task Independent Data Mining.**

E.Poovammal et al<sup>[14]</sup> worked on to design micro data sanitization technique to preserve privacy against proximity and divergence attack and also to preserve the utility of the data for any type of mining task. This approach applies a graded grouping transformation on numerical sensitive attribute and a mapping table based transformation on categorical sensitive attribute<sup>[14]</sup>.

**O. Data Privacy through Optimal K-anonymization.**



Roberto J. Bayardo et al <sup>[15]</sup> worked on de-identification procedure which is known as K-anonymization. They implemented data management strategies that avoid repeatedly sorting the entire dataset for markedly reduced node evaluation times. A feature of K-anonymity is its preservation of data integrity <sup>[15]</sup>.

### III. CONCLUSION AND FUTURE WORK

We found so many techniques, methods and algorithms for privacy preserving for data. Still we have to work on that privacy preserving data mining techniques and try to develop customize algorithm and get accurate result and secure data from others.

### REFERENCES

1. Mathiya, Bhavin J., and Vinodkumar L. Desai. "Apache Hadoop Yarn Parameter configuration Challenges and Optimization." *Soft-Computing and Networks Security (ICSNS), 2015 International Conference on.* IEEE, 2015.
2. Evfimievski, Alexandre, et al. "Privacy preserving mining of association rules." *Information Systems* 29.4 (2004): 343-364.
3. Tomar, Archana, Vineet Richhariya, and Mahendra Ku Mishra. "AImproved PRIVACY PRESERVING ALGORITHM US-ING ASSOCIATION RULE MINING IN CENTRALIZED DA."
4. Bhuyan, Hemanta Kumar, Maitri Mohanty, and Smruti Rekha Das. "Privacy Preserving for Feature Selection in Data Mining Using Centralized Network." (2012).
5. sGoethals, Bart, et al. "On private scalar product computation for privacy-preserving data mining." *Information Security and Cryptology–ICISC 2004.* Springer Berlin Heidelberg, 2005. 104-120.
6. Chaudhuri, Kamalika, and Claire Monteleoni. "Privacy-preserving logistic regression." *Advances in Neural Information Processing Systems.* 2009.
7. Vaidya, Jaideep, and Chris Clifton. "Privacy-preserving k-means clustering over vertically partitioned data." *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM, 2003.
8. Oliveira, Stanley RM, and Osmar R. Zaiane. "Privacy preserving frequent itemset mining." *Proceedings of the IEEE international conference on Privacy, security and data mining-Volume 14.* Australian Computer Society, Inc., 2002.
9. Agrawal, Shipra, and Jayant R. Haritsa. "A framework for high-accuracy privacy-preserving mining." *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on.* IEEE, 2005.
10. Evfimievski, Alexandre, Johannes Gehrke, and Ramakrishnan Srikant. "Limiting privacy breaches in privacy preserving data mining." *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems.* ACM, 2003.
11. Zheleva, Elena, and Lise Getoor. "Preserving the privacy of sensitive relationships in graph data." *Privacy, security, and trust in KDD.* Springer Berlin Heidelberg, 2008. 153-171.
12. Chen, Xia, Maria Orłowska, and Xue Li. "A new framework of privacy preserving data sharing." *Proceedings of 4th IEEE International Workshop on Privacy and Security Aspects of Data Mining,* IEEE Press. 2004.
13. Sharma, Anand, and Vibha Ojha. "Implementation of cryptography for privacy preserving data mining." *International Journal of Database Management Systems (IJDMMS)* 2.3 (2010): 57-65.
14. Poovammal, E., and M. Ponnaivaikko. "Privacy and utility preserving task independent data mining." *International Journal of Computer Applications* 1.15 (2010): 104-111.
15. Bayardo, Roberto J., and Rakesh Agrawal. "Data privacy through optimal k-anonymization." *Data Engineering, 2005. ICDE 2005. Proceedings. 21st International Conference on.* IEEE, 2005.