MJRET
Open Access
Multidisciplinary Journal of Research in Engineering and Technology
Journal homepage: www.mjret.in
**ISSN:2348 – 6953**

# PRIVACY PRESERVING USING ADDITIVE PERTURBATION BASED ON MULTILEVEL TRUST IN RELATIONAL STREAMING DATA

Ashish. E. Mane, Mrs. Sushma Gunjal
Department of Computer Engineering, Dr. D. Y. Patil School Of Engineering,
Lohegaon, Pune, India
ashishan32@gmail.com , sushma.bhasgi@gmail.com

***Abstract:*** *Data perturbation is one of the very popular model that is used for privacy preserving data mining. The previous privacy preserving solutions were limited to only single level trust, which was not sufficient to preserve the privacy of information. So by expanding the scope from single level trust, here in the proposed system, multilevel trust solution for privacy preservation is applied in which data owner generates the different perturbed copies of same data for data miners of different trust levels. In the proposed system additive perturbation approach is used to generate the perturb copies of the relational streaming data. The data miner may mangle the different perturbed copies at different trust levels to collect the extra information about the original data, this is called the diversity attack. So to prevent from diversity attack Multilevel Trust Privacy Preserving Data Mining (MLT-PPDM) approach is used with the addition of Gaussian noise added to original relational data.*

***Keywords:*** *Privacy Preserving Data Mining (PPDM), MLT-PPDM, Relational Streaming data, Additive Perturbation, Batch Generation.*

## 1. INTRODUCTION

Data Mining is defined as the “Discovery of the models for the data” from large set of the databases in data warehouse. Today the data storage requirements are becoming large, as the large number of data is evolving day by day. As the Large database is required to store the large amount of data, so the privacy of the data is also an important factor. Privacy preserving data mining [2], [3], [4] helps to achieve the aim of data mining by preserving the privacy of sensitive data.

The Data perturbation is one of the very popular and critically interesting technique [2], [3] that is to be employed for preserving the privacy of the sensitive data contained in the dataset. By modifying the intelligently selected portion of attribute-values pairs of its transactions privacy of the sensitive original data is maintained or preserved cleverly. The technique employed makes the unrevealed values inaccurate, thus protecting the sensitive data. Previously there was the single level trust assumption procedure applied for the data miners, in which the data owner used to generate the single perturbed copy of the required data. The single level trust approach established insecurity about sensitive values of data, before data made available to third parties.

Data perturbation approaches divides into two main categories namely probability distribution approach and the value alteration approach. The probability distribution approach alters the data with another sample from the same distribution or by the distribution itself. On the other hand, the value alteration approach perturbs the attributes values directly by some additive or multiplicative noise before it is forwarded to the data miner.

So to overcome the drawbacks and shortcomings of the single level trust scenario approach, the new technique MLT-PPDM- Multilevel Trust Privacy Preserving Data Mining [1] is introduced.

In the MLT-PPDM approach the high level data miner can access the perturbed copies of low trust level. Combining the different perturbed copies generated for the different trust levels, the data miner may try to reconstruct or gain the access to original data, this is called diversity attack. To achieve this the additive perturbation approach is used in which Guassian noise is added.

## 2. RELATED WORK

Privacy preserving data mining was first proposed in [2] and [8]. To address this problem, researchers have since proposed various solutions that fall into two broad categories based on the level of privacy protection they provide. Actually there are two important parts of PPDM approach, they are Secure Multiparty Computation (SMC) [6], the basic idea of this approach is that a computation is secure. If at the completion of the computation, others are not able to know anything except its own input and the results. But it is expensive to use. So another approaches are preferred.

The another part is data perturbation which includes various techniques [1] i) Additive perturbation ii) Matrix Multiplicative perturbation iii) K-anonymity iv) Data swapping v)Micro-aggregation vi) Resampling and vii) Data shuffling.

In the additive perturbation technique [1] the data owner adds certain noise to the attribute values of the original data and generates the perturbed copy, so that it becomes hard for the data miner to recover the original data record. The main concern in this paper is the additive perturbation technique [7]. In existing system the Gaussian noise [1] which is added to generate the perturbed copies of data will be incorporated in our proposed system.

The next data perturbation technique i.e. Matrix Multiplicative technique [5], [8] states the chances of affording multiplicative random projection matrices for constructing a new representation of data. The converted data in the new generated representation is forwarded to the data miner.

K-anonymity approach: The K-anonymity [9] contains two methods i.e., Generalization and Suppression techniques. In generalization method the attribute values are generalized. For example, the date of birth can be generalized in the form of year of birth. In the suppression technique the attribute values which are completely removed from the data decrease the threat of recognition with use of public records, while decreasing the accuracy of applications on the changed data.

Data swapping method maintain the secrecy in datasets that contain categorical variables and it can transform by replacing values of private variables between individual records. In Micro-aggregation data is clustered into small group before publication. The common value of the group restores each value of the individual.

## 3. PROPOSED WORK

We propose the new concept in our proposed system i.e the usage of relational streaming data which is not used in the existing system. All the work will be done on the relational streaming data [10].

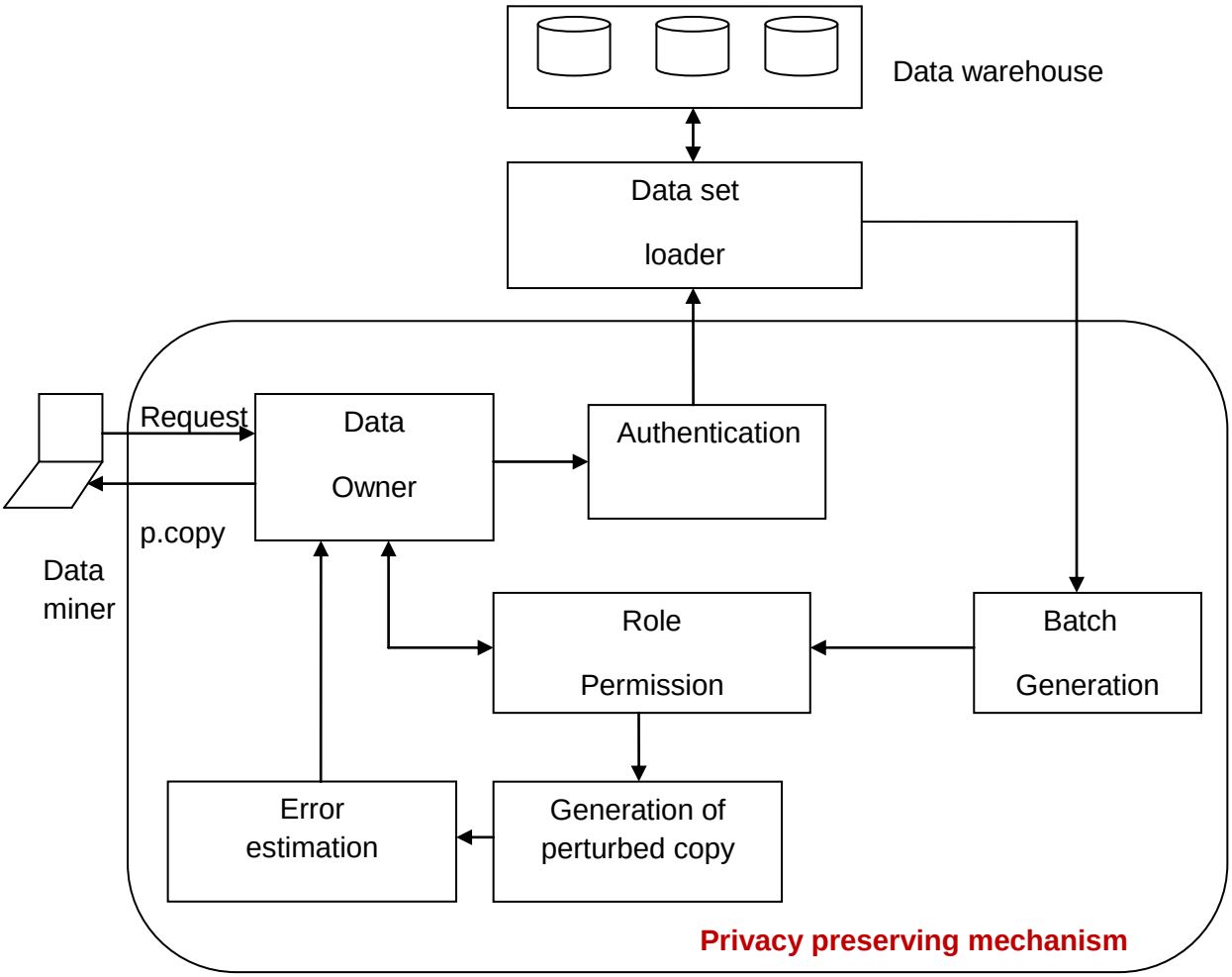


*Fig 1: Proposed System architecture Diagram*

The above figure depicts the working of the proposed system. The proposed system should work as given below :

***Step 1:*** In the first stage data miner sends request to the data owner for required data. The data owner performs the required authentication for the data miner to provide required services to the data miner.

***Step 2:*** The data set is loaded into the dataset loader from the data warehouse successfully. Then the sensitive information is extracted from the selected dataset. The extracted information is send to the Batch Generation module.

***Step 3:*** Then next stage is Batch generation process [1], we propose two batch generation algorithms, the first algorithm generates the 1 to M number of noise in parallel while the another algorithm generates the noise sequentially. After generation of noises, the noise is added to the sensitive information which is extracted from the selected relational streaming dataset. After addition of noise the perturbed copy is generated. Here more than one

perturbed copies are generated based on the data miner request. We are using Additive perturbation approach for the generation of perturbe copy of the relational streaming data.

***Step 4:*** The next stage is Role permissions, in this stage we apply the MLT-PPDM approach and will do the registration for each data miner. Here code is generated for each data miner for identifying the trust level of the data miner. After the determination of the trust level request is send to the data owner. We define three trust levels for data miner namely high, medium and low.

***Step 5:*** After the determination of trust level of data miner, the perturbed copies are to be generated for the data miner based on their respective trust level. For each trust level, request is send to the data owner, the data owner adds the noise according to the trust level of the data miner and generates the perturbed copies by adding the higher or lower amount of noise. The percentage of the noise addition depends on the trust level of the data miner, if the trust level of the data miner is high, less noise is added to the perturbed copy and vice versa i.e if the trust level is low, then more noise is added to the perturbed copy. Thus the different perturbed copies are generated for different trust levels of the data miner.

***Stage 6:*** In the next stage error estimation of perturbed copies will be done and the data owner sends the perturbed copies to the data miner. In this way the original data is not revealed to the data miner by achieving the privacy of the sensitive information.

## 4. CONCLUSION

In this paper we have discussed the privacy preserving techniques for data mining, and the usage of the best applicable data perturbation technique with multilevel trust privacy preserving data mining approach by extending the scope from single level trust approach. We have also discussed the batch generation approach. Here the different processes are to be done on the relational streaming data.

## ACKNOWLEDGEMENT

I like to express my thanks to my guide Prof. Mrs. Sushma Gunjal and to the H.O.D of the department, Prof. Soumitra S Das. To the principal and at last but not least to the P.G coordinator and the departmental staff for their necessary guidance.

## REFERENCES


[1] *Yaping Li, Minghua Chen, Qiwei Li, and Wei Zhang "Enabling Multilevel trust in Privacy preserving data mining", IEEE TRANSACTIONS ON KNOWLEDGE AND ENGINEERING, VOL. 24, NO. 9, SEPTEMBER 2012.*

[2] *D. Agrawal and C.C. Aggarwal, "On the Design and Quantification of Privacy Preserving Data Mining Algorithms," Proc. 20th ACM SIGMOD-SIGACT-SIGART Symp. Principles of Database Systems (PODS '01), pp. 247-255, May 2001.*

[3] *R. Agrawal and R. Shrikant, "Privacy Preserving Data Mining,", Proc. ACM SIGMOD Int'l Conf. Management of Data 2000.*

[4] *Y. Lindell and Benny Pinkas, "Privacy Preserving Data Mining", Proc. Int'l Cryptology Conf. (CRYPTO), 2000.*

[5] *K. Chen and L. Liu, "Privacy Preserving Data Classification with Rotation Perturbation," Proc. IEEE Fifth Int'l Conf. Data Mining,2005.*

[6] *Z. Huang, W. Du, and B. Chen, "Deriving Private Information From Randomized Data," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), 2005.*

[7] *F. Li, J. Sun, S. Papadimitriou, G.Mihaila and I.Stanoi, "Hiding in the Crowd : Privacy preservation on Evolving Streams through Correlation Tracking." Proc.ACM SIGMOD Int'l Conf data eng..(ICDE)2007.*

[8] *[8] K. Liu, H. Kargupta, and J. Ryan, "Random Projection-Based Multiplicative Data Perturbation for Privacy Preserving Distributed Data Mining," IEEE Trans. Knowledge and Data Eng., vol. 18, Jan. 2006.*

[9] *C.C. Aggarwal and P.S. Yu, "A Condensation Approach to Privacy Preserving Data Mining," ProcInt'l Conf. Extending Database Technology (EDBT), 2004.*

[10] *Walid .G. Aref, Arif Ghaffor and nagabhushana Prabhu, "Accuracy constrained privacy preserving A ccess control Mechanism for Relational data", ieee transactions on knowledge and data engineering,vol 26,NO .4 ,April 2014.*