Volume 8, No. 5, May – June 2017
ISSN No. 0976-5697
International Journal of Advanced Research in Computer Science
REVIEW ARTICLE
Available Online at www.ijarcs.info

A Review of Various Multiple Numerical and Categorical Sensitive Attribute for Preserving Privacy

Dharavathu Radha
Department of CS & SE, Andhra University
Visakhapatnam, India

Prof. Valli Kumari Vatsavayi
Department of CS & SE, Andhra University
Visakhapatnam, India

Abstract: Nowadays Privacy Preserving in Data Publishing (PPDP) has rapidly growth into the research contents in data protection field and also it has rapidly grown in publication of personal data. Recent years in data publishing is exceptionally analytical, because, how to economically preserve numerical and categorical sensitive attribute. However, different tender have been designed for privacy preserving to protect numerical sensitive attribute in data publishing, Like k-anonymity, *l*-diversity, t-closeness, (epsilon, m)-anonymity another methods are implemented for protecting the privacy of data provider. In this paper, we propose a review of various multiple numerical and categorical sensitive attribute for preserving privacy and survey current existing techniques. Yet, we discuss the future instructions of privacy preserving in data publishing, that is we suggest modern technique for "A new enhanced slicing method for both numerical and categorical sensitive attribute via advanced clustering algorithm" that improving the privacy with less information loss, membership, attribute , identity disclosure and error ratio and also calculate distance between two attributes for categorical attribute. This technique suitable for both categorical and numerical sensitive attributes.

Keywords: multiple numerical and categorical sensitive attribute; privacy preserving; k-anonymity; Advanced clustering algorithm.

I. INTERDUCTION

Now a day's society is allowing especially quality of data well as person-certain information computer information as computer knowledge, network connectivity and disk storage space. Privacy preservation is a serious concern in publication of personal data. Analyzing personal records requires the data to be published while at the same time the individual privacy is protected which has become an issue of rising importance for the time being an anonymization concept is a familiar generalization that swaps quasi-identifier attributes with those which are vague but logical throughout. Micro-data play a significant part in data analytic and technical, the propagation and partitioning of micro-data will jeopardize everyone's isolation. In appearance of the demanding risk, few researches have been proposed as a fix of this delicate case, which focus at realising the balance of data utility and statistics privacy when publishing dataset. The continuing research is called multiple numerical sensitive attribute. In the past few years, cognoscenti have taken up the challenge and attempted several of researches. Many possible methods are presented for different privacy preserving scenario, which solve the issues in PPDP effectively. New methods and theory come out continuously in experts' effort to complete privacy preserving. Therefore, some anonymity models have been proposed to protect individual's privacy for microdata publish recently. We are using the privacy in different sectors like bank sectors, health care centers and industry.

I.1 Privacy Preserving Data Publishing (PPDP)

Usually, the method of Privacy Preserving Data Publishing (PPDP) has two phases, data collection and data publish phase. It refers to three kinds of roles in the process who are data owner, data publisher and data recipient. The association of data collection and data publishing scenario in PPDP is shown in figure 1. In the data collection phase, data publisher collects dataset from data owner. Then, in the data publishing phase, data publisher sends the processed dataset to data recipient. It is necessary to mention that raw dataset from data owner cannot be immediately sent to data recipient. The dataset should be processed by data publisher before being sent to data recipient.

Fig. 1: The relationship of Data collection and Data publishing scenario in PPDP [30]

- Contribution

We introduce a survey of multiple numerical sensitive attribute for the published data in the various institutions, like hospitals, industry etc. This paper is organized as follow: Section 1 contains Interdiction, section 2 contains related work, section 3 contains some review of k-anonymity model for privacy protection, section 4 contains operation on anonymity, section 5 contains problem statement, section 6 contains proposed method and section 7 contains conclusion.

II. RELATED WORK

k-anonymity [1] is a simple and effective method to protect privacy in micro-data, which requires that each tuple has at least *k* indistinguishable tuples with respect to quasi-

identifier(QID) in the released data. But it cannot resist homogeneity attack and background knowledge attack, so some other enhanced anonymity models have been proposed, such as *l*-diversity [2] and *t*-closeness [3]. Several techniques have also been proposed to implement the above anonymity models. Generalization [4-5] is a typical one to implement anonymity model, whose idea is to replace real value of quasi-identifier with less specific but semantically reliable value. Generalization distorts real data, which is disadvantageous to data mining. Anatomy [6] is also a fine method to anonymize micro-data, whose thought is to issue all the quasi-identifier and sensitive attributes directly in two separate tables. However, releasing the QID-attribute directly may suffer from a higher breach probability than generalization, to overcome these drawbacks, Tao et al. [7] showed ANGEL, a new anonymization method that is as efficient as generalization in privacy protection, which can preserve better data utility. Leela et al. [8] reliable to preserve privacy in re-publication of dynamic micro-data after inclusions or deletions.

Slicing, that anonymizes micro-data to partitioning horizontally and vertically, which is presented by Li et al. [9]. Neha et al. [10] concluded that slicing preserves data utility better than generalization, in addition, it also prevents membership disclosure. All of above works target on micro-data with single sensitive attribute. This technique will lead to substantially low data utility when they are immediately used for micro-data with multiple sensitive attributes. Right now, there is only a small number of works concentrated on micro-data with multiple sensitive attributes. Yang et al. [11] suggested a **M**ultiple **S**ensitive **B**ucketization (**MSB**) method, but this method is only suitable to deal with micro-data with less sensitive attributes, and it is appropriate for attributes less than three only.

"Privacy-Preserving Data Publishing for Multiple Numerical Sensitive Attributes" [12] nearly every one of the privacy preserving model concentrate that there is one numeric sensitive attribute and many categorical sensitive attribute. There are many applications where multiple numerical sensitive attribute is applying the PPDM technique on them reveals sensitive information. This presented a MNSACM method that eliminates the threat of proximity breach to numerical sensitive data. This method is applicable to only once periodical of a static data set. Jianmin Han [13] presented "SLOMS: A Privacy Preserving Data Publishing Method for Multiple Sensitive Attributes Micro-data", this work shows that the multiple sensitive attributes into more than a few tables and bucketizes every sensitive attribute table to enforce *l*-diversity. Simultaneously, it generalizes the quasi-identifiers to invoke *k*-anonymity. This method only for accounting the analogues for each pair of sensitive attributes and avoids membership disclosure**.**

Toor [14] presented "An Advanced Clustering Algorithm (ACA) for Clustering Large Data Set to Achieve High Dimensionality" this work shows that the main idea of the algorithm is to split the data structures into different subset to keep the labels of the cluster. The distance of all the data objects to the nearest cluster center is calculated while each iteration that can be used in the next iteration. If the computed distance is lesser than or equal to the distance to the old center, then the data object remain in its own cluster that was assigned to it in the prior iteration. Shyamala Susan [15] proposed "Anatomisation with slicing: a new privacy preservation technique for multiple sensitive attributes" shows that the anatomization method minimizes the loss of information and slicing algorithm helps in the preservation of correlation and utility which in turn results in reducing the data dimensionality and information loss. But it is not sufficient for protect the privacy, publishes the QI values in their original forms thus it is easy to find out an individual's record in the published data.

"Secure Multiparty Privacy Preserving Data Aggregation by Modular Arithmetic" [16] Organization need to share their confidential data to third party for integration without revealing the private information. This paper presents a method for protecting privileged information and confidentiality. For this a scheme is develop for obtain multiparty data accumulation with the help of segmental arithmetic model. This method requires low computational time but it does not consider the communication overhead. "D-Mash: A Framework for Privacy-Preserving Data-as-a-Service Mashups" [17] Data-as-a-service mashups allow dynamically integration of data on demand by customers. The problem is that, it show sensitive information of customer when data from multiple parties are integrated. For this, propose a cloud based privacy preserving framework for Data-as-a-service mashup that enable secure collaboration between DAAS provider for generating anonymous dataset to support data mining. This method is not suitable for malicious adversarial model.

III. SOME REVIEW OF K- ANONYMITYMODEL FOR PRIVACY PROTECTION

Sweeney [1] presented *k*-anonymity technique which supposes that person certain data are stocked in a relation of attributes and records. Removing all the explicit identifiers (like, roll number, name) when the process of anonymizing micro-data table. It is becoming this every record in a micro-data table be identical from at least ($k-1$) other records in order to the pre-determined quasi-identifier. Table.1 shows the some review of k-anonymity technique for protecting the privacy.

Table. 1. Review of some PPDP

S. No	Author's	Approach
1.	Sweeny	k-anonymity
2.	R.C.W Wong J. Li et al.	l-deversity
3.	N. Li, T. Li	t-closenes
4.	N Li, T Li et al.	(∝,k)-anonymity
5.	Koudas, et al.	(k, e)-anonymity
6.	T.M. Truta	P sensitive k anonymity
7.	X. Xiao and Y.Tao	m-invariance

Definition 1. Quasi-identifier(QID)

Quasi-identifier of table T, signified as QIT, is a set of attributes in T that can be hypothetically used to relation a record in T to an actual similarity with a substantial feasibility. Furthermore to the quasi-identifier, the table may consist of

publicly anonymous attributes some of which are extremely sensitive. In preference to, a table usually contain three types of attributes: publicly known attributes (i.e., quasi-identifier), sensitive and publicly anonymous (simply, sensitive) attributes, and non-sensitive and publicly anonymous (simply, non-sensitive) attributes. For example, attribute set {Gender, Age, Post-code} in Table 2[18] is a quasi-identifier. Table 2 potentially reveals private information of patients (e.g. young patients with stress and obesity). If the table is joined with other tables, it may reveal more information of patient's disease history. Normally, a quasi-identifier attribute set is understood by domain experts.

Table 2. Qusi-Identifier

QID

Gender	Age	Pcode	Problem
male	middle	4350	stress
male	middle	4350	obesity
male	young	**4351**	stress
female	young	**4352**	obesity
female	old	4353	stress
female	old	4353	obesity

Definition 2. K-Anonymity

Let $T(AT_1,..., AT_n)$ obtain a table and *QIT* be the quasi-identifier associated with it. T is said to assure *k*-anonymity if and only if every continuousness of attributes in T[*QIT*] occur with at least *k* occurrences in T[*QIT*]. The *k*-anonymity method has been significantly considered in recent times as a feasible description of privacy in data publishing because of its relative abstract straightforwardness and shortcomings of *k*-anonymity model as we have [31] *a)* It can't resist a kind of attack, which is assuming that the attacker has background knowledge to rule out some possible values in a sensitive attribute for the targeted victim. That is, *k*-anonymity does not guarantee privacy against attackers using background knowledge. It is also susceptible to homogeneity attack. An attacker can discover the values of sensitive attributes when there is little diversity in those sensitive attributes [2]. Thus some stronger definitions of privacy are generated, such as ℓ -Diversity. *b)* It protects identification information. However, it does not protect sensitive relationships in a data set [19]. *c)* Although the existing *k*-anonymity property protects against identity disclosure, it fails to protect against attribute disclosure [3]. *d)* It is suitable only for categorical sensitive attributes. However, if we apply them directly to numerical sensitive attributes (e.g., salary) may result in undesirable information leakage [20]. *e)* It does not take into account personal anonymity requirements and a *k*-anonymity table may lose considerable information from the microdata which is a valuable source of information for the allocation of public funds, medical research, and trend analysis [21].

- ***ℓ-diversity***

ℓ-diversity [2] provides privacy preserving even when the data publisher does not know what kind of knowledge is possessed by the adversary. The main idea of -diversity is the requirement that the values of the sensitive attributes are well-represented in each group. -Diversity resolved the shortcoming 1 of *k*-anonymity model.

2) t-Closeness [3]

ℓ-diversity [2] principle represents an important step beyond *k*-anonymity in protecting against attribute disclosure. However, it has several shortcomings, such as -diversity may be difficult and unnecessary to achieve and it is insufficient to prevent attribute disclosure. So the paper [3] proposes a new privacy measure is called *t*-Closeness, which requires that the distribution of a sensitive attribute in any equivalence class is close to the distribution of the attribute in the overall table.

- ***(α, k)anonymity***

WONG R C et al. [19] propose an (α, *k*)-anonymity model to protect both identifications and relationships to sensitive information in data and to limit the confidence of the implications from the quasi-identifier to a sensitive value (attribute) to within α. The model avoids the sensitive information is inferred by strong implications.

Definition 3. (α, k)-anonymization [19]:

A view of a table is said to be an (α, *k*)-anonymization of the table if the view modifies the table such that the view satisfies both *k*-anonymity and α-deassociation properties with respect to the quasi-identifier [6].

- ***(k, e)-Anonymity:***

A general framework called (*k*, *e*)-Anonymity is presented by Qing Zhang et al [20]. in order to better capture the need of privacy protection for numerical sensitive attributes and support accurate answering of aggregate queries. It resolves the fault 4 of *k*-anonymity model.

Definition 4. (k, e)-anonymity

A de-identified microdata database D satisfies (*k*, *e*)-anonymity if given D and any given public database P, any association cover that an attacker can derive satisfies:

- The size of the association cover is no less.
- The range of the sensitive attribute values in the association cover is no less than *e* [20]

- ***p-sensitive k-anonymity:***

Traian Marius, et.al [20] introduce *p*-sensitive *k*-anonymity that and protects against both identity and attribute disclosure on the base of extending *k*-anonymity model.

Definition 5. p-sensitive k-anonymity property:

The masked microdata satisfies p-sensitive *k*-anonymity property if it satisfies *k*-anonymity, and for each group of tuples with the identical combination of key attribute values that exists in masked microdata, the number of distinct values for each confidential attribute occurs at least p times within the same group [20]. *P*-sensitive *k*-anonymity protects against attribute disclosure. On this aspect, it is the same with *t*-Closeness.

- ***m-invariance:***

Xiaokui Xiao and Yufei Tao [23] develop a generalization principle m-invariance that effectively limits the risk of privacy disclosure in re-publication. The core of method is that a tuple appears in the microdata at both publication timestamps, two equivalent groups contain the same sensitive

values and certain invariance in all the QI groups that a tuple is generalized to in different snapshots.

IV. OPERATIONS OF ANONYMITY

A) Generalization:

After removing the identifiers from the data it partitions tuples into buckets. Thus a QID values in each bucket are less specific but semantically consistent values so that tuples in the same bucket cannot be distinguished by their QID values. Generalization for k-anonymity losses substantial amount of information. This is due to three reasons. They are Records in the same bucket must be close to each other so that generalizing the records would not loss to much information while in the high dimensional data most data points have similar distance with each other.

Thus the generalization to satisfy k-anonymity for small k. In a generalized data the data utility can be reduced due to the uniform distribution assumption make in the every value of the generalized table. In generalization each attribute is generalized separately thus correlation between different attributes are lost. This will leads to the inherent problem of generalization that prevent the effective analysis of attribute correlations. There are roughly four types of generalization with difference in scope and principle which are full-domain generalization, subtree generalization, cell generalization and multidimensional generalization. 1)**Full-domain generalization** [24] is proposed in early research of PPDP, it has the smallest search space in four types of generalization, while it leads to large data distortion. The key of full-domain generalization is that the value of quasi-identifier must be generalized to the same level. **Subtree generalization** [25, 26], its boundary is smaller than full-domain generalization. When a node in taxonomy tree structure generalizes to its parent node, all child nodes of the parent node need to be generalized to the parent node. **Multidimensional generalization** [27, 28,29] emphasizes different generalization for different combination of values of quasi-identifiers.

B) Bucketization [13]:

In Bucketization SAs are separated from the QIs by doing the random permutation on the SA values in each bucket. Thus an anonymized data consist of a set of buckets with a permuted sensitive attribute values. Bucketization does not prevent membership disclosure because it publishes the QI values in their original forms thus it is easy to find out an individual's record in the published data. Bucketization requires the clear separation of SA and QI attributes and it breaks the attribute correlation between them.

C) Suppression:

Suppression [30] is used to prevent the membership disclosure in the k-anonymity thus it be an assignment technique of placing *, # and & for the attribute values instead of their original values. This suppression technique is mostly used in the quasi identifier fields to preserve the individuals. If we use the suppression technique in the sensitive attribute field then its leads to the loss in data utility. Thus if we use in the quasi identifier then there is no loss in the data utility.

V. 1 PRELIMINARIES

Supposed multiple organizations want to integrate their data. They have the data table in the form of T (x_1, x_2, x_3,…, x_n, sa_1, sa_2, sa_3,…., sa_m) on integration they want to release the integrated data to the public categorized in following ways.

- ***Explicit Identifier:***

It is a set of attributes, such as name etc. containing information that explicitly identifies record owners. These attributes are removed before data release.

- ***Quasi-identifier (QID):***

A quasi-identifier of table, denoted as QT , is a set of attributes in T that can be potentially used to link a record in T to a real-world identity with a significant probability.

- ***Multiple Sensitive Attribute:***

Given a multiple sensitive attributes table *T*, if all of sensitive attributes in *T* satisfy single sensitive attribute *l*-diversity, then *T* satisfies multiple sensitive attributes *l*-diversity.

- ***Advanced Clustering Algorithm (ACA):***

Amanpreet Kaur Toor [14] presented the advanced clustering algorithm, this algorithm to overcome the shortcomings of the SOM algorithm, this paper presents an Advanced Clustering Algorithm method. The main idea of the algorithm is to split the data structures into different subset to keep the labels of the cluster. The distance of all the data objects to the nearest cluster centre is calculated during each iteration that can be used in the next iteration. If the computed distance is smaller than or equal to the distance to the old center, then the data object remain in its own cluster that was assigned to it in the prior iteration. Therefore, there is no need to calculate the distance from the data object to the other k-1 clustering centers. By this methodology ACA saves the computational time to the k-1 cluster centers. Otherwise, we must calculate the distance from the current data object to all k cluster centers and find the nearest cluster center. It assigns this point to the adjacent cluster center and then individually record the distance to its cluster center. Because in each iteration some data points still remain in the original cluster. It means that some parts of the data points will not be calculated and saving total time of calculating the distance. So ACA is enhancing the efficiency of the clustering.

VI. PROPOSED METHOD

In our proposed work new enhanced slicing algorithm obtain protecting the privacy vertical and horizontal partitioning for multiple numerical and categorical sensitive attribute . As this work focuses on QID and multiple sensitive attribute, first of all generalize the QID attribute table to 3-anonymity, and then slice the sensitive attribute table to 3-anonymity. For this we use differential privacy model, finally, publish the multi set anomynity table.

For this work we use advanced clustering algorithm, this algorithm is to split the data structures into different subset to keep the labels of the cluster. The distance of all the data objects to the nearest cluster center is calculated during each iteration that can be used in the next iteration. If the computed distance is smaller than or equal to the distance to the old center, then the data object remain in its own cluster that was assigned to it in the prior iteration.

- **New enhanced slicing algorithm**

The slicing algorithm achieves preservation of privacy through horizontal and vertical partitioning. As this work focuses on multiple SA, SA that are related are grouped together based on their correlation. At that juncture SA are sufficiently clustered and results in different tables of SA making use of advanced clustering algorithm. And in the subsequent phase tuples are partitioned horizontally by means of MFA and *l*-diversity is checked in for each sensitive tuple. Every ST inserts the correlated attributes along with its group membership within a new column group ID. In the same way, the partitioning of QID is done under k-anonymity and then the new QIT contains all of its exact QID values along with its group membership within a new column group ID. The partitioning technique removes the dimensionality of the data that ensures this work to be able to deal with any number of sensitive attributes. Finally, the SA in each group is shuffled and thereafter linked with a common group id, in such a manner that the sensitive value corresponding to an individual can be found by an intruder with the probability of at the most 1/*l*. A larger *l* leads to a much stronger privacy.

VII. CONCLUSION

Today information is shared by the almost all the companies, organizations. Privacy preserving data publishing is used for this information sharing, while preserving the private information of the individual. In this survey, we reviewed various privacy preserving models and proposed a method by which multiple numerical and categorical sensitive attribute can aggregate their data. The important goal of this work is to preserve the privacy of the multiple numerical and categorical sensitive attribute and new enhanced slicing method helps in preserving membership disclosure, suppression ratio and information loss. It reduces the time further through increased processor speed and memory.

VIII. REFERENCES

[1] L. Sweeney, "*k*-anonymity: a model for protecting privacy", International Journal on Uncertainty, Fuzziness and Knowledge based Systems, 2002, pp. 557-570.

[2] A.Machanavajjhala, J.Gehrke, and D.Kifer, et ℓal, " -diversity: Privacy beyond *k*-anonymity", In Proc. of ICDE, Apr.2006.

[3] N. Li, T. Li, and S. Venkatasubramanian, "t-Closeness: Privacy Beyond *k*-anonymity and l-Diversity", In Proc. of ICDE, 2007, pp. 106-115.

[4] Samarati P, Sweeney L. Generalizing data to provide anonymity when disclosing information. *In: Proc. of the 17th ACM-SIGMOD-SIGACT-SIGART Symposium on the Principles of Database Systems*. 1998, pp. 188.

[5] Samarati P, Sweeney L. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. *Tech. rep., SRI International*. March, 1998.

[6] Xiao X, Tao Y. Anatomy: Simple and effective privacy preservation. In: Proc. Of the 32nd International Conference on Very Large Data Bases. Seoul: VLDB Endowment, 2006, pp. 139−150.

[7] Yufei Tao, Hekang Chen, Xiaokui Xiao, Shuigeng Zhou, Member, IEEE Computer Society, and Donghui Zhang. ANGEL: Enhancing the Utility of Generalization for Privacy Preserving Publication. *IEEE Transaction on Knowledge and Data Engineering.2009*, vol. 21.No.7. pp.1073-1087.

[8] Generalization for Privacy Preserving Re-publication of Dynamic Datasets. *International Journal of Computer Applications*, vol. 13, issue 6, 2011, pp. 42-49.

[9] Tiancheng Li, Ninghui Li, Jia Zhang, Ian Molloy. Slicing: A New Approach for Privacy Preserving Data Publishing. *Proc. IEEE Transactions on Knowledge and Data Engineering*, vol.24, No. 3, March 2012.

[10] Neha V. Mogre, Girish Agarwal, Pragati Patil. A Review On Data Anonymization Technique For Data Publishing[J]. *International Journal of Engineering Research & Technology*, vol. 1, issue 10, December 2012.

[11] Yang Xiao-chun, Wang Ya-zhe, Wang Bin. Privacy Preserving Approaches for Multiple Sensitive Attributes in Data Publishing [J]. *Chinese journal of computers*, 2008, 31(04), pp. 574-587.

[12] Qinghai Liu, Hong Shen et al.,*"Privacy-Preserving Data Publishing for Multiple Numerical Sensitive Attributes"* IEEE conference on tsinghua science and technology Volume 20, Number 3, June 2015.

[13] J. Han, F. Luo, J. Lu, and H. Peng, *"SLOMS: A Privacy Preserving Data Publishing Method for Multiple Sensitive Attributes Microdata"*. JOURNAL OF SOFTWARE, VOL. 8, NO. 12, DECEMBER 2013, 8 (12): 3096-3104.

[14] Amanpreet Kaur Toor, Amarpreet Singh, "An Advanced Clustering Algorithm (ACA) for Clustering Large Data Set to Achieve High Dimensionality" , *International Journal of Applied Information Systems (IJAIS) – ISSN : 2249-0868 Foundation of Computer Science FCS, New York, USA Volume 7– No. 2, April 2014.*

[15] Susan, V. Shyamala, and T. Christopher. "Anatomisation with slicing: a new privacy preservation approach for multiple sensitive attributes." *SpringerPlus* 5.1 (2016): 1-21.

[16] Arijit Ukil, Jaydip Sen proposed *"Secure Multiparty Privacy Preserving Data Aggregation by Modular Arithmetic"* 1st International Conference on Parallel, Distributed and Grid Computing (PDGC - 2010).

[17] Mahtab Arafati et al. proposed *"D-Mashup:A Framework for Privacy-preserving Data-as-a-Service mashups"* IEEE 7th International Conference on cloud.

[18] Jiuyong Li.et al, "Achieving *k*-Anonymity by Clustering in Attribute Hierarchical Structures**"** *DaWak. LNCS 4081, Springerverlag, Berlin, Heidelberg*, 2006, pp. 405-416.

[19] WONG R C, LI J, FU A W, et a1, "(α, k)-anonymity : an enhanced *k*-anonymity model for privacy-preserving data publishing", Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM Press, New York, 2006, pp. 754-759.

[20] Qing Zhang, Kaunda's N, et al, "Aggregate query answering on anonymized tables", In Proc. of ICDE, April, 2007, pp. 116-125.

[21] Xiao X, Tao Y, "Personalized privacy preservation", Proceedings of ACM Conference on management of Data (SIGMOD). ACM Press, New York: 2006, pp. 785-790.

[22] TRUTA T M , VINAY B, "Privacy protection: *p*-Sensitive *k*-anonymity property," Proceedings of the 22nd on Data Engineering Workshops, IEEE Computer Society, Washington Dc, 2006.

[23] Xiao X, Tao Y, "M-invariance: towards privacy preserving republication of dynamic datasets", In Proc. of SIGMOD, ACM Press, New York, 2007, pp. 689-700.

[24] P. Samurai, Protecting respondents identities in microdata release, IEEE Trans. On Knowledge and Data Engineering, **13**, (2001).

[25] V. Iyengar, Transforming data to satisfy privacy constraints, In ACM SIGKDD, (2002).

[26] Xuyun Zhang, Chang Liu, Surya Nepal, Jinjun Chen, An efficient quasi-identifier index based approach for privacy preservation over incremental data sets on cloud, Journal of Computer and System Sciences, 542-555 (2013).

[27] Lefevre K., Dewitt D. J., Ramakrishnan R., Incognito: efficient full-domain k-annymity, In Proceedings of ACM SIGMOD, 49-60 (2005).

[28] Xu J.,WangW., Pei J.,Wang X., Shi B., Fu A.W. C., Utilitybased anonymization using local recoding, In Proceedings of the 12th ACM SIGKDD Conference, (2006).

[29] Lefevre K., Dewitt D. J., Ramakrishnan R, Mondrian multidimensional k-anonymity, In Proceedings of the 22nd IEEE International Conference on Data Engineering (ICDE), (2006).

[30] Yang Xu, Tinghuai Ma "A Survey of Privacy Preserving Data Publishing using Generalization and Suppression" Appl. Math. Inf. Sci. 8, No. 3,1103-1116 (2014).

[31] Yan Zhao, et. al, "A Survey on Privacy Preserving Approaches in Data Publishing" 2009 First International Workshop on Database Technology and Applications.