ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Improving Privacy And Data Utility For High-Dimensional Data By Using Anonymization Technique

P.Nithya[1], V.Karpagam[2]

PG Scholar, Department of Software Engineering, Sri Ramakrishna Engineering College, Coimbatore, India[1]

Associate Professor, Department of Information Technology, Sri Ramakrishna Engineering College, Coimbatore, India[2]

ABSTRACT: Privacy Preserving is one of the significant methods in data mining to hide the sensitive information. Anonymization techniques like generalization and bucketization have been used for privacy preserving. The main problem with generalization is it is not applicable for high-dimensional data and bucketization technique does not avoid membership disclosure. Slicing is one of the novel techniques in which the data is partitioned horizontally and vertically. This reduces the dimensionality of the data and it is able to handle high dimensional data better when compared to generalization and bucketization. In slicing, every attribute is in exactly one column. It provides better privacy but there is loss of data utility. Overlapping slicing is a novel technique that allows duplicating an attribute in more than one column so that more attribute correlations is achieved for better data utility. For protecting membership information, a more effectual tuple grouping algorithm is proposed and continuous attributes are handled. Further to improve the privacy, noise enabled slicing method is used.

KEYWORDS: Data mining, Privacy preservation, data anonymization, data security

I. INTRODUCTION

Data mining is the process of analyzing the data and extracting the useful information. Data mining is one of the numbers of analytical tools for examining data It permits users to analyze the data from different angles and classified it. In data mining privacy is an important issue. It is very challenging to preserve the privacy information of the users. The importance of the privacy preserving is to hide the sensitive information so that they cannot be discovered through data mining techniques. Privacy preserving publishing of micro data has been studied expansively in recent years. Micro data includes records that contain information about an individual entity like person, household and an organization. The popular anonymization techniques are generalization and bucketization. In these techniques the attributes are divided into three groups: some attributes are identifiers which can exclusively identify an individual such as name or social security number. Some of the attributes are QI (QI) in which an adversary potentially identifies an individual. Some attributes are sensitive attributes that are unknown to an attacker and it is considered as sensitive information. In generalization and bucketization, the common process is to remove identifiers from the data and partitions tuples into buckets. In the generalization method, it transforms the QI values so that tuples in the same bucket cannot be distinguished by their QI values. In the bucketization method, the sensitive attributes are separated from the QI by randomly permuting the SA values in each bucket. In the generalization K-anonymity is used and l-diversity is used for bucketization.

Generalization for K-anonymity loses considerable amount of information when it handles high-dimensional data. Therefore generalization for K-anonymity suffers from the curse of dimensionality. Bucketization has good data utility even though it has some limitations. Bucketization does not prevent membership disclosure because bucketization

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

publishes the Quasi-Identifier values in their original forms. An attacker can discover whether an individual has a record in the published data or not.

To enhance the privacy and to improve data utility a novel anonymization technique called slicing is used. In this technique, the data set is partitioned both vertically and horizontally. In the vertical partitioning, attributes are grouped according to the associations among the attributes. Each and every column includes a subset of attributes that are highly correlated. Horizontal partitioning is accomplished by grouping tuples into buckets. At the end, within each bucket, to split the linking between dissimilar columns the values in each column are randomly permutated. In slicing, to consider each attribute is in precisely one column. So, an overlapping Slicing method is introduced in which the attributes are duplicated in more than one column. Further to improve the privacy, Noise enabled slicing method is used. In this method, by using the chi-squared and Pearson based correlation coefficient the correlations between pairs of attributes and sensitive attributes is calculated and according to the correlations clustering the attributes.

II. RELATED WORK

Pierangela Samaratib suggested anonymization technique for protecting the micro data [1].when microdata is released, anonymity of the data must be protected. It should satisfy k-anonymity which prevents the re-identification of individual's records in the data while integrity of the data must not be compromised by using generalization and suppression techniques. Benjamin C. M. Fung et.al proposed to preserve the information by using top-Down specialization method[2].A Top-Down Specialization (TDS) approach is provided to simplify a table to assure the anonymity requirement while preserving its usefulness to classification. TDS approach generalizes the table by specializing it iteratively starting from the most general state. In every step, a general value is dedicated into a specific value for a categorical attribute otherwise an interval is split into two sub-intervals for a continuous attribute. It can discard data records that cannot be further specialized. In addition, it preserves both data utility and privacy.Gabriel Ghinita et.al suggested an anonymization of personal data for high dimension [3]. It used privacy preserving requirement such as k-anonymity and l-diversity while maximize the data utility. It handles the sparse high dimensional data efficiently by capturing the correlation within the data.Daniel Kifer et.al suggested a method to evaluating utility [4]. In data publishing, the restricting disclosure needs a careful balance between privacy and utility. To prevent attacks the K-anonymity and l-diversity is required for privacy requirements to hide the sensitive information about the users.Qing Zhang et.al suggested novel privacy intent to better privacy protection for several sensitive attributes [5]. Permutation based anonymization has been used for answering the aggregation queries accurately. In which we randomly permute the association between sensitive attribute and QI instead of generalizing the QI. Privacy protection for numerical sensitive attributes has been proposed for hide the sensitive attributes. By giving the similar grouping of tuples, the permutation based anonymization methods can group queries more accurately when compared to the generalization based methods. Charu C. Aggarwal suggested k-anonymity for privacy preserving in data mining. Several techniques are used for privacy preserving data in the multidimensional data attributes [6]. Anonymization is one of the technique where a record is released only if it is identical from the k other entities in tae data. K-anonymity in which any QI present in the released table must appear in at least k-records. Bee-Chung Chen et.al suggested a general framework for privacy preserving by using the external knowledge [7]. In the data publishing the privacy is a significant problem. With the help of external knowledge adversary is able to infer the sensitive data. Robust techniques skylinecheck and skylineanonymize have been used for checking whether a release candidate is safe or not.

Frank McSherry et.al proposed sensitive based method analysis for the privacy preservation. In the sensitivity based method, examine the sensitivity of the particular data analysis functions that includes histograms, contingency tables and the covariance matrices that have very high-dimensional output and also provide the sensitivities that are independent of the dimension [8]. There are two types of classifications in the perturbation method have been used.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

III. PRIVACY PRESERVATION TECHNIQUES

Slicing

To preserve the sensitive data, a new data anonymization technique is used. Slicing is an anonymization method in which the dataset is partitioned into horizontally and vertically. By grouping the attributes based on the associations among the attributes the vertical partitioning is accomplished. By grouping tuples into buckets the horizontal partitioning is accomplished. At the end, within each and every bucket, the values in each column are arbitrarily sorted to split the linking between dissimilar columns. Based on the privacy requirement of l-diversity slicing is effectual to prevent attribute disclosure. L-diverse slicing is initiated that make sure that an attacker cannot know the sensitive attribute values of any individual person with a probability larger than1/l. To satisfy l-diversity a proficient algorithm is used for evaluating the sliced table. This algorithm partitions attributes into columns and applies column generalization and partitions tuples into buckets. Finally, to illustrate the intuition behind membership disclosure and represent how the slicing avoids membership disclosure.

Modified Slicing

To improve the data utility, the modified slicing method is used. An overlapping slicing is an extension of slicing where an attribute is in more than one column. This makes more attributes correlations. In this method, tuples are generalized to assure some negligible frequency requirement. To point out that column generalization is not a necessary phase in this algorithm. As exposed by Xiao and Tao, bucketization method presents the similar level of privacy protection as generalization, with respect to attribute disclosure. Even though column generalization is not a required phase, it can be useful in several aspects.

Column generalization is needed for the membership confession protection. If a column value is exclusive in a column, a tuple with this unique column value can only have one matching bucket. The main dilemma is that this unique column value can be recognized. In this case, it would be helpful to apply column generalization to make sure that each column value appears with at least some frequency. So we generally focus on the tuple partitioning algorithm.

Algorithm

Attribute partitioning: To handle high-dimensional data, slicing is an effectual method.

By partition the attributes into columns this method diminishes the dimensionality of the data. It enables the slicing to handle high dimensional data. So that highly correlated attributes are in the same column. The highly correlated attributes are conserved so that to preserve the correlation among the attributes for high data utility. Firstly, to evaluate the correlations between the pairs of attributes and then to cluster the attributes according to the correlations by using the chi-squared and pearson based correlation coefficient.

The partitioning is accomplished by two steps:

Through the chi-squared correlation the vertical partitioning is done. Pearson correlation coefficient is utilized for evaluating correlations between two continuous attributes whereas mean-square contingency coefficient is a chi-square measure of correlation between two categorical attributes. Selecting the mean-square contingency coefficient two attributes A_1 and A_2 with domains $\{v_{11}, v_{12} \ldots v_{1d1}\}$ and $\{v_{21}, v_{22} \ldots v_{d2}\}$ correspondingly.

$$\Phi^2(A_1,A_2) = \frac{1}{\min\{d_1,d_2\}-1}\sum_{i=1}^{d_1}\sum_{j=1}^{d_2}\frac{(f_{ij}-f_i f_j)^2}{f_i \cdot f_j}$$

Through pearson correlation coefficient the vertical partitioning is accomplished. Pearson correlation coefficient is utilized to measure associations between two continuous attributes. It is a measure of the correlation between the two variables X and Y, providing a value between +1 and -1 comprehensive. This coefficient is applied to a sample is frequently represented by the symbol called r and it may be represented as sample correlation coefficient. To attain the formula for r by substituting estimates of the covariance and variances based on a sample value.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

The formula is

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

This equivalent expression provides the correlation coefficient as the mean of the products of the standard scores. According to the sample of paired data (X_i, Y_i) the sample pearson correlation coefficient is represented by,

$$\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{X_i - \bar{X}}{s_X}\right)\left(\frac{Y_i - \bar{Y}}{s_Y}\right)$$

$$Where \quad \frac{X_i - \bar{X}}{s_X}, \bar{X}, and\ s_X$$

are denoted as the standard score, sample mean and sample standard deviation.

Attribute Clustering

After the computation of correlations of each pair of attributes, by utilizing the clustering to partition the attributes into columns. In this algorithm, each and every attribute is a point in the clustering space. By utilizing pearson and chi-squared based on the coefficient and then the original attribute correlation coefficient is based on the attribute coefficient.Pearson based l-diversity: Pearson based l-diversity for finding the correlation between attributes is dissimilar from clustering. By utilizing pearson correlation coefficient to calculate the attribute correlation, and the horizontal partitioning is done based on the l-diversity and at the end, the vertical partitioning is done based on attribute correlation coefficient.

Step 1: To evaluate the associations between two continuous attributes first to define pearson correlation coefficient.
Step 2: Horizontal partitioning based on l-diversity
Step 3: Attribute clustering for vertical partitioning is based on attribute correlation coefficient to evaluate correlations between the sensitive attribute SA and each attribute, I identifier is defined below:
The distance between two attributes in the clustering space is defined as,

$$d(A_1, A_2) = 1 - \Phi^2(A_1, A_2)$$ Between 0 and 1.

Chi-squared based l-diversity: By using chi-squared based l-diversity for finding the correlation between attributes is dissimilar from normal attribute clustering. First to compute the attribute correlation using chi-squared and then horizontal partitioning is accomplished based on l-diversity slicing and lastly vertical partitioning is done based on attribute correlation coefficient.

Step 1: To define chi-squared correlation coefficient is used for evaluating correlations between two continuous attributes
Step 2: Horizontal partitioning based on l-diversity
Step 3: Attribute clustering for vertical partitioning is based on attribute correlation coefficient for calculates correlations between the sensitive attribute SA and each attribute, identifier is defined below:
The distance between two attributes in the clustering space is defined as

$$d(A_1, A_2) = 1 - \Phi^2(A_1, A_2)$$ Between 0 and 1.

Dimensionality Checking

By using the greedy partitioning algorithm, the degree of dimensionality flexibility can be attained. Mondrian top-down greedy algorithm is utilized for the dimensional strategies in the overlapping slicing. There is no generalization is helpful to the tuples in the T. In the first step the dimensional regions are defined in order to cover the domain space. In the second step the overlapping slicing functions are created by using the sliced summary functions from the region.

Algorithm: Dimensional (partition)
If (no allowable high dimensional cut for partition)
Return Φ : partition $\rightarrow$ sliced summary

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Else
dim← choose_dimension()
fs ← frequency_set(partition,dim)
splitVal ← find_mediod(fs)
lhs ← {t ∈ partition:t.dim≤ splitVal}
rhs ← {t ∈ partition:t.dim> splitVal}
return Dimension (rhs) ∪ Dimension(lhs)

Tuple partitioning: In this phase, tuples are partitioned into buckets. The Mondrian is utilized for partitioning tuples into buckets. This algorithm maintains two data structures; 1) a queue of buckets 2) a set of sliced buckets. Primarily, queue includes only one bucket which contains all tuples and there is empty in the sliced buckets (line 1). In each and every iteration (lines 2 to 7), the algorithm eliminates a bucket from Q and splits the bucket into two buckets. If the sliced table after the split satisfies '-diversity (line 5), then the algorithm puts the two buckets at the end of the queue Q (for more splits, line 6). Or else, we cannot split the bucket to any further extent and the algorithm puts the bucket into SB (line 7). While Q becomes vacant, to compute the sliced table.

Algorithm tuple-partition (T,l)

1. Q= {T}; SB=Φ
2. While Q is not empty
3. Remove the first bucket B from Q; Q=Q-B;
4. Split B into two buckets B_1 and B_2 as Mondrian
5. if diversity –check (T,Q∪ { B_1, B_2 }
6. else
7. SB=SB∪ {B}
8. return SB.

The main part of the tuple-partition algorithm is to check whether a sliced table satisfies 'l-diversity (line 5) the diversity-check algorithm.

Algorithm diversity-check (T, T^*,ℓ)

1. for each tuple t∈ T, L[t] =Φ
2. for each bucket B in T^*
3. record f(v) for each column value v in bucket B
4. for each tuple t
4. for each tuple t∈ T
5. calculate p(t, B) and find D(t, B)
6. L[t]=L[t] ∪{$< p(t,B), D(t,B) >$}
7. for each tuple t∈ T
8. Calculate p(t, s) for each s based on L[t].
9. if p (t, s) ≥ 1/ℓ return false
10. return true

Tuple Grouping: To group attributes $A_1, \ldots, A_p$ into clusters, we build our information-theoretic attribute clustering algorithm by converting the popular k-means algorithm into what we call the k-modes algorithm by replacing: 1) the concept of the term "mean," which represents the center of a cluster of entities, by the concept of mode which is the attribute with the highest multiple interdependence within an attribute group and 2) the distance measure used in k-means

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

by the interdependence redundancy measure between attributes. We can then formulate the k-modes algorithm in the following.

Initialization: To presume that the number of clusters, k, in which k is an integer greater than or equal to 2, is given. Of the p attributes, we arbitrarily choose k attributes, each of which represents a candidate for a mode $\eta_r, r \in \{1, \ldots k\}$.r Formally, we have $\eta_r = A_i, r \in \{1, \ldots k\}, i \in \{1, \ldots p\}$ to the mode of C_r and $\eta_r \neq \eta_s$ for all s $\in \{1, \ldots k\} - \{r\}$.

Assignment of every attribute to one of the clusters: For each attribute, A_i, i ∈ {1, …, p}, and each cluster mode, η_r, r ∈ {1, …, k}, we calculate the interdependence redundancy measure between A_i and η_r, $R(A_i : \eta_r)$. To allocate A_i to Cr if $R(A_i : \eta_r) \geq R(A_i : \eta_s)$ for all s ∈ {1, …, k} – {r}.Calculation of mode for every attribute cluster: For each cluster, Cr, r ∈ {1… k}, we set $\eta_r = A_i$ if $MR(A_i) \geq MR(A_j)$ for all A_i, $A_j \in Cr$, $i \neq j$.

Termination: Until η_r for the clusters does not vary the step 2 and 3 are repeated. On the other hand, this algorithm also terminates when the pre-specified number of iterations is attained.

Noise enabled Slicing: Membership disclosure protection is provided by the slicing. The bucketization releases each tuple's combination of values in their original form and most individuals can be uniquely identified using the values, the attacker can decide the membership of an individual in the original data by examining whether the individual's combination of values occurs in the released data. Slicing provides protection against membership disclosure because attributes are partitioned into different columns and correlations among different columns within each bucket are busted.

For analyzing the result of the above slicing method and improve the clustering result we need to changes in the horizontal partitioning and analysis the result by adding noise data to original data table and finally perform the vertical partitioning for different cases for both person based and chi square based slicing .

Pearson based Noise enabled slicing: In this step Pearson based Noise enabled slicing first step attribute partitioning of the horizontal is done by using EM clustering, after the EM clustering performed then perform person correlation coefficient for above attributes, then add 10 % noise to sensitive attributes to the origin table return from step2 .Finally vertical partitioning based on the case A, case B.

Step 1: EM Clustering for horizontal partitioning
Step 2: Pearson correlation coefficient for attributes from horizontal partitioning
Step 3: Adding 10% noise to sensitive attribute (SA).
Step 4: Vertical partitioning based on case A and case B.

Chi-square based Noise enabled slicing: In this step chi-square based noise enabled slicing first step attribute partitioning of the horizontal is done by using EM clustering, after the EM clustering performed then perform person correlation coefficient for above attributes, then add the noise sensitive attributes to the origin table return from step2 .Finally vertical partitioning based on the case A, case B

Step 1: EM Clustering for horizontal partitioning
Step 2: Chi-square correlation coefficient for attributes from horizontal partitioning
Step 3: Adding 10% noise to sensitive attribute (SA).
Step 4: Vertical partitioning based on case A and case B.

IV. EXPERIMENTAL RESULTS

Experimental results which evaluates the performance of the existing and the proposed system. For that we take the Netflix prize data set which includes 100,480,507 ratings of 17,770 movies given by 480,189 Netflix subscribers. Every rating has the subsequent design: (user ID, movie ID, rating, date), where rating is an integer in {0, 1, 2, 3, 4, and 5} with 0 being the lowest rating and 5 being the highest rating. In this section the existing and the proposed system is compared in terms of computational efficiency and accuracy. To compare slicing with generalization and bucketization in terms of computational efficiency.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

Computational Efficiency

Fig. 1a shows the computational time as a function of data cardinality where data dimensionality is fixed as 15 (i.e., we use (subsets) of the OCC-15 data set). Fig. 1b shows the computational time as a function of data dimensionality in which the data cardinality is set as 45,222. The results show that the slicing algorithm scales well with both data cardinality and data dimensionality.

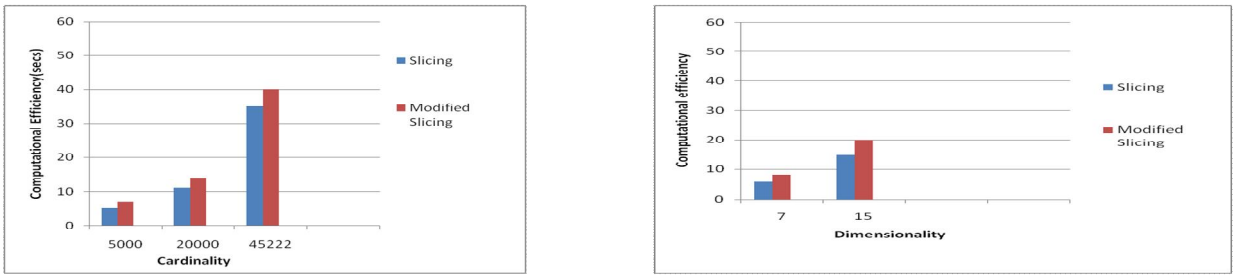

Figure 1. Computational Efficiency (a) Cardinality (b) Dimensionality

Accuracy

The accuracy is compared for the existing and the Modified slicing method. In this experiment, to evaluate the classification accuracy. When compared to the slicing method the accuracy is high in the modified slicing.

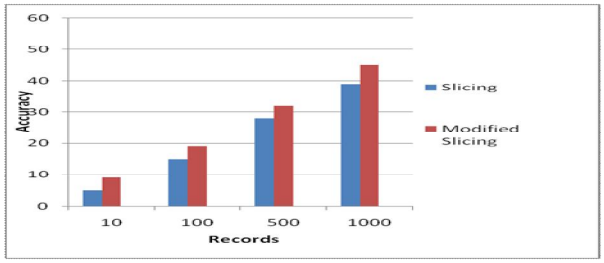

Figure 2. Accuracy

V. CONCLUSION AND FUTURE WORK

A new approach called slicing for privacy preserving in micro data publishing. By using pearson and chi-squared based correlation coefficient the correlations for each pair of attributed are computed. In this method, each attribute is a point in the clustering space. By utilizing the Pearson and chi-squared based correlation coefficient the distance between two attributes in the clustering space is computed and then the original attribute correlation coefficient is based on the attribute coefficient. Slicing preserves better utility and privacy than generalization and bucketization.

To increase the data utility, an overlapping slicing method is used in which duplicates an attribute in more than one column. In addition to that, improve the clustering result we need to changes in the horizontal partitioning and analysis the result by adding noise data to original data table and finally perform the vertical partitioning for different cases for both Pearson based and chi square based slicing. For future work, a number of anonymization techniques have been designed; it remains an open problem on how to use the anonymized data. Another direction is to design data mining tasks using the anonymized data computed by various anonymization techniques.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization) Vol.2, Special Issue 1, March 2014

Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department of CSE, JayShriram Group of Institutions, Tirupur, Tamilnadu, India on 6th & 7th March 2014

REFERENCES

[1] P. Samarati, "Protecting Respondent's Privacy in Micro data Release," IEEE Trans. Knowledge and Data Eng., vol. 13, no. 6, pp. 1010-1027, Nov./Dec. 2001.
[2] B.C.M. Fung, K. Wang, and P.S. Yu, "Top-Down Specialization for Information and Privacy Preservation," Proc. Int'l Conf. Data Eng. (ICDE), pp. 205-216, 2005.
[3] G. Ghinita, Y. Tao, and P. Kalnis, "On the Anonymization of Sparse High-Dimensional Data," Proc. IEEE 24th Int'l Conf. Data Eng. (ICDE), pp. 715-724, 2008.
[4] D. Kifer and J. Gehrke, "Injecting Utility into Anonymized Data Sets," Proc. ACM SIGMOD Int'l Conf. Management of Data (SIGMOD), pp. 217-228, 2006.
[5] N. Koudas, D. Srivastava, T. Yu, and Q. Zhang, "Aggregate Query Answering on Anonymized Tables," Proc. IEEE 23rd Int'l Conf. Data Eng. (ICDE), pp. 116-125, 2007.
[6] C. Aggarwal, "On k-Anonymity and the Curse of Dimensionality," Proc. Int'l Conf. Very Large Data Bases (VLDB), pp. 901-909, 2005.
[7] B.-C. Chen, K. LeFevre, and R. Ramakrishnan, "Privacy Skyline: Privacy with Multidimensional Adversarial Knowledge," Proc. Int'l Conf. Very Large Data Bases (VLDB), pp. 770-781, 2007.
[8] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating Noise to Sensitivity in Private Data Analysis," Proc. Theory of Cryptography Conf. (TCC), pp. 265-284, 2006.