International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 8, August 2017)

Recent Privacy Preserving Data Mining Techniques: A Survey

Meenakshi Bansal[1], Dinesh Grover[2], Dhiraj Sharma[3]
[1]*Research Scholar,* [2]*Professor, I K Gujral PTU, Jalandhar, India*
[3]*Asst. Professor, Punjabi University, Patiala, India*

***Abstract*— Objectives: The security is the main concern in the field of data mining. Therefore Privacy preserving data mining came into existence to solve this problem. It helps to maintain the trade of between usability and privacy of the sensitive information. This paper provides in depth study of various rule hiding techniques.**

***Methods/Statistical analysis:* Since the advent of associate rule hiding problem in 1999, many techniques have been developed which are based on increasing or decreasing the support or confidence of the rules to specify its significance. But sensitivity of the rule cannot only be decided based upon its support and confidence. Therefore research community discovered other rule hiding techniques that undertake others factors along with minimizing the side effects on the original database. In this paper a comprehensive survey of recent Privacy Preserving Data Mining approaches have been done which are recently being used. Findings: Needs of different techniques are identified and their pros & cons are listed in a comparative manner. Finally this paper concludes with the discussion of useful future directions that can help researchers to identify areas where further research is needed.**

***Applications:* Comparative study of various PPDM techniques discussed in this paper can be used for statistical analysis and Knowledge Discovery Data Mining (KDDM). It also facilitate research in many areas like marketing, medicines, crime investigation.**

***Keywords*— Association Rule Hiding, Data Mining, Privacy Preserving Data Mining, Sensitive Information.**

I. Introduction

Privacy preserving data mining is one of the advance techniques which help the user to share the mined data among multiple users along with maintaining the secrecy of the sensitive information; it maintains the accuracy of data. Various Association Rule Hiding techniques are used to solve the problem. In late nineties[1] PPDM began and till now, there is great improvement in this field. A large number of techniques and approaches have been developed in this field of research. Nevertheless PPDM is a challenging task. There are many issues like privacy, trust, data quality (DQ) and malicious data mining intrusion detection system that must be taken care in context of crucial online databases, thereby making this task more complex[2].

II. Need Of PPDM

With the increase in business competition, large internet users and rapid development of e-commerce data is growing at very fast rate. These data contain users' private and sensitive information. Data mining algorithms when implemented on this data, releases both sensitive and non-sensitive information. Sharing of this mined data between multiple parties (Banks, Hospitals and Business organizations) can cause the leakage of sensitive information. Therefore need of PPDM has emerged. PPDM is the concept of preventing sensitive information from disclosure. Actual threat to people is that their private information should not be leaked without their knowledge. It makes them reluctant of sharing their data with others but PPDM provides the solution to the problem [3]. Main objectives due to which PPDM came into existence are[4]: to reduce business threat, increase data security, information retrieval and data sanitization. Figure 1 illustrates the procedure of privacy preserving data mining.

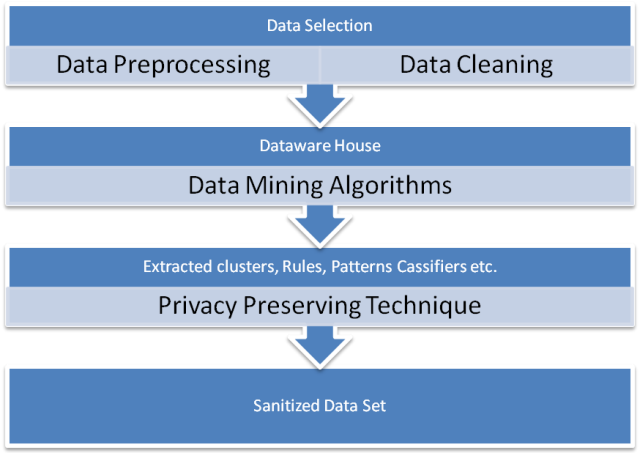

Figure 1: Framework for PPDM

III. Dimensions Of PPDM

Privacy preserving data mining has been classified by four dimensions[5]: first dimension is over distribution of data i.e. data is centralized or distributed.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 8, August 2017)

Second dimension is regarding the modification of original data used as input for data mining task. Third dimension is based on type of data mining algorithm. Fourth dimension depends upon the technique used for sanitizing the output of data mining application for privacy preservation. Following all the above dimensions original data is changed into synthetic data which can reduce the usability of the data. So PPDM causes trade off between privacy and accuracy of the data. Researchers from all over the world are investigating on the techniques which can provide better privacy while maintaining usability and accuracy of the data.

IV. Various Types Of PPDM

Data mining helps to mine knowledge from large amount of data. Main threat faced by data mining is data privacy when data is shared among multiple users. To solve this problem various PPDM techniques have been proposed which modified the data to perform privacy preservation. Various PPDM techniques proposed for providing privacy in data mining. These approaches are classified based on two categories: i) Data Hiding Approaches and ii) Knowledge hiding approaches. In data hiding various techniques like perturbation, generalization sampling or transforming data by applying cryptography techniques are used on raw data. While in Knowledge hiding techniques like heuristic approach randomization, exact approach, k-anonymity are applied to hide the sensitive data from the results obtained after applying various data mining algorithms like association rule mining, clustering, and classification etc. Classification among these techniques is done based upon the inference one can have of the original data from the modified one, loss of information and estimation of data accuracy.

Randomization is a technique which modifies the original attributes of data by adding noise to the data. Modified data maintain all the statistical values but it is hard to get back the original data. The masked records are known as perturbed records[6]. Various methods of adding the noise is additive strategy and multiplicative strategies. Additive strategy is a technique to add random noise to the non-random data to get the perturbed data. While in multiplicative strategy records are multiplied by random vectors and then transformed to preserve the inner record distance of the data. Main drawback of this technique is increase in privacy with decrease in utility of data as the original data can never be reconstructed. This technique is also vulnerable to Known input-output attack and known sample attack[7].

There are various other attacks against randomization technique e.g attacker can use noise filtering technique to get back the original values, PCA based analysis and spectral noise reduction technique can be used[8] . In spite of these attacks randomization is used in various data mining applications like classification, association rule hiding, swapping etc because of its simplicity and it is also not required to know about other records in the data, so it can be applied during the time of data collection. Applying randomization provides only distribution of the data set without giving the individual records. Distributions provided by this technique are along individual dimension while many data mining algorithms are dependent upon multivariate dimensions. So new techniques must be developed which can work upon multivariate dimensions. Challenging task of increasing dimensionalities is density estimation which becomes inaccurate. Main limitation of this technique is that all the records are treated equally irrespective of their location. As a result outliers are more acceptable to attacks as compare to the records in denser region. To solve this noise should be added to all the records in data but it leads to the loss of utility of data. Based upon the perturbation of the records randomization is of two types: additive perturbation and multiplicative perturbation. In the former technique noise is added randomly to the records. Aggregate distribution can be calculated over these randomized records by designing specific data mining algorithm. In multiplicative perturbation random projection or random rotation techniques are used in order to perturb the record.

When sensitive attributes in the table are combined with some of the attributes in the public database then it becomes possible to identify the identity of an individual such public database attributes are known as Quasi-Identifiers (QI). To eliminate the problem of anonymization, technique has been developed. Using this technique granularity of data is reduced by applying generalization and suppression method[9]. This technique is further categorized as K-Anonymity, L-diversity and T-Closeness.

K-anonymity algorithm was first developed by[10] which states that protection can be achieved if the information of the individual in the data base cannot be distinguished from at least k-1 individuals whose information is also there in the released databases. It can be achieved by reducing the granularity of representation of the pseudo identifiers by using generalization or suppression technique. E.g there are two databases on is of voter id and seconds is of Medical record.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 8, August 2017)

Voter id database contains various attributes like individual name, its address, city zip code, gender and DOB. Combination of zip code, gender and DOB from voter id database with medical database one can easily known the medical record of a particular individual. Problem in k-anonymity is that if the size of dataset increases its utility decreases[11].

K- Anonymity is also vulnerable to certain attacks like background knowledge, homogeneity attack so *L- Diversity* came as extension to k- anonymous model which is based on adding intra group diversity for sensitive attributes in anonymous technique. L- Diversity model state that each sensitive attribute has L-diverse values in each equivalence class[12]. L- Diversity model has one major disadvantage is that it can prevent exact attack but not probabilistic attack because all the sensitive values are treated in a same way without considering their data distribution. Due to this, feasible representation of l-diverse data is a problem.

In [13]observed that when the overall distribution of a sensitive attribute is skewed L-diversity doesn't prevent attribute linkage attack. To prevent skewness attack author[13] proposed a privacy model called *t-closeness* which requires that the distribution of the sensitive attributes in any group on QID to be close to the distribution of the attributes in the overall table (global distribution) i.e difference between these distributions should not be more than the threshold value *t.*

Another type of PPDM is *distributed privacy preserving technique.* It is used when individual entity wants to maintain their privacy by only sharing the summative results over the entire data sets distributed across multiple sites. This approach was first used by[14] in data mining. Depending upon the distribution of data it is of two types horizontal (each party owns the same set of attributes but different subsets of records) and vertical (Several data holders own different sets of attributes on the same set of records and want to publish the integrated data on all attributes). Need of this type of technique arises due to certain situations, firstly when people want to share the results than the data, secondly it is difficult to send large databases to a central site and thirdly the heterogeneity of data is increasing day by day therefore combining those sources is a difficult task[15]. So this type of privacy technique has been implemented with the objective of sharing the results of multiple sites without sharing their private data. Problems under this technique can be classified based upon three basic parameters: outcome of data mining algorithm e.g cluster, association rule etc., type of distribution, privacy conditions on sharing of data.

Main drawback of this technique is the complexity of the algorithm because in this technique other than the size of data transmitted and memory size number of messages and number of communication rounds required by an algorithm are also considered.

Next PPDM technique is *downgrading application effectiveness.* In this technique instead of input data output applications like association rule mining, classification, clustering or query processing are available to other party. It results in violation of privacy. Here this technique helps in providing privacy by modifying data or output applications. Number of techniques has been developed e.g association rule hiding, query auditing, classifier and cluster downgrading.

Condensation Approach: Authors[16] in this approach proposed that large numbers of records are contracted into multiple groups. Records in each group are indistinguishable from each other. Certain statistics is calculated over each group. Minimum size of each group is *k.* This will represent the indistinguishable level. Privacy of data depends upon the value of *k* .Greater the value of *k* higher is the privacy of the data**.** Main advantage of this technique among other perturbation methods is that it calculated data distribution on multidimensional unlike other techniques in which each dimensions were treated independently without considering the correlation among different dimensions.

V. Comparison Of Various PPDM Approaches

All the above explained PPDM techniques are listed in Table 1. below with their specific pros and cons

TABLE I
Factors and Their Pros and Cons of Different PPDM Techniques

Sr. No	Type of PPDM	Factors	Pros and Cons of approach
1	Randomization	Adding noise to the original data	*Pros:* this approach is simple and doesn't require the knowledge of other records in the data. *Cons:* it treats all the records equally irrespective of their local density. As a result outliers are more prone to attack as compare

			to the records in dense region.
2	K- anonymity	Based on the anonymization of attribute with at least k-1 other attributes	*Pros:* it protects against identity disclosure but insufficient for attribute disclosure *Cons:* it is vulnerable to homogeneity and background knowledge attack.
3	L-diversity	To add intra group diversity in a table	*Pros*: in this model as the data set increases utility also increases. It also prevents attribute disclosure[17]. *Cons:* permissible to skewness attack and similarity attack. It also does not consider semantic meaning of sensitive values.
4	t- closeness	It makes the distance of distribution of sensitive attributes in equivalence class and distribution of other attributes in over all tables lesser than the threshold value.	*Pros:* in t-closeness, removing an outlier may smooth a distribution and bring it closer to the overall distribution and provides high data utility. *Cons:* in this model co-relation between different attributes is lost because all attributes are generalized separately. Utility of data will be lost if threshold value of t will be chosen very small.
5	Distributed PPDM	Based upon the partition of data horizontally or vertically	*Pros:* possessed advantage over other methods in respect to performance and privacy. *Cons:* complexity of algorithms increases as number of computation rounds and number of messages increases.
6	downgrading application effectiveness	Depends upon output data mining application like association rule hiding, classifiers, clustering, query based etc.	
7	Condensation	It contracts the original data into multiple groups and calculates the statistics of each group. Unlike other techniques which modify original data for perturbation, here pseudo data is used.	*Pros:* condensation approach works on multidimensional records which does not need the modification of already existing data mining for using in this approach. This preservers the Inter-attribute correlation of the data very effectively. *Cons:* this approach cause major information loss as multiple records are condensed into single statistical group.

VI. Classification Of Association Rule Hiding For PPDM

Association rule hiding refers to the process of modifying the novel database in such a way that some sensitive association rules vanish without poorly affecting the data and the non sensitive rules. The problem of association rule hiding was first probed in 1993[18].

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 8, August 2017)

The intention of the proposed Association rule hiding algorithm for PPDM is to conceal certain information so that it cannot be discovered through association rule mining algorithm. Association rule mining is to find out association rules that satisfy the predefined minimum support and confidence from a given database. The problem is usually decomposed into two sub problems.

1. To find those itemsets whose occurrences exceed a predefined threshold in the database; those itemsets are called frequent or large itemsets.
2. To generate association rules from those large itemsets with the constraints of minimal confidence.

Support of a rule X→Y is the percentage of transactions of the transaction database that contain item XUY. Support for the rule (X→Y) can be calculated by using the formula given in (1)[19]

Support (x→y)= $|x\ Uy|/N$ .....(1)

Where N is total number of transactions in transactional database.

Confidence of a rule X→Y is the percentage of transactions in the transaction database that contain X also contain Y. the confidence of rule (X→Y) can be calculated by using following formula (2)[19]

Confidence (x→y)=$|x\ Uy|/x$.....(2)

Association rule hiding algorithms are categorized mainly into five types: exact approach, border based approach, heuristic approach, reconstruction approach and cryptographic approach. Heuristic approach is considered as the fastest and most efficient among all the five PPDM techniques.

A. *Heuristic Based Approach*

Heuristic Based Approaches can be further divided into two groups based on data modification techniques: data distortion techniques and data blocking techniques. Authors[20] proposed Data-Distortion, based on data perturbation or data transformation, and in particular, to change a selected set of 1-values to 0-values (delete items) or 0-values to 1- values (add items) if we consider the transaction database as a two-dimensional matrix. It is aimed to reduce the support or confidence of the sensitive rules below the user pre-defined security threshold. [21]Proposed another data modification approaches for association rule hiding known as Data blocking. Instead of making data distorted (part of data is altered to false), blocking approach is implemented by replacing certain data items with a question mark "?".

The introduction of this special unknown value brings uncertainty to the data, making the support and confidence of association rule become too uncertain intervals respectively. At the beginning, the lower bounds of the intervals equal to the upper bounds. As the number of "?" in the data increases, the lower and upper bounds begin to separate gradually and the uncertainty of the rules grows accordingly. When either of the lower bounds of a rule's support interval and confidence interval gets below the security threshold, the rule is deemed to be concealed.

B. *Border Based Approach*

This approach hides sensitive association rule by modifying the borders in the lattice of the frequent and the infrequent itemsets of the original database. The itemsets which are at the position of the borderline separating the frequent and infrequent itemsets forms the borders. The algorithms in this approach differ in the methodology they follow to enforce the new, revised borders, in the modified database. Border based approach used the theory of borders[22] The first frequent itemset hiding methodology that is based on the notion of the border was proposed by[23, 24]. It maintains the quality of database by greedily selecting the modifications with minimal side effect. Then[25, 26] presented more efficient algorithms based on border theory.

C. *Exact Approach*

This approach contains non-heuristic algorithms which formulates the hiding process as a constraints satisfaction problem or an optimization problem which is solved by integer programming. These algorithms can provide optimal hiding solution with ideally no side effects. [27]Proposed an exact algorithm for association rules hiding which tries to minimize the distance between the original database and its sanitized version. [28]Proposed an exact border based approach to achieve optimal solution as compared to previous approaches.

D. *Reconstruction Approach*

In this approach first frequent item set is extracted from non frequent item set and privacy protected data is released. The new released data is then reconstructed from the sanitized knowledge base. This approach, first perform data perturbing and then reconstruct the database. Basically this approach reconstructs the database in a manner that all sensitive information has been hidden. This method cannot guarantee to find a consistent one within a polynomial time.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 8, August 2017)

E. Cryptographic approach

In this technique multiple parties want to share their data which is distributed across multiple sites. This can be done by encrypting the sensitive information by using certain cryptography algorithm. Different protocols are designed to protect the information from the adversaries but sometime even in the presence of these protocols the adversaries attempt to collect certain information about the user. Depending upon different adversary behaviour there are 2 types of adversaries: *semi honest adversary* and *malicious adversary*[29].

Main motive of the participating parties is to compute the aggregate result on the data at different sites while maintaining the privacy of data. Based upon the distribution of data it is of two types. Horizontal partitioned data and vertically partitioned data. In horizontally partitioned data records of the individual is distributed in different locations while in vertically partitioned data attributes of the individual are kept at different locations. Main algorithm used for this technique is Secure Multiparty Computation (SMC). In distributed privacy preserving there are four techniques based on cryptography technique. Those are Secure Sum, Secure Set Union, Secure Size of Set Intersection, and Scalar Product. In secure multiparty computation these techniques are used for different purposes. Secure sum is used when multiple parties wish to calculate secure sum on the shared data. Secure set union technique is used when multiple participating parties share the frequent itemsets, association rules or clusters extra without revealing the owner of the data. Secure size of set intersection is used when each party owns some data from the common data set. This technique helps to find the size of intersection of the parties over the common database. Scalar product technique provides solution to various data mining problems by calculating the scalar product of two vectors securely.

VII. Comparison Of Association Rule Hiding Techniques

All the above explained ARH techniques are listed in Table 2. below with their need and specific pros and cons

TABLE II
Comparative Analysis of ARH Techniques PPDM Techniques

Sr. No	ARH Approach	Need	Advantages and disadvantages
1	Heuristic based Approach[20]	It is aimed to reduce the support or confidence of the sensitive rules below the user pre-defined security threshold.	*Adv*: It is fast efficient and scalable algorithm for sanitizing the sensitive information in the original database. *Disadv:* algorithms may suffer from undesirable side effects that lead them to identify approximate hiding solutions.
2	Border Based approach	Hides sensitive association rule by modifying the borders in the lattice of the frequent and the infrequent itemsets of the original database.	*Adv:* It maintains the quality of database by greedily selecting the modifications with minimal side effect. *Disadv:* it still rely on the heuristic algorithms for modifications in original databases. Therefore it is unable to find optimal hiding solution.
3	Exact Approach	Hiding process is considered as constraint satisfaction problem which is solved by integer	*Adv:* Provides optimal hiding solution with no side effects. *Disadv*: it has high time complexity

		programming.	due to integer programming solver.
4	Re-construction based approach	Using this technique original data is perturb for hiding the sensitive data.	*Adv:* produces lesser side effects as compare to heuristic approach. *Disasv:*
5	Cryptography approach	Using this technique sensitive information is encrypted into certain unreadable form by using various cryptography algorithms.	*Adv:* Cryptographic approach provides security in multiparty computations where data is distributed among different parties but does not provide security for the output of the computation. *Disadv:* This approach is not scalable when number of parties involved in communication increases and database is very large.

VIII. CONCLUSIONS

The objective of this paper is to make the researcher familiar with some of the past of PPDM techniques, current state-of-art and the future scope of these techniques. The literature survey carried in this paper provides researcher better understanding about the growth of this field and the issues related to PPDM. In this paper, classification of common PPDM techniques is done based upon different requirements.

PPDM is applicable in different Data Mining fields like classification, clustering, association rule hiding etc. more focus is emphasized on PPDM for association rule hiding developed in the latest decades. A list of pros and cons of these approaches along with the need of each technique will help the researcher to know the usefulness of each technique. Brief description of other techniques is also provided. In addition, all these techniques have been compared in a tabular form providing more useful information about these approaches. Finally, researchers have shown new future scope that will help them in improving PPDM techniques so that this research field progresses continuously.

REFERENCES

[1] Agrawal R, Imielinski T and Swami A (1993). Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp 207-216.

[2] Jadav K, Vania J and Patel D. A Survey on Association Rule Hiding Methods. International Journal of Computer Applications. 2013 Nov, 28(13), pp. 20-25.

[3] Lindell Y, Pinkas B. Privacy preserving data mining. Journal of cryptology. 2002 Jul, 15(3), pp. 177-206.

[4] Verykios VS, Bertino E, Fovino IN, Provenza LP, Saygin Y, Theodoridis Y. State-of-the-art in privacy preserving data mining. ACM Sigmod Record. 2004 Mar, 33(1), pp.50-7.

[5] Aggarwal CC, Philip SY. A general survey of privacy-preserving data mining models and algorithms. Proceedings of Privacy-preserving data mining, Springer, US, 2008, pp. 11-52.

[6] Liew CK, Choi UJ, Liew CJ. A data distortion by probability distribution. ACM Transactions on Database Systems (TODS), 1985 Sep, 10(3), pp. 395-411.

[7] Aggarwal CC and Philip SY. A survey of randomization methods for privacy-preserving data mining. Proceedings of Privacy-Preserving Data Mining Springer, US, 2008, pp. 137-56.

[8] Kargupta H, Datta S, Wang Q and Sivakumar K. Random-data perturbation techniques and privacy-preserving data mining. Knowledge and Information Systems, 2005 May, 7(4), pp. 387-414.

[9] Machanavajjhala A, Gehrke J, Kifer D, Venkitasubramaniam M. L-diversity: Privacy beyond k-anonymity. Proceedings of the 22nd International Conference, 2006, pp. 24.

[10] Samarati P. Protecting respondents identities in microdata release. IEEE transactions on Knowledge and Data Engineering, 2001 Nov, 13(6), pp.1010-27.

[11] Sweeney L. K-ANONYMITY: A Model For Protecting Privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002 Oct, 10 (5), pp. 557-70.

[12] Keyvanpour M and Moradi SS. Classification and evaluation the privacy preserving data mining techniques by using a data modification-based framework. International Journal on Computer Science and Engineering, 2011 Feb, 3(2), pp. 862-70.

[13] Li N, Li T and Venkatasubramanian S. T-Closeness: Privacy beyond k-Anonymity and l-Diversity. Proceedings of ICDE Conference, Turkey, 2007, pp. 106-115.

[14] Lindell Y and Pinkas B. Privacy preserving data mining. Proceedings of Advances in Cryptology – CRYPTO Springer-Verlag, 2001, pp. 36–54.

[15] Clifton C. Privacy preserving distributed data mining. Proceedings of 13th European Conference on Machine Learning, 2001, pp. 19-23.

[16] Aggarwal CC, Philip SY. A condensation approach to privacy preserving data mining. Proceedings of Advances in Database Technology-EDBT Springer Berlin Heidelberg, 2004, pp. 183-199.

[17] Kumar PMV and Karthikeyan M. L-diversity on K-anonymity with External Database for Improving Privacy Preserving Data. International Journal of Computer Applications. 2012 Sept, 54(14), pp. 7-13.

[18] Agrawal R, Imielinski T and Swami A. Mining association rules between sets of items in large databases. Proceedings of the ACM SIGMOD International Conference on Management of Data, 1993, pp. 207-216.

[19] Sirole T and Choudhary J. A Survey of Various Methodologies for Hiding Sensitive Association Rules. International Journal of Computer Applications. 2014 Jan, 96(18), pp. 12-15.

[20] Atallah M, Bertino E, Elmagarmid A, Ibrahim M and Verykios V. Disclosure limitation of sensitive rules. Proceeding of Knowledge and Data Engineering Exchange (KDEX'99), 1999, pp. 45-52.

[21] Saygin Y, Verykios VS and Clifton C. Using unknowns to prevent discovery of association rules. ACM SIGMOD Record. 2001 Dec, 30(4), pp.45-54.

[22] Mannila H and Toivonen H. Level wise search and borders of theories in knowledge discovery. Data Mining and Knowledge Discovery. 1997 Sept, 1(3), pp. 241–258.

[23] Sun X and Yu P S. A border–based approach for hiding sensitive frequent itemsets. Proceedings of the 5th IEEE International Conference on Data Mining (ICDM), 2005, pp. 426– 433.

[24] Sun X and Yu P S. Hiding sensitive frequent itemsets by a border–based approach. Journal of Computing Science and Engineering. 2007 Sept, 1(1), pp.74–94.

[25] Moustakides G V and Verykios V S. A max–min approach for hiding frequent itemsets. Proceedings of the 6th IEEE International Conference on Data Mining (ICDM), 2006, pp. 502–506.

[26] Moustakides G V and Verykios V S. A maxmin approach for hiding frequent itemsets. Data and Knowledge Engineering. 2008 Apr, 65(1), pp.75–89.

[27] Gkoulalas-Divanis A, and Verykios VS. An integer programming approach for frequent itemset hiding. Proceedings of the 15th ACM international conference on Information and knowledge management, 2006, pp: 748-757.

[28] Gkoulalas-Divanis A, and Verykios VS. Exact Knowledge Hiding through Database Extension. IEEE Transactions on Knowledge and Data Engineering. 2009 May, 21(5), pp. 699–713.

[29] Pinkas B. Cryptographic techniques for privacy-preserving data mining. ACM SIGKDD Explorations Newsletter. 2002 Dec, 4(2), pp.12-19.