

Title of the Project : **AN EFFICIENT TECHNIQUE FOR PRIVACY PRESERVING DATA MINING**

Funding Agency : University Grants Commission (UGC) – MRP, New Delhi

Approval No. : F. No. 37-630/2009(SR) dated 11.01.2010

Period of Execution : 2010 – 2012

Principal Investigator : Dr. S. Vijayarani  
Assistant Professor  
Department of Computer Science  
Bharathiar University  
Coimbatore – 641 046



## **EXECUTIVE SUMMARY**

Data mining has been a popular research area which extracts hidden information from the large data repositories. In recent years, a progress in the computer hardware industry has led to boost the ability to store and record a lot of personal data about clients and individuals. This has created ways in which the personal information may be misused for various purposes. In order to overcome these problems, many techniques have recently been proposed to execute the data mining tasks in a privacy-preserving way. Privacy preserving data mining is a innovative research trend in data mining where data mining algorithms are analyzed for the side effects. The primary intention of researchers is to propose and develop novel and efficient security techniques, algorithms and methods for protecting the confidential data and privacy of people.

In this project, two important privacy problems, such as micro data protection and association rule hiding are considered. Micro data protection is one of the sub-disciplines of statistical disclosure control. Micro data protection methods can generate the protected micro data set by masking the original data. Masking techniques are classified into perturbative and non-perturbative. Perturbative masking techniques alter an attribute value by a new value. The confidential attributes of the data set are

distorted before publication. Non-perturbative masking techniques do not alter the data instead they produce partial suppressions or reductions of detail in the original dataset.

The main objective of the micro data protection research is to develop techniques which efficiently modify the confidential data items and also perform the data mining tasks in a secured manner. In this micro data protection research project, four new perturbative masking techniques, namely, *data transformation*, *bit transformation*, *bit++* and *bit--* are proposed for protecting the data items of the confidential numeric attributes of the micro data. The performances of the proposed masking techniques are compared with the existing perturbative masking techniques such as *additive noise*, *micro aggregation* and *rounding*. Based on the experimental results, it can be seen that the proposed *Data Transformation* perturbative masking technique had given better results than the existing techniques.

Association rules are inferences that hold in a transactional database under certain user-specified parameters that account for their importance. Significant association rules supply knowledge to the data miner as they efficiently recapitulate the data, while finding any concealed relations among items that hold in the data. Association rule hiding discusses, different hiding techniques to change the database because it is necessary to protect sensitive association rules and ensure that the techniques will not disturb the non-sensitive rules. The job of association rule hiding is to accurately alter or change the original data so that any association rule mining algorithms that may be applied to the sanitized version of the data will be incapable to mine the sensitive rules under certain parameter settings. The sanitized data will be able to mine all the non-sensitive rules that appeared in the original dataset under the same or higher parameter settings and no other fake rules will be generated.

In this association rule hiding research work, two new hiding techniques are proposed, namely, *Hiding Technique based on Tabu Search* and *Hiding Technique based on Ant Colony Optimization*. These techniques are used for modifying the sensitive items for protecting the sensitive association rules. The performances of these hiding techniques are compared. The factors used for measuring the performance of the proposed techniques are hiding failure, misses cost, artifactual error and efficiency. Based on the experimental results, it can be observed that, the proposed *Hiding Technique based on Tabu Search* has produced better results than the ant colony optimization technique.

## **Publications:**

1. S.Vijayarani, Dr. A.Tamilarasi, "*K-Anonymity Techniques – A Review*", International Journal of Computer Science and Applications (IJCSA) – ISSN 0974-0767. Issue 1, pp 170-177, IJCSA 2010.
2. S.Vijayarani, Dr. A.Tamilarasi, "*An Efficient K-Anonymity Technique*", CiiT International Journal of Data Mining Knowledge Engineering (IJDMKE), ISSN 0974-9683, Vol 2, No 7, pp 117-122, July 2010.
3. S.Vijayarani, Dr. A.Tamilarasi, "*Evaluating Clustering Performance for the Protected Data using Perturbative Masking Techniques in Privacy Preserving Data Mining*", International Journal of Computer and Network Security (IJCNS), PISSN : 2076-2739, EISSN : 2076-9199, Vol.2, No.10, pp 145-149, Oct 2010.
4. S.Vijayarani, Dr. A.Tamilarasi, "*Bit Transformation Perturbative Masking Technique for Protecting Sensitive Information in Privacy Preserving Data Mining*", International Journal of Database Management Systems (IJDMMS), ISSN : 0975-5705 (Online); 0975-5985 (Print), Vol.2, No.4, pp 107-114, Nov 2010
5. S.Vijayarani, Dr. A.Tamilarasi, "*Data Transformation Technique for Protecting Private Information in Privacy Preserving Data Mining*", Advanced Computing: An International Journal (ACIJ), Vol.1, No.1, ISSN : 2229 - 6727 [Online] ; 2229 - 726X, [Print], pp 1-9, November 2010.
6. S.Vijayarani, Dr.A.Tamilarasi, R.SeethaLakshmi, "*Privacy Preserving Data Mining Based on Association Rule- A Survey*", International Conference on Communications and Computational Intelligence (INCOCCI 2010), Kongu Engineering College, Erode on Dec 27-29, 2010. Conf.proc. published by Scitech Pub. ISBN No:978-81-8371-369-6. P.No:99-103.IEEE Xplore. INSPEC Accession Number: 11887500.
7. S.Vijayarani, Dr.A.Tamilarasi, M.Sampoorna, "*Analysis of Privacy Preserving K-Anonymity Methods and Techniques*", International Conference on Communications and Computational Intelligence (INCOCCI 2010), Kongu Engineering College, Erode on Dec 27-29, 2010. Conference proceedings published by Scitech Publishers, ISBN No.: 978-81-8371-369-6. P.No:540-545. IEEE Xplore. INSPEC Accession Number: 11887481
8. S.Vijayarani, Dr. A.Tamilarasi, "*An Efficient Masking Technique for Sensitive Data Protection*", International Conference on Recent Trends in Information Technology (ICRTIT, 2011), Madras Institute of Technology, Chennai, June 3- 5, IEEE Catalog no. CFP1122P-CDR, ISBN No.978-1-4577-0589-2
9. S.Vijayarani, Dr. A.Tamilarasi, "*Confidential Numeric Data Protection in Privacy Preserving Data Mining*", International Journal of Computational Intelligence and Informatics, Vol. 1: No.2, July-September 2011, ISSN: 2231 – 0258.