Privacy Preserving Data Mining: Techniques, Classification and Implications - A Survey

Alpa Shah
MCA Department
Sarvajanik College of Engineering and Technology,
Surat, India

Ravi Gulati
Department of Computer Science,
Veer Narmad South Gujarat University,
Surat, India

ABSTRACT
Privacy has become crucial in knowledge based applications. Proper integration of individual privacy is essential for data mining operations. This privacy based data mining is important for sectors like Healthcare, Pharmaceuticals, Research, and Security Service Providers, to name a few. The main categorization of Privacy Preserving Data Mining (PPDM) techniques falls into Perturbation, Secure Sum Computations and Cryptographic based techniques. There exist tradeoffs between privacy preservation and information loss for generalized solutions. The authors of the paper present an extensive survey of PPDM techniques, their classification and give a preliminary implication of technique to be used under specific scenarios.

Keywords
PPDM, Perturbation, Cryptography, SMC, Randomization, Condensation, Anonymization

1. INTRODUCTION
Recent years have seen unprecedented growth in applicability of Computer Science in day-to-day activities. Organizations, community and individuals show an augmented trend of storing their data electronically. The huge amount of data collected can be used for analyzing trends of markets and individual or society. Data mining activities involve extracting knowledge from this massive pool of data. The sensitive information about the individuals may be disclosed creating ethical or privacy issues. Many individual therefore don't share their data publicly, creating data unavailability. Privacy of individual should not be compromised under any case. PPDM has gained popularity so as to address the privacy concerns while data mining is being carried out. The authors of this paper attempts to provide a comprehensive literature survey based on the techniques, classification and the scenarios of their implications on various techniques applied for PPDM.

The flow of this survey paper is as follows:

1. Section II provides the fundamentals of need for privacy.
2. Section III provides classification of PPDM techniques based on centralized and distributed scenarios.
3. Section IV explores various studies related to privacy issues.
4. Section V compares the available techniques for PPDM.

2. NEED FOR PRIVACY
With modern world getting digitized, there is an increase in electronic data. It is important to analyze socio-economic trends of the individuals of the society. Privacy concern is important when data disclosure is taken into account. Say for an example, Medical data contains sensitive data as it contains information about the patients' diseases. It is important to privatize this data before making it available for data mining. In medical scenarios, it is important to preserve the mining model with effective privacy; else it will lead to inaccurate predictions that are improper. Personal specific details must not be disclosed which may otherwise be considered unethical. Privacy can be defined as prevention of unwanted disclosure of information when data mining is performed on aggregate results. Privacy must be addressed at all the levels while mining is carried out.

Privacy and security both are impediment for data mining task. A clear demarcation between security and privacy requirements of published data is essential. [44] provides an address for identifying the importance of security and privacy in data mining. An extensive literature survey on PPDM is also performed by authors of [43] [45]. In this paper, the authors first distinguish between privacy and security in context of Census data. The remaining section provides an introduction to privacy policies and issues that are taken care by various governing bodies within India and other countries.

2.1 How are Privacy and Security Different?
Privacy and security are two terms used interchangeably under different contexts. But both are related to each other and at the same time entirely separate issues.

- The three fundamentals of security are *Confidentiality*, *Integrity* and *Availability* [28]. In context of Census data, security can be termed as the facility for controlling person-specific access information, protect it from unauthorized disclosure, modification, loss or destruction of his information. Security can be accomplished through controls based on operational and technical knowhow.
- In contrast privacy is very specific. It can be termed as a right of an individual to keep his/her personal information from being disclosed. Privacy can be accomplished through policies and procedures. Person's personal information which may lead to his identification may not be disclosed under ethical grounds.

PPDM is extensively studied by researchers to address these issues for privacy. The security aspects can be taken care by enforcing vigorous methods for protection of sensitive data.

2.2 Privacy Concerns

Privacy is considered as an important aspect of preserving information without information loss. The perspective of privacy differs based on the data in use and the way in which it is used. Many methods like attribute removal, anonymization, randomization, aggregation on numeric values are applied on data sets to provide privacy. These methods incur information loss in some situations too. Cryptographic techniques involve additional computational overhead. Secure sum computations require the feasibility of basic combinatorial circuit which computes the functions on data [11]. It has been shown by authors [23] that when number of parties scales high, such computations lead to exponential computational and communication cost. As of now, no generic solutions are available to address all privacy issues with respect to all the scenarios of applicability. Research has been focused on finding efficient protocols for specific problems only. They balance privacy, data utility and computational feasibility at a good level. Still data utility and information loss is trade-offs when effective data mining is conducted with respect to privacy measures.

2.3 C. Privacy Policies

Privacy breaches must be addressed by researchers at a highest priority. Privacy can be said to have been breached when an individual's exact privacy information can be directly linked with him. Identifying all types of breaches is very difficult. Hence the privacy providers must confront to some standards and policies provided by HIPAA of US, Data Protection Act of UK. Federal Health Insurance Portability and Accountability (HIPAA) have stringent privacy policies for medical privacy. In India, Information Technology Rules, 2011 under the Information Technology Act, 2000, has been notified. Reasonable security practices and policies for sensitive personal data have now been enforced in India effectively.

3. PRIVACY PRESERVING DATA MINING [PPDM]

Before data mining tasks are carried out, several methods must be applied to protect the privacy of individuals. Privacy preserving data mining is the branch which includes the studies of privacy concern when mining is applied. Various methods like data hiding, masking, suppression, aggregation, perturbation, anonymization, SMC are studied in literature with regards to PPDM. Next section, describes the classification of PPDM techniques based on current research findings.

Based on the location of computation carried out for mining results, PPDM techniques can be classified as described in Figure 1. The mining can be entrusted to a trusted third party who collects all sensitive data. Another scenario is when the individual parties privatize their data before mining process is carried out. The classification thus can be broadly categorized as: Central/Commodity Server and Distributed. The implementation of various techniques related to Fuzzy and Neural Networks is still rudimentary and is discussed in brief here. Authors in [6] [20] have discussed the implementations of Rough Sets, Genetic algorithms in direction of PPDM. A new research direction in Genetic Algorithms and its implementation with PPDM is also open.

3.1 Central Commodity Server Scenario

In this scenario, a trusted third party Central Commodity Server plays an important role. Each of the contributing parties entrust the Central Commodity Server the task of preserving the privacy of individual contributing parties. Before publishing the data, all the contributors transfer their data to the server. The mining task is independently carried out by the server. The mining is directly carried out by the server and the number of users is scalable. Generally solutions present in literature do not allow scalability to the number of users. The server must privatize the data prior to mining. The task of data mining is independent to the users that contribute the data in nature and avail more flexibility in terms of aggregating the datasets. Datasets may be horizontally or vertically partitioned in case of central trusted commodity server scenario. Anonymization and Perturbation are the best suitable methods under central server scenarios.

3.2 Distributed Scenario

An altogether different mechanism implies the privacy of the individual contributors at their end. The contributing parties prior to publishing the data sanitize the data and privatize it. The mining can be performed by the data owners and their aggregate results are then used for finding the effective association rules. Most of this type of scenarios have very specific goals and are based on heavy computation techniques like Secure Multiparty Computation (SMC) and Cryptographic techniques. The contributing parties can also generate perturbed copies and based on the level of privacy requirement generate perturbed versions of original dataset. As stated previously too, still the literature poses study on protocols which are based on specific applications only. Research to find generic protocols that deal with wide range of applications is still elusive which focuses on data mining fundamentals of classification, clustering, pattern matching, association rule hiding and others.

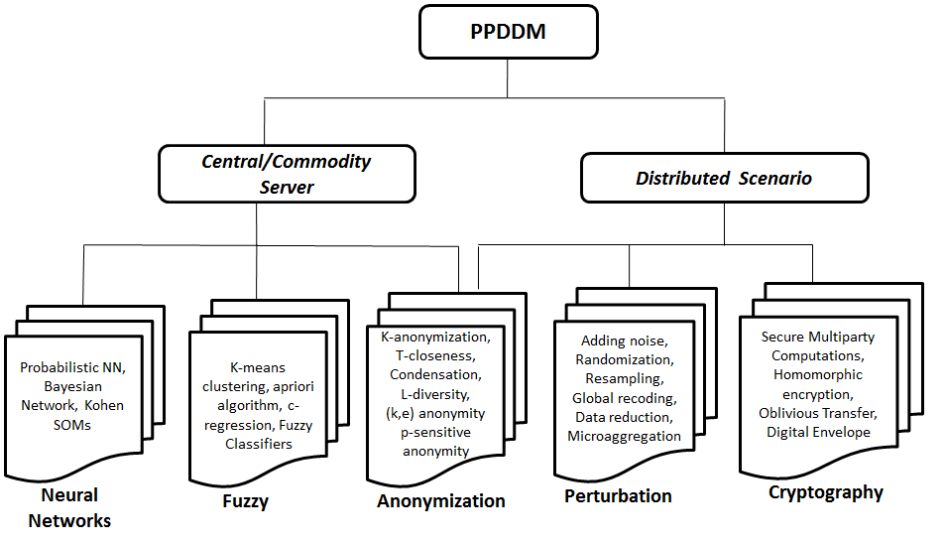

Figure 1: PPDM Classification Hierarchy

4. PPDM TECHNIQUES

A detailed literature survey is now presented based on the classification hierarchy discussed above. As shown in Figure 1, the major classification for PPDM is based on Anonymization, Perturbation, Cryptography Fuzzy and Neural Networks.

Anonymization Based

At certain times the data is required to be published in its original form publicly. The data may not be encrypted and perturbed, but still some sort of precaution should be taken before releasing the data in terms of anonymization. This is a kind of generalization of some attributes that protects against identity disclosure. Anonymization can be achieved by methods like generalization, suppression, data removal, permutation, swapping etc [36]. k-anonymity method is treated as the conventional anonymization method and many studies are based on k-anonymity. Improved methods like l-diversity, t-closeness, km -anonymization, (α,k) anonymity, p-sensitive k-anonymity, (k,e) anonymity, are described in [40], which are also studied in literature. Their work provides a detailed survey of anonymization methods and also illustrates drawbacks in k-anonymity.

Quasi-Identifier is a combination of person specific sensitive attribute (say for example, age, disease and pin-code for census data). The authors in [13][16] have proved that the removal of the quasi-identifier from dataset do not ensure data protection, still k – anonymity method is better choice for publishing data. A simple approach is to generalize fields which are part of quasi identifier. Say for an example, age can be categorized in groups. Authors in their work [5] suggest a novel approach which uses a bottom-up method to group and then anonymize quasi-identifiers. Another work in [3] suggests a task-based technique which satisfactorily balances both the privacy and utility trade-offs. Mining is done after the algorithm in [3] is applied which hides the sensitive data effectively. Anonymizing quasi-identifiers and sensitive attributes in datasets pose an information loss which is not desirable for mining. The authors of [22] focus on medical datasets and try to address the issues related to privacy requirements.

Anonymization methods are also useful for addressing specific problems. Authors in [31] have used k-anonymity based method for optimal feature set partitioning. [34] emphasises cluster analysis for preserving the sensitivity of data. Authors in [33] have proposed data reconstruction approach which achieves k-anonymity protection in predictive data mining. The potentially identifiable attributes are first mapped using aggregation for numeric data and swapping is done for nominal data. A technique based on genetic algorithm is applied to the masked data for finding a better subset from it. The subset is replicated to generate published dataset which satisfies the k-anonymity constraint.

Condensation is a statistical approach which constructs constrained clusters in a dataset and then generates pseudo data from statistics of these clusters [23]. Clusters of non-homogeneous size are constructed from whole data, such that, each record lay in a group whose size is at least equal to its anonymity level. After this pseudo data is generated from each group, and synthetic dataset is created with similar aggregate distribution as that of the original dataset. Condensation is effectively used for solving the classification problem. An additional layer of protection is provided with pseudo data making it difficult for adversaries. Also, aggregate behaviour of data is preserved with condensation, making it useful for data mining tasks.

4.1 Perturbation Based

Perturbation techniques employ a mechanism to distort data prior to data mining. A perturbed copy can be locally created by the individual contributor by adding noise. Once the local perturbed copy is generated the miner can reconstruct the perturbed version to obtain the original data distribution. The authors in [1] have tried to add Gaussian noise to generate perturbed version of dataset for decision tree classification. In same lines, authors in [2] have proposed an individually adaptable perturbation model. A multilevel privacy can be specified by the users. This opens a new venture in field of privacy preserving – Multi-level Trust PPDM(MLT PPDM). Based on the privacy settings a contributor specifies, the perturbed version of dataset will be generated. The authors have successfully proved with experiments the correctness of their approach for satisfying personal privacy. Another work [35] offers the flexibility to the data owners to generate perturbed copies for arbitrary trust levels on demands.

Perturbation methods can be classified into probability distribution category and fixed data perturbation. Probability

distribution allows adding noise based on some known distribution pattern like Gaussian. The data distortion techniques like addition of noise, from some known distribution, randomization and condensation are applied. Perturbation methods are well suited in both central commodity based computing as well distributed scenarios [41]. A different type of perturbation called Geometric Data Perturbation (GDP) is based on service oriented framework and is discussed in [18]. In Literature [23] a perturbation based technique which builds a classifier for the original dataset from the perturbed training dataset by skipping the steps of reconstructing the original data distribution is discussed

Randomization is a data perturbation technique where the data distortion is masked by random data. Warner in his study has introduced this technique of statistics to solve the survey problem. Authors in [32] have proposed a mechanism to scramble data in a manner that the central repository won't be able to judge whether the information can be classified as true or false. With large number of users, aggregate information can be estimated with accuracy. This information can be used for decision-tree classification as the latter is based on aggregate values of a dataset.

4.2 Cryptography Based

If the parties distributed across multiple sites are legally prohibited from sharing their datasets, a mining model to be built must be able to maintain the privacy of contributing parties. Author in [25] have discussed the efficiency and have demonstrated their relevance for PPDM. Examples to demonstrate secure sum computation of data mining algorithms are also discussed. Previous categories of PPDM allow disclose of data beyond the control of the data collection. Authors in [14] have addressed the problem of reconstructing missing values by building a data model where the parties are distributed and data is horizontally partitioned. A cryptographic protocol based on decision-tree classification is described by them. A survey on cryptographic techniques for PPDM is studied by authors of [49]. Distributed environment where the sharing is constrained either under legal or privacy policy issues use the cryptographic techniques. Oblivious transfer is used as building block for constructing an efficient PPDM model by authors in [11]. The problem of distributed ID3 is addressed by authors in [12]. The implementations of these protocols consist of computationally intensive operations and generally consist of hard wired circuits.

Secure Multiparty Computation is a technique in which computations are done beforehand on the basis of certain rules in statistical disclosure limitation. Basically there are three broad types of techniques under SMC: homomorphic encryption, circuit evaluation and secret sharing scheme. Both semi-honest and malicious adversaries are addressed by SMC protocols. A semi-honest adversary abides the protocol specification righteously but may try to learn facts by supplying incorrect information to the protocol. Most of the applications under SMC are built which address the semi-honest adversaries. Authors in [7] have proposed a SMC based model for malicious adversaries. The authors have proposed a framework that assigns liability for privacy to the responsible parties. Authors in [42] have made an analysis to support the accuracy and efficiency of SMC based protocols. [27] provides a privacy preserving framework based on SMC using Gaussian mixture models. Authors of [30] have devised a protocol for based on encryption which will protect the privacy at each contributor end. Authors in [9][36] have introduced a cryptographic approach for privacy preservation for classification problem. Authors in [15] have devised a method for privacy preservation based on homomorphic encryption for association rule mining.

Another form of cryptographic application is Pseudonomization. Here, the links between the personal and his medical information are broke by anonymizing. Directly the information pertaining to personal identification is not removed from the dataset, but a pseudonym is generated and replaced. This information cannot be retrieved without compromising a secret shared previously. [11] proposes encryption based technique for building pseudonyms. The pseudonyms are generated at the distributed site by the contributor parties.

4.3 PPDM based on Fuzzy Algorithms

PPDM based on Fuzzy algorithms allow achieving anonymization without significant loss of information. The algorithms merge similar records into clusters. Each cluster formed is distinct from other clusters and the records of each cluster are not distinguishable from those of other clusters. A technique k-means clustering for anonymizing using Fuzzy logic is proposed in [54]. The record in cluster k is anonymized to make it indistinguishable from remaining k-1 clusters. [47] have suggested a modified apriori algorithm based on Fuzzy data in order to identify and then privatize sensitive rules in distributed scenarios. The method proposed by them for association rule hiding is efficient in terms of information hiding with fewer side effects. Authors in [52] have used a fuzzy-based c-regression method to generate microdata (synthetic data). Trusted third party commodity servers are then entrusted with task of statistical computation with minimum risk of information loss.

4.4 Neural Network based

Neural network is a mathematical model or computational model based on biological neural networks. Neural Network based PPDM is studied in literature to achieve privacy of individual contributing parties without compromising information loss.[50] [24] proposes a probabilistic neural network committee for peer-to-peer data mining by selecting best of weight-based peer member. Authors in [48] have used Kohen Self Organizing Feature Maps that maintains the privacy of data and outliers with minimum disclosure probability and probability loss. Authors in [53] construct a Bayesian network for Learning Distribution of data. The algorithm performs accurately for binary and non-binary discrete data. [46] proposes a protocol for Bayesian networks on vertically partitioned data with negligible overhead. The protocol proposed by them provides better performance, ensures complete privacy and is accurate.

5. COMPARISON OF PPDM TECHNIQUES

Table 1 presents a pilot comparison of various PPDM techniques to justify the optimal technique best suited for each scenario. It also illustrates the methods that are employed by different techniques. The table iterates and summarizes the discussion in previous sections on PPDM techniques.

Table 1: Comparison of PPDM Techniques

Techniques	Methods Employed	Scenarios	Data Mining Tasks			
			Classification	Clustering	Association Rule Mining	Outlier detection
Anonymization	Generalization Suppression, Permutation	Central Commodity	✓	✓	✓	✓
Condensation	Aggregation	Central Commodity	✓			
SMC	Cryptographic	Distributed	✓	✓	✓	✓
Pseudonymization	Cryptographic	Distributed	✓	✓	✓	✓
Perturbation	Adding Noise, Data Swapping, Global recoding, Microaggregation	Both	✓	✓	✓	✓
Randomization	Adding Noise, Scrambling, Resampling	Both	✓			
Fuzzy based	Clustering, Microaggregation, c-regression	Central Commodity	✓	✓	✓	
Neural Network Based	Bayesian Network, Probabilistic NN	Central Commodity	✓	✓	✓	✓

6. CONCLUSIONS

Privacy is indispensible for data mining tasks. It is challenging to protect the privacy while the computation tasks are carried on. A trade-off between utility of information and privacy always exists. An optimum solution that reduces the computational overheads and balances information loss is still a topic of research. The authors in the paper have tried to classify the PPDM techniques available in the literature and showed its implications best suited under various scenarios. Currently no such technique that provides the best solutions under different scenarios exists. A study to find a new technique altogether or combination of these techniques best suited is an open research area still. Different frameworks are explored in [8][17][23][38][40] which can be still further enhanced to provide better results.

7. REFERENCES

[1] Agrawal and Srikant, "Privacy Preserving Data mining", Proceedings of the ACM SIGMOD International Conference on Management of data, 2000.

[2] Li Liu, Murat Kantarcioglu and Bhavani Thuraisingham, "The applicability of the perturbation based privacy preserving data mining for real-world data", Data & Knowledge Engineering 65 (2008) 5–21.

[3] E. Poovammal and M. Ponnavaikko, "Task Independent Privacy Preserving Data Mining on Medical Dataset", International Conference on Advances in Computing, Control and Telecommunication Technologies, 2009.

[4] Marina Blanton, "Achieving Full Security in Privacy-Preserving Data Mining", IEEE International Conference on Privacy, Security, Risk, and Trust, and IEEE International Conference on Social Computing, 2011.

[5] Tiancheng Li, Ninghui Li, "Towards Optimal k-anonymization", Data & Knowledge Engineering, 2008 Elsevier. 303

[6] E .Poovammal and Dr. M. Ponnavaikko, "An Improved Method for Privacy Preserving Data Mining", IEEE International Advance Computing Conference (IACC 2009) Patiala, India, 6-7 March 2009.

[7] Jiang, Clifton and Kantarcıoğlu, "Transforming Semi-Honest Protocols to Ensure Accountability", Data & Knowledge Engineering, 2008 Elsevier.

[8] Bhavani Thuraisingham, "Privacy constraint processing in a privacy-enhanced database management system", Data & knowledge Engineering, 2005.

[9] XunYi, YanchunZhang, "Privacy-preserving naive Bayes classification on distributed data via semi-trusted mixers", Information Systems 34 (2009) 371–380.

[10] Jian Wang, Yongcheng Luo, Yan Zhao, Jiajin Le, "A Survey on Privacy Preserving Data Mining", First International Workshop on Database Technology and Applications, 2009.

[11] Yun Ding and Karsten Klein, "Model-Driven Application-Level Encryption for the Privacy of E-Health Data", International Conference on Availability, Reliability and Security, 2010.

[12] Yehuda Lindell, Benny Pinkas, "Privacy Preserving Data Mining", http://www.pinkas.net/PAPERS/id3-final.pdf.

[13] Samarati P, "Protecting respondent's privacy in Microdata release", IEEE Transactions on Knowledge and Data Engineering, 13:1010–1027

[14] Geetha Jagannathan, Rebecca N. Wright, "Privacy-Preserving Imputation of Missing Data", Data & Knowledge Engineering, 2008 Elsevier.

[15] Justin Zhan, Stan Matwin, Li Wu Chang, "Privacy-preserving collaborative association rule mining", Journal of Network and Computer Applications 30 (2007) 1216–1227.

[16] Sweeney L, "k-anonymity: A model for protecting Privacy", International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10(5):557–570.

[17] Jimmy Secretan, Michael Georgiopoulos, Anna Koufakou, Kel Cardona, "APHID: An architecture for private, high performance integrated data mining", Future Generation Computer Systems 26 (2010) 891_904.

[18] Keke Chen and Ling Liu, "Privacy-Preserving Multiparty Collaborative Mining with Geometric Data Perturbation", IEEE Transactions on Parallel and Distributed Systems, Vol. 20, No. 12, December 2009.

[19] Jitao Zhao and Ting Wang, "A General Framework for Medical Data Mining", International Conference on Future Information Technology and Management Engineering, 2010.

[20] R. Mukkamala and V.G. Ashok, "Fuzzy-based Methods for Privacy-Preserving Data Mining", Eighth International Conference on Information Technology: New Generations, 2011.

[21] F. Emekci, O.D. Sahin, D. Agrawal, A. El Abbadi, "Privacy preserving Decision tree learning over multiple parties", Data & Knowledge Engineering 63 (2007) 348–361.

[22] Yan ZHU and Lin PENG, "Study on K-anonymity Models of Sharing Medical Information", 1-4244-0885-7/07/$20.00 ©2007 IEEE.

[23] Li Liu, Murat Kantarcioglu and Bhavani Thuraisingham, "Privacy Preserving Decision Tree Mining from Perturbed Data", Proceedings of the 42nd Hawaii International Conference on System Sciences – 2009.

[24] Samet, S. ; Miri, A., 2009, Privacy-Preserving Bayesian Network for Horizontally Partitioned Data International Conference on Computational Science and Engineering, 2009. CSE '09. (Volume:3), pp: 9-16

[25] Benny Pinkas, "Cryptographic techniques for privacy preserving data mining",

[26] Jinfei Liu, Jun Luo and Joshua Zhexue Huang, "Rating: Privacy Preservation for Multiple Attributes with Different Sensitivity Requirements", 11th IEEE International Conference on Data Mining Workshops, 2011.

[27] Madhusudana Shashanka, "A Privacy–Preserving Framework for Gaussian Mixture Models", IEEE International Conference on Data Mining Workshops, 2010.

[28] José Luis Fernández-Alemán,Inmaculada Carrión Señor, Pedro Ángel Oliver Lozoya, Ambrosio Toval, "Methodological Review-Security and Privacy in electronic health records: A systematic literature review", Journal of Biomedical Informatics(2013).

[29] Yan Zhao, Ming Du, Jiajin Le, Yongcheng Luo, "A Survey on Privacy Preserving Approaches in Data Publishing", First International Workshop on Database Technology and Applications, 2009.

[30] Xun Yi, Yanchun Zhang, "Privacy-preserving distributed association rule mining via semi-trusted mixer", Data & Knowledge Engineering 63 (2007) 550–567.

[31] Nissim Matatov, Lior Rokach, Oded Maimon, "Privacy-preserving data mining: A feature set partitioning approach", Information Sciences 180 (2010) 2696–2720.

[32] Majid Bashir Malik, M. Asger Ghazi, Rashid Ali, "Privacy Preserving Data Mining Techniques: Current Scenario and Future Prospects", Third International Conference on Computer and Communication Technology, 2012.

[33] Dan Zhu, Xiao-Bai Li, Shuning Wu, "Identity disclosure protection: A data reconstruction approach for privacypreserving data mining", Decision Support Systems 48 (2009) 133–140.

[34] Benjamin C. M. Fung, Ke Wang, Lingyu Wang, Patrick C.K. Hung, "Privacy-preserving data publishing for cluster analysis" , Data & Knowledge Engineering 68

[35] Yaping Li, Minghua Chen, Qiwei Li, and Wei Zhang, "Enabling Multilevel Trust in Privacy Preserving Data Mining", IEEE Transactions On Knowledge And Data Engineering, Vol. 24, No. 9, September 2012.

[36] Asmaa H.Rashid and Prof.dr. Abd-Fatth Hegazy, "Protect Privacy of Medical Informatics using K-Anonymization Model", IEEE Explore

[37] Alper Bilge, Huseyin Polat, "A comparison of clustering-based privacy- preserving collaborative filtering Schemes", Applied Soft Computing 13 (2013) 2478–2489.

[38] Gerardo Canfora, Elisa Costante, Igino Pennino, Corrado Aaron Visaggio, "A three–layered model to implement data privacy policies", Computer Standards & Interfaces 30 (2008) 398–409

[39] Weijia Yang, Sanzheng Qiao, "A novel anonymization algorithm: Privacy protection and knowledge preservation", Expert Systems with Applications 37 (2010) 756–766.

[40] Sergio Martínez, David Sánchez, Aida Valls, "A semantic framework to protect the privacy of electronic health records with non-numerical attributes", Journal of Biomedical Informatics 46 (2013) 294–303.

[41] R. Vidya Banu, N .Nagaveni, "Evaluation of a perturbation-based Technique for privacy preservation in a multiparty clustering scenario", Information Sciences 232 (2013) 437–448.

[42] Sin G Teo, Vincent Lee, Shuguo Han, "A Study of Efficiency and Accuracy of Secure Multiparty Protocol in Privacy-Preserving Data Mining", 26th International Conference on Advanced Information Networking and Applications Workshops, 2012.

[43] Alpa K. Shah, Ravi Gulati, "Contemporary Trends in Privacy Preserving Collaborative Data Mining– A Survey", Proceedings in IEEE International Conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO), 2015

[44] Alpa K. Shah, Ravi Gulati, "Privacy, Collaboration and Security – Imperative Existence in Data Mining" VNSGU Journal of Science and Technology Vol 4 ,No 1, July 2015, Pg. 44-49, 0975-5446

[45] Jisha Jose Panackal1 ,Dr Anitha S Pillai, "Privacy Preserving Data Mining: An Extensive Survey", in Proceedings of Proc. of Int. Conf. on Multimedia

Processing, Communication and Info. Tech., MPCIT, 2013.

[46] Tsiafoulis, S.G. Zorkadis, V.C., 2010, A Neural Network Clustering Based Algorithm for Privacy Preserving Data Mining, International Conference on Computational Intelligence and Security (CIS), 2010, pp: 401-405

[47] SathiyaPriya, K.; Sadasivam, G.S.;Celin, "A new method for preserving privacy in quantitative association rules using DSR approach with automated generation of membership function", World Congress on Information and Communication Technologies (WICT), 2011, pp: 148-153

[48] Zhiqiang Yang ; Wright, R.N. 2005, Improved Privacy-Preserving Bayesian Network Parameter Learning on Vertically Partitioned Data, 21st International Conference on Data Engineering Workshops, 2005. Pp:1196

[49] Alpa K. Shah, Ravi Gulati," A Survey on Cryptographic Techniques for Privacy Preserving Data Mining", IIJDWM, Mining Vol 2 Issue1 Feb 2012 pp: 8-12

[50] Wang Hongmei ; Zhao Zheng ; Sun Zhiwei, 2005, Privacy preserving Bayesian network structure learning on distributed heterogeneous data,.11th Pacific Rim International Symposium on Dependable Computing, 2005. Proceedings, DOI: 10.1109/PRDC.2005.49

[51] Syed Zahid Hassan and Brijesh Verma, "A Hybrid Data Mining Approach for Knowledge Extraction and Classification in Medical Databases", Seventh International Conference on Intelligent Systems Design and Applications.

[52] Cano I., Torra V, "Generation of synthetic data by means of fuzzy c-Regression" . IEEE International Conference on Fuzzy Systems, 2009. FUZZ-IEEE, pp: 1145 – 1150

[53] Kokkinos, Y., Margaritis, K., 2013, Distributed privacy-preserving P2P data mining via probabilistic neural network committee machines, Fourth International Conference on Information, Intelligence, Systems and Applications (IISA), 2013, pp: 1-4

[54] Honda, K. ; Kawano, A. ; Notsu, A. ; Ichihashi, H., 2012, "A fuzzy variant of k-member clustering for collaborative filtering with data anonymization", Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on, pp: 1-6