ISSN:2229-6093

Methods and Techniques to Protect the Privacy Information in Privacy Preservation Data Mining

N.Punitha
Research Scholar
Department of Computer Science,
PSGR Krishnammal College for Women
Coimbatore, India.
harinitha07@gmail.com

R.Amsaveni
Assistant Professor
Department of Computer Applications,
PSGR Krishnammal College for Women
Coimbatore, India.

***Abstract*- Due to the explosive development of Internet browser's can extract desired data from large amount of database. Data mining is a collection of technique helps the miner can retrieve exact knowledge from it. Privacy Preservation has been one of the greater concerns in data mining; there are many methods and techniques for privacy Preservation data mining. This paper discussed very keen points of various Privacy Preservation data mining algorithms and analyzing techniques and concludes the advantages and disadvantages, and a way to direct the problems.**

***Keywords*- Privacy, Security, Accuracy, Privacy Data Mining, k-Anonymity model and l-diversity, Randomization, Secure Multiparty, Computation**

1. INTRODUCTION

Data mining consists of number of techniques for manufacture automatically and interestingly to retrieve the information from the large amount of database which consists of sensitive information also. It requires data preparation which can uncover information which may compromise confidentiality and privacy obligations. Efficient data mining technique has increases the disclosure risks of sensitive data. The general way for this to happen because of data aggregation. Data aggregation is used for when the data are accrued, possibly from various sources and put together, so that they can be analyzed. But the data mining by it result the preparation of data before, for the purposes of analysis. To make up a publicly available scheme secure , must ensure not only the private sensitive data to be fit out, but also to build sure that certain inference channels have to prevented as well. An individual's privacy comes under a struggle when the data previously composed, that data may roots the data miner, or newly created data sets, are able to make out specific individuals, especially when initially the data were anonymous. Security to sensitive data against unauthorized access has been a long term goal for the database security research community. Hence, Privacy preservation data mining is novel research direction in data mining. It composed of number of effective method and techniques to make sure that might result in information loss, side effects, improve accuracy, utility and efficiently.

2. CLASSIFICATION OF PRIVACY DATA MINING

Data Hiding	Data Perturbation	Value Distortion	Additive Perturbation
			Multiplicative Perturbation
			Data Microaggregation
			Data Anonymization
			Data Swapping
			Other Randomization Techniques
		Probability Distribution	Sampling Method
			Analytical Method
	Secure Multi-Party Computation (SMC) / Cryptographic Protocols		
	Distributed Data Mining (DDM)		
Rule Hiding	Association Rule Hiding		Data Perturbation
			Data Blocking
	Classification Rule Hiding		Parsimonious Downgrading

Table 2.1. A brief overview of privacy preserving data mining techniques.

2.1 Privacy Preservation Data Mining

Privacy has been gaining more attention to handle the terrorism, the government needed to examine, using data mining technology, more information about individuals to detect unusual disease outbreaks, financial fraudulent behaviors, network intrusions, etc. While all of these applications of data mining can benefit our society, there is also a negative side to this technology because it could be a threat to the

individuals' privacy. Overcome the "limitations" of data mining techniques including areas like data security and privacy preserving data mining, which are actually active and growing research areas

2.2 Data Distribution

The PPDM algorithm can be divided into two major categories, Centralized and distributed data. In a centralized database, data are stored in a single database, in distributed data can further classified into horizontal a vertical data distributions. In Horizontal data distribution from different records of the same data attributes are resided in different places. In vertical data distributor different attributes of the same record of data are resided in different places most research occurred on a centralized data base. Applying PPDM algorithm to a distributed database privacy concerns, communication cost is too expensive

2.3 Purpose of PPDM

The PPDM algorithms main purpose is hiding is data hiding and rule hiding. In Data Hiding the sensitive data from original database like indentify name and address are linked, directly or indirectly to an individual person are hided. In rule hiding the sensitive data (rule) from original database after applying data mining algorithm is removed. Most of the PPDM algorithms hide sensitive patterns by modifying data hiding.

2.4 PPDM Algorithm

The PPDM algorithm are specifically on the tasks of classification, association rule and clustering classification is the process of finding a set of models that describe and distinguish data classes or concepts, for the purpose of the model is used for prediction the class of objects whose label is unknown clustering analysis concerns the problem of separating a data set in one group which are similar top each other and are different as possible in other group.

2.5 PPDM Techniques

PPDM techniques used by four categories Sanitation, it can remove or modify items for a database to reduce the support of some frequently used items sets that sensitive patterns are not to be mined. Blocking it can replace certain attributes the data with a question mark. According to this the minimum support and confidence level will be altered into a minimum interval. In distort, the support and the confidence of a sensitive rule lie below the middle the two and the confidentiality of data is expected to be protected and also known as data perturb action or data randomization, where individual data records are modified from original data, and reconstructed from randomized data. This technique aims to design for distortion methods after which the true value of any individual record is difficult to ascertain, but unchanged for danger data. In generalization transforms and replaces each record value with a correspondery generalized value.

3.Methods of PPDM

Develop data mining methods without increasing the risks of mi the data used to generate those methods. Typically, such methods reduce the granularity of representation in order to reduce the privacy. It results in some loss of information. This is the natural trade off between privacy and information loss. Most of the methods form of transformation on the original data in order to perform the privacy reservation. The transformed dataset must be able for mining and also privacy requirements without losing the benefit of mining.

3.1Randomization method

The randomization technique uses data distortion methods in order to add some noise to the original data. The recovery of individual values from the records is longer by adding noise to the original., but only aggregate distribution cab be recovered. The randomization method gives an appropriate balance between privacy preservation and knowledge discovery. There are kinds of perturbation are possible with the randomization method. Additive perturbation, randomized noise is added to the data records. The data distributions can be recovered from the randomized records. Multiplicative perturbation, the random projection of random rotation techniques projection or random rotation techniques are used in order to perturb the records. This method include random noise based perturbation and Randomized response scheme. Hence it results efficient and high information method.

3.2 The anonymization method

Anonymization method uses generalization and suppression techniques to make individual record be indistinguirlable among a group records. The anonymity model is developing. The concept left in k.anonymity model is that many attributer in the data can after be considered quasi – identifiers. It makes conjuction with public records in order to finiquely indetify the records. The l-diversity model was designed to handle since protecting identities to the level of k.individuals is not the same as protecting the corresponding sensitive value, especially only homogeneity of sensitive values within a group. The advanced methods m, invariance, Personalized,

anonymity, C (CBK) Anonymity, nibble, P- sensitive – anonymity, (a,k) – anonymity l- diversites, t-closenees and so on. The anyonymization method can ensure that the transformed data is accurate, but information loss is quit bit, extent.

3.3 The cryptograph method

The cryptograph method is based on distributed database. In many cases, individual entities may wish to desive aggregate results from data sets which are partitioned across there entities such partitioning may be horixontal (when the records are distributed across multiple entities). Or vertical (when the attributer are distributed across multiple entities. So the individual entities may not desire to share their entire data sets, they may consent to limit the information to share, by the use of variety of protocol. Hence the overall effect of such methods is to maintain privacy for each individual entity, while deriving aggregate results over the entire data. This method can ensure that the transformed data is exact and secure, but much low in efficient.

4.Method of randomization

The randomization method provides an effective yet simple way of preventing the user from learning sensitive data, which can be easily implemented at data collection phase for privacy preserving data mining, because the noise added to a given record is independent of the behavior of other data records. When the randomization method is carried out, the data collection process consists of two steps [3]. The first step is for the data providers to randomize their data and transmit the randomized data to the data receiver. In the second step, the data receiver estimates the original distribution of the data by employing a distribution reconstruction algorithm. The model of randomization is shown in Figure 1.

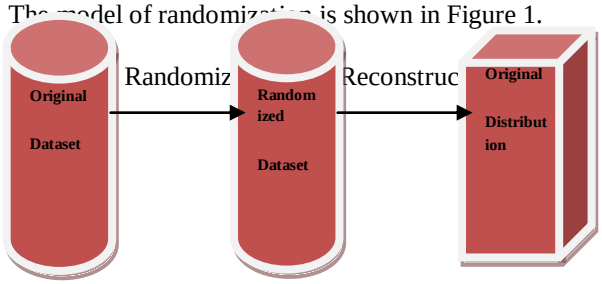


Figure 1. The Model of Randomization.

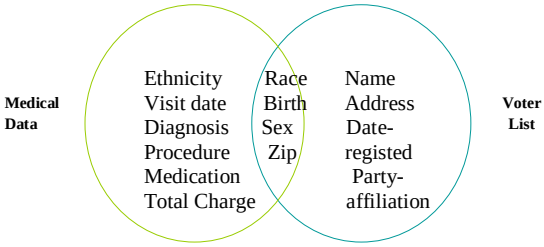
Representative randomization methods include random-noise-based perturbation and Randomized Response scheme. Agrawal and Srikant proposed a scheme for privacy preserving data mining using random perturbation and discussed how the reconstructed distributions may be used for data mining [4]. In their randomization scheme, a random number is added to the value of a sensitive attribute. For example, if $i\ x$ is the value of a sensitive attribute, $i\ i\ x \square r$, rather than $i\ x$, will appear in the database, where $i\ r$ is a random noise drawn from some distribution. It is shown that given the distribution of random noises, reconstructing the distribution of the original data is possible. Subsequently, Evmievski et al. proposed an approach to conduct privacy preserving association rule mining [5]. Kargupta et al. [6] proposed a random matrix-based spectral filtering technique to recover the original data from the perturbed data. Huang et al. further proposed two other data reconstruction methods: PCA-DR and MLE-DR in [7]. In addition, several distribution reconstruction algorithms have been proposed in correspondence to different randomization operators [8-10]. The basic idea of most algorithms is to use Bayesian analysis to estimate the original data distribution based on the randomization operator and the randomized data. For example, the expectation maximization (EM) algorithm [8] generates areconstructed distribution that converges to the maximum likelihood estimate of the original distribution. The Randomized Response (RR) was firstly proposed by Warner [11]. The RR scheme is a technique originally developed in the statistics community to collect sensitive information from individuals in such a way that survey interviewers and those who process the data do not know which of two alternative questions the respondent has answered. In data mining community, Rizvi and Haritsa presented a MASK scheme to mine association rules with secrecy constraints [12]. Du and Zhan proposed an approach to conduct privacy preserving decision tree building [13]. Guo et al. addressed the issue of providing accuracy in terms of various reconstructed measures in privacy preserving market basket data analysis [14]. The randomization method is a simple technique which can be easily implemented at data collection time. It has been shown to be a useful technique for hiding individual data in privacy preserving data mining. The randomization method is more efficient. However, it results in high information loss.

5.Method of anonymization

With the rapid growth in database, networking, and computing technologies, a large amount of personal data can be integrated and analyzed digitally, leading to an increased use of data mining tools to infer trends and patterns. This has raised universal concerns about protecting the privacy of individuals. The data records are often made available by simply removing key identifiers such as the name and social-security numbers from individual records. However, the combinations of other record attributes (usually named as quasi-identifier) can be

used to exactly identify individual records. For example, attributes such as race, birth, sex, and zip are available in public records such as voter list. When these attributes are also available in a given data set such as medical data, they can be used to infer the identity of the corresponding individual with high probability by linking operation, as is shown in

A Sample of Linking Attack



In order to preserve privacy, Sweeney [15] proposed the k-anonymity model which achieves k-anonymity using generalization and suppression, so that, any individual is indistinguishable from at least k-1 other ones with respect to the quasi-identifier attributes in the anonymized dataset. For example, Table 2 is an anonymous table of Table 1. Generalization involves replacing (or recoding) a value with a less specific but semantically consistent value. For example, the date of birth could be generalized to a range such as year of birth, so as to reduce the risk of identification. Suppression involves not releasing a value at all. It is clear that such methods reduce the risk of identification with the use of public records, while reducing the accuracy of applications on the transformed data in recent years; numerous algorithms have been proposed for implementing k-anonymity via generalization and suppression. Bayardo and Agrawal [16] presented an optimal algorithm that starts from a fully generalized table and specializes the dataset in a minimal k-anonymous table. LeFevre et al. [17] described an algorithm that uses a bottom-up technique and a priori computation. Fung et al. [18] presented a top-down heuristic to make a table to be released k-anonymous.

As to the theoretical results, Sweeney [19] proved the optimal k-anonymity is NP-hard and provided approximation algorithms for optimal k-anonymity. However, Machanavajjhala et al. [20] pointed out that the user may guess the sensitive values with high confidence when the sensitive data is lack of diversity, and introduced the l-diversity method. Subsequently, several models such as p-sensitive k-anonymity [21], (a, k)-anonymity [22], t-closeness [23], and M-invariance [24], etc. were proposed in the literature in order to deal with the problem of k-anonymity. The k-anonymity methods mainly focus on a universal approach that exerts the same amount of preservation for all individuals, without catering for their concrete needs. The consequence may be offering insufficient protection to a subset of people, while applying excessive privacy control to another subset. Motivated by this, Xiao and Tao [25] presented a new generalization framework based on the concept of personalized anonymity. Their technique performs the minimum generalization for satisfying everybody's requirements, and thus, retains the largest amount of information from the original data. In addition, the existing k-anonymity solutions based on generalization and suppression techniques suffer from high information loss and low usability mainly due to reliance on pre-defined generalization hierarchies or full order imposed on each attribute domain. References [26-29] provided a kind of new algorithm based on clustering technique, which reduced greatly the amount of information loss resulting from data generalization for implementing data anonymization.

K-Anonymity data mining is however a recent research area and many issues are still to be Investigated, such as, the combination of k-anonymity with other possible data mining techniques; the investigation of new approaches for detecting and blocking k-anonymity violations. The anonymization method can ensure that the transformed data is true, but it also results in information loss in some extent.

Table 1.**Original Data**

Name	Race	Birth	Sex	Zip	Disease
Alice	Blank	1965-3-18	M	02141	Flu
Bob	Blank	1965-5-1	M	02142	Cancer
David	Blank	1966-6-10	M	02135	Obesity
Helen	Blank	1966-7-15	M	02137	Gastritis
Jane	White	1968-3-20	F	02139	HIV
Paul	White	1968-4-1	F	02138	Cancer

Table 2.**Anonymization of table1**

Race	Birth	Sex	Zip	Disease
Blank	1965	M	0214*	Flu
Blank	1965	M	0214*	Cancer
Blank	1966	M	0213*	Obesity
Blank	1966	M	0213*	Gastritis
White	1968	F	0213*	HIV
White	1968	F	0213*	Cancer

The encryption method for distributed privacy preserving data mining

6.Method of Cryptography

The growth of Internet has triggered tremendous opportunities for distributed data mining, where people jointly conducting mining tasks based on the private inputs they supplies. These mining tasks could occur between mutual un-trusted parties, or even between competitors, therefore, protecting privacy becomes a primary concern in distributed data mining setting. Distributed privacy preserving data mining algorithms require collaboration between parties to compute the results or share no-sensitive mining results, while provably leading to the disclosure of any sensitive information. In general, distributed data mining involves two forms: horizontally partitioned data and vertically partitioned data. Horizontally partitioned data means that each site has complete information on a distinct set of entities, and an integrated dataset consists of the union of these datasets. In contrast, vertically partitioned data has different types of information at each site; each has partial information on the same set of entities. Most privacy preserving distributed data mining algorithms are developed to reveal nothing other than the final result. Kantarcioglu and Clifton [30] studied the privacy-preserving association rule mining problem over horizontally partitioned data. Their methods incorporate cryptographic techniques to minimize the information shared, while adding little overhead to the mining task. Lindell et al. [31] researched how to privately generate ID3 decision trees on horizontally partitioned data. The problemof privately mining association rules on vertically partitioned data was addressed in [32, 33]. Vaidya and Clifton [34] first studied how secure association rule mining can be done for vertically partitioned data by extending the Apriori algorithm. Du and Zhan [35] developed a solution for constructing ID3 on vertically partitioned data between two parties. Vaidya and Clifton [36] developed a Naive Bayes classifier for privacy preservation on vertically partitioned data and [37] proposed the first method for clustering over vertically partitioned data. All these methods are almost based on the special encryption protocol known as Secure Multiparty Computation (SMC) technology. SMC originated with Yao's Millionaires' problem [38]. The basic problem is that two millionaires would like to know who is richer, with neither revealing their net worth. Abstractly, the problem is to simply compare two numbers, each held by one party, without either party revealing its number to the other. The SMC literature defines two basic adversarial models:

Semi-Honest Model

Semi-honest adversaries follow the protocol faithfully, but can try to infer the secret information of the other parties from the data they see during the execution of the protocol.

Malicious Model

Malicious adversaries may do anything to infer secret information. They can

abort the protocol at any time, send spurious messages, spoof messages, collude with other (malicious) parties, etc.SMC technology used in distributed privacy preserving data mining areas mainly consists of a set of secure sub-protocols, such as, secure sum, secure comparison, dot product protocol, secure intersection, secure set union and so on. In the following, we will briefly describe the basic idea of two kinds of secure sub-protocols used in horizontally partitioned and vertically partitioned setting.

Secure Sum

Secure Sum can securely calculate the sum of values from different sites. Assume that each site *i* has some value *i v* and all sites want to securely compute $S = v_1 + v_2 + \ldots + v_n$, where *i v* is known to be in the range [0..m]. For example, in horizontally partitioned association rule mining setting, we can securely calculate the global support count of an itemset by the secure sum sub-protocol.

Dot Product Protocol

To present, many secure dot product protocols have been proposed. The

problem can be defined as follows: Alice has a n-dimensional vector $(x_1, x_2, \ldots, x_n) = X$, while Bob has a n-dimensional vector $(y_1, y_2, \ldots, y_n) = Y$. At the end of the protocol, Alice should get $a + b = r$ $(a = X \cdot Y + r)$ where *b r* is a random number chosen

from uniform distribution that is known only to Bob, and n X Y x y x y , , x y 1 1 2 2 . For example, using the dot product protocol we can securely calculate the global support count of an itemset whose items are located at different sites in vertically partitioned setting. The encryption method can ensure that the transformed data is exact and secure, but it is much low efficient. Moreover, most existing work on very efficient privacy preserving data mining only provides the protocols against semi-honest adversaries. An important area for future research is to develop efficient mining protocols that remain secure and private even if some of the parties involved behave maliciously.

7.Recent trend of privacy

Many social networks being analyzed today are generated from sources with privacy concerns. A number of network centrality measures have been introduced to better quantify various .Here propose an approximation of a social network that allows for certain centrality measures to be calculated while hiding information about the full network. Our approximation is not a perturbed graph, but rather a generalize trie structure containing a network hop expansion set for each node in the graph. We show that a network with certain topological structures, candidate nodes in each equivalence class. The storage of our graph approximation naturally clusters nodes of the network with similar graph expansion structure and therefore, can also be used as the basis for identifying 'like' nodes in terms of similar structural position in the network. For branches of the trie that are not private enough, we introduce heuristics that locally merges segments of the trie to enforce k-node anonymity.

A fingerprint authentication system for the privacy protection of the fingerprint template stored in a database is introduced here. The considered fingerprint data is a binary thinned fingerprint image, which will be embedded with some private user information without causing obvious abnormality in the enrollment phase. In the authentication phase, these hidden user data can be extracted from the stored template for verifying the authenticity of the person who provides the query fingerprint. A novel data hiding scheme is proposed for the thinned fingerprint template. This scheme does not produce any boundary pixel in the thinned fingerprint during data embedding. Thus, the abnormality caused by data hiding is visually imperceptible in the marked-thinned fingerprint.

Compared with using existing binary image data hiding techniques, the proposed method causes the least abnormality for a thinned fingerprint without compromising the performance of the fingerprint identification.

8.Conclusion

This paper carries out various approaches for privacy preservation data mining and analysis techniques and method what are existing. All the proposed methods are just approximate to achieve the goal of privacy upon some extend. Firstly, new algorithm with better approximation ratio and/or time complexity in this framework needs to be under development , still introduce considerable information loss with high-dimensional metric space involved. Drawback can be try to solve in next research paper .

REFERENCES

[1] Privacy, Security, and Data Mining, pp.1-8, 2002.] Han Jiawei, M. Kamber, and Data Mining: Concepts and Techniques, Beijing: China Machine Press, pp.1-40, 2006.

[2] V.S.Verykios, E.Bertino, I.N.Fovino, L.P.Provenza, Y.Saygin, Y.Theodoridis, "State-of-the-art in Privacy Preserving Data Mining", New York, ACM SIGMOD Record, vol.33, no.2,Pp.50-57, 2004.

[3] N. Zhang, "Privacy-Preserving Data Mining", Texas A&M University, pp.19-25, 2006.

[4] R. Agrawal, R. Srikant, "Privacy-Preserving Data Mining", ACM SIGMOD Record, New York, vol.29, no.2, pp.439-450,2000.

[5] A. Evfimievski, R. Srikant, R. Agrawal, J. Gehrke, "Privacy Preserving Mining of Association

[6] Rules", Information System, vol.29, no.4, pp.343-364, 2004.

[7] H. Kargupta, S. Datta, Q. Wang, K. Sivakumar, "On the Privacy Preserving Properties of Random Data Perturbation Techniques", In Proceedings of the 3rd International Conference on Data Mining, pp.99-106, 2003.

[8] Z. Huang, W. Du, B. Chen, "Deriving Private Information from Randomized Data", In Proceedings of the ACM SIGMOD Conference on Management of Data, Baltimore, Maryland,USA, pp.37-48, 2005.

[9] D. Agrawal, C.C. Aggarwal, "On the Design and Quantification of Privacy Preserving Data Mining Algorithms", In Proceedings of the 20th ACM SIGMOD-SIGACTSIGART Symposium on Principles of Database Systems, pp.247-255, 2001.

[10] A. Evfimievski, R. Srikant, R. Agrawal, J. Gehrk, "Privacy Preserving Mining of Association Rules", In Proceedings the 8th ACM SIGKDD International Conference on Knowledge Discovery in Databases and Data Mining, pp.217-228, 2002.

[11] S. Rizvi, J. Haritsa, "Maintaining Data Privacy in Association Rule Mining", In Proceedings the 28th International Conference on Very Large Data Bases, pp.682-693, 2002.

[12] S. L. Warner, "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias", J. Am. Stat. Assoc., vol.60, no.309, pp.63-69, 1965.

[13] S.J. Rizvi, J.R. Haritsa, "Maintaining Data Privacy in Association Rule Mining", In Proceedings the 28th VLDB conference, pp.1-12, 2002.

[14] W. Du, Z. Zhan, "Using Randomized Response Techniques for Privacy Preserving Data Mining", In Proceedings 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.505-510, 2003.

[15] Guo, S. Guo, X. Wu, "Privacy Preserving Market Basket Data Analysis", In Proceedings the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases,Pp.103-114, 2007.

[16] L. Sweeney, "k-Anonymity: A Model for Protecting Privacy", International Journal of Uncertainty, Fuzziness and Knowledge-based Systems, vol.10, no.5, pp.557-570, 2002.

[17] R. Bayardo, R. Agrawal, "Data Privacy Through Optimal k-Anonymization", In Proceedings the 21st International Conference on Data Engineering, pp.217-228, 2005.

[18] K. Lefevre, J. Dewittd, R. Ramakrishnan, "Incognito: Efficient Full-Domain k-Anonymity", In Proceedings of the 2005 ACM SIGMOD International Conference on Management of Data, Pp.49-60, 2005.

[19] B. Fung, K. Wang, P. Yu, "Top-down Specialization for Information and Privacy Preservation", In Proceedings of the 21st IEEE International Conference on Data Engineering, pp.205-216, 2005.

[20] L. Sweeney, "Achieving k-Anonymity Privacy Protection Using Generalization and Suppression", International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, vol.10, no.5, pp.571-588, 2002.

[21] A. Machanavajjhala, J. Gehrke, D. Kifer, "l-Diversity: Privacy Beyond k-Anonymity", ACM Transactions on Knowledge Discovery from Data, pp.24-35, 2007.

[22] T. Truta, B. Vinay, "Privacy Protection: p-Sensitive k-Anonymity Property", In Proceedings of the 22nd International Conference on Data Engineering Workshops, pp. 94-103, 2006.

[23] R.C.Wong, J.Y.Li, A.W. Fu, "(a, k)-Anonymity: An Enhanced k-Anonymity Model for Privacy-Preserving Data Publishing", In Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.754-759, 2006.

[24] N.H. Li, T.. Li, "t-Closeness: Privacy beyond k-Anonymity and l-Diversity", In Proceedings of the 23rd International Conference on Data Engineering, pp.106-115, 2007.

[25] X.K. xiao, Y.F. Tao, "M-Invariance: Towards Privacy Preserving Re-Publication of Dynamic Datasets", In Proceedings of the ACM Conference on Management of Data (SIGMOD), pp.689-700, 2007.

[26] X.K. Xiao, Y.F. Tao, "Personalized Privacy Preservation", In Proceedings of the ACM Conference on Management of Data (SIGMOD), pp.229-240, 2006.

[27] G. Loukides, J.H. Shao, "An Efficient Clustering Algorithm for k-Anonymisation", International Journal of Computer Science and Technology, vol.23, no.2, pp.188-202, 2008.

[28] J.L. Lin, M.C. Wei, "Genetic Algorithm-Based Clustering Approach for k-Anonymization",International Journal of Expert Systems with Applications, vol.36, no.6, pp.9784-9792, 2009.

[29] L.J. Lu, X.J. Ye, "An Improved Weighted-Feature Clustering Algorithm for k-Anonymity", In Proceedings of the 5th International Conference on Information Assurance and Security, Pp.415-419, 2009.

[30] Z.H. Wang, J. Xu, W. Wang, B.L. Shi, "Clustering-Based Approach for Data Anonymization", Journal of Software,vol.21, no.4, pp.680-693, 2010.

[31] M. Kantarcioglu, C. Clifton, "Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data", IEEE Transactions on Knowledge and Data Engineering, vol.16, no.9, pp.1026-1037, 2004.

[32] Lindell, Yehuda, Pinkas, "Privacy preserving data mining", In Proceedings of the Advances in Cryptology–CRYPTO, pp.36–54, 2000.

[33] J. Vaidya, C. Clifton, "Privacy Preserving Association Rule Mining in Vertically Partitioned Data", In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.639-644, 2002.

[34] I. Ioannidis, A. Grama, M.J. Atallah, "A Secure Protocol for Computing Dot-Products in Clustered and Distributed Environments", In Proceedings of the 31st International Conference on Parallel Processing, pp.379-384, 2002.

[35] J. Vaidya, C. Clifton, "Privacy Preserving Association Rule Mining in Vertically Partitioned Data", In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.639–644, 2002.

[36] W.L. Du, Z.J. Zhan, "Building Decision Tree Classifier on Private Data", In Proceedings of the IEEE International Conference on Data Mining Workshop o

[37] J. Vaidya, C. Clifton, "Privacy Preserving Naive Bayes Classifier for Vertically Partitioned Data", In Proceedings of the 2004 SIAM International Conference on Data Mining, pp.522–526, 2004.

[38] J. Vaidya, C. Clifton, "Privacy-Preserving k-Means Clustering over Vertically Partitioned Data", In Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.206–215, 2003.

[39] Yao, C. Andrew, "How to Generate and Exchange Secrets", In Proceedings of the 27th IEEE

[40] Symposium on Foundations of Computer Science, pp.162-167, 1986.