An Efficient Approach to Publish The Data of MultipleSensitive Attributes Using Modified (l,m,d) Algorithm

Ms.Bantupalli Renuka Sai [1]

Ms.Bantupalli Renuka Sai [1] Assistant Professor, computer science and Engineering, Raghu Engineering College , Andhra Pradesh, India

ABSTRACT

Directly publishing data may violate the confidentiality of the individuals. So, it is necessary to take care while publishing the data. To obtain useful informationfrom the published data it is important to consider all the sensitive data. Publishing multiple sensitive attributes may also provide scope to specify the individual from the published table. There are different anonymization techniques to publish the data of multiple sensitive attributes. The main objective of the proposed work is used to increase the utility levels on the published data. Proposedapproach uses bucketization technique to bucketize the tuples and apply different privacy thresholds to resist from different types of privacy attacks and publishes the data. Here, to bucketize the tuples it uses the modified (l, m, d) algorithm.

Key words: *privacy preservation, bucketization, data utility,privacy constraints.*

1.INTRODUCTION

Data plays a dominant role while publishing. When data is published it contains the information about the individuals. Published data contains the Sensitive information about the individuals, so it is important to preserve the privacy of a particular individual. Directly, publishing data may violate the confidentiality of an individual. So, in order to get rid ofsuch types of associations PPDP is adopted. PPDP (Privacypreserving Data Publishing is the important essential area topredict the data from the future outlook to publish the data. It also means that providing the privacy to the individual involved in the published data by considering the sensitive information. It is possible to predict the useful data and bringthe better associations without violating the utility of a data.The main objective of the PPDP is to provide the better the privacy levels and utility levels to the published data by using different techniques. The main inconsonant requirement for PPDP is both privacy preservation and data utility where it should be taken into consideration while publishing the data. Data can be published either static or dynamic. Our paper deals with dynamic publications.

Generally, all the research organizations like hospitals, government agencies, LIC companies, etc…release their data to the public. With the help of that data, the researches provide better associations without violating the confidentiality of the personal information. This published data may contains the quasi data and sensitive data, so it is necessary to anonymizes the data before publishing to violate the privacy of an individual. Both quasi and sensitivedata is used to obtain the things for the future outlooks to publish the data. There were many different anonymization techniques to publish the data. By using, these techniques it is possible to publish the future things. Some of the different anonymization techniques were K-Anonymity, L-diversity, T-closeness and Slicing.

Here, typically the original data contains four types of attributes. There were Explicit Identifier, Quasi Identifier, Sensitive Attributes, Non- Sensitive Attributes.

Explicit Identifier: It is uniquely defined and it is removedfrom the table.

Quasi Identifier: it is not uniquely defined, with the help ofsome of the quasi we can achieve the goal.

Sensitive Attributes: it contains the specific or private or confidential information about the individuals.

Non-Sensitive Attributes: it can be known for public without any concern.

Generally, the published data can be anonymizes by generalization or suppression. These are the most common anonymity operations used. Generalization brings out good result when compared to the suppression. It is necessary to consider the sensitive attributes based on the sensitive categories of a sensitiveness of a sensitive attributes. It is not good to consider all the sensitive attributes.

The main priority should be given to sensitive attributes while publishing the data. Basically, the sensitive attributes consists of both high sensitive attribute values and low sensitive attributes values. Exposing the high sensitive values of a sensitive attribute will acquire more privacy compression of a person involved in the dataset while compared to the low sensitive attribute values of a sensitive attribute. Division of sensitive attribute can be executed based on the number of high sensitive values while remain in the sensitive attribute. Compared to the low sensitive attributes, the high sensitive attributes contains plenty number of high sensitive values and obviously, the low sensitive attributes contains less number of sensitive values.

Extensive experiments have progressed more priority to the single sensitive attribute. By proposing different anonymization and different bucketization techniques on the published the data to resist from different types of privacy attacks like record linkage attack, back ground knowledge attack, probabilistic attack, table linkage attack, etc….By considering only the single sensitive attribute it may not leads to good expected outputs while publishing. Considering the same sensitive level may not execute good expected results for the future hopes. So, it is need to consider the multiple sensitive attributes on the published data for the future out looks.

Here, our proposed model initially splits the data based on the sensitiveness of the sensitive category of the sensitive attributes and applies the modified (l, m, d) bucketization techniques to bucketize the tuples and imposes the privacy model in order to resist from different types of privacy attacks. This proposed approach mainly concentrates on multiple sensitive attributes.

Existing system concentrates on single sensitive attribute. It uses SLPPA (Sensitive Label Privacy Preservation with Anonymization) technique to publish the data by using two process: Table division and Group division. Table division performs the entropy to calculate the weights of the quasi attributes and apply correlation and partition the tables. In group division, it performs (alpha, beta, gamma and delta) privacy model to resist from attacks but this model does not resist from all types of attacks.

A. Motivation in Privacy Preserving:

This section explains the different methodologies and anonymization techniques and their drawbacks:

a) **K-Anonymity:** The information present in each individual present in the released table should not be different from at least K-1 records.

For eg: When an attacker is trying to identify a particular person in the released table but you know only information like age and data of birth and they were k persons in the table having same age and date of birth then it may leads to homogeneity attack and back ground knowledge attack.

b) **L-diversity:** In each equivalence class, it should have at least well l represented sensitive values of a sensitive attribute then such type of table is called L-diversity. This technique mainly suffers from the similarity attack.

c) **T-closeness:** The threshold value t should be less than the ratio of the distance between the distribution of a sensitive attribute in an equivalence class and the distribution of the entire table. This technique mainly uses the earth move distance to protect the published data.

2. LITERATURE SURVEY

COMPLETE OVERVIEW ON PRIVACY PRESERVATION

a) Privacy preservation:

It is an important scenario which is facilitating the Sensitive information or the private data of an individual from the different privacy breaches on the published micro data.

b) Need for privacy preservation:

It is a definite need to preserve the privacy of a particular individual. Based on the Sensitive attributes, the different privacy breaches can be reduced. By considering the single sensitive attributes (SSA), it is impossible to get an expected results from all the corners of the attacker when compared tothe MSA (multiple sensitive attributes).

c) How to preserve the privacy:

The information gathered from different prosperities of a records aimed for a period of time for accumulation which isdone by data recipient and this was carried out by different scientists or people to conduct different experiments on a published micro data.

In data publishing model, there were mainly three parties included there were: individual user, data publisher and public user. Individual users send their data to the publisher and here, the publisher keeps the individuals data and publishes the data to the public users which is usually named as micro data. The privacy of an individual may be leaked when the public users were untrusted or unfaithful. So, in order to come across these consequences it is important to preserve the data while publishing.

“PPDP” means Privacy preserving data publishing. It uses different anonymization techniques to preserve the data. It anonymizes the data by considering the individuality of a data of a record holders. The entire work completely strives to hide the sensitive information of a particular individual by using different PPDM (privacy preserving data mining) algorithms to get rid of different privacy attacks for future outlook. Some of the different methods or algorithms were K-anonymity, T-closeness, L- diversity, etc….. Were included to preserve the data.

The following concerns were used by different researchers by considering the multiple sensitive attributes for privacy constraints. Tames at.el [1] progressed an approach on singlesensitive attributes. This work extended both the models of K-Anonymity and L-diversity on multiple sensitive attributes to prevent the linking attacks and also it anonymizes the data. To protect the privacy among the multiple sensitive attributes it also uses the concept of distortion. It mainly specifies the extension of lower degree of diversity for an attributes with very tiny distinct values.

Yu Lin at.el [2] proposed an approach decomposition on multiple sensitive attributes. Here, it uses the concept of data distortion by adding the random noise. Initially, it considers one primary sensitive attribute by selecting one sensitive attribute. This, progressed work desires a new privacy attack and it also clears by giving two solutions which one of it is decomposition and (l1,l2,l3,…..ld) diversity model. This work is only suitable for static releases. It gives various directions for both numerical and categorical sensitive attributes.

Das and Bhattacharya [3] demonstrated an approach decomposition+ for the generation of a data set by using an L-diversity over multiple sensitive attributes. It was mainly proposed to overcome one of the drawback partition. It is mainly suitable for dynamic releases and also for low dimensional data sets.

P.Usha at.el [4] introduced a model on multiple sensitive attributes based privacy preserving data mining using K-anonymity. Existing approaches concentrates on a technique homogeneous anonymization, where it anonymizes the single sensitive attributes. To overcome the information loss and data utility, the clustering based non homogeneous anonymization was proposed. Here, based on the sensitivity levels of clusters, the non-homogeneous anonymization techniques like generalization and suppression is applied to identify the identified quasi attributes of each cluster. It achieves high degree of data utility and also high degree of data integrity.

Luting Chen [5] at.el demonstrated a new anonymity model (l,m,d) where it is going to resist the similarity attack for multiple sensitive attributes. Initially, this model uses in a hierarchical tree where to analyze and compute similarities which are semantically among the values of a sensitive attributes. This model cannot achieve the consumption of time and information loss.

O. Temuujin at.el [6] discussed an approach is mainly for dynamic dataset. It is mainly developed for preserving the privacy for the datasets which are dynamically evolved. This model uses L-diversity anonymization for the defined limitations on the existing. It is introduced as a new algorithm i.e. cuckoo filter algorithm. It is in a structure of a new probabilistic data mainly used to express the data in an appropriate results and also to improve the efficiency of a data processing. It is not applicable for static data sets. It is mainly applicable for any updating, additions or any deletions on the datasets. It is also failed to concentrate on the stronger anonymization models like T-closeness.

Yuelei Xiao at.el [7] proposed a model privacy preserving data publishing for multiple sensitive attributes based on security level. This approach doesn’t consider the different values of a sensitive attribute but it considers the sensitive attributes having different sensitivity requirements. A new model L si -diversity model is introduced for multiple sensitive attributes. Based on this algorithm there were 3 greedy algorithms are introduced there were MBF (Maximal Bucket First), MMDCF (Maximal Multi-Dimensional Capacity First) and MSLF (Maximal Security Level First). Based on the security levels on sensitive attribute, the sensitivity requirement is defined. This experiment is done for two unshielded non empty dimensional buckets when they have same maximal sensitive attribute level. But, it does not reach when the buckets are selected randomly and also it does not achieve data integrity level.

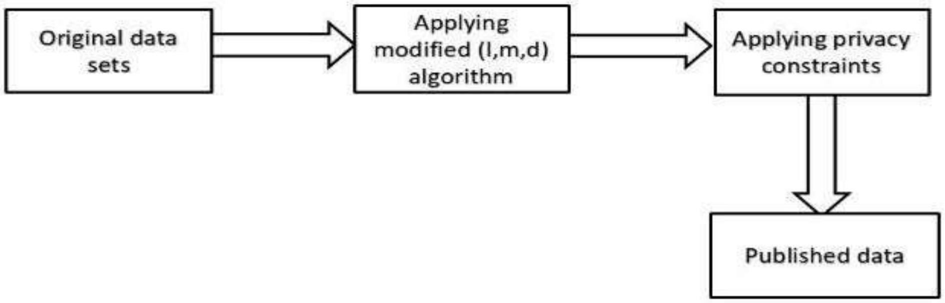
Razaullah Khan at.el [8] demonstrated a privacy preserving on multiple sensitive attributes against finger print correlation attack satisfying C – diversity .This paper considers the both dimensionality in quasi identifiers and also multiple sensitive attributes. It mainly concentrates on back ground knowledge where it includes the finger print correlation knowledge, quasi knowledge and non- membership knowledge. Finger print correlation attack is considered as a strong privacy attack. Militating model proposed an approach (c, k) anonymizaton for privacy prevention in multiple sensitive attribute. It mainly acts linking between generalized table and sensitivity table and it also connected to one to many. The proposed algorithm (c, k) anonymization performs two functionalities. Initially, it calculates the weight based on the categorization of multiple sensitive attributes and also creates the finger print bucket for all the datasets. It doesn't achieve the concept of data integrity as well as the data containing the huge information loss.

Hui Zhu at.el [9] discussed a model (l, k) - diversity privacy model for sequential publication with high utility. It is mainly used for sequential publication. This privacy model is totally based on the concept of both generalization and segmentation by the record anonymity to L-diversity and also individual anonymity by satisfying K- anonymity. This model is mainly applied to the data sets which allows multiple records for same individual which has a low information loss. It is needed to be extended the algorithm which has multiple records for multiple individual and also needed to improve the data utility rate also.

3.Proposed System:

The proposed new bucketization algorithm is designed in a correct standard to obtain the better good results in all directions. Many of the existing system concentrates on the single sensitive attributes, but sometimesit may not lead good associations among the published data. Proposed system concentrates on the multiple sensitive attributes. To get better associations, proposed system concentrates on the multiple sensitive attributes along with the quasi.

Proposed System Architecture:


Original data sets
Applying modified (l,m,d) algorithm
Applying privacy constraints
Published data

Proposed system involves three modules:

a. Bucketization
b. Checking privacy constraints
c. Data publishing

Data Sets: Adult datasets are collected from the UCI machine learning which consists of 32,562 instances. Data set consists of the marital status, occupation, age, etc…Here,the Education, relationship, salary, disease were considered as the sensitive attributes.

https://archive.ics.uci.edu/ml/datasets/Adult

3.1 New Bucketization Algorithm:

Proposed system uses new bucketization algorithm named modified (l,m,d).This algorithm mainly considers all the sensitive attributes based on the sensitiveness.

Here, l determines the total number of distinct sensitivevalues of a sensitive bucket
m specifies the total number of sensitive attributes.
d specifies in each attribute of the bucket should satisfies 'd' distinct sensitive category

This proposed (l, m, d) algorithm considers the both semantic levels and sensitive levels of the sensitive values. Here, Bucketization is the process of dividing the tuples into different buckets based on the sensitiveness of a sensitive attribute. Based on the sensitivity level, divide the sensitive attributes into both high sensitive values and low sensitive values. When placing the sensitive values into the bucket, place both high sensitive values and low sensitive values of that semantic category into the bucket. In a similar way we consider the specific semantic category into the bucket.

3.1 Obtaining Required Privacy Constraints:

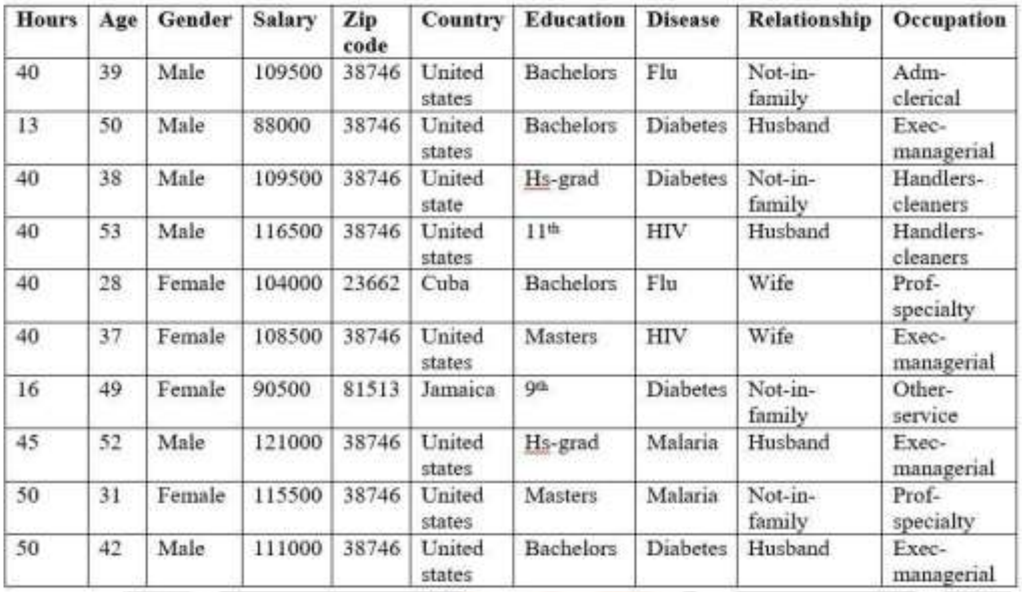
New bucketization algorithm imposes privacy thresholds on only high sensitive values of each sensitive attribute for all the sensitive buckets. This approach overcomes the high sensitive values only when thefrequency exceeds the specified thresholds of the concernedsensitive attributes. This approach suppress only the high sensitive values up to the specified thresholds levels of the sensitive attributes concerned, and do not consider the low sensitive attributes neither it exceeds nor the specified thresholds levels if the sensitive attribute. Due to this era, it is possible to obtain the better utility rates on the low sensitive attributes and maximum high utility rates on the high sensitive attributes.

To display the running example, it is assumed thatthe privacy thresholds of the sensitive attributes are occupation is 2, disease is 2, relationship is 3 and education is 3. Masters, some college and doctorate were considered as a high sensitive attributes in education sensitive attribute. Asper the table 1, the frequency of masters and bachelors is 3and there is no need to suppress the values of education. Inoccupation attribute the frequency of the exec – managerial is 3.So apparently, the one value of the exec - managerial issuppressed as shown in the table 2.In relationship attribute the frequency of the not- in-family is 4, it exceeds the one value of threshold level so it is necessary to suppress the values of not-in-family of relationship attribute. In disease attribute, the frequency of flu is 2 and diabetes is 3 so it is not necessary

3.1 Creation of multiple tables:

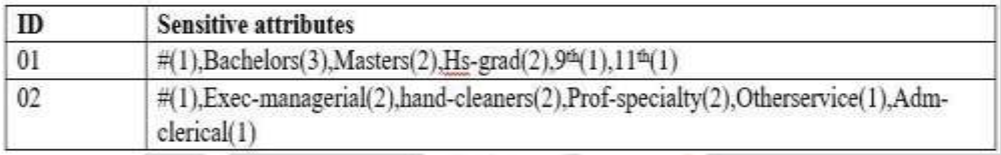
The proposed modified (l,m,d) algorithm model assigns the buckets into each partition of a sensitive bucket. Here, the correlations are applied among the all sensitive attributes. The sensitive attributes which are highly correlated or more correlated for them provide the bucket idto each sensitive bucket partitions SID's(sensitive bucket identifier) to each sensitive partitions as shown in the table 4 and 5.Here, to provide the link between the values of both quasi and sensitive values concatenation is applied. Later it concatenates the correlated quasi attributes with SID's of sensitive buckets as shown in the table 5.This model also applies the random permutations on each quasi bucket of a quasi-table to create the final published table as shown in table 6. It also used to prevent the data from different types of linking attacks. Finally, this model published the data in terms of different sensitive tables and quasi tables. This approach publishes table 4, table 5 and table 6.

Table 1: Original Dataset



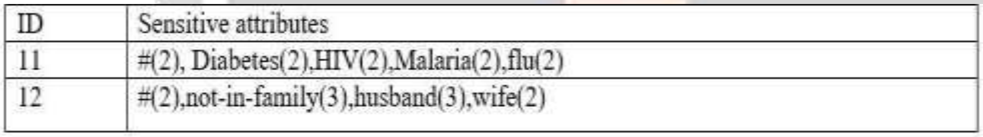
Hours	Age	Gender	Salary	Zip code	Country	Education	Disease	Relationship	Occupation
40	39	Male	109500	38746	United states	Bachelors	Flu	Not-in-family	Adm-clerical
13	50	Male	88000	38746	United states	Bachelors	Diabetes	Husband	Exec-managerial
40	38	Male	109500	38746	United state	Hs-grad	Diabetes	Not-in-family	Handlers-cleaners
40	53	Male	116500	38746	United states	11th	HIV	Husband	Handlers-cleaners
40	28	Female	104000	23662	Cuba	Bachelors	Flu	Wife	Prof-specialty
40	37	Female	108500	38746	United states	Masters	HIV	Wife	Exec-managerial
16	49	Female	90500	81513	Jamaica	9th	Diabetes	Not-in-family	Other-service
45	52	Male	121000	38746	United states	Hs-grad	Malaria	Husband	Exec-managerial
50	31	Female	115500	38746	United states	Masters	Malaria	Not-in-family	Prof-specialty
50	42	Male	111000	38746	United states	Bachelors	Diabetes	Husband	Exec-managerial

Table 2: Grouping of SA's at explicit threshold level for education and occupation.



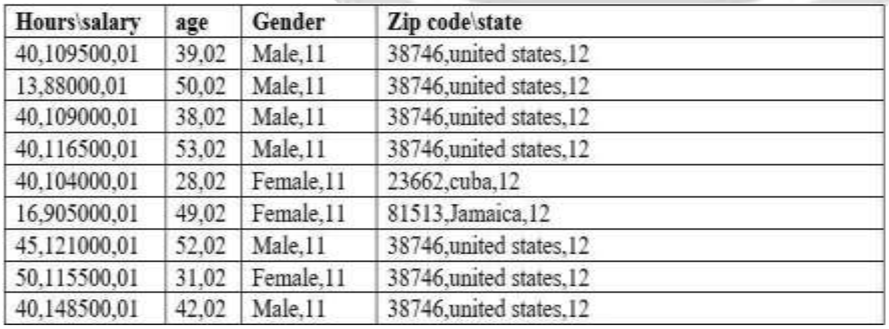
ID	Sensitive attributes
01	#(1),Bachelors(3),Masters(2),Hs-grad(2),9th(1),11th(1)
02	#(1),Exec-managerial(2),hand-cleaners(2),Prof-specialty(2),Otherservice(1),Adm-clerical(1)

Table 3: Grouping of SA's at explicit thresholds level fordisease and occupation.



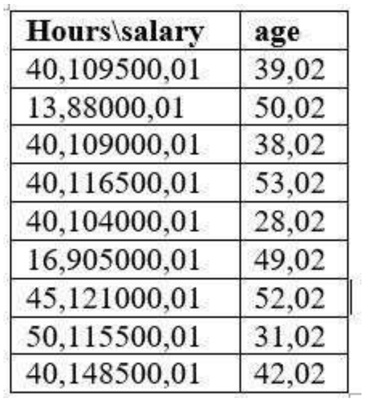
ID	Sensitive attributes
11	#(2), Diabetes(2),HIV(2),Malaria(2),flu(2)
12	#(2),not-in-family(3),husband(3),wife(2)

Table 4: concatenated quasi attributes on their correlation

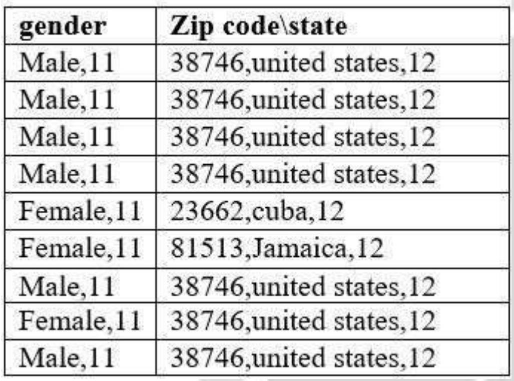


Hours\salary	age	Gender	Zip code\state
40,109500,01	39,02	Male,11	38746,united states,12
13,88000,01	50,02	Male,11	38746,united states,12
40,109000,01	38,02	Male,11	38746,united states,12
40,116500,01	53,02	Male,11	38746,united states,12
40,104000,01	28,02	Female,11	23662,cuba,12
16,905000,01	49,02	Female,11	81513,Jamaica,12
45,121000,01	52,02	Male,11	38746,united states,12
50,115500,01	31,02	Female,11	38746,united states,12
40,148500,01	42,02	Male,11	38746,united states,12

Table 5: Separate sensitive table for highly correlatedsensitive attributes

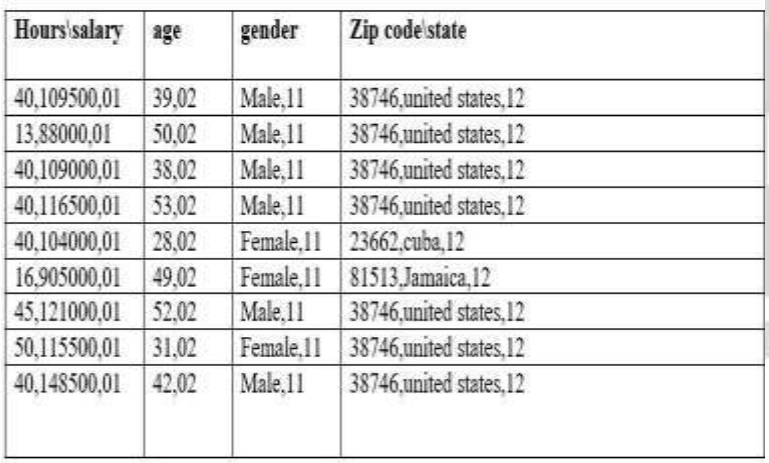


Hours\salary	age
40,109500,01	39,02
13,88000,01	50,02
40,109000,01	38,02
40,116500,01	53,02
40,104000,01	28,02
16,905000,01	49,02
45,121000,01	52,02
50,115500,01	31,02
40,148500,01	42,02



gender	Zip code\state
Male,11	38746,united states,12
Male,11	38746,united states,12
Male,11	38746,united states,12
Male,11	38746,united states,12
Female,11	23662,cuba,12
Female,11	81513,Jamaica,12
Male,11	38746,united states,12
Female,11	38746,united states,12
Male,11	38746,united states,12

Table 6: Permutated QID's and SID's of SAs



Hours\salary	age	gender	Zip code\state
40,109500,01	39,02	Male,11	38746,united states,12
13,88000,01	50,02	Male,11	38746,united states,12
40,109000,01	38,02	Male,11	38746,united states,12
40,116500,01	53,02	Male,11	38746,united states,12
40,104000,01	28,02	Female,11	23662,cuba,12
16,905000,01	49,02	Female,11	81513,Jamaica,12
45,121000,01	52,02	Male,11	38746,united states,12
50,115500,01	31,02	Female,11	38746,united states,12
40,148500,01	42,02	Male,11	38746,united states,12

4.1 Bucketization Algorithm :

modified (l,m,d) algorithm

Step 1: Begin
Step 2: take a dataset.

Step 3: find total number of tuples from data set.

Step 4: Bucket size= total number of tuples from datasets/ total number of distinct sensitive values of all sensitive values

Step 5: find the distinct values count of all sensitive attributes in a table and assign a minimum count to l.

Step 6: Next assign a two digit value to each distinct valuesof all SA.

{Consider the first digit as 0 or 1}

[0 or 1 represents sensitive level and another second digit number represent a semantic level]

Step 7: Select the sensitive attribute which has highest threshold and create a stack for all the unique sensitive attributes having highest thresholds in a table.

Step 8: Now bucketize the tuples

Step 9: place the records into the bucket

Step 10: if bucket is fill, place the records into another bucket

Step 11: here, each sensitive attribute of a bucket should satisfy l value or else swap the distinct records from the other stacks.

Step 12: Now check the sensitivity and dissimilarity of all distinct values in each bucket

Step 13: each sensitive attribute of a bucket should satisfy d else swap the records from other stacks and repeat the loop for the sensitive attributes of a bucket.

Step 14: end

4.2 Applying privacy constraints

Step 15: apply the sensitive thresholds on each sensitiveattribute of a bucket based on their sensitiveness

Step 16: if the frequency count of the sensitive attribute exceeds the thresholds of a sensitive attribute then it is necessary to suppress the sensitive attribute to the thresholdsof the concerned sensitive attribute.

Step 17: Update the sensitive values count of each sensitiveattribute.

Step 18: Repeat the above steps for all the sensitive buckets.

4.3 Creation of multiple tables

Step 19: find the correlation among the sensitive attributes

Step 20: create a separate sensitive table for highlycorrelated sensitive attributes

Step 21: Assign a bucket id to each sensitive bucket partition

Step 22: Concatenate the IDs of sensitive bucket partitionswith the quasi values of the quasi buckets.

Step 23: Apply random permutations on all the quasi valuesof the quasi buckets.

Step 24: End.

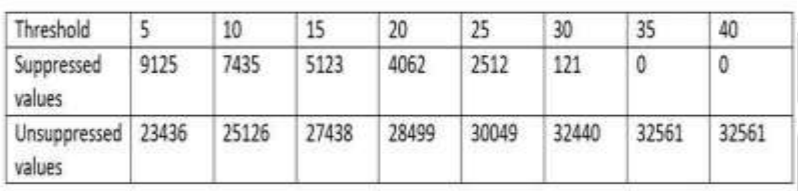
4.PERFORMANCE EVALUATION

The main objective of performance evaluation is to distinguish the quality of the data to the previous anonymous version and the original version. Each and every step of the anonymization operations performed can measured or identified by using some search metrics. Here, the performance of the proposed model is measured using the loss metrics.

5.1 Loss Metrics (LM):

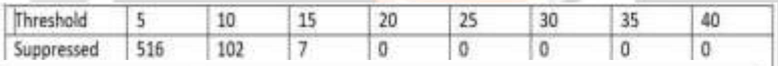
This parameter is used to measure the loss of all values of each sensitive attribute only when they are either suppressed or generalized. The disease sensitive attribute at different thresholds and table 9 indicates the total number of suppressed and unsuppressed values of occupation sensitive attribute at different thresholds and table 10 determines the total number of unsuppressed and suppressed values of the education sensitive attribute at different thresholds.

Finally figure 1. Represents the overall loss metric values of all the sensitive attributes.

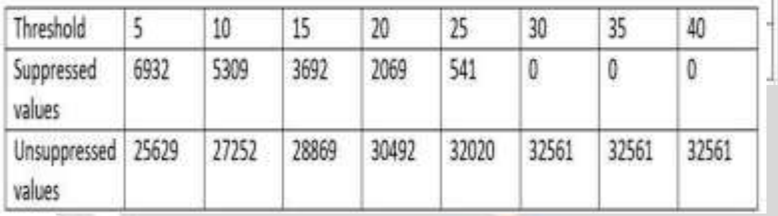


Threshold	5	10	15	20	25	30	35	40
Suppressed values	9125	7435	5123	4062	2512	121	0	0
Unsuppressed values	23436	25126	27438	28499	30049	32440	32561	32561

Table 7: LM of suppressed and unsuppressed values of arelationship attribute



Threshold	5	10	15	20	25	30	35	40
Suppressed	516	102	7	0	0	0	0	0



Threshold	5	10	15	20	25	30	35	40
Suppressed values	6932	5309	3692	2069	541	0	0	0
Unsuppressed values	25629	27252	28869	30492	32020	32561	32561	32561

Table 8: LM for suppressed and unsuppressed values of adisease attribute

Table 9: LM for suppressed and unsuppressed values of anoccupation attribute

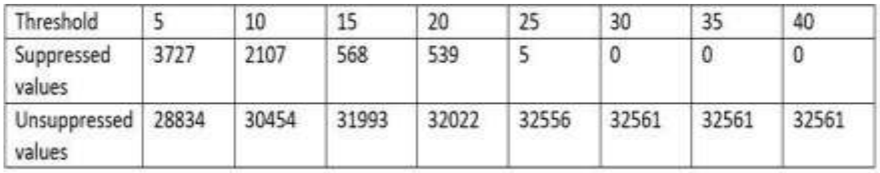
Loss metrics= N- 1 / |c|-1 As per the equation, N indicates the total suppressed values in an attribute domain.
|c| specifies the attribute domain size.

5.2 Working area and dataset:

Modified (l,m,d) algorithm was executed on the windows 10 environment of 16-GHz Pentium processor with 8GB RAM. It is implemented using python and MYSQL. Data set has been extracted from the UCI machine learning repository where it consists of 32561 instances of tuples.

5.3 Loss metrics:

Proposed model provides a best results when compared to the previous traditional models. Table 7 specifies the loss metrics values of suppressed and unsuppressed values of the relationship attributes at differentspecified thresholds at 5, 10, etc…. Table 8 specifies the lossmetric values of suppressed and unsuppressed values of the relationship and disease attributes



Threshold	5	10	15	20	25	30	35	40
Suppressed values	3727	2107	568	539	5	0	0	0
Unsuppressed values	28834	30454	31993	32022	32556	32561	32561	32561

Table 10: LM for suppressed and unsuppressed values ofan education attribute.

5. SCREEN SHOTS

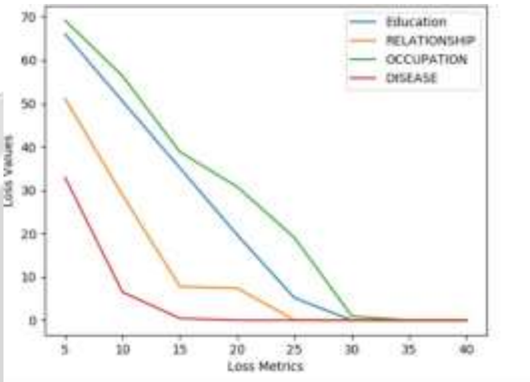


Figure 2: overall LM for all sensitive attributes

From figure 2 we can draw a conclusion that here disease and occupation were considered as high sensitive attributes and education and relationship were considered as a low sensitive attributes that's why here we can apply required number of results by applying at different thresholds on different types of sensitive attributes.

Comparison Graphs of suppressed values of all sensitive attributes

X-axis = privacy thresholds

Y-axis=percentage of loss suppressed values

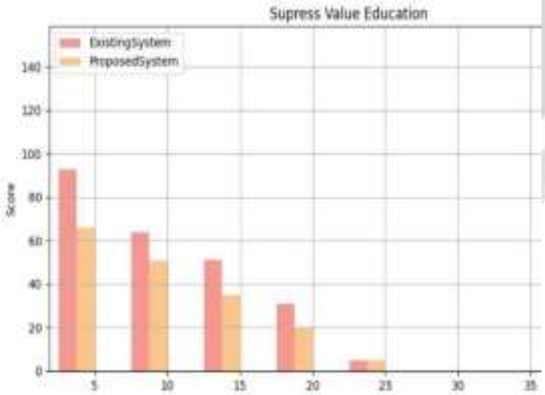


Fig 5: suppresses value graph of sensitive attribute education

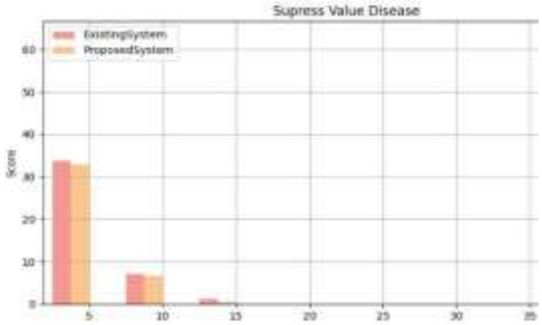


Fig 6: suppress value of sensitive attribute disease

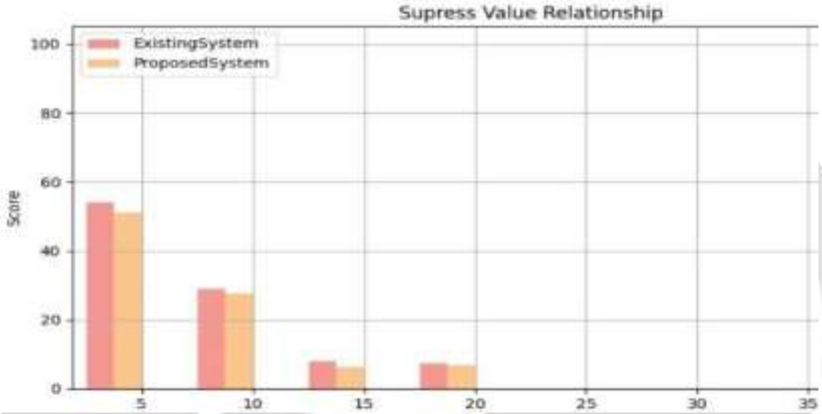


Fig 7: suppress value of sensitive attribute relationship

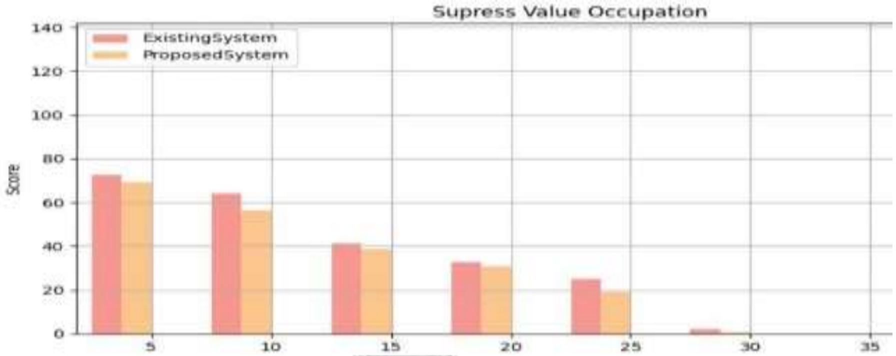


Fig 8: suppress value of sensitive attribute occupation

6. CONCLUSIONS

This proposed system concentrates on the publication of multiple sensitive attributes. This project explore a new scheme to provide the privacy of individuals involved in the published data by introducing modified (l,m,d) algorithm. This scheme uses modified (l,m,d) algorithm to assign the records into the buckets and applied different privacy thresholds to resist the published data from different types of privacy attacks. The proposed system acquires less suppression rate when compare to the existing system. This proposed system leads to acquire good associations when compared to the previous traditional models. In the future work it can consider the sensitive requirements of the sensitive attributes and can extend the algorithm for the published data. This publication is done only for the categorical attributes. It can also be done by extending the numerical variables.

7. REFERENCES

1) Lin Yao, Zhenyu Chen, Xin Wang, Dong Liu and Guowei Wu, Sensitive Label Privacy Preservation With Anatomization for Data Publishing, Journal,2019.Knowledge and Information System , Volume 23, no:1, pp.83,97

2) Odsuren Temuujin, Jinhyunahn, Dong-Hyukim,"efficient L-diversity algorithm for preserving privacy of dynamically published datasets. IEEE Access 2019. vol.7,pp.11,Issue Number 122879.

3) N.V.S.LakshmipathiRaju,M.N.Seetaramanath,P.Srinivasa Rao, A Novel Dynamic KCi-Slice Publishing Prototype for Retaining Privacy and Utility of Multiple Sensitive Attributes". International Journal of Information Technology and Computer Science (IJITCS)2019, Vol. 11, No. 4 pp.18-32, DOI:10.5815/ijitcs.2019.04.03.

4) Lakshmipathi Raju, N.V.S., Seetaramanath, M.N., Srinivasa Rao, P., An enhanced dynamic KC-slice model for privacy preserving data publishing with multiple sensitive attributes by inducing sensitivity. Journal of King Saud University-Computer and Information Sciences 2018.

5) Onashoga, S.A., Bamiro, S.A., Akinwale, A.T. and Oguntuase, J.A., "KC-Slice: A dynamic privacy-preserving data publishing technique for multi sensitive attributes". Information security journal: A Global Perspect,2017. Vol. 26, No. 3, pp. 121-135,
DOI: 10.1080/19393555.2017.1319522.

6) Susan, V. S. and Christopher, T.,, "Anatomization with slicing: a new privacy preservation approach for multiple sensitive attributes". *Springer Plus, 2016.*Vol. 5, No. 1, pp.964-984, DOI: 10.1186/s40064-016-2490-0.

7) P.Usha, R.Shriram, S.SatishKumar, Multiple Sensitive attributes based Privacy Preserving Data Publishing Data Mining Using k-Anonymity.InJournal,2014. Knowledge and Information System , pp. 433-437

8) Liu, F., Jia, Y., Han, W., 2012. A new k-anonymity algorithm towards multiple sensitive attributes. In 2012 IEEE 12th International Conference on Computer and Information Technology (CIT). IEEE, pp. 768–772.

9) https://archive.ics.uci.edu/ml/datasets/Adult

10) Junjie Jia, Luting Chen, (l,m,d)-Anonymity : A Resisting Similarity Attack Model for Multiple Sensitive Attributes. In Journal, 2009. Knowledge and InformationSystem, 654205992@qq.com,2499549816@qq.com.

11) N.Li, T.Li and S.Venkata Subramanian, T-closeness: Privacy beyond k-anonymity and l-diversity, Journal, 2007. Knowledge and Information System, pp. 106–115.

12) A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkata Subramanian, l-diversity: Privacy beyond k-Anonymity, 22nd International Conference, 2006. p. 24.