Multiple-Published Tables Privacy-Preserving Data Mining: A Survey for Multiple-Published Tables Techniques

Abou_el_ela Abdo Hussein
Department of Computer Science
Faculty of Science and Arts,
Shaqra University
Shaqra, KSA

Nagy Ramadan Darwish
Sciences, Institute of Statistical
Studies and Research, Cairo
University,
Cairo, Egypt

Hesham A. Hefny
Department of Computer and
Information Sciences, Institute of
Statistical Studies and Research,
Cairo University,
Cairo, Egypt

***Abstract*—With large growth in technology, reduced cost of storage media and networking enabled the organizations to collect very large volume of information from huge sources. Different data mining techniques are applied on such huge data to extract useful and relevant knowledge. The disclosure of sensitive data to unauthorized parties is a critical issue for organizations which could be most critical problem of data mining. So Privacy preserving data mining (PPDM) has become increasingly popular because it solves this problem and allows sharing of privacy sensitive data for analytical purposes. A lot of privacy techniques were developed based on the k-anonymity property. Because of a lot of shortcomings of the k-anonymity model, other privacy models were introduced. Most of these techniques release one table for research public after they applied on original tables. In this paper the researchers introduce techniques which publish more than one table for organizations preserving individual's privacy. One of this is (α, k) – anonymity using lossy-Join which releases two tables for publishing in such a way that the privacy protection for (α, k)-anonymity can be achieved with less distortion, and the other one is Anatomy technique which releases all the quasi-identifier and sensitive values directly in two separate tables, met l-diversity privacy requirements, without any modification in the original table.**

Keywords—Data mining; privacy; sensitive attribute; quasi-identifier; Anatomy

I. Introduction

One of the most important aspects of data applications is data mining. Data mining technique intelligently and automatically extracts information or knowledge from a very large volume of data.

One of the disadvantages of data mining is the disclosure of sensitive individual data to unauthorized parties which are a critical issue for organizations. So Privacy Preserving Data Mining (PPDM) is playing very important role in both applications and research; it publishes much more accurate data while maintaining privacy information. Each record in released data corresponding to one individual and has a number of attributes, which can be divided into three categories:

*1) **Identity attributes** (e.g., SSN and Name) whose values can uniquely identify an individual;*

*2) **Quasi-identifier** (QI-group) attributes (e.g., age, Zip code and gender) whose values can potentially identify an individual;*

*3) **Sensitive attributes** (e.g., income and disease) which indicate confidential and sensitive information of individuals.*

Several Privacy-Preserving data mining techniques have been published most of them depending on k-anonymity. *The anonymization techniques* (e.g. *k-anonymity)* aim at using techniques of generalization and suppression to make the individual record indistinguishable from a group of records. The motivating factor behind the k-anonymity approach is that many attributes in the data can often be considered quasi-identifiers that are used with public records to uniquely identify the records. Because of k-anonymity has some shortcomings, many advanced methods have been proposed, such as *p*-sensitive k-anonymity, (*α, k*)-anonymity, *l*-diversity, *t*-closeness, *M*-invariance, Personalized anonymity, and so on. Although the anonymization method can ensure that the transformed data is true, it also results in information loss to some extent [1]. Also, there is a technique called Anatomy technique that releases all the quasi-identifier and sensitive values directly in two separate tables. In this paper, researchers focus only on those techniques that publish more than one table for the purposes of data mining. In next section, researchers introduce k-anonymity technique and both generalization and suppression concepts. In section three both multiple-published tables techniques, Anatomy and (α, k) – anonymity using lossy-Join ending with a comparison between them are introduced, and last section introduces paper conclusion.

II. Related Research Areas

Numerous algorithms have been proposed for implementing k-anonymity via generalization and suppression. First the researchers introduce K-anonymity Technique proposed by L. Sweeney in next sub-section, then generalization and suppression concepts are introduced in last sub-section.

A. K-anonymity Technique

K-anonymity classified the attributes into three classes as mentioned before [2]. Table I. introduces the three classes of attributes where, ***Identity attributes*** (e.g., Name), ***Quasi-***

identifier (*QI-group*) attributes (e.g., gender, age and Zip code), ***Sensitive attributes*** (e.g., Diagnosis). K-anonymity technique anonymizes *QI-group* to prevent the attacker using link attack to infer the privacy of individuals. *Quazi-identifiers* can be used to re-identify individual using linking attack as given in below example.

TABLE I. CLASSIFICATION OF ATTRIBUTES FOR K-ANONYMITY

Identifier attribute	Quasi-identifier			Sensitive attributes
Name	Gender	Age	Zip code	Diagnosis
Ali	Male	25	423101	Depression
Mohsen	Male	27	423508	HIV

The two tables, Table II. contains Medical data set and Table III. Contains voter list which are available publically. To avoid the identification of records in microdata, the traditional approach is to de-identify records by removing the ***identity attribute*** (e.g., Name). But removing the ***identity attribute*** does not solve the problem because by linking Zipcode, Age and Sex of medical table (Table II.) with voter list table (Table III.) intruder can disclose that Omar is sick with cancer and in this way the privacy of individual is disclosed. This is happened because the combination of quazi-identifiers value is unique in medical data set, if published data in such a way that there is no unique combination for quazi-identifiers then this type of re-identification cannot occurs. This can be done using anonymizing tables.

TABLE II. MEDICAL DATA SET

ID	Zip code	AGE	SEX	DIAGNOSIS
1	423065	29	M	Heart Disease
2	422036	32	F	Flu
3	423245	38	M	Cancer
4	422035	37	F	HIV
5	423012	47	M	Headache
6	423432	53	F	Viral

Sweeney [1] proposed the k-anonymity model in order to prevent linking attacks using quasi-identifiers, where some of the *QI* fields are generalized or suppressed. A table is said to satisfy k-anonymity if every record in the table is indistinguishable from at least k-1 other records to every set of quasi-identifier attributes. The table is called a k-anonymous table if, for every combination of attributes of the *QI*s, there are at least k records that share those values. This ensures that individuals cannot be uniquely identified using linking attacks. Table IV. shows a 2-anonymous view corresponding to Table II. The sensitive attributes (Diagnosis Result) is stayed without change in this example.

TABLE III. VOTER LIST

NAME	Zipcode	AGE	SEX
Mohamed	423234	49	M
Ahmed	466987	35	M
Ali	423223	28	M
Rawan	424435	41	F
Omar	423245	38	M
Iman	423446	33	F

TABLE IV. 2-ANONYMOUS VIEW OF TABLE II.

ID	Zipcode	AGE	SEX	DIAGNOSIS
1	423***	>25	M	Heart Disease
2	423***	>25	M	Cancer
3	422***	3*	F	Flu
4	422***	3*	F	HIV
5	423***	>40	*	Headache
6	423***	>40	*	Viral

Numerous techniques implementing k-anonymity have been proposed using generalization and suppression [3]. Generalization involves modifying (or recoding) a value with a less specific but semantically consistent value. Suppression involves not publishing a value at all. An algorithm that exploits a binary search on the domain generalization hierarchy to find minimal k-anonymous table have been proposed by Samarati [4]. A. Machanavajjhala [5] proposed l-diversity technique in 2006 to solve k-anonymity problem. It tries to put constraints on minimum number of distinct sensitive values seen within an equivalence class , T-closeness technique present by S. Venkatasubramanian in 2007 [6] to overcome attacks possible on l-diversity like similarity attack[7], Bayardo and Agrawal [8] presented technique that starts from a fully generalized table and specializes the dataset in a minimal k- anonymous table. R. Wong, J. Li, A. Fu, K. Wang [9] proposed an (α, k)-anonymity technique to protect both identifications and relationships to sensitive information in data in the literature in order to deal with the problem of k-anonymity. Fung et al. [10] presented a top-down approach to make a table satisfied k-anonymous. LeFevre et al [11] introduces technique that uses a bottom-up technique. Pei [12] discusses the approaches for multiple constraints and incremental updates in k-anonymity. However the traditional k-anonymity techniques take consider that the all values of the sensitive attributes are sensitive and need to be protected. The previous models lead to excessively generalize and more information loss in publishing data.

B. Generalization and Suppression

Generalizing an attribute is a simple concept idea. A value is replaced by a less specific, more general value that is faithful to the original [1, 13, 14, and 15]. Generalization involves replacing (or recoding) a value with a less specific but semantically consistent value. Generalization could be achieved through global recoding or local recoding. In global recoding, the domain of the quasi identifier values are mapped to generalized values for achieving k-anonymity, which means that all k-tuples have the same generalized attribute value.

In local recoding generalization scheme, any two or more regions can be merged as long as the aggregated attribute value such as satisfies the anonymity requirement, which means that each k-tuple could have its own generalization attribute value. The limitation of the global recoding is; the domain values are over generalized resulting in utility loss where as in local recoding, the individual tuple is mapped to a generalized tuple.

The information loss of the global recoding is more than the local recoding approach. Comparison between global and local recoding is in table V.

TABLE V. COMPARISON BETWEEN GLOBAL AND LOCAL RECODING

Generalization method	**Global Recoding**	**Local Recoding**
Generalized value	The same value	Different values
Information loss	More information loss	Less information loss
Utility level	More utility loss	Less utility loss
Domain values	Over generalized	Suitable generalization
Generalization kind	Global generalized	Local generalized
Privacy level	High level of privacy	Lower level of privacy
Generalization level	Higher generalization level	Minimum generalization level

While Generalization replaces the actual *QI* values with more general ones (e.g., replaces the city name with the state name); Suppression involves not releasing a value at all. Suppression is the most common practice in related works on such data. There is a generalization hierarchy (e.g., city name→ state name → country name). On the other hand Suppression excludes some *QI* attributes or entire records (known as outliers) from the microdata. Comparison between generalization and suppression in table VI.

TABLE VI. COMPARISON BETWEEN GENERALIZATION AND SUPPRESSION

Method	Generalization	Suppression
Generalized value	Releases general value	Not releasing value at all
Information loss	Less information loss	More information loss
Utility level	Less utility loss	More utility loss
Privacy level	Lower privacy level	High privacy level
Common method	Less common	More common practice

Both Generalization and Suppression Architecture could be explained by figure 1.

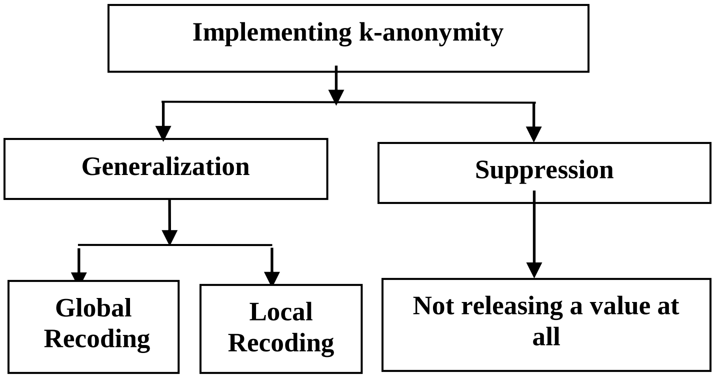

Fig. 1. generalization and suppression

III. MULTIPLE-PUBLISHED TABLES

In this section the researchers introduce techniques which publish more than one table for organizations preserving individual's privacy. One of this is (α, k) – anonymity using lossy-Join which releases two tables for publishing and the other one is Anatomy technique which releases all the quasi-identifier and sensitive values directly in two separate tables. Next subsections introduce these two techniques in details.

A. *(α, k) – anonymity using lossy-Join*

The Lossy Join Approach

Lossy join of multiple tables is useful in privacy-preserving data publishing [9]. The mean idea is that if two tables with a join attribute are released, the join of the two tables can be lossy and this lossy join helps to maintain the private information. In this paper, authors use the idea of lossy join to derive a new technique for achieving privacy preservation purpose. Let us see Table VII. an (0.5, 2) - anonymization. From this table, temp table could be generated as shown in Table VIII.

For each equivalence class *E* in the anonymized table, there is a unique identifier (*ID*) to *E* and also to all tuples in *E*. Then, the correspondence *ID* to each record in the original raw table could be attached and form a new table called Temp. From the Temp table, two separate tables could be generated, Tables IX.(a) and IX.(b). The two tables share the attribute of *ClassID*. If these two tables are joined using the *ClassID*, the join is lossy and it is not possible to obtain the table Temp after the join. The resulted table is given in Table X.

From the lossy join, each individual is linked to at least 2 values in the sensitive attribute. Therefore, the required privacy of individual can be maintained.

Also, in the joined table, for each individual, there are at least 2 persons that are linked to the same bag *B* of sensitive values, so they are not distinguishable.

For example, the first record in the raw table (*QID* = (clerk, 1975, 4350)) is linked to bag {*HIV*, *flu*}. The second record (*QID* = (manager, 1955, 4350)) is also linked to the same bag *B* of sensitive values. This is the goal of k-anonymity for the protection of sensitive values.

TABLE VII. AN (0.5, 2)-ANONYMOUS DATA SET

Job	**Birth**	**Post Code**	**Illness**
Clerk	1975	4350	HIV
manager	1955	4350	flu
clerk	1955	5432	flu
factory worker	1955	5432	fever
factory worker	1975	4350	flu
technical supporter	1940	4350	fever

TABLE VIII. TEMP TABLE

Job	**Birth**	**Post Code**	**Illness**	**ClassID**
Clerk	1975	4350	HIV	1
manager	1955	4350	flu	1
clerk	1955	5432	flu	2
factory worker	1955	5432	fever	2
factory worker	1975	4350	flu	3
technical supporter	1940	4350	fever	3

B. *Anatomy*

Anatomy releases two different tables *QI* (*Quisi-identifier)* attributes table and *SI* (*Sensitive)* attributes table instead of publishing one single table with the generalized values. Anatomy [16] releases all *QIs* and *SI* directly in two separate tables, which met *L*-diversity privacy requirement, so there is no need to modify the original table. Anatomy avoids the drawbacks of generalization as in next example. Assume that hospital intents to publish patients' medical records as in Table XI., referred to as the microdata.

The sensitive Attribute is Disease, so the hospital must ensure that no intruder can correctly infer any patient disease with confidence. Age, Sex, and Zipcode are the quasi-identifier (*QI*) attributes, which could be utilized in combination to infer the identity of an individual, which disclose privacy.

TABLE IX. (A): NSS TABLE

Job	**Birth**	**Post Code**	**ClassID**
Clerk	1975	4350	1
manager	1955	4350	1
clerk	1955	5432	2
factory worker	1955	5432	2
factory worker	1975	4350	3
technical supporter	1940	4350	3

(B): SS TABLE

ClassID	**Illness**
1	HIV
1	flu
2	flu
2	fever
3	flu
3	fever

Consider an intruder who has the personal details (i.e., age 25 and Zipcode 11500) of Ali, and knows that Ali has been hospitalized before. In Table XI., since only record 1 matches Ali's QI-values, the adversary knows that Ali has pneumonia. To avoid this problem, generalization [4, 17, 18, and 5] divides records into *QI-groups*, and transforms their *QI-values* into less specific forms, so that records in the same *QI-group* cannot be distinguished by their *QI-values*. Table XII. is a generalized version of Table XI. (e.g., the age 25 and Zipcode 11500 of record 1 have been replaced with intervals [19, 20] and [10001, 60000], respectively). Here, generalization produces two QI-groups, including records 1-4 and 5-8, respectively. As a result, even if an intruder has the exact QI values of Ali, s/he still does not know which record in the first QI-group belongs to Ali.

Two notions, k-anonymity and l-diversity, have been proposed to measure the degree of privacy preservation. A (generalized) table is k-anonymous [4, 17, 18] if each QI-group involves at least k records (e.g., Table XII. is 4-anonymous). However, even with a large k as shown in *l-diversity* [5], k-anonymity may still allow an intruder to infer the sensitive value of an individual with high confidence. So, *l-diversity in* [5] provides stronger privacy preservation.

TABLE X. SS TABLE

Job	Birth	Post Code	Illness	ClassID
Clerk	1975	4350	HIV	1
manager	1955	4350	HIV	1
Clerk	1975	4350	flu	1
manager	1955	4350	flu	1
clerk	1955	5432	flu	2
factory worker	1955	5432	flu	2
clerk	1955	5432	fever	2
factory worker	1955	5432	fever	2
factory worker	1975	4350	flu	3
technical supporter	1940	4350	flu	3
factory worker	1975	4350	fever	3
technical supporter	1940	4350	fever	3

TABLE XI. THE MICRODATA

Tuple ID	Age	Sex	Zipcode	Disease
1(Ali)	25	M	11500	pneumonia
2	29	M	13200	dyspepsia
3	33	M	59300	dyspepsia
4	55	M	12700	pneumonia
5	60	F	54600	flu
6	59	F	25200	gastritis
7(Hoda)	60	F	25100	flu
8	58	F	31000	bronchitis

Specifically, a table is ***l***-diverse if, in each *QI-group*, at most 1/***l*** of the records possesses the most frequent sensitive value1. For instance, Table XII. is 2-diverse because, in each *QI-group*, at most 50% of the records have the same value of Disease. As mentioned earlier, the intruder (targeting Ali's medical record) knows that Ali's record must be in the first

QI-group, where two records are associated with pneumonia, and two with dyspepsia. Hence, the adversary can only make a probabilistic conjecture: Ali could have either disease with the same probability.

TABLE XII. A 2-DIVERSE TABLE

Tuple ID	Age	Sex	Zipcode	Disease
1	[21, 60]	M	[10001, 60000]	pneumonia
2	[21, 60]	M	[10001, 60000]	dyspepsia
3	[21, 60]	M	[10001, 60000]	dyspepsia
4	[21, 60]	M	[10001, 60000]	pneumonia
5	[21, 60]	F	[10001, 60000]	flu
6	[21, 60]	F	[10001, 60000]	gastritis
7	[21, 60]	F	[10001, 60000]	flu
8	[21, 60]	F	[10001, 60000]	bronchitis

Anatomy technique has been proposed to overcome the disadvantages of generalization which often losses considerable information in the microdata. Anatomy captures the exact *QI*-distribution and releases two tables, a quasi-identifier table (*QIT*) and a sensitive table (*ST*), which separate *QI*-values from sensitive values. For example, Tables XIII.(a) and XIII.(b) demonstrate the *QIT* and *ST* obtained from the microdata Table XI., respectively [16].

First, the microdata partitioned the records into different QI-groups, based on a certain strategy. Here, following the grouping in Table XII., records 1-4 into *QI*-group number 1and records 5-8 into QI-group number 2 of Table XI.

Second, the quasi-identifier table (*QIT*) has been created. Specifically, for each record in Table XI., the *QIT* (Table XIII.(a) includes all its exact *QI-values*, together with its group membership in a new column Group-ID. However, *QIT* does not have any Disease value.

Finally, it is possible saying that ST (Table XIII.(b) maintains the Disease statistics of each *QI*-group.

Anatomy preserves privacy because the *QIT* does not indicate the sensitive value of any record, which must be randomly guessed from the ST. To explain this, consider again the adversary who has the age 25 and Zip code 11500 of Ali. Hence, from the *QIT* (Table XIII.(a), the adversary knows that record 1 belongs to Ali, but does not obtain any information about his disease so far. Instead, s/he gets the id 1 of the QI-group containing record 1. Judging from the ST (Table XIII.(b), the adversary realizes that, among the 4 records in QI-group 1, 50% of them are associated with pneumonia (or dyspepsia) in the micro data. Note that s/he does not gain any additional information, regarding the exact diseases carried by these records. Hence, s/he could only expect that Ali could have contracted pneumonia (or dyspepsia) with 50% probability.

TABLE XIII. THE ANATOMIZED TABLES

(a) The quasi-identifier table (QIT)

row #	Age	Sex	Zipcode	Group-ID
1(Ali)	25	M	11500	1
2	29	M	13200	1
3	33	M	59300	1
4	55	M	12700	1
5	60	F	54600	2
6	59	F	25200	2
7(Hoda)	60	F	25100	2
8	58	F	31000	2

(b) The sensitive table (ST)

Group-ID	Disease	Count
1	Dyspepsia	2
1	Pneumonia	2
2	Bronchitis	1
2	Flu	2
2	Gastritis	1

Researchers introduce Comparison between Anatomy [16] and (α, k) – anonymity using lossy-Join [9] in table XIV.

TABLE XIV. COMPARISON BETWEEN ANATOMY AND (Α, K) – ANONYMITY USING LOSSY-JOIN

Technique	**Anatomy**	**(α, k) – anonymity using lossy-Join**
No. of tables	Two tables	Two tables
l diverse	Achieve *l* diversity	Achieve *l* diversity
Information Loss	No Information Loss	There is Information Loss
Data Utility	More Data Utility	Less Data Utility

The Architecture for both Multi-published tables' techniques is represented in figure 2.

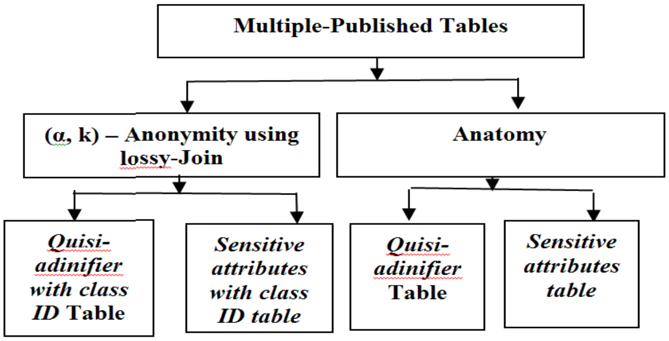
Multiple-Published Tables
(α, k) – Anonymity using lossy-Join
Anatomy
Quisi-adinifier with class ID Table
Sensitive attributes with class ID table
Quisi-adinifier Table
Sensitive attributes table

Fig. 2. multi-published tables archetictcher

IV. Conclusion and Future Work

This paper introduces a survey for most common privacy preserving data mining techniques that using Multiple-Published Tables PPDM & PPDP and explains their effects on Data Privacy. Both Anatomy and (α, k) – anonymity using lossy-Join [9] are used for security of respondents identity and decreases linking attack. It is observed that using generalization and suppression in (α, k) – anonymity using lossy-Join technique on those attributes lead to reduce the precision of publishing table. (α, k) – anonymity using lossy-Join also causes data lose because suppression emphasize on not releasing values which are not suited for k factor although it maintaining privacy. The idea of (α, k) – anonymity using lossy-Join is that if two tables with a join attribute are published, the join of the two tables can be lossy that helps to maintain the private information. On the other hand anatomy technique applied on sensitive tables reduces information loss, because it releases all the quasi-identifier and sensitive values directly in two separate tables without applying any suppression or even any generalization leads to data utility maintaining. The idea of Anatomy preserving privacy is that *QIT* does not indicate the sensitive value of any record, which is randomly guessed. Future work can include defining a new privacy technique for multiple sensitive attributes and researchers will focus to publish attributes without suppression using generalization boundaries technique that used to achieve k-anonymity maintaining individual privacy without influence data utility.

References

[1] L. Sweeney, "Achieving k-Anonymity Privacy Protection Using Generalization and Suppression", International Journal on Uncertainty, Vol. 10, No. 5, pp. 571–588, 2002.

[2] Abou el ela A. Hussien, Nermin Hamza, Ashraf A. Shahen and Hesham A. Hefny, "A survey of privacy preserving data mining algorithms", Yanbu Journal of Engineering and Science, Vol. 5, October, 2012.

[3] Bhavana Abad (Khivsara), Kinariwala S.A., " A Novel approach for Privacy Preserving in Medical Data Mining using Sensitivity based anonymity", International Journal of Computer Applications , Vol. 42, No.4, March, 2012.

[4] P. Samarati. "Protecting respondents identities in microdata release", IEEE Transactions on Knowledge and Data Engineering, Vol. 13, No 6, 2001.

[5] Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "L-Diversity: Privacy beyond k-anonymity", In Proceedings of the IEEE ICDE, 2006.

[6] N. Li, T. Li, S. Venkatasubramanian, "T-Closeness: Privacy Beyond k-Anonymity and l-Diversity", ICDE, pp. 106-115, 2007.

[7] Abou el ela A. Hussien, Nermin Hamza, Hesham A. Hefny, "Attacks on Anony-mization-Based Privacy-Preserving: A Survey for Data Mining and Data Publishing", Journal of Information Security jis, vol 4, pp.101-112, April, 2013.

[8] R. Bayardo and R. Agrawal, "Data privacy through optimal k-anonymity", In Proceedings of the 21st International conference on Data Engineering (ICDE), pp. 217-228, Tokyo, Japan, 2005.

[9] Raymond Chi-Wing Wong, Yubao Liu, Jian Yin, Zhilan Huang, AdaWai-Chee Fu1, and Jian Pei, " (α, k)-anonymity Based Privacy Preservation by Lossy join", Lecture Notes in Computer Science,Vol. 4, pp. 733-744, 2007.

[10] B. Fung, K. Wang, P. Yu, "Top-down Specialization for Information", Conference on Data Engineering (ICDE), pp. 205-216, 2005.

[11] K LeFevre, D DeWitt, R Ramakrishnan. Incognito," Efficient full domain k-anonymity", Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 49-60, Baltimore, Maryland, 2005.

[12] J Pei, J Xu, Z. B Wang, W Wang, K Wang, " Maintaining k- anonymity against incremental updates", Proceedings of the 19th International Conference on Scientific and Statistical Database, 2007.

[13] K. Venkata Ramana and V.Valli Kumari, "Graph Based Local Recoding For Data Anonymization", International Journal of Database Management Systems (IJDMS), Vol.5, No.4, August, 2013.

[14] Jian Xu Wei Wang Jian Pei Xiaoyuan Wang Baile Shi Ada Wai-Chee Fu3 "Utility-Based Anonymization Using Local Recoding", KDD'06, Philadelphia, Pennsylvania, USA. pp.20-23, August, 2006.

[15] Manolis Terrovitis _ Nikos Mamoulis _ Panos Kalnis," Local and Global Recoding Methods for Anonymizing Set-valued Data", Research Center Mathematical and Computer Sciences and Engineering,King Abdullah University of Science and Technology, 2011.

[16] Xiaokui, Xiao Yufei Tao, "Anatomy: Simple and Effective Privacy Preservation", VLDB, September, Seoul, Korea, 2006.

[17] P. Samarati and L. Sweeney, "Generalizing data to provide anonymity when disclosing information". In PODS, page. 188, 1998.

[18] L. Sweeney, k-Anonymity: "A Model for Protecting Privacy", International Journal on Uncertainty, Vol. 10, No. 5, pp. 557–570, 2002.

[19] X. Xiao and Y. Tao. "Personalized privacy preservation", SIGMOD, 2006.

[20] 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.