Incremental Diversity for Preserving Privacy of Student Data

Prof. Veena Gadad, Prof. Sindhu Rajendran, Hardik Hiraman Pawar, Tanmay S Lal, S Mohammed Ashiq, Mohammed Raza

Department of Computer Science and Engineering,

RV College of Engineering,

Bangalore.

ABSTRACT

Myriads of privacy preserving algorithms are present but very few aim to take into consideration the privacy for multiple sensitive attributes. Many pre-existing algorithms aim to segregate the table into sensitive and non-sensitive tables or completely mask or generalize the information. This paper combines the best of both worlds and develops an innovate algorithm termed as incremental diversity.

Incremental diversity algorithm is able to diversify the data for multiple sensitive attributes and also produce lesser quantity of residue records. Incremental diversity mainly chooses a primary sensitive attribute having a certain set of favorable characteristics such the presence of a greater number of unique values as compared to other sensitive attributes, many parents in its semantic hierarchical tree, and also more varying frequency for each sensitive value in the dataset.

The inference obtained after careful examination of the graphs is that incremental diversity has much better time complexity, generates lesser number of residue records for a given value of k, but it comes at the cost of having lesser diversity than (l, e) diversity.

The algorithm can be suitably adjusted to balance the trade-off between diversity and residue records produced for a given dataset depending on the algorithm's use-case and the degree of sensitivity that is desired. Privacy can be easily strengthened by choosing the most suitable primary sensitive attribute of the microdata.

1. INTRODUCTION:

If the data is published in a highly hidden and inconspicuous way in order to protect privacy, the researchers who need the data cannot extract much information. For example, consider the table showing data of students:

Table 1.1: Sample student data

Month of birth	Hall ticket number	Percentage of marks	College name
February	5243	81%	P
February	1673	92%	Q
February	5882	88%	Q
July	3227	61%	R

Let A be one among the students whose data is displayed. An attacker can easily get to know A's college name if he knows A's hall ticket number. Imagine, the above data is published in the manner subsequently mentioned in order to protect privacy of students.

Table 1.2: Modified student data

Month of birth	Hall ticket number	Percentage of marks	College name

January - April	1000 -8000	60 - 100%	P
January - April	1000 - 8000	60 - 100%	Q
January - April	1000 - 8000	60 - 100%	Q
May - August	1000 - 8000	60 - 100%	R

From the above table (Table 1.2), attacker cannot get to know A's college name. **S**uppose a researcher or data analyst wants to know how many students from college P have scored more than 80%, then the researcher cannot get to know from the above data. Now, with the help of above table, two tables can be made.

Table 1.3: Quasi-identifier table (QIT)

Month of Birth	Hall ticket number	Percentage of marks	Group ID
February	5243	81%	1
February	1673	92%	1
February	5882	88%	1
July	3227	61%	2

Table 1.4: Sensitive table (ST)

Group ID	College name	Count
1	P	1
1	Q	2
2	R	1

The researcher can find that one student from college P has scored more than 80% as Group ID 1 corresponds to P and Q and count of P is one (seeing second table), so by observing first table, 81%, 92% and 88% corresponds to either P or Q. As all three scores are above 80%, it can be understood that one student from college P has scored above 80% and also college name (P) is hidden. This method is called Anatomy [1] (discussed in Background and Previous Works Section).

2. BACKGROUND AND PREVIOUS WORKS

The previous researches in the process of publishing the privacy preserved data is mainly focused on single sensitive attribute and which cannot be implemented for the multiple sensitive attributes. The model of k-anonymity [30] was based on partitioning the given microdata table in a manner such that each equivalence class of the microdata table should contain at least k number of records which cannot be distinguishable from the other set of records. The drawback of this model is inefficient resistance to homogenous attacks. The improved model called p-sensitive model for multiple sensitive attribute was developed. This model assigns the Group ID to equivalence classes along with masking and generalizing the quasi-identifier group. The drawback is that it did not consider the frequency of multiple sensitive attributes [11]. A better model (l-diversity) was proposed in order to consider the frequency of attributes into account. This model describes the method of grouping the data such that there should be at least *l* different sensitive values. If the attacker wants to identify the individual with sensitive values, then he must have at least *(l-1)* sensitive values with high probability of occurring. The drawback of this model is that it applies only to single sensitive attribute [16]. The modified version of the previous model was implemented in order to consider multiple sensitive attributes called l-m-d anonymity. l-m-d anonymity deals with method of anonymizing the microdata table by creating equivalence classes along with generalizing and masking. The drawback of l-m-d anonymity is that it considers the process of generalizing the quasi-identifier [13]. The best method for anonymizing the data called anatomy was proposed. deals with the problem of generalization by dividing the microdata table into the two tables namely quasi-identifier table and the sensitive attribute table. In the quasi-identifier table, the quasi-identifiers pertaining to an individual is displayed along with their Group ID. The sensitive attribute table consists of the Group ID, sensitive attribute, and the count of the sensitive value. The process of anatomy was strong and effective but there was a need for diversifying data [27]. In the process of converting the data into more diversified form, it is important to identify the primary and secondary sensitive attribute. The sensitive attribute containing higher and more unique sensitive values is considered as primary sensitive attribute [6]. The method of calculating the diversity index for the anonymized table was finally implemented. The degree of diversity is an indication of how diversified the table is [14].

3. BASIC DEFINITIONS & KEYWORDS

Microdata table: A raw data obtained from any survey or census (for example: details of all students in a class)

Quasi Identifier attribute (QI): An attribute which in combination with other attributes can identify an individual uniquely. (For example: gender, age etc.)

Sensitive attribute (SA): An attribute the individual wants to maintain secret or private (for example: his marital details, bank account number etc.)

Equivalence class: A table of records consisting of above-mentioned attributes i.e., quasi-identifier attribute and sensitive attribute which is obtained from microdata table. It is basically a subclass obtained from microdata table.

Semantic hierarchical tree: A representation in the form of a tree which describes the relationship between various values or attributes (similar to family tree of a family).

4. ALGORITHM

Input: Microdata Table (M_1), k, l, e, Total no. of records (N)

Output: QIT ($q_1, q_2, q_3, \ldots, q_n$), ST ($s_1, s_2, s_3, \ldots, s_n$), (l, e) Masked Microdata Table (MMT)

Steps:

1. Standardization of Microdata and storage in a nested dictionary format.
2. Assignment of the Group ID (or more specifically, the equivalence class number) to each record with the formula given below.

$$Group\ ID = \lfloor n - 1 \rfloor + 1, \qquad n \in [1, N]$$

$$where, \qquad \lfloor\ \rfloor = rounding\ down\ function$$

$$n \in \text{record number uniquely associated with the corresponding record}$$

$$N = Total\ no.of\ records$$

3. Diversification of records by any one of the five different conditions, on the basis of which records are removed from the microdata table, added to the temporary dictionary, and added back to the modified microdata table (discussed in Tables and Formulae Section).
4. Creation of a separate dictionary (called the temporary dictionary) to store the records that do not fit in the selected equivalence class of the microdata table on the basis of the conditions associated.
5. Picking and placing of the records from the temporary dictionary into the new microdata dictionary based on the pre-selected parameters along with the extra condition that the size of the existing equivalence class should be strictly lesser than k (k-anonymity [30]).
6. Increase in the diversity of Secondary, Tertiary and Quaternary Sensitive Attributes by swapping their sensitive values which are repeating in a particular equivalence class with non-repeating values from the temporary dictionary.
7. Segregation of the modified microdata table into Quasi-Identifier Table (QIT) and Sensitive Table (ST).
8. Masking and generalisation of the attributes with ease & efficiency (formulas discussed in Tables & Formulae Section).

5. TABLES AND FORMULAE

The five diversification conditions are:

I. Based on the unique appearance of a primary sensitive attribute with few parents in its semantic hierarchical tree (Eg: Marital Status): In a particular equivalence class, a value of the chosen primary sensitive attribute appears only once, i.e., the sensitive value is unique for a specific equivalence class.

II. Based on the unique appearance of a parent in the semantic hierarchical tree consisting of a fewer number of parents for the primary sensitive attribute: It is similar to the former algorithm with the key difference being that instead of the sensitive value, the parent of the chosen primary sensitive attribute appears only once in the group, i.e., it is unique for a specific equivalence class. Example: If a record in an equivalence class has the value of "Widow" for the primary sensitive attribute "Marital Status", the parent of "Widow" being "Unmarried", no other record containing Marital Status value as Separated, Divorced or Never-Married is allowed to be added into the equivalence class.

III. Based on the appearance of a parent in the semantic hierarchical tree consisting of a fewer number of parents for the primary sensitive attribute at most twice: The parent of the chosen primary sensitive attribute is allowed to appear twice in a particular equivalence class.

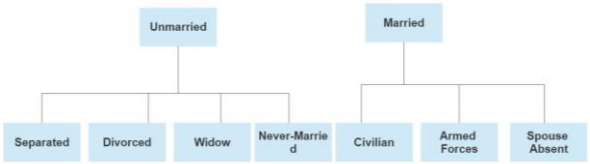

Figure 5.1: Semantic Hierarchical Tree for Marital Status

IV. Based on the unique appearance of a primary sensitive attribute with more varying frequency of each value in the dataset (Eg: Relationship): It is an offshoot of the first condition where only the primary sensitive attribute is changed from one having fewer parents in its semantic hierarchical tree (Marital Status) to the attribute having more variations in its values (Relationship).

V. Based on the unique appearance of a parent in the semantic hierarchical tree consisting of high number of parents for the primary sensitive attribute: A primary sensitive attribute with a higher no. of parents in its semantic hierarchical tree is chosen (Disease) and the parent uniquely appears in an equivalence class. For example, if in an equivalence class there is the presence of a disease for a particular record, let's say, Cardiomyopathy, belonging to the Circulatory System Disorder Category (Parent), then no other record with the disease belonging to the same parent (Circulatory System Disorder) will be allowed to be placed in that equivalence class.

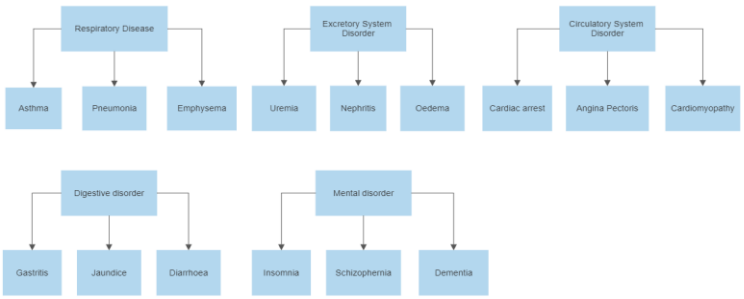

Figure 5.2: Semantic Hierarchical Tree for Disease

Quasi Identifier Table (QIT) & Sensitive Table (ST)

Table 5.3: Quasi – identifier table

DISPLAYING QIT

Age	Gender	Zip Code	Education	Employment	Marital Status	Marital Parent	Relationship	Race	Salary	Disease Parent	Group ID
39	Male	77516	Bachelors	State-gov	Never-married	Unmarried	Not-in-family	White	<=50K	Respiratory disease	1
50	Male	83311	Bachelors	Self-emp-not-inc	Married-civ-spouse	Married	Husband	White	<=50K	Mental disorder	1
38	Male	215646	HS-grad	Private	Divorced	Unmarried	Not-in-family	White	<=50K	Circulatory_system disorder	1
53	Male	234721	11th	Private	Married-civ-spouse	Married	Husband	Black	<=50K	Excretory_system disorder	2
28	Female	338409	Bachelors	Private	Married-civ-spouse	Married	Wife	Black	<=50K	Circulatory_system disorder	2
31	Female	45781	Masters	Private	Never-married	Unmarried	Not-in-family	White	>50K	Digestive disorder	2
49	Female	160187	9th	Private	Married-spouse-absent	Married	Not-in-family	Black	<=50K	Digestive disorder	3
52	Male	209642	HS-grad	Self-emp-not-inc	Married-civ-spouse	Married	Husband	White	>50K	Mental disorder	3
37	Female	284582	Masters	Private	Married-civ-spouse	Married	Wife	White	<=50K	Circulatory_system disorder	3
42	Male	159449	Bachelors	Private	Married-civ-spouse	Married	Husband	White	>50K	Excretory_system disorder	4
37	Male	280464	Some-college	Private	Married-civ-spouse	Married	Husband	Black	>50K	Digestive disorder	4
30	Male	141297	Bachelors	State-gov	Married-civ-spouse	Married	Husband	Asian-Pac-Islander	>50K	Respiratory disease	4
23	Female	122272	Bachelors	Private	Never-married	Unmarried	Own-child	White	<=50K	Digestive disorder	5
32	Male	205019	Assoc-acdm	Private	Never-married	Unmarried	Not-in-family	Black	<=50K	Excretory_system disorder	5
40	Male	121772	Assoc-voc	Private	Married-civ-spouse	Married	Husband	Asian-Pac-Islander	>50K	Respiratory disease	5

Table 5.4: Sensitive table (ST)

DISPLAYING ST

Disease	Group ID
Emphysema	1
Insomnia	1
Cardiac arrest	1
Nephritis	2
Cardiomyopathy	2
Diarrhoea	2
Jaundice	3
Insomnia	3
Cardiac arrest	3
Oedema	4
Gastritis	4
Emphysema	4
Jaundice	5
Nephritis	5
Asthama	5

Table 5.5: Count table

DISPLAYING Disease COUNT

Group ID	Disease	Count
1	Emphysema	1
1	Insomnia	1
1	Cardiac arrest	1
2	Nephritis	1
2	Cardiomyopathy	1
2	Diarrhoea	1
3	Jaundice	1
3	Insomnia	1
3	Cardiac arrest	1
4	Oedema	1
4	Gastritis	1
4	Emphysema	1
5	Jaundice	1
5	Nephritis	1
5	Asthama	1

Table 5.6: Masked microdata table

DISPLAYING MASKED MICRODATA

Age	Gender	Zip Code	Education	Employment	Marital Status	Marital Parent	Relationship	Race	Salary	Disease	Disease Parent	Group ID
(30 - 39)	M/F	77***	Bachelors	State-gov	Never-married	Unmarried	Not-in-family	White	<=50K	Emphysema	Respiratory disease	1
(50 - 59)	M/F	83***	Bachelors	Self-emp-not-inc	Married-civ-spouse	Married	Husband	White	<=50K	Insomnia	Mental disorder	1
(30 - 39)	M/F	215***	HS-grad	Private	Divorced	Unmarried	Not-in-family	White	<=50K	Cardiac arrest	Circulatory_system disorder	1
(50 - 69)	M/F	234***	11th	Private	Married-civ-spouse	Married	Husband	Black	<=50K	Nephritis	Excretory_system disorder	2
(20 - 39)	M/F	338***	Bachelors	Private	Married-civ-spouse	Married	Wife	Black	<=50K	Cardiomyopathy	Circulatory_system disorder	2
(30 - 49)	M/F	45***	Masters	Private	Never-married	Unmarried	Not-in-family	White	>50K	Diarrhoea	Digestive disorder	2
(40 - 69)	M/F	160***	9th	Private	Married-spouse-absent	Married	Not-in-family	Black	<=50K	Jaundice	Digestive disorder	3
(50 - 79)	M/F	209***	HS-grad	Self-emp-not-inc	Married-civ-spouse	Married	Husband	White	>50K	Insomnia	Mental disorder	3
(30 - 59)	M/F	284***	Masters	Private	Married-civ-spouse	Married	Wife	White	<=50K	Cardiac arrest	Circulatory_system disorder	3
(40 - 49)	M/F	159***	Bachelors	Private	Married-civ-spouse	Married	Husband	White	>50K	Oedema	Excretory_system disorder	4
(30 - 39)	M/F	280***	Some-college	Private	Married-civ-spouse	Married	Husband	Black	>50K	Gastritis	Digestive disorder	4
(30 - 39)	M/F	141***	Bachelors	State-gov	Married-civ-spouse	Married	Husband	Asian-Pac-Islander	>50K	Emphysema	Respiratory disease	4
(20 - 39)	M/F	122***	Bachelors	Private	Never-married	Unmarried	Own-child	White	<=50K	Jaundice	Digestive disorder	5
(30 - 49)	M/F	205***	Assoc-acdm	Private	Never-married	Unmarried	Not-in-family	Black	<=50K	Nephritis	Excretory_system disorder	5
(40 - 59)	M/F	121***	Assoc-voc	Private	Married-civ-spouse	Married	Husband	Asian-Pac-Islander	>50K	Asthama	Respiratory disease	5

Formulae incorporated for masking and generalising the chosen sensitive attributes:

Generalising Age:

$$Lower\ Limit\ (L)\ =\ age - age\%10$$

$$Upper\ Limit\ (U)\ =\ L\ +\ 10 * ((group\_id - 1)\%3\ +1) - 1$$

$$Where, \quad \%\ represents\ the\ modulo\ operation$$

$$group\_id\ corresponds\ to\ the\ Group\ ID\ of\ a\ particular\ record\ in\ the\ equivalence\ class$$

$$The\ age\ is\ returned\ as\ the\ string\ "(L\ -\ U)"$$

This a type of cyclic masking with step size of 3.

Generalising Gender: “M/F” is used.

Masking Zip Code: Only the last three digits are hidden with (*). (Eg: 560***)

Masking Employment, Race or Salary: Asterisk symbol (*) is used.

The various performance parameters to evaluate and analyse the algorithm for varying values of the inputs (no. of records and k) is calculated with the following formulas:

$$Code\ Runtime\ =\ (end_{time} - start_{time})\ x\ 1000$$

$$Where, \quad start_{time}\ and\ end_{time}\ are\ initialised\ at\ the\ beginning\ and\ end\ of\ the\ program\ respectively\ with\ the$$

$$function\ time.time()\ which\ returns\ the\ number\ of\ seconds\ passed\ since\ epoch$$

$$\text{Residue Percentage} = \frac{No.of\ records\ in\ Residue\ Dictionary}{Total\ no.of\ records\ in\ the\ original\ Microdata\ Table} * 100\%$$

$$\text{Diversity Percentage} = \frac{\sum_{EQ=1}^{MGID} \left(\frac{\sum Diversity\ of\ NMA}{No.of\ NMA}\right)}{MGID}$$

$$Where, \quad MGID\ =\ Maximum\ Group\ ID$$

$$EQ = Equivalence\ Class\ No.$$

$$NMA = Non\ Masked\ Attribute$$

The diversity for each NMA is calculated as follows:

Diversity for each NMA = $\frac{Number\ of\ unique\ values\ of\ attribute}{Total\ no.of\ values\ of\ attribute}$

6. EXPERIMENTS

Here Age, Gender and Zip code are the Quasi Identifiers (QI).
5 different conditions are used for plotting multiple line graphs.

1) Marital Status:
Primary Sensitive - Marital Status
Secondary Sensitive - Education
Tertiary Sensitive - Employment
Quaternary Sensitive - Race

2) Marital Semantic Tree (One):
Primary Sensitive - Marital Parent
Secondary Sensitive - Education
Tertiary Sensitive - Employment
Quaternary Sensitive - Race

3) Marital Semantic Tree (Two):
Primary Sensitive - Marital Parent
Secondary Sensitive - Education
Tertiary Sensitive - Employment
Quaternary Sensitive - Race

4) Relationship:
Primary Sensitive - Relationship
Secondary Sensitive - Education
Tertiary Sensitive - Employment
Quaternary Sensitive - Race

5) Disease Semantic Tree:
Primary Sensitive - Disease Parent
Secondary Sensitive - Education
Tertiary Sensitive - Employment
Quaternary Sensitive - Race

Numerous graphs for various parameters and constant conditions are plotted

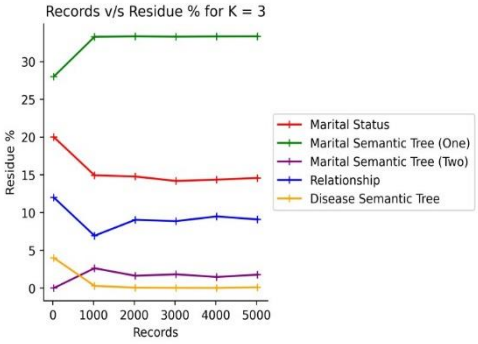

Fig 6.1
Records v/s Residue % for K=3

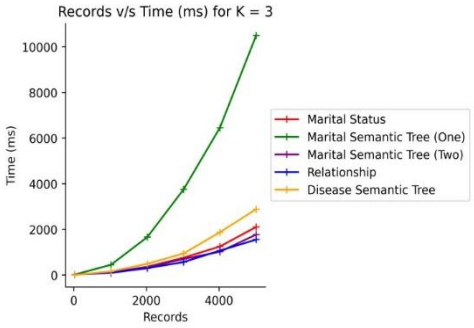

Fig 6.2
Records v/s Time (ms) for K=3

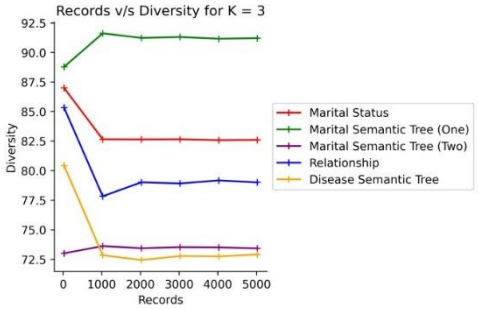

Fig 6.3
Records v/s Diversity for K=3

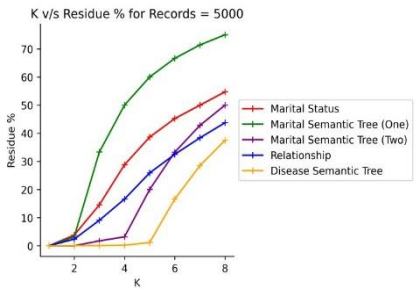

Fig 6.4
K v/s Residue % for Records=5000

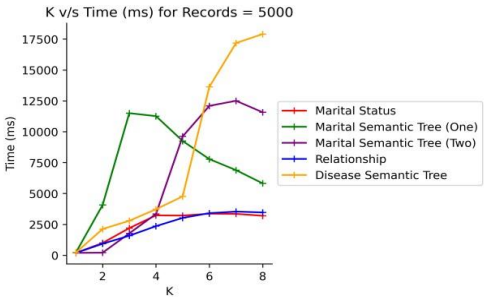

Fig 6.5
K v/s Time (ms) for Records=5000

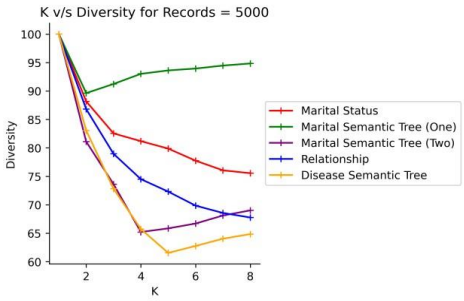

Fig 6.6
K v/s Diversity for Records=5000

Comparison of (l, e) Diversity and Incremental Diversity: -

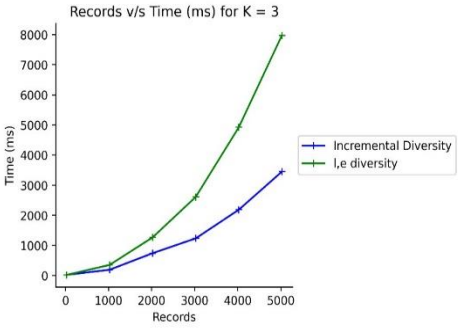

Fig 6.7
Records v/s Time (ms) for K=3

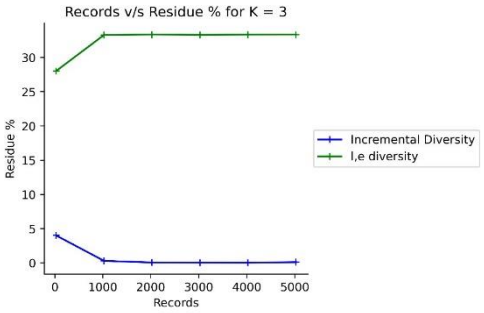

Fig 6.8
Records v/s Residue % for K=3

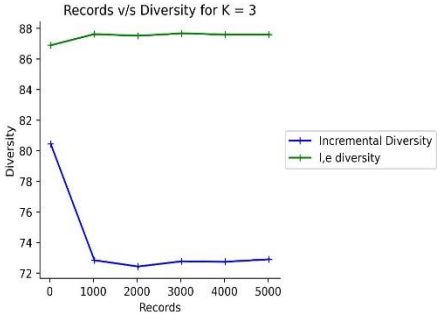

Fig 6.9
Records v/s Diversity for K=3

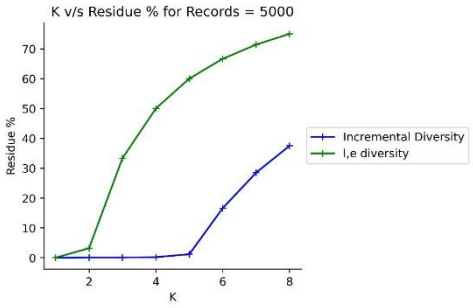

Fig 6.10
K v/s Residue % for Records=5000

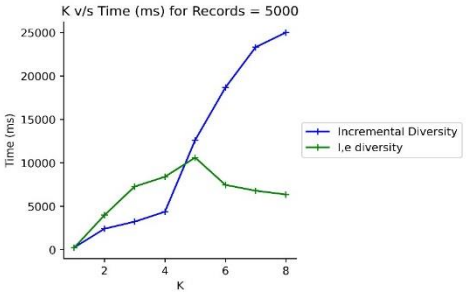

Fig 6.11
K v/s Time (ms) for Records=5000

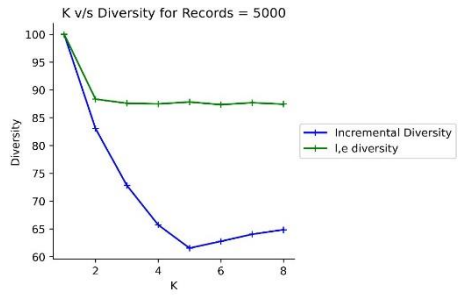

Fig 6.12
K v/s Diversity for Records=5000

Software Tools used are: -

1) Python Language

2) CSV Files	Contains Dataset
3) Math Module	Mathematical Operations
4) Time Module	Time taken for Execution
5) Matplotlib.pyplot Module	Plot Graphs
6) Tabulate Module	Tabulate Function
7) Copy Module	Deepcopy() function

7. RESULTS AND DISCUSSIONS

1. No. of Records v/s Residue %: From Fig 6.1, the best possible case for arrangement of records will be based on unique Diseases (condition 5) in each equivalence group as the residue % for it is very less for larger records since number of unique Diseases are more than the number of unique Marital Parents.

2. No. of Records v/s Time (ms): In Fig 6.2, since the uniqueness of the Marital Parent is very low, large number of residues will be obtained in the residue list. Therefore, more time is spent iterating through the residue list multiple times to replace records.

3. No. of Records v/s Diversity: In Fig 6.3, for condition 2, the diversity increases as the number of records increases because most of the records will be sent to residue list hence less repetition in the modified microdata table and more diversity.

4. K v/s Residue %: In Fig 6.4, there is a steady increase in residue % for all conditions because when the K value increases there will be more repetition which leads to higher residue %.

5. K v/s Time (ms): From Fig 6.5, the time taken for execution increases as the value of K increases, since the program iterates through larger residuals and then swaps the datasets leading to increase in the time taken for execution.

6. K v/s Diversity: In Fig 6.6, the diversity is very high for smaller values of K because the no. of unique values is more relative to K values, so less repetition.

Comparison between (l, e) diversity algorithm and incremental diversity algorithm

(l, e) diversity algorithm employs the use of a primary sensitive attribute with lesser no. of parents in its semantic hierarchical tree (Eg: Marital Status) having one common parent for the sensitive value in the equivalence class.

Incremental diversity algorithm makes use of a primary sensitive attribute with more no. of parents in its semantic hierarchical tree (Eg: Disease) allowing only non-repeating parents to be present in each equivalence class while also performing the incremental diversification for secondary, tertiary and quaternary sensitive attributes. It outperforms (l, e) diversity in terms of faster time performance and overall decrease in residue records percentage. From the Fig 6.7 and Fig 6.8 we can infer that incremental diversity produces lesser residue records and runs faster despite performing diversification for multiple sensitive attributes. Fig 6.9, producing lesser residue records is a double-edged sword which leads to lesser diversity in the records present in the equivalence classes of the table which is a drawback of the incremental diversity algorithm as compared to (l, e) diversity algorithm.

8. Conclusion

Choosing Primary Sensitive Attribute**:** The choice of the right primary sensitive attribute can make or break the diversification and privacy of a table. There will be a trade-off between residue records produced and diversity of the final private table. For example, Disease as the primary sensitive attribute (condition 5) produces lesser no. of residue records but at the same time diversity of the records is hampered. On the other hand, Marital Status as the primary sensitive attribute (condition 1) produces highly diverse table but it comes at the cost of more no. of residues produced and the loss of precious data. It depends on the use case as to whether diversity needs to be sacrificed in order to produce lesser residue records and prevent huge data loss or, the diversity of the data is of utmost importance and it is okay to overlook the production of enormous residue records.

REFERENCES

[1] Boyu Li, Yanheng Liu, Minghai Wang, Geng Sun and Bin Li. Local anatomy for personalized privacy protection, 2021.

[2] Widodo, Murien Nugraheni and Irma Permata Sari. Simple Distribution of sensitive values for multiple sensitive attributes in privacy preserving data publishing to achieve anatomy, 2021.

[3] Khan Razaullah, Tao Xiaofeng, Anjum Adeel, Haider Sajjad and Saif ur Rehman Malik. Privacy Preserving for Multiple Sensitive Attributes against Fingerprint Correlation Attack Satisfying c-Diversity, 2020.

[4] Yuelei Xiao and Haiqi Li. Privacy Preserving Data Publishing for Multiple Sensitive Attributes Based on Security Level, 2020.

[5] Jassma N Vanasiwala and Nirali R Nanavati. Privacy Preserving Data Publishing of multiple Sensitive Attributes by using Various Anonymization Techniques, 2020.

[6] W Widodo and A Wahyudin, A preliminary phase on anatomyzing multiple sensitive attributes by determining main sensitive attribute, 2020.

[7] Widodo Widodo, Eko K. Budiardju, Wahyu Catur Wibowo and Harry Achson. An approach for distributing sensitive values in k – anonymity, 2019.

[8] Rong Wang, Yan Zhu, Tung-Shou Chen and Chin-Chen Chang. Privacy-Preserving Algorithms for Multiple Sensitive Attributes Satisfying t-Closeness, 2018.

[9] Ozgu Can. Personalised Anonymity for micro data release, 2018.

[10] Edwin M. Knorr, Giulio Valentino Dalla Riva and Orlin Vakarelov. Anatomy of a new data science course in privacy, ethics and security, 2018.

[11] Widodo and Wahyu Catur Wibowo. A distributional model of sensitive values on p – sensitive in multiple sensitive attributes, 2018.

[12] Wantong Zheng, Zhongyue Wang, Tongtong Lv and Yong Ma. K – Anonymity Algorithm based on improved clustering, 2018.

[13] Junjie Jia and Luting Chen. (l,m,d) – Anonymity : A Resisting similarity attack model for multiple sensitive attributes, 2017.

[14] Haiyuan Wang, Jianmin Han and Jivi Wang. (l,e) – diversity A privacy preserving model to resist semantic similarity attack, 2014.

[15] Hua Jin, Shang – Cheng Liu and Shi-Guang Ju. Privacy preserving for multiple sensitive attributes based on l – coverage, 2014.

[16] Michael Kern. Anonymity: A Formalization of privacy – l – diversity, 2013.

[17] Sunyong Yoo, Moonshik Shin and Doheon Lee. An Approach to Reducing Information Loss and Achieving Diversity of Sensitive Attributes in k-anonymity Methods, 2012.

[18] Xiang Min Ren, Jing Yang, Jian Pei Zhang and Zong Fu Jia. Uncertain Data Privacy Protection Based on K-Anonymity via Anatomy, 2012.

[19] Vijayarani Mohan. K-Anonymity Techniques – A Review, 2010.

[20] Khaled El Emam and Fida Kamal Dankar. Protecting Privacy Using k-Anonymity, 2008.

[21] Alina Campan and T.M. Truta. Data and Structural k-Anonymity in Social Networks, 2008.

[22] Ashwin Machanavajjhala, Daniel Kifer and Johannes Gehrke. L -diversity, 2007.

[23] V. Ciriani, S. De Capitani di Vimercati, S. Foresti and P. Samarati. k-Anonymity, 2007.

[24] Xioakui Xiao and Yufei Tao. Anatomy: Privacy and Correlation Preserving Publication, 2006.

[25] T.M. Truta and Bindu Vinay. Privacy Protection: p-Sensitive k-Anonymity Property, 2006.

[26] M. Nergiz and Chris Clifton. Thoughts on k-Anonymization, 2006.

[27] Xioakui Xiao and Yufei Tao. Anatomy: Simple and Effective Privacy Preservation, 2006.

[28] Ashwin Machanavajjhala, Johannes Gehrke, Daniel Kifer and Muthuramakrishnan Venkitasubramaniam. ℓ-Diversity: Privacy Beyond k-Anonymity, 2005.

[29] Wei Jiang and Chris Clifton. Privacy-Preserving Distributed k-Anonymity, 2005.

[30] L. Sweeney. Achieving k-Anonymity Privacy Protection Using Generalization and Suppression, 2002.

[31] Latanya Sweeney. k- anonymity: A model for protecting privacy, 2001. W Widodo and A Wahyudin, A preliminary phase on anatomyzing multiple sensitive attributes by determining main sensitive attribute, 2020.

[32] Widodo Widodo, Eko K. Budiardju, Wahyu Catur Wibowo and Harry Achson. An approach for distributing sensitive values in k – anonymity, 2019.

[33] Rong Wang, Yan Zhu, Tung-Shou Chen and Chin-Chen Chang. Privacy-Preserving Algorithms for Multiple Sensitive Attributes Satisfying t-Closeness, 2018.

[34] Ozgu Can. Personalised Anonymity for micro data release, 2018.

[35] Edwin M. Knorr, Giulio Valentino Dalla Riva and Orlin Vakarelov. Anatomy of a new data science course in privacy, ethics and security, 2018.

[36] Widodo and Wahyu Catur Wibowo. A distributional model of sensitive values on p – sensitive in multiple sensitive attributes, 2018.

[37] Wantong Zheng, Zhongyue Wang, Tongtong Lv and Yong Ma. K – Anonymity Algorithm based on improved clustering, 2018.

[38] Junjie Jia and Luting Chen. (l,m,d) – Anonymity : A Resisting similarity attack model for multiple sensitive attributes, 2017.

[39] Haiyuan Wang, Jianmin Han and Jivi Wang. (l,e) – diversity A privacy preserving model to resist semantic similarity attack, 2014.

[40] Hua Jin, Shang – Cheng Liu and Shi-Guang Ju. Privacy preserving for multiple sensitive attributes based on l – coverage, 2014.

[41] Michael Kern. Anonymity: A Formalization of privacy – l – diversity, 2013.

[42] Sunyong Yoo, Moonshik Shin and Doheon Lee. An Approach to Reducing Information Loss and Achieving Diversity of Sensitive Attributes in k-anonymity Methods, 2012.

[43] Xiang Min Ren, Jing Yang, Jian Pei Zhang and Zong Fu Jia. Uncertain Data Privacy Protection Based on K-Anonymity via Anatomy, 2012.

[44] Vijayarani Mohan. K-Anonymity Techniques – A Review, 2010.

[45] Khaled El Emam and Fida Kamal Dankar. Protecting Privacy Using k-Anonymity, 2008.

[46] Alina Campan and T.M. Truta. Data and Structural k-Anonymity in Social Networks, 2008.

[47] Ashwin Machanavajjhala, Daniel Kifer and Johannes Gehrke. L -diversity, 2007.

[48] V. Ciriani, S. De Capitani di Vimercati, S. Foresti and P. Samarati. k-Anonymity, 2007.