Volume 3, Issue 11, November 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

An Exemplary Study of Privacy Preserving Association Rule Mining Techniques

Dharmendra Thakur
M.Tech. Scholar, P.C.S.T., BHOPAL
India

Prof. Hitesh Gupta
C.S Dept, P.C.S.T., BHOPAL
India

Abstract- ***The Privacy preserving Data mining (PPDM) has become an important issue in recent years because of abundance of sensitive information on the internet . In this paper, we present taxonomy and a survey of recent approaches that have been applied to PPDM .Subsequently; we present taxonomy on Association Rule hiding algorithms. Association Rule mining and sensitive rule hiding is one of the hottest research areas in PPDM. Association rule hiding refers to the process of modifying the original databases in such a way that certain sensitive association rule disappear without seriously affecting the data and the non-sensitive rules. After survey of different techniques of association rule hiding, a detailed presentation of metrics used to evaluate the performance of those approaches is also given. Finally, we conclude our study by heuristic based algorithms and enumerating interesting future directions in this research body.***

Keywords- Privacy-preserving data mining (PPDM), Association rules, Support, Confidence, Sanitized database, Perturbation.

I. INTRODUCTION

In recent years, data mining has been viewed as a threat to privacy because of the widespread proliferation of electronic data maintained by corporations. This has lead to increased concerns about the privacy of the underlying data. Privacy preserving data mining technique gives novel way to solve this problem. In recent years, a number of techniques have been proposed for modifying or transforming the data in such a way so as to preserve privacy. The Association rule hiding approach is one of the widely used approach. The problem of association rule hiding was first probed in 1999 [2]. Association rule hiding problem can be defined as: convert the original database into sanitized database so that data mining techniques will not be able to mine sensitive rules from the database while all non sensitive rules remain visible. It is known that each strong rule extracts from frequent itemsets. To prevent sensitive rules (determined by the experts) being mined in the process of association rule mining, many methods are developed [3-11], all of which are based on reducing the support and confidence of rules that specify how significant they are. The rest of this paper is organized as follows. In section II, we will review the basic concepts of PPDM and different studies performed in the area of PPDM under five categories. We discuss the association rule mining in section III. In section IV, we will discuss privacy preserving association rule mining. We will focus on metrics that are used for measuring side-effects resulted from privacy preserving process in section V. In section VI, we will discuss a heuristic based algorithms. Finally, we conclude in section VII.

II. PRIVACY PRESERVING DATA MINING (PPDM)

Data mining is the process of gathering information about the user specific data, also called knowledge discovery, on internet. The problem with data mining output is that it also discloses some information, which is considered to be private and personal. Effortless access to such personal data causes a peril to individual privacy. Official statistics, Health information, and E-commerce are some key concern for privacy. Privacy preserving data mining technique gives novel way to solve this problem. The main purpose of privacy preserving data mining is to design competent frameworks and algorithms that can extract relevant knowledge from a large amount of data without revealing of any sensitive information. It protects sensitive information by providing sanitized database of original database on the internet or a process is used in such a way that private data and private knowledge remain private even after the mining process. It is PPDM due to which the benefits of data mining be enjoyed, without compromising the privacy of concerned individuals.
PPDM Techniques can be classified over five dimensions [13].The first dimension is related to distribution of data i.e. Centralized or Distributed. The second dimension refers to the modification of original values of data that are to be released for data mining task. Modification is carried out using perturbation, blocking, aggregation, merging, swapping or sampling or any combination of these. The third dimension is that of data mining algorithms. The data mining algorithm are applied on the transformed data to get useful nuggets of information that were hidden previously. The fourth dimension refers to whether the raw data or aggregated data should be hidden. The fifth and the final dimension refer to the techniques that are used for protecting privacy. Based on these dimensions, different PPDM techniques may be classified into following five categories [13-15]

1. Anonymization based PPDM
2. Randomized Response based PPDM
3. Condensation approach based PPDM
4. Cryptography based PPDM
5. Perturbation based PPDM

A. Anonymization based PPDM

Actually, when quasi identifiers [set of attributes that could potentially identify a record] are linked to publicly available data, identity of individual can be predicted with higher probability. Such attacks are called as linking attacks. Anonymization approach conceal identity or/and sensitive data about record owners using generalization and suppression in anonymized dataset [16-17]. Replacing a value with less specific but semantically consistent value is called as generalization and suppression involves blocking the values. Such data when released for mining, reduces the risk of identification when combined with publicly available data. But, besides, accuracy of the applications on the transformed data is reduced. Limitation of the *k*-anonymity model is that it may be very hard for the owner of a database to determine which of the attributes are available or which are not available in external tables.

TABLE 2.I
K-ANONYMOUS DATA

Age	**Weight**	**Name**
35	50	Ramesh
60	55	Shweta
65	50	Sham

(a) Original Data

Age	**Weight**	**Name**
[35,45]	[50,65]	Ramesh
[35,45]	[50,65]	Shweta
[55,65]	[50,65]	Sham

(b) K-Anonymous Data

B. Randomized Response based PPDM

Randomized response is statistical technique to solve a survey problem. In Randomized response [14], the data is jumbled in such a way that the central place cannot tell with probabilities better than a pre-defined threshold, whether the data from a customer contains truthful information or false information. The information received from each individual user is snarled and if the number of users is significantly large, the aggregate information of these users can be predictable with good amount of accuracy. The data collection process in randomization method is carried out using two steps [14]. During first step, the data providers randomize their data and transmit the randomized data to the data receiver. In second step, the data receiver reconstructs the original distribution of the data by employing a distribution reconstruction algorithm. The randomization response model is shown in fig. 2.1

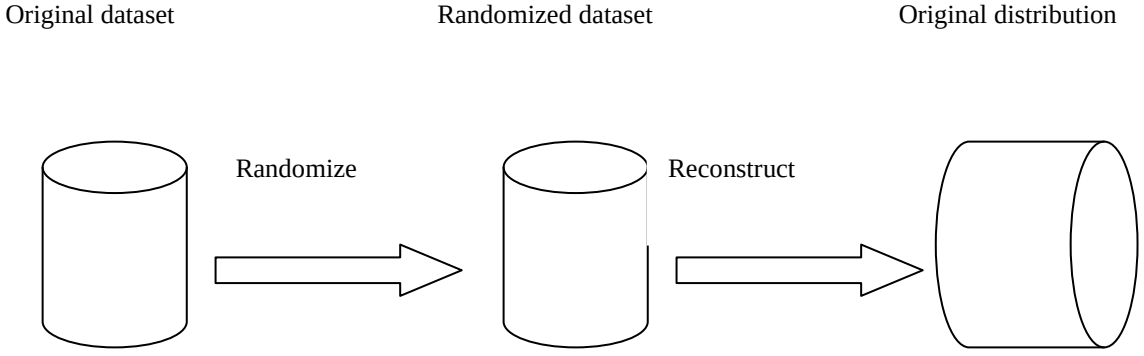

Fig.2.1 Randomization response model

The randomization method can be implemented at data collection time. It does not need a trusted server to contain all the original records in order to perform the anonymization process [18]. The weakness of a randomization response based PPDM technique is that it treats all the records equal irrespective of their local density. This leads to a problem where the outlier records become more prone to adversarial attacks than to records in more dense regions in the data [20]. One solution to this is to unnecessarily adding noise to all the records in the data. But, it reduces the utility of the data for mining purposes as the reconstructed distribution may not yield results in consistency of the purpose of data mining. The randomized response approach has extended its strengths to a number of data mining problems. In [19], Agrawal and Srikant have discussed the use of randomized response approach for classification. A number of other techniques [21-22] have also been proposed that work well over a variety of different classifiers.

C. Condensation approach based PPDM

Condensation approach constructs constrained clusters in dataset and then generates fake data from the statistics of these clusters [23]. It is called as condensation because of its approach of using condensed statistics of the clusters to generate fake data. It constructs groups of non-homogeneous size from the data, such that it is guaranteed that each record lies in a group whose size is at least equal to its secrecy level. Subsequently, fake data is generated from each group so as to create a synthetic data set with the same aggregate distribution as the original data [18]. This approach can be effectively used for the problem of classification [23]. The use of fake 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 [18].This approach helps in better privacy preservation as compared to other techniques as it uses fake data rather than modified data. Moreover, it works even without redesigning data mining algorithms since the fake 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 large amount of information is lost because of the condensation of a larger number of records into a single statistical group entity [14].

D. Cryptography based PPDM

Cryptographic techniques are ideally meant for such scenarios where multiple parties collaborate to compute results or share non sensitive mining results and thereby avoiding disclosure of sensitive information [14]. Cryptographic techniques find its utility in such scenarios because of two reasons [14]: First, it offers a well defined model for privacy that includes methods for proving and quantifying it. Second a vast set of cryptographic algorithms and constructs to implement privacy preserving data mining algorithms are available in this domain. Although cryptographic techniques ensure that the transformed data is exact and secure but this approach fails to deliver when more than a few parties are involved [24].

E. Perturbation based PPDM

Perturbation has a long back history, being used in statistical disclosure control as it has an inherent property of simplicity, efficiency and ability to preserve statistical information [25]. In perturbation the original values are replaced with some synthetic data values so that the statistical information computed from the perturbed data does not differ from the statistical information computed from the original data to a larger extent. The perturbed data records do not correspond to real-world record owners, so the attacker cannot perform the sensitive linkages or recover sensitive information from the published data. In perturbation approach, records released is synthetic i.e. it does not correspond to real world entities represented by the original data. Therefore the individual records in the perturbed data are meaningless to the human recipient as only statistical properties of the records are preserved. Perturbation can be done by using additive noise or data swapping or synthetic data generation. Since the perturbation method does not reconstruct the original values but only the distributions, new algorithms are to be developed for mining of the data [14].

III. ASSOCIATION RULE MINING

Let I = {I1, I2,…, Im } be a set of items[2]. Let *D* be a database of transactions where each transaction *T* is a set of items such that T $\subseteq$ I. Each transaction is associated to an identifier, call TID. A transaction *T* is said to contain *A* if and only if $A \subseteq T$. An association rule is an implication of the form $A \Rightarrow B$, where $A \subseteq I$, $B \subseteq I$, and $A \cap B = \phi$ .The rule $A \Rightarrow B$ holds in the transaction set *D* with support *s*, where s is the percentage of transactions in *D* that contain $A \cup B$. The rule $A \Rightarrow B$ has confidence *c* in the transaction set *D*. That is,

$$Sup\ (A \Rightarrow B) = P\ (A \cup B) = \frac{|A \cup B|}{|D|} \qquad (1)$$

$$Conf\ (A \Rightarrow B) = P\ (B/A) = \frac{|A \cup B|}{|A|} \qquad (2)$$

Where $|A|$ is named as the support count of the set of items *A* in the set of transactions *D*, as denoted by *sup_ count (A).* *A* occurs in a transaction T, if and only if $A \subseteq T$. Rules that satisfy both a minimum support threshold (*min_ sup)* and a minimum confidence threshold (*min_conf)* are called strong. A set of items referred to as an itemset. An itemset that contains *k* items is a *k* -itemset. Itemsets that satisfy *min_ sup* is named as frequent itemsets. All strong association rules result from frequent itemsets.

IV. PRIVACY PRESERVING ASSOCIATION RULE MINING

Privacy preserving association rule mining needs to prevent disclosure not only of confidential personal information from original or aggregated data, but also to prevent data mining techniques from discovering sensitive knowledge. It is known that each strong rule extracts from frequent itemsets. To prevent sensitive rules (determined by the experts) being mined in the process of association rule mining, many methods are developed [3-11], all of which are based on reducing the

support and confidence of rules that specify how significant they are. In order to achieve this goal, transactions are modified by removing some items, or inserting new items depending on the hiding strategy. In the following, we will discuss some general methods for hiding sensitive rules. The conceptual framework for association rule hiding is shown in the fig.4.1:

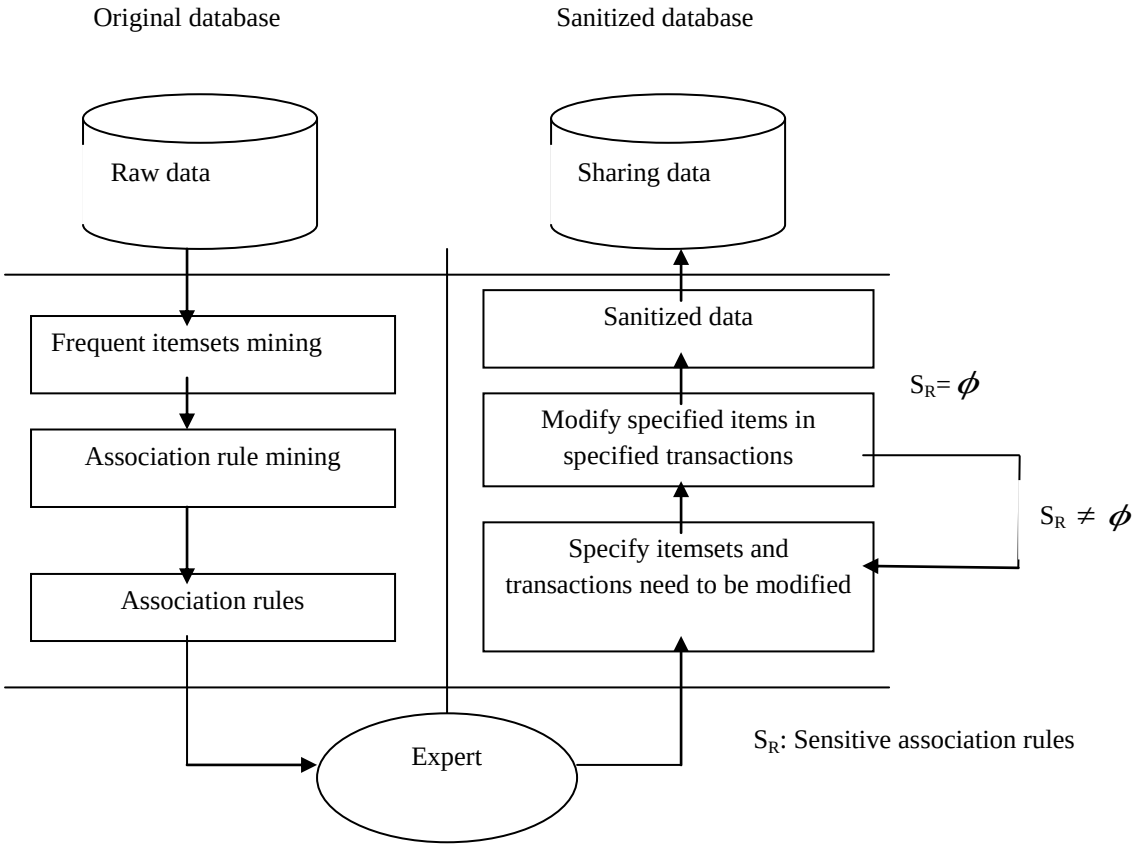

Fig. 4.1 Conceptual framework for association rule hiding

Let *R* be a set of rules extracted from the transaction database with the *min_ sup* and the *min_conf*. Let S_R be a set of sensitive rules determined by the experts. In order to hide a rule $A \Rightarrow B$, on the basis of (1) and (2) in section III, we can either decrease the support of the itemset $A \cup B$ below the *min_ sup* threshold, or decrease the confidence below the *min_conf* threshold while giving as little harm as possible to the remaining non-sensitive rules to keep the data quality as high as possible.

A. Classes Of Association Rule Hiding Algorithms

Association rule hiding algorithms can be divided into three distinct classes, namely *border-based* approaches, *exact* approaches and *heuristic* approaches.

1) Border-based Approaches: These approaches consider the task of sensitive rule hiding through modification of the original borders in the lattice of the frequent and the infrequent patterns in the dataset. In these schemes, the positive and the negative borders in the lattice of all itemsets are computed first and then focus on preserving the quality of the computed borders during the hiding process. The quality of the borders directly affects the quality of the sanitized database that is produced, which can be maintained by greedily selecting those modifications that lead to minimal side-effects.

2) Exact Approaches: Exact approaches are typically capable of providing superior solutions but at a high computational cost. They achieve this by formulating the sanitization process as a constraint satisfaction problem and by solving it using an integer/linear programming solver. Thus, the sanitization of the dataset is performed as an atomic operation which avoids the local minima.

3) Heuristic Approaches: These approaches involve efficient, fast algorithms that selectively sanitize a set of transactions from the database to hide the sensitive knowledge. Due to their efficiency and scalability, the heuristic approaches have been the focus of attention for the vast majority of researchers in the knowledge hiding field. We will discuss this approach in detail in section VI of this paper.

V. METRICS AND PERFORMANCE ANALYSIS

The performance of two hiding algorithms can be compared on the basis of following Metrics. These Metrics determine the most efficient algorithm between the two.

A. Hiding Failure (HF)

It is the percentage of the sensitive data that remain exposed in the sanitized dataset. It is defined as the fraction of the restrictive association rules that appear in the sanitized database divided by the ones that appeared in the original dataset. Formally,

$$\text{HF} = \frac{|S_R(D')|}{|S_R(D)|}$$

Where $|S_R(D')|$ is number of the sensitive rules discovered in the sanitized dataset D', $|R_R(D)|$ is the number of sensitive rules appearing in the original dataset D. Ideally, the hiding failure should be 0%.

B. Misses Cost (MC)

It is the percentage of the non-sensitive data that are hidden as a side-effect of the sanitization process. It is computed as follows:

$$\text{MC} = \frac{|S'_R(D)| - |S'_R(D')|}{|S'_R(D)|}$$

Where $|S'_R(D)|$ is the size of the set of all non-sensitive rules in the original database D and $|S'_R(D')|$ is the size of the set of all non-sensitive rules in the sanitized database D'.

C. Artifactual Patterns (AF)

It is the percentage of the discovered patterns that are artifacts. It is computed as follows:

$$\text{AF} = \frac{|P'| - |P \cap P'|}{|P'|}$$

where P is the set of association rules discovered in the original database D and P' is the set of association rules discovered in D'.

D. Dissimilarity (Diss)

It is the difference between the original and the sanitized datasets by comparing their histograms, where the horizontal axis contains the items in the dataset and the vertical axis corresponds to their frequencies. It is calculated as follows:

$$\text{Diss}(D, D') = \frac{1}{\sum_{i=1}^{n} f_D(i)} \times \sum_{i=1}^{n} \left[f_D(i) - f_{D'}(i) \right]$$

where $f_X(i)$ represents the frequency of the i^{th} item in the dataset X, and n is the number of distinct items in the original dataset D.

E. Side-Effect Factor (SEF)

It is the amount of non-sensitive association rules that are removed as an effect of the sanitization process. It is defined as follows:

$$\text{SEF} = \frac{|P| - (|P'| + |R_P(D)|)}{|P| - |R_P|}$$

F. Recovery Factor (RF)

This measure expresses the possibility of an adversary to recover a sensitive rule based on the non-sensitive ones. The recovery factor of a pattern takes into account the existence of its subsets. If *all* the subsets of a sensitive rule can be recovered from the sanitized dataset, then the recovery of the rule itself is possible, thus it is assigned an RF value of 1; otherwise RF = 0. However, this measure is not certain since, for instance, an adversary may not learn an itemset despite knowing its subsets. Bertino et al. [6] propose a set of measures that are directly related to the performance of a hiding algorithm as far as external parameters are concerned.These "process performance" measures are clustered into four categories, as follows:

1) Efficiency: This category consists of measures that quantify the ability of a privacy preserving algorithm to efficiently use the available resources and execute with good performance. Efficiency is measured in terms of CPU-time, space requirements (related to the memory usage and the required storage capacity) and communication requirements.

2) Scalability: This category consists of measures that evaluate how effectively the privacy preserving technique handles increasing sizes of the data from which information needs to be mined and privacy needs to be ensured. Scalability is measured based on the decrease in the performance of the algorithm or the increase of the storage requirements along with the communications cost (if in a distributed setting), when the algorithm is provided with larger datasets.

3) Data Quality: The data quality of a privacy preservation algorithm depends on two parameters. There are the qualities of the dataset after the sanitization process, and the quality of the data mining results when applied to this dataset, compared to the ones attained when using the original dataset. Among the various possible measures for the quantification of the data quality, the most preferable are: (i) *accuracy*, which measures the proximity of a sanitized value to the original one and is closely related to the information loss resulting from the hiding strategy, (ii) *completeness*,which is used to evaluate the degree of missed data in the sanitized database and (iii) *consistency*, which is related to the relationships that must continue to hold among the different fields of a data item or among data items in a sanitized database.

4) Privacy Level: This category consists of measures that estimate the degree of uncertainty according to which, the protected information can still be predicted. Measures, such as the information entropy, the level of privacy and the *J*-measure [6], are some among the possible metrics that one can apply to quantify the privacy level attained by a hiding scheme

VI. OVERVIEW OF HEURISTIC APPROACH FOR HIDING THE SENSITIVE RULE

Heuristic approaches hide sensitive association rules by directly modifying, or we say, sanitizing the original database DB, and get the released database DB' directly from DB. Heuristic Based Approaches can be divided into two groups based on data modification techniques: data distortion techniques and data blocking techniques.

A. Data-Distortion

It is based on data perturbation or data transformation, and in particular, the procedure is 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.

***1) Support-based Algorithm*:** INPUT: a set of rules to hide S_R, the source database *D*, the *min_ sup* threshold, the *min_conf* threshold; OUTPUT: the database *D* transformed so that the rules in S_R cannot be mined.

FOREACH rule *r* IN S_R DO
{
WHILE (conf *(r)* ≥*min_conf* AND *sup(r)* ≥*min_ sup)*
{
T = {t∈ *D* | *t* supports *r}*
// sort T in ascending order of size of the transactions and
Choose the one with the lowest size
t = choose transaction (*T)*
// choose the item of *r* with the minimum impact on the
(| *r* | − 1)-itemsets
j = choose_item(r*)*
// set to 0 the bit of *t.list _ of _items* that represents item
j
Set to 0(j, *t.list _ of _items)*
sup(r) = □*sup(r)*-1
conf (r)= □*sup(r) / sup(pre(r))*
}
S_R = □ S_R -*r*
}

***2) Confidence-based Algorithm*:** We decrease the confidence of the rule (a) by increasing the support of the rule antecedent *A*, through transactions that partially support both *A* and *B*, (b) or by decreasing the support of the rule consequent *B*, in transactions that support both *A* and *B.* The following is the specific algorithm process.
INPUT and OUTPUT is same as Algorithm 1

FOREACH rule *r* IN S_R DO
{
WHILE (conf *(r)* ≥*min_conf)*
{
T = {t∈ *D* | *t* supports *r}*
// sort T in ascending order of size of the transactions and choose the one with the lowest size
t = choose transaction (*T*)
// choose the item of *pre(r) with* the minimum impact on the (| *pre(r)* | − 1)-itemsets
j = choose_item (pre*(r))*
// set to 0 the bit of *t.list _ of _items* that represents item *j*
Set to 0(j, *t.list _ of _items)*

sup(r) =□sup(r) -1
conf (r)=□sup(r) / sup(pre(r))
}
S_R =□ S_R -*r*
}

Incorporating (1) and (2), we can decrease the support of the itemset $A \cup B$ below the *min_ sup* threshold or decrease the confidence below the *min_conf* threshold by decreasing the support of either the rule antecedent A , or the rule consequent B .

B. Data-Blocking

It is another data modification approach for association rule hiding. 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.

Between these two categories of approaches, the distortion-based are the ones commonly adopted by the overwhelming majority of researchers. SPM[26] (Super programming Model), Cluster-based ARH[27], ILARH[28] (Intersection Lattice Association Rule Hiding), DSSRC[29] (Decrease Support of R.H.S. item of Rule Clusters), MDSRRC [1] (Modified Decrease Support of R.H.S. item of Rule Clusters) are some heuristic based data distortion algorithms to hide the sensitive association rules.

VII. CONCLUSION AND FUTURE WORK

In this survey, we have discussed the basic of PPDM and its different approaches. Subsequently, association rule hiding approaches and metrics for performance comparison of those approaches are discussed. Before we conclude our study we have provided an overview of heuristic approaches. These approaches involve efficient, fast algorithms to hide the sensitive knowledge. Due to their efficiency and scalability, the heuristic approaches have been the focus of attention for the vast majority of researchers in the knowledge hiding field. Different algorithms can be designed and developed by taking ideas from existing algorithms and compare their efficiency using metrics.

REFERENCES

[1] Nikunj H. Domadiya, and Udai Pratap Rao, "Hiding Sensitive Association Rules to Maintain Privacy and Data Quality in Database," IEEE, 2012, *3rd IEEE International Advance Computing Conference (IACC)*, 2013.

[2] Atallah, M., Bertino, E., Elmagarmid, A., Ibrahim, M., and Verykios, V.S. Disclosure limitation of sensitive rules. In: Scheuermann P, ed.Proc. of the IEEE Knowledge and Data Exchange Workshop (KDEX'99). IEEE Computer society, 1999. 45-52.

[3] M. Atallah, E. Bertino, A. Elmagarmid, M. Ibrahim3, V. Verykios, "*Disclosure Limitation of Sensitive Rules*," Proc. IEEE Knowledge and Data Engineering Workshop, Chicago, Illinois, 1999, pp. 25-52.

[4] E. Dasseni, V. S. Verykios, A. K. Elmagarmid, and E. Bertino,"Hiding *Association Rules by Using Confidence and Support*," Proc. the 4th Information Hiding Workshop, Pittsburg, PA, 2001, LNCS2137, pp. 369-383.

[5] S. R. M. Oliveira, and O. R. Zaiane, "*Privacy Preserving Frequent Itemset Mining*" Proc. IEEE ICDM Workshop on Privacy, Security, and Data Mining, Maebashi, Japan, 2002, pp. 43-54.

[6] S. R. M. Oliveira and O. R. Zaiane, "*Algorithms for Balancing Privacy and Knowledge Discovery in Association Rule Mining*," Proc. the 7th International Database Engineering and Applications Symposium, Hong Kong, China, 2003, pp. 54-63.

[7] Y. Wu, C.M. Chiang, and A.L.P. Chen, "*Hiding Sensitive Association Rules with Limited Side-effects*," IEEE Transactions on Knowledge and Data Engineering, vol. 19, Jan. 2007, pp. 29-42.

[8] S.R.M. Oliveira, O.R. Zaiane, and Y. Saygin, "*Sharing Secure Association-Rule*," Advances in Knowledge Discovery and Data Mining,Springer Berlin, Heidelberg,Vol. 3056,2004, pp. 74-85.

[9] V.S. Verykios, A.K. Elmagarmid, E. Bertino, Y. Saygin, E. Dasseni, "*Association Rule Hiding*," IEEE Transactions on Knowledge and Data Engineering, vol. 16, Apr. 2004, pp, 434-447.

[10] R. Agrawal, and R. Srikant, "*Privacy-preserving data mining*," ACM SIGMOD Record, New York, vol. 29, Feb. 2000, pp.439-450.

[11] Y. Saygin, V. S. Verykios, and C. Clifton, "*Using Unknowns to Prevent Discovery of Association Rules*," ACM SIGMOD Record, New York, vol. 30, Apr. 2001, pp. 45-54.

[12] Saygin, Y., Verykios, V.S., and Clifton, C. Using unknowns to prevent discovery of association rules. SIGMOD Record, 2001, 30(4):45-54.

[13] S.V. Vassilios , B. Elisa, N.F. Igor, P.P. Loredana, S. Yucel and T. Yannis, 2004, "*State of the Art in Privacy Preserving Data Mining*" Published in SIGMOD Record, 33, 2004, pp: 50-57.

[14] 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.

[15] Wang P, *"Survey on Privacy preserving data mining*", International Journal of Digital Content Technology and its Applications, Vol. 4, No. 9, 2010.
[16] Sweeney L, *"Achieving k-Anonymity privacy protection using generalization and suppression"* International journal of Uncertainty, Fuzziness and Knowledge based systems, 10(5), 571-588, 2002.
[17] Sweeney L, *"k-Anonymity: A model for protecting privacy"* International journal of Uncertainty, Fuzziness and Knowledge based systems, 10(5), 557-570, 2002.
[18] Aggarwal C, Philip S Yu, *"A General Survey of Privacy-Preserving Data Mining Models and Algorithms"*, Springer Magazine, XXII, 11-52, 2008.
[19] R. Agrawal and R. Srikant. "*Privacy Preserving Data Mining*", ACM SIGMOD Conference on Management of Data, pp: 439-450, 2000
[20] Y. Lindell and B. Pinkas, "*Privacy Preserving Data Mining*", Journal of Cryptology, 15(3), pp.36-54, 2000.
[21] Zhang P., Tong Y., Tang S., Yang D., *"Privacy-Preserving Naïve Bayes Classifier"*, Lecture Notes in Computer Science, Vol 3584, 2005.
[22] Zhu Y., Liu L., *"Optimal Randomization for Privacy- Preserving Data Mining"*, ACM KDD Conference, 2004.
[23] Aggarwal C, Philip S Yu, *"A condensation approach to privacy preserving data mining"*, EDBT, 183-199, 2004.
[24] Clifton C, Kantarcioglu M, Vaidya J, Xiaodong L, Michael Y, *"Tools for Privacy Preserving Distributed Data mining"*, SIGKDD Explorations letters Vol. 4, Issue 2, December 2002.
[25] Benjamin C M Fung, Ke Wang, Rui Chen, Philip S Yu, *"Privacy Preserving Data Publishing: A Survey of recent developments"*, ACM Computing Surveys, Vol. 42, No. 4, Article 14, June 2010.
[26] Dejiang Jin and Sotirios G. Ziavras. A Super-Programming Approach for Mining Association Rules in Parallel on PC Clusters.IEEE Transactions on Parallel and Distributed Systems , Vol. 15, No. 9, September 2004. 783-794
[27] Kshitij Pathak, Narendra S Chaudhari, and Aruna Tiwari ,"Privacy Preserving Association Rule Mining by Introducing Concept of Impact Factor," IEEE, 2011, *7th IEEE Conference on Industrial Electronics and Applications (ICIEA),2012.*
[28] Hai Quoc Le, and Somjit Arch-int, "A Conceptual Framework for Privacy Preserving of Association Rule Mining in E-Commerce," IEEE, 2011, *7th IEEE Conference on Industrial Electronics and Applications (ICIEA),2012.*
[29] C. N. Modi, U. P. Rao, and D. R. Patel, "Maintaining privacy and data quality in privacy preserving association rule mining," 2010 Second Iternational conference on Computing, Communication and Networking Technologies, pp. 1–6, Jul. 2010.

