

Privacy Preservation Using L-Diversity

A.S. Praveenkumar¹, S. Kousalya², K. ArulMani³, Dr. K. Thirukumar⁴

Department of Computer Science and Engineering,

Dr. Mahalingam College of Engineering and Technology, Pollachi, Coimbatore, India^{1,2,3}

Associate professor, Department of Computer Science and Engineering,

Dr. Mahalingam College of Engineering and Technology, Pollachi, Coimbatore, India⁴

Abstract: An increasing number of organizations maintain collections of data about individuals. Hospitals keep medical records of their patients, commerce companies collect information of their clients and web service companies keep track of the preferences of their users. Publication of these data can be useful for research, epidemic studies, commerce development, statistical analysis, etc. It is difficult to keep all the data open on internet, in some cases, to keep confidence and in safe manner where it contains some important or sensitive details. The removal of directly identifying information (such as Name, Disease) from the published records is not enough to guarantee individuals privacy. The personal data can be misused, for a variety of purposes. Maintaining the privacy for high dimensional database has become difficult. A potential attacker could infer a record's identity by linking public external information sources (voters registration lists, phone number catalogues, etc.) with a combination of other attributes, like age, gender and postal code, which are not generally unique per person. The main goal is to focus on Privacy and utility of the shared data. l - Diversity is one of the methods to preserve the data which is very sensitive and confidential. Keeping more sensitive data in dataset helps us to preserve the database more safe and secure manner. The advantage of using l-diversity provides a greater distribution of sensitive attributes within the group, thus increasing data protection. This method protects against attribute disclosure, which is an enhancement of k-anonymity technique.

Keywords: Sensitive Attribute, Privacy Preserving Data, Adult Dataset, l-Diversity Model, Quasi Identifier, Utility, Sensitive Value

I. INTRODUCTION

As the technologies are rapidly expanding, so are the various cyber-crimes such as internet phishing where the confidential information is violated and this raises concerns of preserving data privacy and security among users and enterprises globally. Besides, the use of social network sites, electronic healthcare systems, online trading, etc. has generated a large number of datasets that constitutes big data. The analysis performed on patients' data causes privacy and security concerns at stages such as data collection, storage, and processing. Thus, there is a high demand for Privacy Preserving Data Publishing for protected data sharing via the internet. To protect the privacy of the data proprietor, multitude de-identification and anonymization techniques are applied before the data is released to the public or for secondary use.

Due to the rapid growth in information technologies, companies at the present time collect and store huge amounts of information in their databases. Typically, such information is stored in the form of tables and each record is corresponding to an individual. Every record has a number of attributes which can be divided into three categories: 1. Explicit identifiers which can clearly identify individuals. 2. Quasi Identifying attributes whose values when taken can easily identify individual identities. 3. Sensitive Attribute which are considered sensitive and need not be disclosed [3].

Publishing data about individuals without revealing sensitive information about them is an important problem. In recent years, a new definition of privacy called k-anonymity has gained popularity. In a k-anonymized dataset, each record is indistinguishable from at least k-1 other records with respect to certain "identifying" attributes. Two simple attacks that a k-anonymized dataset has some subtle, but severe privacy problems First, an attacker can discover the values of sensitive attributes when there is little diversity in those sensitive attributes. Second, attackers often have background knowledge, and k-anonymity does not guarantee privacy against attackers using background knowledge [7]. Detailed analysis of these two attacks and propose a novel and powerful privacy definition called l-diversity. In addition to building a formal foundation for l-diversity, an experimental evaluation that l-diversity is practical and can be implemented efficiently.

A. Objective

The purpose of this project is to improve the privacy of data which provides a greater distribution of sensitive attributes within the group. Also to focus on privacy and utility of the shared data. There is always a trade-off between privacy and utility, focus more on privacy the utility will be low and vice versa.

B. Overview

In introduction described about privacy preservation and its application where it has been used. Then brief description about objective. Related works have been listed. And existing method is proposed where the flow diagram, methodology, dataset and expected results are listed.

II. RELATED WORKS**A. Privacy-preserving data publishing**

Anonymization refers to the privacy preserving in data publishing approach that seeks to hide the identity and/or the sensitive data of record owners, assuming that sensitive data must be retained for data analysis. Clearly, explicit identifiers of record owners must be removed. Even with all explicit identifiers being removed, an individual's name in a public voter list was linked with his record in a published medical database through the combination of zip code, date of birth, and sex. Each of these attributes does not uniquely identify a record owner, but their combination, called the quasi identifiers often singles out a unique or a small number of record owners [1].

In the above example, the owner of a record is re-identified by linking his quasi identifier. To perform such linking attacks, the attacker needs two pieces of prior knowledge: the victim's record in the released data and the quasi-identifiers of the victim. Such knowledge can be obtained by observation. For example, the attacker noticed that his boss was hospitalized, and therefore knew that his boss's medical record would appear in the released patient database. Also, it was not difficult for the attacker to obtain his boss's zip code, date of birth, and sex, which could serve as the quasi identifiers in linking attacks. Having a better understanding of the privacy problem from different perspectives can help realize successful applications of privacy-preserving technology.

B. Privacy-Preservation Methods

Nowadays, a variety of everyday life activities such as medical examinations, credit-card purchases, social-network interactions, internet browsing, web searching, etc., result to the collection of individuals' personal information. Data distribution could be either open, for example uploaded on a webpage on the internet, or restricted to a third organization or research group. In either case, the data holder must ensure the privacy of individuals whose personal information is included in the released dataset [2].

The replacement of a value of an attribute by a more general value or interval that contains the original value is called a generalization [48, 49, 53]. For arithmetic attributes the use of or arithmetic intervals is common (ex. Age = 25 replaced by [20, 30]). The removal of a piece of information from the released dataset is called suppression. A whole record, or individual attributes can be suppressed in the anonymization process. The suppression of an attribute value can be perceived as a generalization to the root level, as this replacement reveals no information of the replaced value.

C. Anonymization in Privacy Preserving Data Mining

Anonymization method aims at making the individual record be indistinguishable among a group records by utilizing techniques of generalization and suppression. Anonymization refers to a methodology where character or/and delicate data about record holders are to be covered up. It even accepts that delicate data should be retained for analysis [4]. There are four sort of quality of fundamental type of data:

Explicit Identifiers is a situated of properties containing information that recognizes a record manager explicitly, for example, name, percentage and so forth. Quasi Identifiers is a situated of properties that could potentially recognize a record manager when combined with publicly available data. Sensitive Attributes is a situated of properties that contains touchy individual particular information, for example, illness, salary and so forth. Non-Sensitive Attributes is a situated of properties that makes no problem if revealed even to conniving gatherings [4].

D. l-Diversity: Privacy beyond k-anonymity

To avoid attack on k-anonymous dataset l-diversity framework that gives stronger privacy guarantee. k-anonymity can create group that leak information due to lack of diversity in sensitive Attribute k anonymity does not protect against at attack based on background knowledge. Tool for reasoning privacy, theoretical principal of privacy, difficulties that need to be overcome the practical definition of privacy, in privacy principle there is positive disclosure & negative disclosure. Adversary can correctly identifies value of sensitive attribute with high probability in positive disclosure. Adversary can correctly eliminate value of sensitive attribute. In uninformative principle little additional knowledge beyond background knowledge [11]

Block is l -diverse if it contains at least well represented values of sensitive attributes. Implementing privacy preserving data publishing table preserve privacy then generalization of table also preserve privacy. Sets of attributes (like gender, date of birth, and zip code) that can be linked with external data to uniquely identify individuals in the population are called quasi-identifiers. To counter linking attacks using quasi-identifiers, definition of privacy called k -anonymity [8].

E. l - Diversity by Generalization Algorithm

In Relational database, to avoid identity disclosure one record in table has same Quasi identifier at least $k-1$ record, to implement l -diversity. Minimal generalization used when data are not generalized more than necessary to provide k -anonymity. Algorithms for minimal generalization that satisfy prefer criteria. Quasi identifier a set of attributes in private table can linked with external information to re identify respondent to whom information refer [12].

Each release data must be such that combination of values of quasi identifier match at least k individual. In generalization five digits zip codes generalized to 4 digits & 3 digits. Value in private table can be substituted upon release with generalized value. Suppression is removing data from table so it can be released. An algorithm for minimal generalization is proposed in which lowest height among all generalization in return. It's having smaller number of generalization step. Both generalization and suppression technique important for protection of data

III. METHODOLOGY

In recent years, data mining has been viewed as a threat to privacy because of the widespread proliferation of electronic data maintained by corporations. This may lead to increased concerns about the privacy of the underlying data. In recent years, a number of techniques have been proposed for modifying or transforming the data in such a way so as to preserve privacy. A survey on some of the techniques used for privacy-preserving data mining may be found.

The k -anonymity model was developed because of the possibility of indirect identification of records from public databases. This is because combinations of record attributes can be used to exactly identify individual records. In the k -anonymity method, reduce the granularity of data representation with the use of techniques such as generalization and suppression. This granularity is reduced sufficiently that any given record maps onto at least k other records in the data. The l -diversity model was designed to handle some weaknesses in the k -anonymity model since protecting identities to the level of k -individuals is not the same as protecting the corresponding sensitive values, especially when there is homogeneity of sensitive values within a group.

A. Existing System

At initially, taken input data as adult dataset and it has fourteen attributes. The data which are taken have some errors or missing data in it. Metadata from the real dataset is loaded into the execution environment. The loaded dataset needs to be preprocessed by the pre-processing module. The module truncates the most unwanted symbols from the dataset. Now preprocessed data set is ready. Then, identifying the sensitive value and going to be anonymized the data. By applying l -diversity, the given datasets are divided into buckets and the values are interchanged within their field. Some other fields are changed in the output (For example: Age=35 will be displayed as Age=30-40). Then the values are preserved and privacy is implied.

1. Data Preprocessing

In real world data are generally incomplete, noisy, inconsistency. In the adult data set, there are some missing values and errors. It will be rectified through preprocessing. Some preprocessing steps must be applied on the anonymized data before it can be used for workload tasks. Metadata from the real dataset is loaded into the execution environment. The loaded dataset needs to be preprocessed by the pre-processing module. The module truncates the most unwanted symbols from the dataset. Preprocessed dataset needs to be used for anonymizing. Grouping values can be done by clustering and also Normalization can be done by attribute values to fall within the specified range.

2. l -diversity Method

l - Diversity is a form of group based anonymization that is used to preserve privacy in data sets by reducing the granularity of a data representation. This reduction is a tradeoff that results in some loss of effectiveness of data management or mining algorithms in order to gain some privacy. The l -diversity model is an extension of the k -anonymity model which reduces the granularity of data representation using techniques including generalization and suppression such that any given record maps onto at least $k-1$ other records in the data. The l -diversity model adds the promotion of intra-group diversity for sensitive values in the anonymization mechanism.

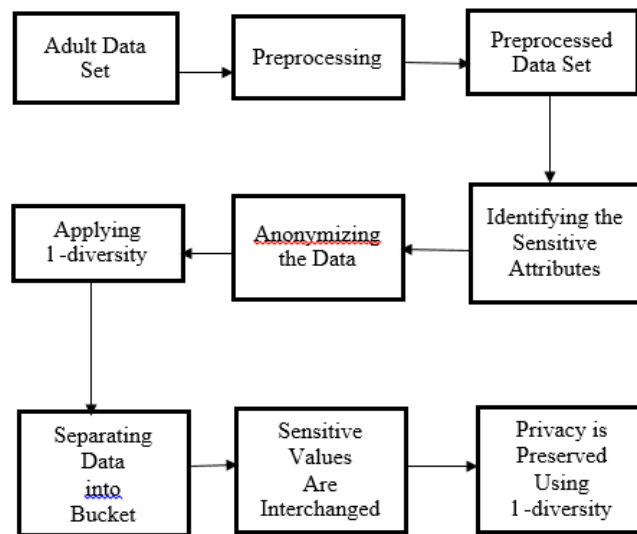


Figure 1 Block Diagram of Existing System

2.1 Tuple Partitioning Algorithm

In the tuple partitioning phase, tuples are partitioned into buckets. We modify the Mondrian algorithm for tuple partition. The below algorithm gives the description of the tuple-partition algorithm. The algorithm maintains two data structures: (1) a queue of buckets Q and (2) a set of sliced buckets SB . Initially, Q contains only one bucket which includes all tuples and SB is empty (line 1).

Algorithm tuple-partition (T, ℓ)

1. $Q = \{T\}; SB = \emptyset$.
2. while Q is not empty
3. remove the first bucket B from Q ; $Q = Q - \{B\}$.
4. split B into two buckets $B1$ and $B2$, as in Mondrian.
5. if diversity-check ($T, Q \cup \{B1, B2\} \cup SB, \ell$)
6. $Q = Q \cup \{B1, B2\}$.
7. else $SB = SB \cup \{B\}$.
8. return SB .

In each iteration (line 2 to line 7), the algorithm removes a bucket from Q and splits the bucket into two buckets. If the sliced table after the split satisfies ℓ -diversity (line 5), then the algorithm puts the two buckets at the end of the queue Q (for more splits, line 6). Other- wise, we cannot split the bucket anymore and the algorithm puts the bucket into SB (line 7). When Q becomes empty, we have computed the sliced table. The set of sliced buckets is SB (line 8). The main part of the tuple-partition algorithm is to check whether a sliced table satisfies ℓ -diversity (line 5)

2.2 Diversity-check Algorithm

The below algorithm gives a description of the diversity-check algorithm For each tuple t , the algorithm maintains a list of statistics $L[t]$ about t 's matching buckets. Each element in the list $L[t]$ contains statistics about one matching bucket B : the matching probability $P(t, B)$ and the distribution of candidate sensitive values $D(t, B)$. The algorithm first takes one scan of each bucket B (line 2 to line 3) to record the frequency $f(v)$ of each column value v in bucket B .

Algorithm Diversity-check (T, T^*, ℓ)

1. for each tuple $t \in T, L[t] = \emptyset$.
2. for each bucket B in T
3. record $f(v)$ for each column value v in bucket B .
4. for each tuple $t \in T$
5. calculate $p(t, B)$ and find $D(t, B)$.
6. $L[t] = L[t] \cup \{hp(t, B), D(t, B)\}$.
7. for each tuple $t \in T$
8. calculate $p(t, s)$ for each s based on $L[t]$.
9. if $p(t, s) \geq 1/\ell$, return false.
10. return true.

Then the algorithm takes one scan of each tuple t in the table T (line 4 to line 6) to find out all tuples that match B and record their matching probability $P(t, B)$ and the distribution of candidate sensitive values $D(t, B)$, which are added to the list $L[t]$ (line 6).

Table 1 The Original table

Age	Nationality	Occupation
25-30	India	Doctor
30-35	London	Business
20-25	China	Engineer
30-35	Russia	Doctor
45-50	India	Business
45-50	America	Sales-Rep
55-60	Russia	Professor
50-55	London	Driver
30-35	America	Professor
35-40	Africa	Accountant
35-40	India	Business
35-40	Singapore	Engineer

At the end of line 6, we have obtained, for each tuple t , the list of statistics $L[t]$ about its matching buckets. The sliced table is ℓ -diverse if for all sensitive value s , $p(t, s) \leq 1/\ell$ (line 7 to line 10)

The table 1 consists of three attributes such as Age, Nationality and Occupation. In the table 1, age and nationality are quasi identifiers and occupation is kept as sensitive attribute. The size of the bucket kept here is 3. Using 1-diversity method, quasi identifier Age value is kept within the range (For example, if age =32 it is kept as 30-35) nationality remains same. Then the sensitive information occupation is shuffled within the bucket and the data is preserved. In table 2, age is kept between some range (30-35) and the sensitive attribute occupation is shuffled within the bucket and the privacy is obtained.

Table 2 Example for 3-diverse

Quasi identifier		Sensitive attribute
Age	Nationality	Occupation
28	India	Doctor
30	America	Engineer
35	Africa	Doctor
32	London	Business
46	India	Sales-Rep
49	America	Driver
56	Russia	Business
53	London	Professor
22	China	Accountant
30	Russia	Professor
39	India	Engineer
36	Singapore	Business

IV. EXPERIMENTS

ℓ -diversity framework that gives stronger privacy guarantees. There are several avenues for future work. First, extend initial ideas for handling multiple sensitive attributes, and to develop methods for continuous sensitive attributes. Second, although privacy and utility are duals of each other, privacy has received much more attention than the utility of a published table.

The identified sensitive data is interchanged within the given bucket for preserving privacy. One of the quasi identifiers is taken as some specific range (Example: Age=32, is kept in the range 30-35). The remaining quasi identifiers are proceeded depending upon the attribute. Consider adult dataset which is a set of demographic records of 32,561 people. In the Adult dataset (Table 3), there is a 15 attributes and one attribute as a sensitive attribute out of 15 attributes.

Table 3 Description of the Adult dataset

Attributes	Description
Age	Continuous
Work class	Categorical
Fnlwgt	Continuous
Education	Categorical
Education-num	Continuous
Marital status	Categorical
Occupation	Categorical
Relationship	Categorical
Race	Categorical
Gender	Categorical
Capital-gain	Continuous
Capital-loss	Continuous
Hours-per-week	Continuous
Native Country	Categorical
Salary	Categorical

As an input adult dataset is taken, where the data is preprocessed The tuples with missing values are eliminated. Occupation is considered as a sensitive attribute(S). Then identified sensitive data have to be preserved. Extensions of k-anonymity method, which can ensure data privacy even without identifying the enemy’s background knowledge to avoid attribute disclosure. And the tuples are split into buckets. Within the buckets, values are interchanged and it is preserved.

An equivalence class is said to have l-diversity if there are at least l “well represented” values for the sensitive attribute. A table is said to have l-diversity if every equivalence class of the table has l-diversity.

V. RESULTS

A. Evaluation Metrics

Privacy-preserving show several changes in terms of the efficiency and effectiveness with respect to the original data set. The following metrics to measure the changes quantitatively: cell overlap, and information loss.

Information Loss: Anonymization that involves data transformation for the sake of privacy preservation generates inevitable information loss. The transformation of values caused by the generalization methods and the weakening of the relationship between attributes caused by the bucketization method are typical types of the information loss. Evaluating information loss is important because the quality of privacy-preserving data analysis depends on information loss. Two different metrics: the loss metric (LM), and media relative error.

LM measures the degree of the generalization. The value of LM indicates the proportion of lost information and varies from 0 to 1. The LM of a numeric attribute is defined as follows: where U_i and L_i are the upper and lower bounds of the generalized interval, respectively, and U and L are the maximum and minimum values of the whole domain, respectively. For example, the LM for an anonymized age {20–29} is computed as follows:

$$LMa(anum) = \frac{U_i - L_i}{U - L}$$

Equation 1 Loss Metric continuous attribute

$$LM a(age) = \frac{29 - 20}{99 - 0} = \frac{9}{99} = 0.09$$

Finally, the LM of a cell is defined as where n is the number of quasi-identifier attributes in the data cube.

$$LM = \frac{1}{n} \sum_{i=1}^n LMa(a_i)$$

Equation 2 Loss Metric of a cell

The relative error is then computed as

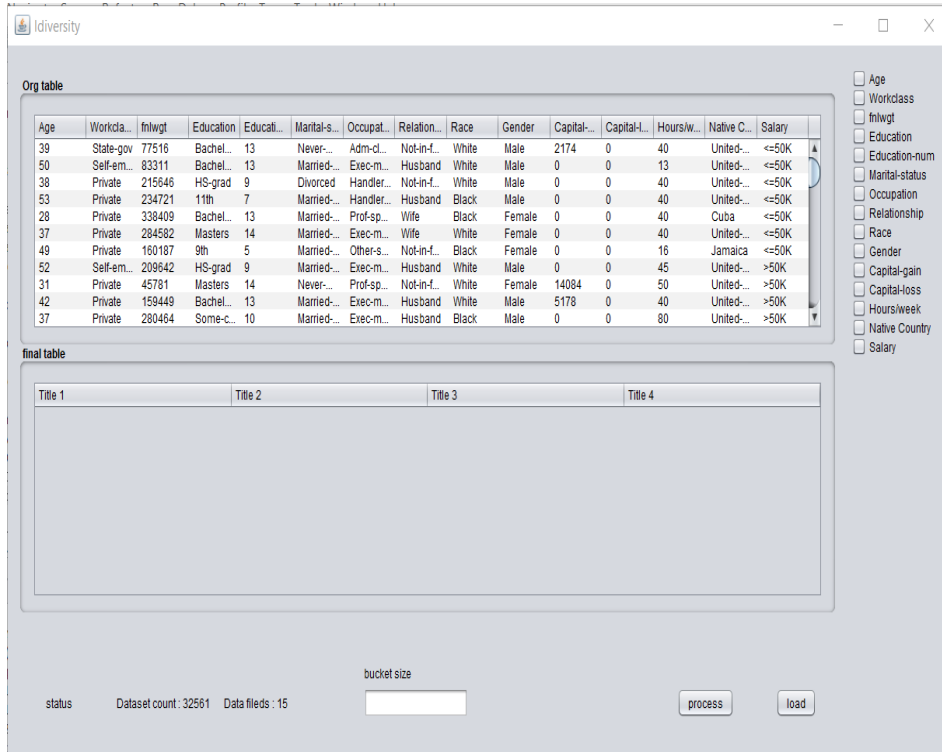
$$RE = \frac{|act - est|}{act} \tag{3}$$

Equation 3 Relative error

where act is the result of the original data cube and est is the result of a privacy-preserving data cube and repeat the random queries 1,000 times for each cuboid and finally obtain the median value of the relative errors. By measuring the

relative error, the information loss of bucketization can also be evaluated. Note that LM cannot evaluate the information loss of data cubes constructed from the bucketization methods, which do not generalize quasi-identifiers.

B. Snapshots



Org table

Age	Workcla...	fnlwtg	Education	Educati...	Marital-s...	Occupat...	Relation...	Race	Gender	Capital...	Capital-H...	Hours/w...	Native C...	Salary
39	State-gov	77516	Bachel...	13	Never...	Adm-cl...	Not-in-f...	White	Male	2174	0	40	United...	<=50K
50	Self-em...	83311	Bachel...	13	Married...	Exec-m...	Husband	White	Male	0	0	13	United...	<=50K
38	Private	215646	HS-grad	9	Divorced	Handler...	Not-in-f...	White	Male	0	0	40	United...	<=50K
53	Private	234721	11th	7	Married...	Handler...	Husband	Black	Male	0	0	40	United...	<=50K
28	Private	338409	Bachel...	13	Married...	Prof-sp...	Wife	Black	Female	0	0	40	Cuba	<=50K
37	Private	284582	Masters	14	Married...	Exec-m...	Wife	White	Female	0	0	40	United...	<=50K
49	Private	160187	9th	5	Married...	Other-s...	Not-in-f...	Black	Female	0	0	16	Jamaica	<=50K
52	Self-em...	209642	HS-grad	9	Married...	Exec-m...	Husband	White	Male	0	0	45	United...	>50K
31	Private	45781	Masters	14	Never...	Prof-sp...	Not-in-f...	White	Female	14084	0	50	United...	>50K
42	Private	159449	Bachel...	13	Married...	Exec-m...	Husband	White	Male	5178	0	40	United...	>50K
37	Private	280464	Some-c...	10	Married...	Exec-m...	Husband	Black	Male	0	0	80	United...	>50K

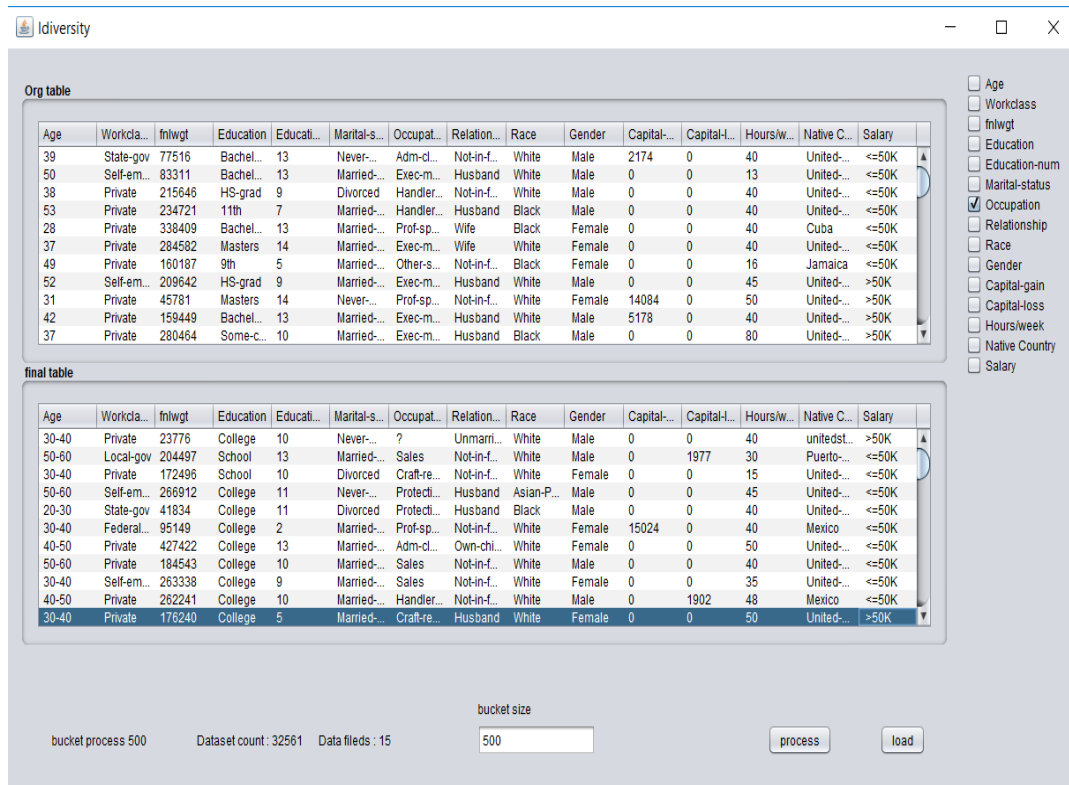
final table

Title 1	Title 2	Title 3	Title 4

bucket size:

status: Dataset count: 32561 Data files: 15

Figure 2 Original Dataset



Org table

Age	Workcla...	fnlwtg	Education	Educati...	Marital-s...	Occupat...	Relation...	Race	Gender	Capital...	Capital-H...	Hours/w...	Native C...	Salary
39	State-gov	77516	Bachel...	13	Never...	Adm-cl...	Not-in-f...	White	Male	2174	0	40	United...	<=50K
50	Self-em...	83311	Bachel...	13	Married...	Exec-m...	Husband	White	Male	0	0	13	United...	<=50K
38	Private	215646	HS-grad	9	Divorced	Handler...	Not-in-f...	White	Male	0	0	40	United...	<=50K
53	Private	234721	11th	7	Married...	Handler...	Husband	Black	Male	0	0	40	United...	<=50K
28	Private	338409	Bachel...	13	Married...	Prof-sp...	Wife	Black	Female	0	0	40	Cuba	<=50K
37	Private	284582	Masters	14	Married...	Exec-m...	Wife	White	Female	0	0	40	United...	<=50K
49	Private	160187	9th	5	Married...	Other-s...	Not-in-f...	Black	Female	0	0	16	Jamaica	<=50K
52	Self-em...	209642	HS-grad	9	Married...	Exec-m...	Husband	White	Male	0	0	45	United...	>50K
31	Private	45781	Masters	14	Never...	Prof-sp...	Not-in-f...	White	Female	14084	0	50	United...	>50K
42	Private	159449	Bachel...	13	Married...	Exec-m...	Husband	White	Male	5178	0	40	United...	>50K
37	Private	280464	Some-c...	10	Married...	Exec-m...	Husband	Black	Male	0	0	80	United...	>50K

final table

Age	Workcla...	fnlwtg	Education	Educati...	Marital-s...	Occupat...	Relation...	Race	Gender	Capital...	Capital-H...	Hours/w...	Native C...	Salary
30-40	Private	23776	College	10	Never...	?	Unmarr...	White	Male	0	0	40	unitedst...	>50K
50-60	Local-gov	204497	School	13	Married...	Sales	Not-in-f...	White	Male	0	1977	30	Puerto...	<=50K
30-40	Private	172496	School	10	Divorced	Craft-re...	Not-in-f...	White	Female	0	0	15	United...	<=50K
50-60	Self-em...	266912	College	11	Never...	Protect...	Husband	Asian-P...	Male	0	0	45	United...	<=50K
20-30	State-gov	41834	College	11	Divorced	Protect...	Husband	Black	Male	0	0	40	United...	<=50K
30-40	Federal...	95149	College	2	Married...	Prof-sp...	Not-in-f...	White	Female	15024	0	40	Mexico	<=50K
40-50	Private	427422	College	13	Married...	Adm-cl...	Own-chi...	White	Female	0	0	50	United...	<=50K
50-60	Private	184543	College	10	Married...	Sales	Not-in-f...	White	Male	0	0	40	United...	<=50K
30-40	Self-em...	263338	College	9	Married...	Sales	Not-in-f...	White	Female	0	0	35	United...	<=50K
40-50	Private	282241	College	10	Married...	Handler...	Not-in-f...	White	Male	0	1902	48	Mexico	<=50K
30-40	Private	176240	College	5	Married...	Craft-re...	Husband	White	Female	0	0	50	United...	>50K

bucket size:

bucket process 500 Dataset count: 32561 Data files: 15

Figure 3 Resultant Dataset

CONCLUSION

Preserving the sensitive information about an individual in the published data is an important aspect in data mining. Though the l -diversity approach avoids the homogeneity attack in the equivalence classes by rearranging the records in such a way that the information loss is less, reordering the records leads to probability inference attack in the equivalence classes since the probability of the sensitive attribute increases in the equivalence class where the record is reordered. Thus, to create an algorithm for l -diversity, simply take any algorithm for k -anonymity and make the following change: every time a table T is tested for k -anonymity and check for l -diversity instead. Since l -diversity is a property that is local to each q block and since all l -diversity tests are solely based on the counts of the sensitive values, this test can be performed very efficiently.

REFERENCES

- [1]. Benjamin C.M. Fung, Ke Wang, Rui Chen, "Introduction to privacy-preserving data publishing: Concepts and techniques". ACM Computing Surveys, Vol. 42, June 2010.
- [2]. Gkountouna, Olga, "A Survey on Privacy Preservation Methods". Technical Report, Knowledge and Database Systems Laboratory, National Technical University of Athens (NTUA), Vol 6, June 2011.
- [3]. Arora, Dilpreet Kaur, Divya Bansal, and Sanjeev Sofat, "Comparative Analysis of Anonymization Techniques", International Journal of Electronic and Electrical Engineering, Vol 7, 773-778, 2014.
- [4]. Presswala, Freny, Amit Thakkar, and Nirav Bhatt. "Survey on anonymization in privacy preserving data mining", International Journal of Innovative and Emerging Research in Engineering, Vol 2, Issue 2, 2015.
- [5]. Kohlmayer, Florian, Fabian Prasser, and Klaus A. Kuhn, "Implementing generalization and suppression for anonymizing biomedical data with minimal information loss", Journal of biomedical informatics, Vol 58, 37-48, September 2015.
- [6]. Rajesh, N., K. Sujatha, and A. Arul Lawrence, "Survey on Privacy Preserving Data Mining Techniques using Recent Algorithms" International Journal of Computer Applications, Vol 133, 30-33, January 2016.
- [7]. Malaisamy, A., and G. M. Nawaz Kadhar. "Clustering Based L-Diversity Anonymity Model for Privacy Preservation of Data Publishing", International Journal of Enhanced Research in Science, Technology & Engineering, Vol 5, 55-66, November 2016.
- [8]. Ashwin Machanavajhala, Johannes Gehrke, Daniel Kifer, Muthuramakrishnan Venkitasubramaniam, " l -Diversity: Privacy Beyond k -Anonymity", International Journal of Computer Applications, Vol 17, December 2017.
- [9]. Soohyung Kima, Hyukki Leeb, Yon Dohn Chungb, "Privacy-preserving data cube for Electronic Medical Records: An experimental evaluation", International Journal of Medical Informatics, Vol 97, 33-42, September 2017.
- [10]. Taric, G. Jelin, and E. Poovammal. "A Survey on Privacy Preserving Data Mining Techniques", Indian Journal of Science and Technology, Vol 10, February 2017.
- [11]. Tiwari, Anisha, and Minu Choudhary. "A Review on K-Anonymization Techniques", Scholars Journal of Engineering and Technology, Sch. J. Eng. Tech., Vol 5, 238-245, 2017.
- [12]. Swati Abhimanyu Nase, "Implement l -diversity By Using Generalization Algorithm", International Research Journal of Engineering and Technology, Vol 4, Issue: 07, July 2017.
- [13]. Keerthana Rajendran, ManojJayabalan, Muhammad Ehsan Rana "A Study on k -anonymity, l -diversity Techniques focusing Medical Data", International Journal of Computer Science and Network Security, Vol 17, December 2017.
- [14]. Yaseen, Saba, Syed M. Ali Abbas, Adeel Anjum, Tanzila Saba, Abid Khan, Naveed Ahmad, Basit Shahzad, and Ali Kashif Bashir. "Improved Generalization for Secure Data Publishing", IEEE Access, Vol 6, 27156 – 27165, March 2018.
- [15]. Teshu Kant Parkar, Yamini Chouhan, Samta Gajbhiye, "A Survey on Anonymization Based Privacy Preserving Models", Journal of Network Communications and Emerging Technologies (JNCET), Vol 8, Issue 3, March 2018.