AUSTRALIAN JOURNAL OF BASIC AND APPLIED SCIENCES

ISSN:1991-8178 EISSN: 2309-8414
Journal home page: www.ajbasweb.com

Anonymization and Reducing Information loss in Incremental Dataset through Grouping and Local Recoding

[1]R. Mahesh and [2]T. Meyyappan

[1]Research ScholarAlagappa UniversityDepartment of Computer science, Karaikudi-630303, Tamilnadu State, India.
[2]Professor, Alagappa UniversityDepartment of Computer science, Karaikudi, Tamilnadu State, India.

Address For Correspondence:
R. Mahesh, Research ScholarAlagappa UniversityDepartment of Computer science, Karaikudi-630303, Tamilnadu State, India.

ARTICLE INFO

Article history:
Received 18 June 2017
Accepted 28 July 2017
Available online 20 August 2017

Keywords:
Data privacy,anonymity,data publishing,privacy preserving

ABSTRACT

In this decade, Massive amount of digital information has created the gateway for gathering knowledge and effective decision-making. Private organizations and government sectors publish the data for business, statistical analysis and research purposes. While publishing the data that represents the individuals' information, for sharing to public, Privacy preservation should be followed. The objective of privacy preservation protects the individual's sensitive information from adversaries, during data publishing. Researchers have contributed the anonymization methods and algorithms for publishing the static dataset. Data collection process execute in different time stamp. Sometimes released information has updated in different time stamp and published in different time series. Hence, Anonymization techniques for static dataset turn into worthless. Although existing methods in the literature perform the anonymization, they do not take care of increasing information loss and ignoring the privacy of data owners because of complexity of process. In addition, Increasing information loss of anonymized dataset twist the data become meaningless. In this paper, Our Proposed method preserves the individual privacy in incremental dataset publishing scenario and also reduce the information loss through grouping record-sets based on relational algebra and applies local re coding technique. Adult dataset with more than 30,000 records are experimented with the proposed method. The results show the significant reduction in information loss over existing methods.

INTRODUCTION

For a decade, we are facing serious problems in preserving the privacy in data publishing although the research contributions are delivering various anonymization techniques. An adversary applies the probability distribution to quasi identifiers (qid) for predicting and extracting the sensitive attribute values. For avoiding the privacy breaches, Various anonymization techniques (Wu, Y., Z. Sun and X. Wang, 2009; Tamir Tassa, *et al*., 2012; Benjamin, C.M., *et al*., 2010; Xiaoxun Sun, *et al*., 2008; Fung, B.C.M., *et al*., 2008) has been applied in data publishing for protecting the sensitive attribute values. K-anonymity is one of the familiar privacy model that preserves the privacy against record linkage attack. The anonymized dataset satisfy the condition that the minimum equivalence group size on quasi identifier is at atleast k. The record owner have several records with same sensitive attribute values. In this case, K-anonymity model fails to ensure the privacy and also it doesn't enable to preserve the privacy against attribute linkage attack. L-diversity privacy model overcomes limitation of k-anonymity. This model works out against attribute linkage attack with satisfy the condition that every quasi identifier group contains atleast *l* well represent values. That means the records having at least I distinct values for the sensitive attribute in each qid group. According to general view of L- diversity, sometimes the frequency of sensitive attribute values is more than others in a group. So this model does not ensure the privacy against probabilistic inference attacks. Many privacy models have been applicable for single static releases of data. In

To Cite This Article: R. Mahesh and T. Meyyappan., Anonymization and Reducing Information loss in Incremental Dataset through Grouping and Local Recoding. ***Aust. J. Basic & Appl. Sci.,*** *11(11): 139-147, 2017*

practice, Datasets modify continuously at different time stamp and publishing to researchers in a periodic manner. In each time stamp, the dataset has been modified or inserted. In this case, the anonymizaton techniques should be preserved individuals privacy in each publishing time. BCF anonymity and LKC privacy methods has performed the privacy protection in continuous data publishing. Increasing efficiency of anonymity leads to anonymized dataset become valueless and purpose of releasing might be failed. The structure of dataset, views of quasi identifiers and timeline modification have fixed the level of privacy issues. The theme of anonymization algorithms may vary in respect to the type of publishing scenario. The motivation for anonymization techniques increases the privacy and reduces the information loss. Both are mutually exclusive. Various algorithms (Al-Hussaeni, Khalil, 2017; Jun Liao, Chaohui Jiang, Chun Guo, 2016; Sreedhar, K.C., *et al*., 2017; Torsak Soontornphand *et al*., 2017) aims to retain the utility of released dataset through global generalization. Even global generalization performs the preservation in better way; it is deviated from assuring the anonymized dataset in valuable. Our proposed method focuses on the local recoding (Byun, J., Y.Sohn and E. Bertino, 2006) and grouping for achieving the preservation of individual privacy and also reduces the information loss of incremental dataset. The experiment results ensure that our proposed method increase the utility of anonymized dataset.

Motivation:

Health care agencies publish a patient dataset for data analysis and research purpose. The agencies strictly follow the privacy law and perform distribution and sharing. But they are unaware of the individual's privacy breaches while publishing the data. For example, the health care agency wants to publish the dataset T_1 (Job, Age, Sex, Disease) of patients, where Disease is sensitive information. Consider the external dataset T_2 (Name, Job, Sex, Age) for the same persons, where the job is the sensitive information.

Table I: Raw Data T_1

Job	Sex	Age	Disease
Engineer	Male	35	Hepatitis
Engineer	Male	38	Hepatitis
Lawyer	Male	38	HIV
Writer	Female	30	Flu
Writer	Female	30	HIV

Table II: Raw data T_2

Name	Job	Sex	Age
Alice	Writer	Female	30
Bob	Engineer	Male	35
Cathy	Writer	Female	30
Doug	Lawyer	Male	38
Emily	Dancer	Female	30

Combining the two datasets on the common attributes Job, age, sex may reveal the identity of the person who is affected by a particular Disease. For Instance, Alice, 30 years old Female Writer is identified as a Flu patient by qid= (Writer, Female, 30). Anonimization techniques (Sweeney, L., 2002; Ninghui Li, *et al*., 2007; Machanavajjhala, A., *et al*., 2006; Sweeney, L., 2002) are helpful to preserve the sensitive information and protect the data owner's privacy. The publishers update the dataset and release it periodically. Therefore anonymization technique and preservation algorithm needs modification for privacy gain. Summary of previous anonymized results should be maintained. It may mislead to achieve the reduction of information loss. K-anonymization has been applied for a static data release. In practice, new data inserts continuously (Fung, B.C.M., *et al*., 2008) and it has been republished in a timely manner (Xiao, X. and Y. Tao, M-invariance, 2007).

The number of frequency values of Sensitive attribute values is more than one or equal to one. This situation is creating the way for privacy breaches. The awareness of sensitive attribute preservation should be given higher priority while applying privacy models. Taken problem environment is differ from previous works. The proposed method overcomes the limitations of existing anonymization principles (Xiaojian Zhang, Xiaofeng, Rui Chen, 2013; Pinshui Wang, Jiandong Wang, 2013; Nergiz, A.E., *et al*., 2013; Xiao, X. and Y. Tao, M-invariance, 2007; Sweeney, L., 2002; Byun, J., Y.Sohn and E. Bertino, 2006) for publishing incremental dataset and also reduce the information loss satisfactory .

Proposed Method:

A. ***Preliminaries:***

Let $T(a_1,a_2,a_3\ldots a_n)$ be a dataset,$R_1,R_2,R_3,R_4...R_n$ are the finest set of records in Table T ie $R \subseteq T$ and then $a_1,a_2,a_3\ldots a_n$ are finite set of attribute 'a' in table T and tuple $t \subseteq T$. Each tuple represents the information of a single person or an object. There are two types of attribute such as Quasi Identifier Attribute and Sensitive attribute. The attributes which are published in another dataset which are called quasi identifiers Q_i where

i=1,2,3..n. Let QI be quasi identifier attributes, $QI \subseteq a$ and sensitive attributes $SA \subseteq a$. Quasi Identifiers could be known to adversaries. Adversaries may be finding the individual's sensitive information (Sensitive attribute) through Quasi Identifiers. Hence Sensitive attributes should be preserved. Let Q_T be Quasi Identifier of table T and A be attribute of table T where $\{Q_i \ldots Q_j\} \subseteq \{A_1 \ldots A_n\}$.

Table III: Raw dataset

Age	Salary	Disease	Age
28	4000	Bronchitis	28
42	8000	H1N1	42
32	10000	Bronchitis	32
29	5000	Bronchitis	29
34	1000	Flu	34
30	7000	Gastritis	30
36	9000	Gastric ulcer	36
48	8000	Gastritis	48
52	11000	Gastritis	52
28	6000	Cancer	28
32	5000	Gastric ulcer	32
23	9000	Gasteric ulcer	23

Let Table III denoted as T. Let T be an published Medical data which contains quasi identifiers Q_T {pincode,age,salary} Q_i(i=1,2…n) and sensitive attribute S.

B. ***Grouping the dataset:***

Generally, the frequency value of each tuple is greater than or equal to one. Frequency of sensitive attribute value influences the efficiency level of anonymization. In this work, each tuple is made to participate in anonymization process. Grouping is important for generalizing all the tuples of Q_i where i=1,2,3..n (Byun, J., *et al*., 2007). Hence this method constructs N number of groups based on N Number of frequency values of sensitive attribute. Here the term group is means to collection of records group by sensitive attribute. Therefore, Specific algorithm is not essential to apply for grouping. Each group contains same sensitive attribute values. The collection of distinct value forms the $n+1^{th}$ group. Let G_i be a group where i= 1,2,3..n,n+1 and $G \subseteq T$

Table IV: (A). GROUP G_1

Age	Salary	Disease
28	4000	Bronchitis
29	5000	Bronchitis
32	10000	Bronchitis

Table IV: (B). GROUP G_2

Age	Salary	Disease
23	3000	Gasteric ulcer
32	5000	Gastric ulcer
18	9000	Gasteric ulcer

Table IV: (C). GROUP G_3

Age	Salary	Disease
30	7000	Gastritis
48	8000	Gastritis
52	11000	Gastritis

Table IV: (D). GROUP G_4

Age	Salary	Disease
28	6000	H1N1
34	1000	Flu
64	8000	Cancer

Table IV(A,B,C,D) shows the groups G_1,G_2,G_3 and G_4 which are created from original dataset in Table III. But group G_4 is completely differs from other groups. It contains a collection of distinct values of sensitive attribute.

C. ***Local Recoding:***

In each group G_i where i=1, 2, 3..n ,n+1, the quasi identifier attribute is selected for a generalization process that is based on the maximum count of unique value. The next lowest value L_i and next biggest integer value M_i are taken for each tuple of quasi identifier and rewritten as a pattern like $L_i <= M_i$ in each group G_i where

i=1,2,3..n,n+1. This generalization process (Mahesh, R., T. Meyyappan, 2013) continues until all the Qi values in each group. The output of the generalization process in group G_3 is given below:

Table V: Generalized The Group G_3

Age	Salary	Disease
30<48	7000<=8000	Gastritis
30<=52	7000>11000	Gastritis
48<=52	8000<=11000	Gastritis

D. ***Group Difference operation:***

Let Gi be a group where i=1, 2,3.. n. While merging all groups like $G_1 \cup G_2$, , $G_2 \cup G_3$, $G_3 \cup G_4$ …. $G_n \cup G_{n+1}$, some tuple values are away from the range of generalized Q_i values. In table VI(A), the generalized tuple value of G_2 '18<=23' is not generalized with other Q_i attribute value in Group G_1. Hence, the sensitive attribute values, 'Gasteric Ulcer' stay alone after group difference operation. Here, Record and attribute linkage are found in anonymized group G_2. To overcome this issue,The M_i value is taken from G_1 , it should be greater than Mi value of selected tuple in G_2 and rewritten as selected tuple Li <= Mi as value of G_1.

Table VI: (A). GENERALIZED GROUP G_2

Age	Salary	Disease
18<=23	3000<=9000	Gasteric ulcer
18<=32	5000<=9000	Gastric ulcer
23<=32	5000<=9000	Gasteric ulcer

Table VI: (B). GENERALIZED GROUP G_1

Age	Salary	Disease
28<=29	4000<=5000	Bronchitis
28<=32	4000<=10000	Bronchitis
29<=32	5000<=10000	Bronchitis

Table VII shows the re-generalized record in G_1 after applying the group difference operation.

Table VII: REgeneralized Group G_1 after difference operation

Age	Salary	Disease
18<=29	4000<=5000	Bronchitis
28<=32	4000<=10000	Bronchitis
29<=32	5000<=10000	Bronchitis

Group difference operation is performed on all groups G_1-G_2, G_2-G_3, G_3-G_4, G_4-G_1.The group difference operation will generates modified anonymized records, if any Q_i tuple value stay away from generalized tuple value of operand groups. Finally, all groups are merged to make a complete anonymized dataset. T'

E. ***Dynamic Data Publishing:***

Publisher republishes the dataset that contains new records and some updated records at new timestamp. Before generalizing the new record R_i where i=1, 2,3.. n, they are merged into the groups in previous anonymized dataset. The new Q_i tuple values are merged into existing generalized Q_i values of group.

Table VIII: (A). new records published at new timestamp

Age	Salary	Disease	Status
30	3000	Bronchitis	Added
35	4000	Bronchitis	Added
45	2000	Gastiritis	Updated age & salary
50	7000	Cancer	Added
62	5000	Cancer	Added
75	8000	Stomach Cancer	Added
48	8000	Gastritis	Deleted
52	11000	Gastritis	Deleted

1) Inserting new records and update existing group:

In table VIII(A), first two new records are added to republish data at new timestamp. New record {age:30,Salary:3000,Disease:Bronchitis} is merged with a second record of Group G_1. In the second record, Q_i age value {35} can't merge with any tuple value of Q_i 'Age' in G_1. So the new record gets into group G_1. If the new tuple 'age' value is greater than maximum Mi value in G_1, it will be generalized with maximum M_i value. If it is less than least Li value, it will be generalized with least Li value. Table VIII(b) shows regeneralized Q_i values with new records in Group G_1.

Table VIII: (B). Regeneralized group G_1 with new records at new timestamp

Age	Salary	Disease
18<=29	4000<=5000	Bronchitis
28<=32	3000<=10000	Bronchitis
29<=32	5000<=10000	Bronchitis
32<=35	2000<=10000	Bronchitis

2) Inserting new records and form new group:

In table VIII(A), the sensitive value of 4th and 5th records are not present in G_1,G_2 and G_3. The new sensitive attribute value is present in G_4 that contains a collection of distinct S_i values. Here, new S_i value forms a new group with existing S_i value in G_4. In table IV(D), the sensitive tuple value 'cancer' is in singular form. In new timestamp, the updated raw data contains similar S_i values. Here, the frequency value of new S_i value is greater than one. Hence, a new group should be created. Generalization is taken over QI values as shown table VIII(C) then duplicate record will be deleted in G_4.

Table VIII: (C). Create new group G_5 with new records at new timestamp

Age	Salary	Disease
50<=62	5000<=7000	Cancer
50<=64	5000<=8000	Cancer
62<=64	7000<=8000	Cancer
50<=62	5000<=7000	Cancer

3) Delete existing records and group:

In the table VIII(A) the two records are deleted from published data. The new S_i values presents in Group G_3. Once the two records are deleted in the published table, The frequency value of S_i 'Gastritis' become one. The group G_3 is erased and the S_i values are added in Group G_4. Table VIII(D) shows the regeneralized group G_4 and the record is moved from G_3 to G_4.

Table VIII: (D). Regeneralized group G_4 at new timestamp

Age	Salary	Disease
28<=30	6000<=7000	H1N1
28<=34	1000<=6000	Gastritis
30<=34	1000<=7000	Flu
28<=30	6000<=7000	H1N1

In every data publishing in timestamp series,the groups participates into union , difference operations and local recoding process.

F. ***Computational Procedure of proposed method:***

Input: Table T which containing Quasi identifier Q_i, sensitive attribute S_i, where i=1,2,3….n

Step 1:

Form Groups G where $G_{i=1..n} \subseteq T$ based on sensitive attribute values S_j(j=1..m) such that similar attribute values belong to n-1 groups and unique values belong to n^{th} group.

Each group contains tuples t_i where i=1,2,3..k, $t_{i=1..k} \subseteq G$.

$t_{i=1..k} \in G_{i=1..n} \in T$

Step2:

For each group G_i where i=1,2,3..n repeat step 3 to5 from 1 to n

Step 3:

Let $L_j = t_j$.

Find the next nearest minimum integer L_i of t_i in Group G_i where i=1,2,3,…n and if found, Let $L_i = t_j$

Step4:

Let $M_j = t_j$.

Australian Journal of Basic and Applied Sciences, 11(11) August 2017, Pages: 139-147

Find the next nearest largest integer M_i of t_i in Group G_i where i=1,2,3,…n and if found, Let $M_j = t_j$

Step5:

Tupe t_i values of quasi identifier Q_i in same group G_i is rewritten as a range value $L_j <= M_j$. The generalization condition is set as $t_j = L_j <= M_j$

Step6: Perform difference operation dv_k (k=1..p)between groups G_j and G_{j+1} (j=1..n-1)

Step7: Between each Group Gi and Gi+1 (i=1… n-1)

If $G_i(L_k) < G_{i+1}(L_p)$ (k=1..x, p=1..y)

Set $G_i(L_k)$ = Next-Least $G_{i+1}(L_p)$

Else

Set $G_{i+1}(L_p)$ = Next-Least $G_i(L_k)$

If $G_i(M_k) < G_{i+1}(M_p)$ (k=1..x, p=1..y)

Set $G_i(M_k)$ = Next-Biggest $G_{i+1}(M_p)$

Else

Set $G_{i+1}(M_p)$ = Next-Biggest $G_i(M_k)$

.

Step8: Merging all groups G_i where i=1,2,3..n and get anonymized dataset T*

1) Computational method for incremental dataset:

Let T* be anonymized dataset which containing Generalized tuples GT_i, sensitive attribute S_i, where i=1,2,3….n

Let U be new records which containing tuples t_i and sensitive attribute values S_i, where i=1,2,3….n

Step a: Insert new records and update existing group

Form Groups GT where $GT_{i=1..n} \subseteq T^*$ based on sensitive attribute values S_i(i=1..m) such that similar attribute values belong to n-1 groups and unique values belong to n^{th} group.

If $t_i > GT_i(L_k)$ and $t_i < GT_i(M_k)$ and $GT(s_i)=U(S_i)$,

Generalize incremented data set U by repeating step 3 to 5 for all Q_i

Step b: Insert new records and form new group

If $U_i(S_i) <> GT_i(S_i)$ and $COUNT(U_i(S_i)) > 1$

add $U_i(S_i)$ to GT (i=1..n)

n=n+1

Else If $U_i(S_i) = GT_n(S_i)$ and $COUNT(U_i(S_i)) > 1$

add $U_i(S_i)$ to GT (i=1..n)

n=n+1

Remove i^{th} row which contains $U_i(S_i)$ values from GT_n

end

Step c: For each deleted Record i

Remove ith row from data set U

Repeat step b)

Experimental Results:

The proposed method is tested with adult data set on a machine with 2.66 GHz processor and 2GB of main memory. The adult data set was collected from the UCI machine learning repository. Java coding in Netbeans7.2.41 environment is employed for implementation

Table 1X: Metadata About Adult Dataset

Sno	Attributes	Distinct Values	Type
1	Age	74	Quasi Identifier
2	Education-Num	16	Quasi Identifier
3	Hours per week	94	Quasi Identifier
4	Country	41	Quasi Identifer
5	Occupation	14	Sensitive Attribute

In the experiment, Age,Education-Num,Hours per Week status and Country are considered as Quasi Identifier attributes and occupation is considered as a Sensitive attribute value.

The performance of the proposed algorithm is assessed in terms of information loss metrics and execution time. In experimental analysis, performance of the proposed method is compared with three existing methods (Nergiz, A.E., *et al*., 2013; Pinshui Wang, Jiandong Wang, 2013). There are three existing methods. The first method is l-diversity based full-domain generalization which re-anonymizes the entire dataset (denoted by l-

diversity algorithm I)(Nergiz, A.E., *et al*., 2013). The Second method employs the same l-diversity algorithm but on incremental dataset (denoted by l- diversity algorithm II) (Fung, B.C.M., *et al*., 2007). The third one is incremental l-diversity algorithm (denoted by l- diversity algorithm III) (Pinshui Wang, Jiandong Wang, 2013).

The number of records is gradually increased from 5000 to 35000 with increment of 5000 records. The performance metrics are compared in terms of information loss and execution time. The information loss is compared through the derived formula proposed by author (Xiao, X. and Y. Tao, 2006).

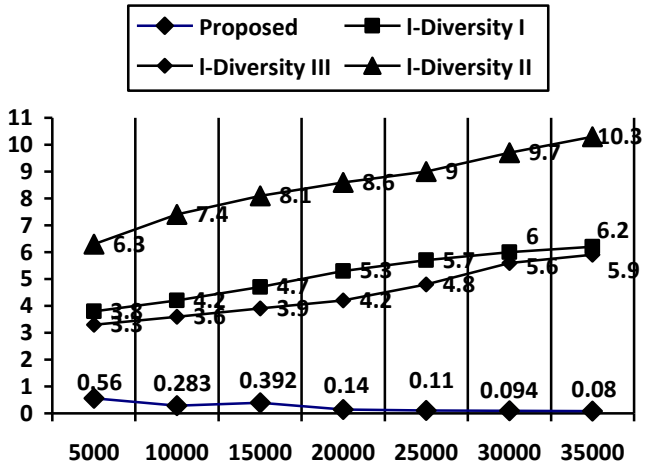

Fig. 1: Information Loss

Fig. 1 shows the information loss metric values of the proposed method compared with the existing three methods. The incremental dataset is added with 5000 records in each data publishing. It is evident from the figure 1, the information loss of the proposed method is gradually decreasing with increase in the size of the dataset. Frequency value of quasi identifier attribute value is increased. When record size increases in every time stamp. Therefore, in every timestamp, the process of anonymization is minimized while comparing existing methods. The grouping process involves in to ensuring the preservation of privacy of updated and inserted records and also published records in earlier timestamps. All the three existing methods anonymize the dataset from the scrap for every increase in the size of the dataset. However, the proposed method performs anonymization incrementally. The proposed method employs local recoding over previously anonymized dataset at new timestamp. The information loss in the anonymized data set is reduced much compared to existing methods.

The comparison of Execution time engages to prove the efficiency of resource time consumption. In Fig. 2 shows the execution time (ms) for the three anonymizing methods for incremental dataset. Our proposed method consumes less time in first two time stamp (5000 to 15000 records).Afterwards, the metric value is not much higher and lower rate. The execution time of the proposed method is dramatically lesser than that of two other existing methods (l- diversity algorithm II and l- diversity algorithm III) from the beginning stage. Anonymization process is applied over previously anonymized dataset in time series. Hence, with reduced generalization operation reduces the overall execution time compared to existing methods.

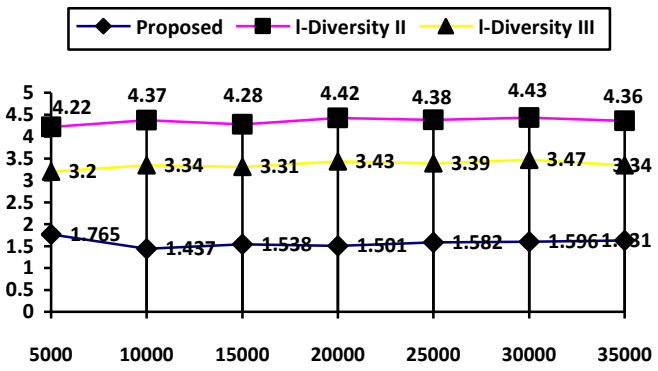

Fig. 2: Execution Time in msecs

Conclusion And Future Work:

In this paper, In the current data publishing age, voluminous data about individual are published in every domain. Individuals' privacy should not be compromised and preserved against attack models. Many researchers have contributed various techniques for preserving the privacy of static data. Furthermore, the earlier contributions focus on the building methods for improving the efficient preservation. So that reduction of information loss hasn't come in to the account of outcome of existing methods. In this paper, the proposed method achieves the preservation of individual privacy and decrease the information loss to explore the anonymized dataset as meaningful. The proposed method has applied into incremental dataset publishing model and involve in the experimental analysis. The outcome results of proposed method ensure the reduction of information loss considerably compared to existing methods. The further research is in progress to find out methods for various data publishing scenarios such as multiple views data release and sequential release with new attributes. Furthermore, the direction of future research is to analyze the variety and volume of data that how much privacy issue impact render in data publishing.

REFERENCES

Machanavajjhala, A., J. Gehrke, D. Kifer and M. Venkitasubramaniam, 2006. ℓ-diversity: Privacy beyond k-anonymity, In Proc. 22nd Intnl international Conference on data engineering. (ICDE) 24.

Nergiz, A.E., C. Clifton, Q.M. Malluhi, 2013. Updating Outsourced Anatomizced Private Databases",EDBT 13 Proceedings of the International Conference on Extending Database Technology, Genoa, Italy, pp: 179-190.

Al-Hussaeni, Khalil, 2017. Preserving Privacy in High-Dimensional Data Publishing,Thesis in Department of Electrical and Computer Science Engineering, Concordia University,Montreal Quebec,Canada.

Fung, B.C.M., Ke Wang and P.S. Yu, 2007. Anonymizing classification data for privacy preservation, IEEE Transactions on Knowledge and Data Engineering(TKDE), pp: 711-725.

Fung, B.C.M., Ke Wang, Ada Wai-Chee Fu, 2008. Jian Pei,2008,Anonymity for Continous Data Publishing, Proceedings of the International Conference on Extending Database Technology, pp: 264-275.

Benjamin, C.M., K.E. Fung, Wang, Ada Wai-Chee Fu and Philip S. Yu, 2010. Introduction to Privacy-Preserving Data Publishing Concepts and techniques, ISBN:978-1-4200-9148-9

Byun, J., A. Kamra, E. Bertino and N.H. Li, 2007. Efficient k-anonymization using grouping techniques, Proceedings of the 12th International Conference on Database Systems for Advanced Applications,Springer-Verlag Berlin Heidalberg,LNCS,4443,2007,188-200

Byun, J., Y.Sohn and E. Bertino, 2006. Secure anonymization for incremental datasets, Proceedings of the 3rd VLDB Workshop on Secure Data Management, pp: 48-63.

Jian-Xu, Wei Wang, Jian Pei, Baile Shi, Ada Wai-Chee Fu, 2006. "Utility-Based Anonymization Using Local Recoding",KDD06,83-106

Jun Liao, Chaohui Jiang, Chun Guo, 2016. Data privacy protection based on sensitive attributes dynamic update, 4th International Conference on Cloud Computing and Intelligence Systems (CCIS), 978-1-5090-1256-5

Sweeney, L., 2002. An Achieving k-Anonymity Privacy Protection Using Generalization and Suppression", International Journal of Uncertainty, Fuzziness and Knowledge-Based System, pp: 571-588.

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

Mahesh, R., T. Meyyappan, 2013. Anonymization Techniques through record elimination to preserve the privacy of published data, International Conference on Pattern Recognition, Informatics and Mobile Engineering (PRIME), pp: 328-332.

Ninghui Li, Tiancheng Li, Suresh Vengakatasubramaniam, 2007. t-Closeness: Privacy Beyond k-Anonymity and ℓ-Diversity, International Conference on Data Engineering, pp: 106-115.

Pinshui Wang, Jiandong Wang, 2013. L.Diversity Algorithm for Incremental Data Release, Appl. Math.Inf.Sci, pp: 2055-2060.

Sreedhar, K.C., M.N. Faruk, B. Venkateswarlu, 2017. A genetic TDS and BUG with pseudo-identifier for privacy preservation over incremental data sets, Journal of Intelligent & Fuzzy Systems, 32(4): 2863-2873.

Tamir Tassa, ArnonMazza and Aristides Gionis, 2012. k-Concealment: An Alternative Model of k-Type Anonymity, TRANSACTIONS ON DATA PRIVACY 5: 189-222.

Torsak Soontornphand,Nattapon Harnsamut, Juggapong Natwichai, 2017. Privacy Preservation Based on Full-Domain Generalization for Incremental Data Publishing, Information Science and Applications (ICISA, 577-588.

Xiaojian Zhang, Xiaofeng, Rui Chen, 2013. Differentially Private Set-Valued Data Release against Incremental Updates,DASFAA 2013,Springer-verlag Berlin Heidalberg,LNCS 7825: 396-406.

Xiao, X. and Y. Tao, 2006. Personalized privacy preservation, In Proceedings of ACM Conference on Management of Data (SIGMOD'06"), pp: 229-240.

Xiao, X. and Y. Tao, M-invariance, 2007. Towards privacy preserving re-publication of dynamic datasets, Proceedings of the ACM Conference on Management of Data (SIGMOD), pp: 689-700.

Xiaoxun Sun, Hua Wang, Jiuyong Li and Traian Marius Truta, 2008. Enhanced P-Sensitive K-Anonymity Models for privacy Preserving Data Publishing, Transactions On Data Privacy, pp: 53-66.

Wu, Y., Z. Sun and X. Wang, 2009. Privacy preserving k-anonymity for republication of incremental dataset, Proceedings of the WRI World Congress on Computer Science and Information Engineering, pp: 53-60.