Journal Homepage: - www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI: 10.21474/IJAR01/1592
DOI URL: http://dx.doi.org/10.21474/IJAR01/1592
ISSN NO. 2320-5407

RESEARCH ARTICLE

Survey paper on privacy preserving data mining.

Mrs. Revathy Swaminathan and Dr. T. Arun Kumar.
Scope School, VIT University, Vellore.

Manuscript Info

Manuscript History

Received: 16 July 2016
Final Accepted: 13 August 2016
Published: September 2016

Key words:-
privacy preserving, PPDM, Data Hiding, Knowledge Hiding, Hybrid technique.

Abstract

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information, which gets exposed to other parties during data mining process. Privacy preserving data mining technique has been introduced with the aim of protecting sensitive information of an individual. A variety of methodologies has been developed for this privacy preserving data mining such as hiding data, hiding knowledge (Rules), hybrid technique. The idea of PPDM is to hide sensitive information from unauthorized access and at the same time preserving utility of the information. In this paper Iam going to present the review of different privacy preserving techniques which will be used during data mining process to maintain efficiency and utility of the information.

Introduction:-

Data mining is one of the essential steps in the processes of knowledge discovery of databases. Data mining has the feature of extracting information from huge amount of data available in various areas such as customer relationship management, market basket analysis. The mined information can be a patterns, rules, clusters or classification models. Data mining processes which outputs information typically contains sensitive information of an individual which get exposed to several parties. This information when linked with public data it's possible to get sensitive information of an individual. So, PPDM deals with protecting privacy of sensitive data without compromising utility of the information. The main goal of PPDM techniques or algorithm is that mine appropriate information without revealing sensitive

Information to other parties at same time accuracy of data is maintained. The problem has been discussed in multiple communities such as the database community, the statistical disclosure control community and the cryptography community. The key guidelines in the field of privacy-preserving data mining are [1]:

– Privacy-preserving data publishing
– Changing the results of data mining applications to preserve privacy
– Query auditing
– Cryptographic methods for distributed privacy
–Theoretical challenges in high dimensionality

Corresponding Author:- Mrs. Revathy Swaminathan.
Address:- Scope School, VIT University, Vellore.

Ultimate goal of privacy preserving data mining is to develop techniques or algorithms for changing the original data so, that private information or private data remains private after mining from data. In this paper I provide overall techniques or methodologies applied for protecting privacy of an individual without compromising utility of the data. The data which is stored may be either centralized or distributed. In either case we need to preserve privacy of sensitive information while mining for useful information. In distributed, data will be distributed among several parties at different nodes. These distribution of partition may be either horizontal or vertical partition and data stored in each site has to be protected from other parties so separate algorithms and techniques has been applied for this distributed data which is different from centralized where data stored at single site.
Main research goal of privacy preserving data mining (PPDM) is to protect the sensitive information or private knowledge from leaking in the mining process, for the moment obtains the accurate results of data mining. The privacy preserving data mining is divided into two categories:

1. First level of PPDM is focus on protecting the sensitive data such as id, name, address and other sensitive information.
2. Second level of PPDM is focus on protecting the sensitive knowledge which is used during data mining activities.

Many privacy preserving techniques are using some form of conversion to achieve privacy. Privacy preserving is mainly focused on data distortion, data reconstruction and data encryption technology. The implementation of PPDM techniques has become the
Demand of the moment. The aim of this paper is to present the review on privacy preserving techniques which is very helpful while mining process over large data sets with reasonable efficiency and preserve security. Section 2 discusses about the architecture of PPDM, Section 3 discusses about various techniques and methodologies used for preserving data/information/knowledge and last Section4 discuss about conclusion and future aspects of PPDM

PPDM Architecture:-

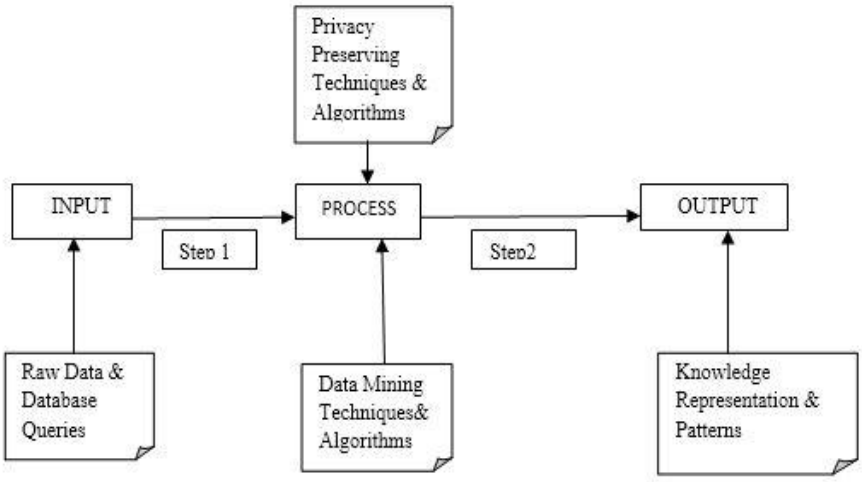


Fig.1 PPDM Architecture

The Architecture for PPDM is shown in fig.1. In the process of discovery of knowledge from databases first the raw data is collected from single or various sites and it will be stored in the respective databases. Then, it gets transformed into a format that suits for analysis ,and gets stored in large data warehouses and next step is application of data mining algorithms and techniques along with privacy preserving techniques and algorithms to ensure privacy. Last step is the output that is generation of information/knowledge.

At step 1,the raw data collected from single or multiple data warehouses or even data marts gets transformed into a format that suits for analysis. Even at this step, privacy issues are needed to be taken care of. Different techniques suitable for privacy issues have been applied but most of them deal with making raw data for analysis purpose.
At step 2, the raw data is subjected to various processes to make the data sanitized so that it can be revealed to all parties of data miners. The privacy techniques applied at this step are blocking, suppression, perturbation, modification, randomization, generalization etc. Then data mining algorithms are applied to the processed data for

knowledge/pattern discovery. Even the data mining algorithms are customized for the idea of protecting privacy without sacrificing the goals of data mining.
At step 3, the output of data mining which is information will be checked for its privacy so that disclosure of sensitive attributes will be avoided. Each of the three steps, application of privacy techniques/algorithms is available but application or combination of any of these can be used.

PPDM Techniques:-

Techniques of PPDM can be broadly classified into three categories which were introduced in the techniques of PPDM paper [2]. They are
1. Data Hiding techniques.
2. Knowledge Hiding techniques.
3. Hybrid techniques.

A. Data Hiding Techniques

This technique sanitizes the data before it gets published for the purpose of data mining task. The input data provided for data mining will be blocked in such a way that sensitive information will not get exposed to other parties. There are different ways of implementing these techniques which are explained in detail.

B. Knowledge Hiding Techniques

This technique hides rules which are generated from the process of data mining task. These techniques are very important because the sensitive rules extracted by data mining process can be used to derive confidential information. There are different ways of implementing these techniques which are explained in detail.

C. Hybrid Technique

Each techniques mentioned above have some advantages and disadvantages. None of the technique is perfect. In hybrid technique, an effort has been made to combine any two techniques mentioned above in order to get one prefect technique. Some of the hybrid techniques are mentioned.

3.1 Data Hiding Technique:-

Data Hiding:-

- A. Data perturbation Technique
  - 1. Randomization
  - 2. Data Swapping
- B. **C**ryptographic Technique (Or) Secure Multiparty Approach
- C. Anonymization Technique
  - 1. Generalization
  - 2. Suppression
- D. Condensation Technique

A. Data Perturbation

Perturbation based PPDM approach introduces random data to get modified before the data get published for the purpose of data mining activities. Perturbation of data is a very easy and effective method for protecting sensitive information of the data from unauthorized users. This approach of perturbation is discussed in [3].there are two types of data perturbation for protecting data namely [4]:
1. Probability distribution approach (or) Data Swapping: This approach takes the data and replaces it from the same distribution sample or from the distribution itself.
2. Value distortion approach (or) Noise addition: This approach perturbs data by adding noise, or other randomized processes.

In perturbation based PPDM, the original values of the data are replaced with some artificial values so that the result computed from the perturbed data does not differ from the result computed from the original data to a larger extent. But the individual records of the perturbed data are of no use as only statistical properties of records are preserved. As the perturbed data records does not match with the original records, the attacker cannot recover the sensitive information from the perturbed data. Perturbation of data can be done by adding noise to the data, swapping the data or replacing values of the original data with artificial values.

Data swapping technique were first introduced by Dalenius and Reiss in 1982, for categorical values modification in the context of secure statistical databases [5]. The main idea of the method was it keeps all original value in the data set, while at the same time makes the record re-identification very complex. This method actually replace the original data set by another one where some original values belonging to a sensitive attributes are exchanged between them. An introduction to existing data swapping technique can be found in [6], [7].Here, records are exchanged in such a way that low-order frequency counts are maintained.

Noise addition approach perturbs data by adding random noise to the original values of numerical attributes This random number is generally drawn from a normal distribution with zero mean or standard deviation. Noise is added in a controlled way so as to maintain variance, co-variance and means of the attributes of a data set. Due to the absence of natural ordering in categorical values, addition of noise in categorical attributes is not straightforward as like addition of noise in numerical attributes. Many techniques proposed for noise addition in data mining. Evfimievski et al. proposed a novel noise addition technique for privacy preserving association rule mining in 2002 [8]. Agarwal and Srikant proposed a noise addition technique in 2000 which is based on addition of random noise to attribute values in such a way that the distributions of data values belonging to original and perturbed data set were very difficult [9]. Du and Zhan presented a decision tree building algorithm which is used to perturb multiple attributes [10]. In 2004 Zhu and lieu [11] proposed a general framework for randomization using a well studied statistical model called mixture model. According to this scheme data are generated from a distribution that depends on some factors including original data as well. Their randomization framework supports privacy preserving density estimation.

B. Cryptographic Technique:-

Cryptography is a technique which is used for data encryption in that sensitive attributes present in the data will be encrypted. Many cryptographic approaches have been proposed in the context of privacy preserving data mining algorithms. Cryptography based approaches like secure Multiparty computation (SMC) are secure at the end of computations. In [12], authors introduced cryptographic technique which is very popular because it provides security and safety of sensitive attributes. There are different algorithms of cryptography available. In that [13] presents four secure multiparty computations based on the methods that can support privacy preserving data mining. SMC is mainly uses in distributed environment. The purpose of SMC is that it is necessary to guarantee the correctness of the calculation, but also to protect their respective input and output data from leaking when two or more participants who are carrying out the cooperation calculation. There are different SMC techniques for different type of computation [2]. Secure Sum is one SMC technique which is used parties to find sum of their local values securely. Secure Set union is another example of SMC technique which is used by parties to securely compute union of all private items owned by parties, without revealing the owner of the item. Secure set sum and secure set union methods can be used for securely mining association rule on horizontally distributed data. An association rule is an implication of the form, X =>Y. The rule X =>Y holds in the transaction set D with confidence c if c percent of transactions in D containing X also contain Y. The rule X => Y has support s in the transaction set D if s percent of transactions in D contain X < Y. In association rule mining, we need to find all association rules in D with support s > st and confidence c > ct, where st and ct are user-defined thresholds for "interesting" rules. By combining Secure Set Union and Secure Sum method this association rule mining can be carried out securely on horizontally distributed data .Each party Pi can find all possible association rules and local confidence, support for that rules using its local data. The rules having support and confidence greater than threshold (ct and st) will be added to the local set of association rule, say LRi (for ith party). Then all sites can give their local rule set as an input to secure set union algorithm and participate in the algorithms to find global set of association rule say GR. After getting GR, each party can calculate local confidence and local support for each rule in global set GR. Next step is to find global support and global confidence for each rule in GR. For this secure sum method can be used. One rule from GR can be selected at a time and for that rule each party can give their local support and confidence as input to secure sum method. The method returns global support and global confidence for selected rule. Association rule having global confidence and support greater than threshold (ct and st) can be added in final output set.

C. Anonymization Technique:-

Anonymization is an approach that reduces the risk of individual identity disclosure whereas the data still remains realistic. The basic form of the data in a table consists of following four types of attributes proposed in [14]:

(i) Explicit Identifiers is a set of attributes containing information that identifies a record owner explicitly such as name, SS number etc.
(ii) Quasi Identifiers is a set of attributes that could potentially identify a record owner when combined with publicly available data.
(iii) Sensitive Attributes is a set of attributes that contains sensitive person specific information such as disease, salary etc.
(iv) Non-Sensitive Attributes is a set of attributes that creates no problem if revealed even to untrustworthy parties.

To protect individuals' identity when releasing sensitive information, data holders often encrypt or remove explicit identifiers, such as names and unique security numbers. However, unencrypted data provides no guarantee for anonymity. In order to preserve privacy, k-anonymity model has been proposed by Sweeney [15] which achieves k-anonymity using generalization and suppression [15], In K-anonymity, it is difficult for an imposter to decide the identity of the individuals in collection of data set containing personal information. Each release of data contains every combination of values of quasi-identifiers and that is indistinctly matched to at least k-1 respondents. Generalization involves replacing a value with a less specific (generalized) but semantically reliable value. For example, the age of the person could be generalized to a range such as youth, middle age and adult without specifying appropriately, so as to reduce the risk of identification. [15] Suppression involves reduce the exactness of applications and it does not liberate any information .By using this method it reduces the risk of detecting exact information. A survey on most of the common attacks techniques for anonymization-based PPDM & PPDP is presented in [16] and their effects on Data Privacy are explained. A new approach for building classifiers using anonymized data by modeling anonymized data as uncertain data is proposed in [17]. In [18], a novel technique called slicing is proposed, which preserves better data utility than generalization and can be used for attribute disclosure protection and membership disclosure protection.

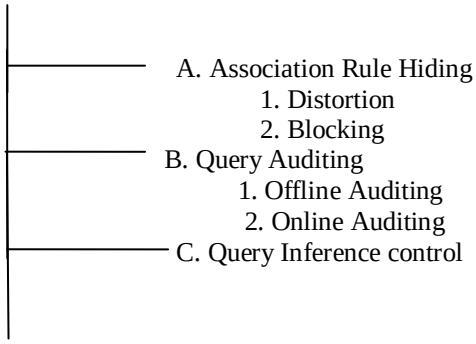
D. Condensation Technique:-

Condensation approach was introduced by Charu C. Aggarwal and Philip [19] which builds constrained clusters in the data set and after that produces pseudo-data. The basic concept of the method is to contract or condense the data into multiple groups of predefined size. For each group, certain statistics are maintained. This approach can be effectively used for the classification problem. The use of pseudo-data provides an additional layer of protection, as it becomes difficult to perform adversarial attacks on synthetic data. Moreover, the aggregate behavior of the data is preserved, making it useful for a variety of data mining problems [20].condensation technique achieves better efficiency in privacy preserving compared to other techniques as it uses only pseudo data rather than modified data which is applied in other techniques. Moreover, it works even without redesigning data mining algorithms since the pseudo data has the same format as that of the original data. It is very effective in case of data stream problems where the data is highly dynamic. At the same time, data mining results get affected as huge amount of information is released because of the compression of a larger number of records into a single statistical group entity [21].

3.2 Knowledge Hiding Techniques:-

In this approach, sensitive knowledge extracted from the Data Mining process is excluded for use. These techniques are as important as data hiding techniques because knowledge extracted in data mining process can be used to derive confidential information. This problem is also commonly called the "database inference problem". Following are different ways used for hiding knowledge.

Knowledge Hiding



A. Association Rule Hiding:-

Association rule hiding is one of the techniques used by PPDM to hide sensitive rules generated by association rule mining. The objective of association rule hiding is that to protect access to sensitive information that can be obtained through non-sensitive data and inference rules. Some of the earliest work on the challenges of association rule mining for database security may be found in [22].

Two broad approaches applied for association rule hiding:

Distortion: In distortion [23], the entry for a given transaction is modified to a different value. Since, we are typically dealing with binary transactional data sets, the entry value is flipped. A formal proof of the NP-hardness of the distortion method for hiding association rule mining may be found in [22]. In [22], techniques are proposed for changing some of the 1-values to 0-values so that the support of the corresponding sensitive rules is appropriately lowered. The utility of the approach was defined by the number of non-sensitive rules whose support was also lowered by using such an approach. This approach was extended in [24] in which both support and confidence of the appropriate rules could be lowered. In this case, 0-values in the transactional database could also change to 1-values. In many cases, this resulted in spurious association rules (or ghost rules) which were an undesirable side effect of the process.

Blocking: In blocking [25], the entry is not modified, but is left incomplete. Thus, unknown entry values are used to prevent discovery of association rules. The broad idea of blocking was proposed in [26]. The attractiveness of the blocking approach is that it maintains the truthfulness of the underlying data, since it replaces a value with an unknown (often represented by '?') rather than a false value. Some interesting algorithms for using blocking for association rule hiding are presented in [27]. The work has been further extended in [25] with a discussion of the effectiveness of reconstructing the hidden rules. Another interesting set of techniques for association rule hiding with limited side effects is discussed in [28]. The objective of this method is to reduce the loss of non-sensitive rules, or the creation of ghost rules during the rule hiding process. In [29], it has been discussed how blocking techniques for hiding association rules can be used to prevent discovery of sensitive entries in the data set by an adversary. In this case, certain entries in the data are classified as sensitive, and only rules which disclose such entries are hidden. An efficient depth-first association mining algorithm is proposed for this task [29].

B. Query Auditing and Inference Control:-

Query Auditing and Inference control is an approach in which sequence of queries will be posted through a public interface to get results from the database. This leads to a danger in which an adversary poses a sequence of queries through which he or she can infer sensitive information about the data. Two broad methodologies designed to reduce the likelihood of sensitive data discovery:

Query Auditing: In query auditing, we deny one or more queries from a sequence of queries. The queries to be denied are chosen such that the sensitivity of the underlying data is preserved. Some examples of query auditing methods include [33, 35].An overview of classical methods for query auditing may be found in [32]. The query auditing problem has an *online* version, in which we do not know the sequence of queries in advance, and an *offline* version, in which we do know this sequence in advance. Clearly, the offline version is open to better optimization from an auditing point of view. The problem of query auditing was first studied in [33]. This approach works for the online version of the query auditing problem. In these works, the sum query is studied, and privacy is protected by using restrictions on sizes and pairwise overlaps of the allowable queries. In [34], a number of variations of the

offline auditing problem have been studied. In the offline auditing problem, we are given a sequence of queries which have been truthfully answered, and we need to determine if privacy has been breached.

Query Inference Control: In this case, we perturb the underlying data or the query result itself. The perturbation is engineered in such a way, so as to preserve the privacy of the underlying data. Examples of methods which use perturbation of the underlying data include [29].Examples of methods which perturb the query result include [30].A classical method for aggregate queries such as the sum or relative frequency is that of random sampling [31]. In this technique, a random sample of the data is used to compute such aggregate functions. The random sampling approach makes it impossible for the questioner to precisely control the formation of query sets.

3.3 Hybrid Technique:-

Privacy preservation is a very huge field. Many algorithms have been proposed in order to secure the data. Hybrid Technique is a new technique through which one can combine two or more techniques to preserve the data. *Sativa Lohiya and Lata Ragha* [36] proposed a hybrid technique in which they used randomization and generalization. In this approach first they randomize the data and then generalized the modified or randomized data. This technique protects private data with better accuracy; also it can reconstruct original data and provide data with no information loss. Murat Kantarcioglu and Chris Clifton also proposed a hybrid technique to combine noise addition and SMC for securely mining of association rules over horizontally partitioned data [37]. In this technique while sharing encrypted rule set to other parties, a little noise is added in the rule in form of false rules. Also in [38], AES encryption technique is combined with Anonymization to provide higher level of security.

Conclusion:-

Data mining is an important technique used in every organization to extract knowledge. But privacy and security concerns may create barrier in data mining activities. These barriers can be removed by applying variety of PPDM techniques and by ensuring security in every data mining task. So In this paper, we presented a survey of broad areas of privacy preserving data mining and the underlying techniques. I also presented the architecture of PPDM and discussed variety of data modification techniques, knowledge hiding techniques and hybrid techniques. Finally, conclusion is that there does not exists single privacy preserving data mining technique that outperforms all the other algorithm on possible criterion like efficiency, utility, complexity and cost . Different algorithm may outperform better than one another on one particular criterion.

References:-

1. Aggarwal, Charu C., and S. Yu Philip. A general survey of privacy-preserving data mining models and algorithms. Springer US, 2008.
2. Mrs. Suchitra Shelke., and Prof. Babita Bhagat. Techniques for Privacy Preservation in Data Mining. IJERT, 2015.
3. J. Liu, J. Luo and J. Z. Huang, "Rating: Privacy Preservation for Multiple Attributes with Different Sensitivity requirements", in proceedings of 11th IEEE International Conference on Data Mining Workshops, IEEE 2011.
4. https://www.techopedia.com/definition/25013/data-perturbation.
5. T. Dalenius and S. P. Reiss. Data Swapping: A technique for disclosure control. Journal of Statistical Planning and Inference, 6(1):73-85, 1982.
6. S. E. Fienberg and J. McIntyre. Data swapping: Variations on a theme by Dalenius and Reiss. Journal of Official Statistics, 21:309-323, 2005.
7. K. Murlidhar and R. Sarathy. Data Shuffling – a new masking approach for numerical data. Management Science, Forthcoming, 2006.
8. A. V. Evfimievski, R. Srikant, R. Agarwal and J. Gehrke. Privacy preserving mining of association rules. In Proc. Of the Eighth ACM SIGKDD International Conference on Knowledge and Data Mining, pages 217-228, 2002.
9. R. Agarwal and R. Srikant. Privacy –preserving data mining. In Proc. Of the ACM SIGMOD Conference of Management of Data, pages 439-450. ACM Press, May 2000.
10. W. Du and Z. Zhan. Using randomized response techniques for privacy preserving data mining. In Proc. of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 515-510, Washington DC, USA, August 2003.

11. Y. Zhu and L. Liu. Optimal randomization for privacy preserving data mining. In Proc. of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 761-766, Seattle, Washington, USA, August 2004.
12. Y. Lindell, B.Pinkas, "Privacy preserving data mining", in proceedings of Journal of Cryptology, 5(3), 2000.
13. P.Samarati,(2001). Protecting respondent's privacy in micro data release. In IEEE Transaction on knowledge and Data Engineering, pp.010-027.
14. A Survey: Privacy Preservation Techniques in Data Mining International Journal of Computer Applications (0975 – 8887) Volume 119 – No.4, June 2015
15. Pingshui WANG," Survey on Privacy Preserving Data Mining", International Journal of Digital Content Technology and its Applications,Volume 4, Number 9, December 2010.
16. A. Hussien, N. Hamza and H. Hefny, 2013 , Attacks on Anonymization-Based Privacy-Preserving: A Survey for Data Mining and Data Publishing,", Journal of Information Security, Vol.4 No. 2, 2013, pp. 101-112. doi:10.4236/jis.2013.42012
17. Inan, A., Richardson, TX, Kantarcioglu, M., Bertino, E., 2009, Using Anonymized Data for Classification, IEEE 25th International Conference on Data Engineering, 2009. ICDE '09. , pp : 429-430
18. Tiancheng Li, Ninghui Li, Jian Zhang, Ian Molloy, "Slicing: A New Approach for Privacy Preserving Data Publishing", IEEE Transactions on Knowledge & Data Engineering, vol.24, no. 3, pp.561-574, March 2012, doi:10.1109/TKDE.2010.236.
19. C. Aggarwal , P.S. Yu, "A condensation approach to privacy preserving data mining", in proceedings of International Conference on Extending Database Technology (EDBT), pp.183–199, 2004. 746.
20. Charu C. Aggarwal, Philip S.Yu "Privacy-Preserving Data Mining Models and algorithm" advances in database systems 2008 Springer Science, Business Media, LLC.
21. Gayatri Nayak, Swagatika Devi, "A survey on Privacy Preserving Data Mining: Approaches and Techniques", International Journal of Engineering Science and Technology, Vol. 3 No. 3, 2127-2133, 2011.
22. Atallah, M., Elmagarmid, A., Ibrahim, M., Bertino, E., Verykios, V.:Disclosure limitation of sensitive rules, *Workshop on Knowledge and Data Engineering Exchange*, 1999.
23. Oliveira S. R. M., Zaiane O., Saygin Y.: Secure Association-Rule Sharing.*PAKDD Conference*, 2004.
24. Dasseni E., Verykios V., Elmagarmid A., Bertino E.: Hiding Association Rules using Confidence and Support, *4th Information Hiding Workshop*,2001.
25. Saygin Y., Verykios V., Clifton C.: Using Unknowns to prevent discovery of Association Rules, *ACM SIGMOD Record*, 30(4), 2001.
26. Chang L., Moskowitz I.: An integrated framwork for database inference and privacy protection. *Data and Applications Security*. Kluwer, 2000.
27. Saygin Y., Verykios V., Elmagarmid A.: Privacy-Preserving Association Rule Mining, *12th International Workshop on Research Issues in Data Engineering*, 2002.
28. Wu Y.-H., Chiang C.-M., Chen A. L. P.: Hiding Sensitive Association Rules with Limited Side Effects. *IEEE Transactions on Knowledge and Data Engineering*, 19(1), 2007.
29. Aggarwal C., Pei J., Zhang B. A Framework for Privacy Preservation against Adversarial Data Mining. *ACM KDD Conference*, 2006.
30. Agrawal R., Srikant R., Thomas D. Privacy-Preserving OLAP. *Proceedings of the ACM SIGMOD Conference*, 2005.
31. Blum A., Dwork C., McSherry F., Nissim K.: Practical Privacy: The SuLQ Framework. *ACM PODS Conference*, 2005.
32. Denning D.: Secure Statistical Databases with Random Sample Queries.*ACM TODS Journal*, 5(3), 1980.
33. Adam N., Wortmann J. C.: Security-Control Methods for StatisticalDatabases: A Comparison Study. *ACMComputing Surveys*, 21(4), 1989.
34. Dobkin D., Jones A., Lipton R.: Secure Databases: Protection againstUser Influence. *ACM Transactions on Databases Systems*, 4(1), 1979.
35. Chin F.: Security Problems on Inference Control for SUM, MAX, andMIN Queries. *J. of the ACM*, 33(3), 1986.
36. Reiss S.: Security in Databases: A combinatorial Study, *Journal ofACM*, 26(1), 1979.
37. S. Lohiya and L. Ragha, "Privacy Preserving in Data MiningUsing Hybrid Approach", in *proceedings of 2012 FourthInternational Conference on Computational Intelligence andCommunication Networks*, IEEE 2012.
38. The IEEE computer society, 2004, "Privacy-Preserving Data Mining:Why, How, and When".
39. Mahesh Dhande, N.A.Nemade and Yogesh Kolhe, 2013, "PrivacyPreserving in K- Anonymization Databases Using AES Technique".