

KAMOA: K-anonymity Model for Overlapped Attributes

Bhakti Maheshwarkar¹, Pawan Patidar²

¹M.Tech Research Scholar ; ²Associate Professor, Lakshmi Narain College of Technology, Indore (M.P)

¹bhaktimaheshwarkar@gmail.com; ²pawanpatidar4u@rediffmail.com

Abstract- *K-anonymity model is mostly used technique of privacy preserving data publishing. In K-anonymity model data is converted into anonymous state. So, that adversary can't be able to disclose sensitive information about the user. Generalization and suppression are most commonly used anonymity technique, but generalization contains some drawbacks i.e. generalization disturbs correlations between attributes.*

In this paper a novel model is proposed which uses generalization technique specially to maintain correlation among overlapped attributes and way to reduce dimensionality of data set. Experimental evaluation section shows efficiency and correctness of proposed model.

Index Term-- *Privacy Preserving Data publishing; Membership Disclosure; K-anonymity; Slicing; Overlapped Attributes; Multiple Sensitive Attributes.*

1. INTRODUCTION

In today's era knowledge based decision making have special place in all sectors such as government bodies, individual and corporations. Every decision taken by government or individual is based on analysis. For analysis previously generated results are important because they shows exact scenario of market. Data mining and it's results become a important part of individuals as well as governing bodies and specially at this point need of privacy preserving data mining arises.

Data mining is a technique which is applied on data which is collected from various sources and the results the knowledge which is useful for individuals.

Data publisher who involved in publishing the knowledge or information is responsible for securing this data from unauthorized access. Once the data is released for public use can't be modified but it can be judged properly before releasing it publically. In this point the need of privacy preserving data mining is arises. Data publisher have to apply privacy preserving data mining technique on data before releasing it publically. There is several privacy preservation data mining algorithm are present ex: l-diversity, t-closeness. This paper is based on mainly K-anonymity model with l-diversity.

K-anonymity model suggest converting attributes in anonymous form. So, that no one can identify the exact values for selective attributes and information disclosure can be controlled. Most popular techniques for k-anonymity model are generalization and suppression. In generalization data is replaced with more general value, and in suppression the value is suppressed or hidden. Generalization is considered an efficient method to achieve K-anonymity. In recent research article [1] generalization drawbacks are mentioned, first, generalization tends to information loss for high dimensional data, second, for generalization data publisher have to assume that every value in a generalization is possible. Third, it disturbs correlations of attributes because generalization is applied on each attribute separately.

More ever these points are based on analysis but anonymity cannot be applied properly without generalization. In this paper new model of K-anonymity is proposed which is emphasis on generalization technique especially in the case of overlapped attributes. Overlapped attributes are those attributes that present in multiple time in a dataset and form a set of attributes to maintain correlations of attributes. Paper also shows a technique to maintain correlations using generalization technique.

The rest of this paper is organized as follows: In section 2, background work related to this topic is discussed. Section 3 proposes new model. Experimental evaluation on this model and comparison with previous model is present in section 4 and section 5 concludes this paper and discusses future research.

2. BACKGROUND WORK

2.1 All about Data, Data Mining and Privacy Preservation Needs:

Data can be considered as a value which belongs to set of items. There is a interconnection between data, information and knowledge. Data contains facts, unorganized figures or values which need to be processed, when this data is processed or some analysis is done on it reflects some information. Without processing data is useless, information is processed data and information reflects need or use of data. When the information is collected and processed in next level it reflects knowledge

about data which is never before predicted. Data mining can be considered as a class of data applications which are used to reveal hidden patterns from group of data. It is also considered as a knowledge discovery but in some articles and books data mining is one phase of knowledge discovery because it is used to search or mine patterns only, when these patterns are again processed they shows knowledge and reflects something new information. Data mining becomes an important factor now a days, because it shows results which are based on analysis. Now a day's peoples are become analysis dependent. This example supports the point properly. Even than it is not a exact example of data mining but it shows how people are able to mine information at their own level. If a person wants to purchase a Smartphone, there are several models are present in market. At first step he select a model of his choice than perform analysis of model at his extant by checking review related to model, ask his friends and others by checking blogs related to model and its price, after that if he feels satisfied then purchase it or select another model, and again the analysis steps are started till his anxiety is not fulfilled.

Similarly, In data mining this process is applied at broad level and results shows some patterns or figures which are useful for all as well as selected divisions. Data which is available for public use can be classified into two parts. One is non-sensitive data which contains general information about products, individuals and others. In this paper we will discussed data specially related to individuals, contains general information example: Zip code, Age, Sex, Job status etc. Non-sensitive data generally released by private organization, Government organization so or individual itself. These data are not considered as a sensitive because it contains very personal information i.e. medical status, criminal record etc. when a private organization is bound to release their private data for government or semi govt. use data publisher who going to release this data have to perform some security check for protecting sensitive data from adversaries.

Data publishers will decide attributes sensitivity i.e. sensitive and non-sensitive data, in this case it is possible that individual feel that their salaries are sensitive data but as per data publisher view it is important factor and should be considered as a non-sensitive data.

Non-sensitive attributes are also considered as quasi-attributes. Privacy preserving data publishing contains some technique which help data publisher to release the data. If proper technique and effective technique is adopted by data publisher it can be prevent data against linking attacks.

General steps applied on data before release are without privacy preserving data mining are:

Step 1: Distinguishment of attributes from dataset, i.e. Key attributes, non-sensitive attributes, sensitive attributes.

Step 2: Remove key attributes from data set.(where using key attributes anyone can be easily identified i.e. name, SSN, Pan card no. etc.)

Step 3: Release Dataset.

Above mentioned steps causes linking attack and removal of this attack introduced privacy preserving data publishing approaches. Table 1 and 2 show a private table and public table which contains all attributes:

Table 2 shows all non-sensitive attributes and this table publically available for all. When Table 1 released for some specific purpose. Data publisher apply general steps: First, Decide attributes: Name: Key Attributes, Sex and Age, Zip code: Non-sensitive attributes, Disease: Sensitive Attributes.

Second remove key attributes i.e. name. Third release data set.

Table 3 shows contents of released table when table 2 and 3 linked by adversary it will discloses the information i.e. rahul is suffering from HIV. Table 2 record no. 6 and table 3 record no. 1 matches the values as adversary known's that rahul is single So, he select S. No. 1 rather than 5.

This type of sensitive information leakage is known as linking attack and this problem arises a need of a model which prevent against this type of attacks.

2.2 K-anonymity Model:

K-anonymity is one of the popular approaches of privacy preserving data publishing. Anonymity is a term taken by a Greek work *anonymia* means nameless state. When anonymity technique is applied on some attributes of dataset attributes becomes in unpredictable states. K-anonymity model was specially proposed for preventing against linking attack. Generalization and suppression are widely used techniques with K-anonymity, where K is a factor which assures that sensitive information

cannot be distinguished from at least K-1 individuals whose information also appears in the release Figure 1 shows K-anonymity Model.

2.2.1 Generalization and Suppression Technique:

Generalization is one of the most popular approaches of K-anonymity. In this approach data publisher replace quasi-identifiers with most general values. So, that adversary fails to identify sensitive value associated with it. Quasi-identifiers are non-sensitive attributes which are present in both tables private as well as public.

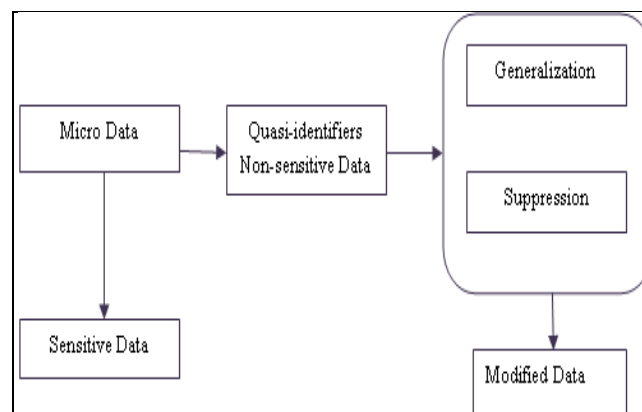


Figure 1: K-anonymity Model

Quasi-identifiers are responsible for linking attack. When generalization is applied on these attributes they become in unpredictable mode. For example: Suppose Zip code value is 456001 if generalization is applied at first level then the new generalized value of zip code is set 45600*. The keyword * is used to show generalization, if adversary get this attribute value 45600*, he needs to predict 0-9 numeric values in the place of *, and for this value dataset shows at least K records which seems similar to this record. As generalization level increases the confusion of adversary is increases, but as generalization level increases the truthfulness and accuracy of data decreases. Minimality principal suggests using low, level of generalization, because the purpose of generalization is to protect data from adversary not hide it from all.

Generalization can be applied on tuple, cell and attribute level. To protect against adversary in this paper we apply generalization at attribute level. Generalization is classified into two ways: value generalization hierarchy and domain generalization hierarchy, although the results of both hierarchies are same they are used to represent generalization assignment especially for attribute generalization.

The highest level of generalization is considered as suppression. In suppression values are not released or suppressed. As similar to generalization suppression can be also applied on cell, tuple or attribute level. In Certain value of attribute another attribute value is depends. In this paper a novel model is proposed which maintains correlations among attributes with the help of generalization, especially when overlapped attributes are present in the dataset. N-SA K-anonymity model [2] the drawbacks of tuple suppression is shown. Removal of key attributes can be considered as attribute suppression. In recent research article [1] drawbacks of generalization is mentioned i.e. generalization causes information loss, if high level of generalization is applied then it causes information loss and proper analysis of data cannot be possible, another drawback of generalization is it disturbs correlations of attributes, i.e. it is possible for

2.3 Overlapped Attributes:

Data set which shows correlations among other attributes, example: People suffering from heart problems generally after age of 55, but if someone is suffering from heart disease at the age of 20, then this data is important for medical department because their information gives an important statics regarding age, disease and reasons of the disease. Another example of correlated attributes is if age and criminal records are not disturbs then it shows the age wise crime tendency of peoples.

In datasets some records are related to each other's and they should not be disturbed but using generalization generally quasi-identifiers such as age or zip code are generalized. In this paper a new model is proposed which supports maintaining correlations among attributes with the help of generalization especially when overlapped attributes are present. When one

attribute present more than one set of attribute to maintain correlations then those attributes are considered as overlapped attributes are from sensitive types of attributes then the overhead of data publisher increases, because the possibility of attacks increases rapidly. There should be some extent of diversity present in newly generated sets to prevent information disclosure. In previously proposed model slicing [1] emphasis on use of slicing technique instead of generalization and uses Bucketization but there are some shortcomings of this model. Table 4 and table 5 which is converted form of table 4 shows some short comings. The major challenges in the previous models are:

1. All attribute set values are true, only attribute set values are interchanged in (Zip code, Disease) field within a bucket.
2. High possibility of homogeneity attack.[4].
3. If adversaries have background knowledge attack then system cannot prevent disclosure. [nm privacy issues]
4. The values are randomly permuted within buckets. So, if adversary has some knowledge for distribution then he can easily access the data.
5. If in bucket only four records are present and 3 of them shows same diseases then the resultant combination shows lack of diversity [5].
6. Possibility of unsorted matching attack.[6].

These points' shows need of new model to overcome these drawbacks.

3. PROPOSED MODEL: K-ANONYMITY MODEL FOR OVERLAPPED ATTRIBUTES

K-anonymity Model for Overlapped Attributes is a novel model to protect sensitive data even in the presence of overlapped attribute. In section 2 overlapped attribute is discussed. This paper also shows the need of generalization technique especially when overlapped attributes are present in data set, because overlapped attributes may cause information disclosure and generalization technique is a best technique to control information disclosure as well as membership disclosure and attribute disclosure [2]. Figure 2 shows architecture of proposed model and figure 3 shows algorithm for model.

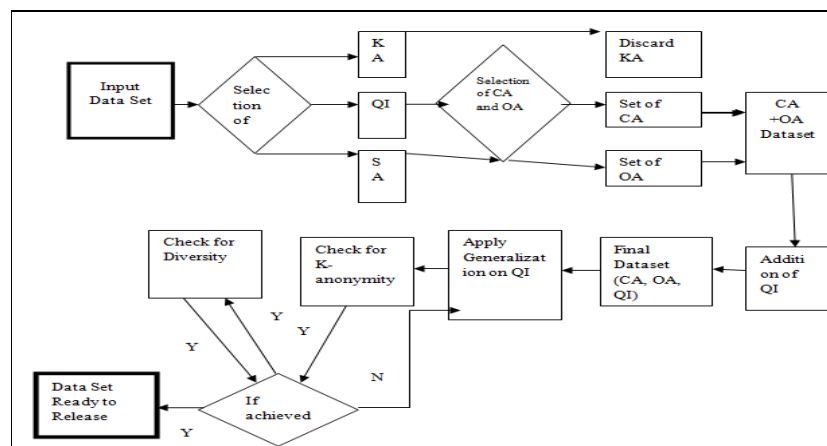


Figure 2: K-anonymity Model for overlapped Attributes

K-anonymity Model for Overlapped Attributes is a model proposed for privacy preserving data publishing. The need of data publishing increases day by day and the factor of privacy also increases in this ratio. Attribute Correlations and Overlapped Attributes are two major factors of KAMOA model which are discussed in section 2.

The main objective of K-anonymity Model for Overlapped Attributes is to design and develop A Model for privacy preserving data publishing where attributes correlations are maintained. if any attribute from dataset shows some dependencies with another attribute set, i.e. age and disease, mostly below 16 age individuals are not suffer from heart disease. This information must be forwarded to medical health department because it shows some important statics about Heart Disease and chances of Heart Disease after 16.

In given example age and disease shows some important relations which should not be generalized or suppressed. Similarly, occupation and disease shows some relation but age and occupation are not related. To maintain this combination we can't store the attribute in one set i.e. (Age, Occupation, and Disease). In this case disease attribute is overlapped in two groups (Age, Disease) and (Occupation, Disease). When any attribute is copied or attached in an attribute set more than once known as overlapped attribute. Proposed Model proposes a model for data set contains overlapped attributes. This model uses generalization technique because in the case of overlapped attributes the possibility of membership disclosure and other severe attack i.e. homogeneity attack, background knowledge attack also increases. Using generalization on QI maintains diversity also. Figure 3 shows algorithm of Proposed Model.

// M master table contains QI quasi-identifiers, KA Key Attributes, SA sensitive attributes, OA overlapped attributes, EQI attribute for generalization, FD Final Dataset set of OA,CA and EQI .

Step 1: Select KA, QI and SA from M.

Step 2: Discard KA.

Step 3: Select OA from CA.

Step 4: Select QI from CA+OA set.

Step 5: Create FD.

Step 6: Apply Generalization on EQI.

Step 7: Check for K-anonymity.

Step 8: if yes, check for diversity.

Step 9: else go to Step 5 then step 6.

Step 10: If achieved go to step 8

Step 11: Else go to step 5 then step 6.

Step 12: Exit.//Data set is ready to release.

Figure 3: Algorithm for K-anonymity Model for overlapped Attributes

4. EXPERIMENTAL EVALUATION

The huge amount of data which is collected by various resources and used for some sort of decision making or knowledge discovery should be arranged in a well manner so that sensitive information's kept secure from attacker's access. Proposed model provide protection against membership disclosure as well as homogeneity attack and background attack when a data set contains overlapped attributes. The algorithm proposed in previous section works in following phases.

Select KA, QI and SA from M: M is a master table which contains data set which supposes to be release. Data publisher who is responsible for data release and may be owner of data divides the data into three sections: Key attributes Quasi-identifiers and Sensitive attribute. Key attributes. Table 6 shows the Private table. Key attribute is used to distinguish individual with others. It shows personal and unique identity of the person. In private table Name is a key attribute. So, data publisher direct remove this attribute from final table. Disease and Salary is a sensitive attribute and data publisher wants to release this information but does not want to show all details i.e. QI related to it. QI shows general information related to individual. In private table QI are: Sex, Age, ZIP, Weight, Qualification, and Occupation.

Select OA, CA from SA: There are two sensitive attributes Disease and salary shown in Table 6 . Proposed model allow data publisher to select attributes category as well as overlapped attribute (OA) and CA from sensitive attributes and Quasi-identifiers. Overlapped attributes belongs to multiple columns with different attributes. Table which is used to show the working of proposed model contains Disease is an overlapped attribute. This model can be work efficiently for other sensitive attribute salary or salary and disease both, because the system is designed in such a way that data publisher is allowed to chose attributes as per his need. Select EQI from QI: Previous model based on slicing shows the drawbacks of

generalization because it disturbs correlations of attributes. KAMOA uses generalization of QI because it is useful to prevent against several attacks. Especially in the case of OA without using generalization diversity is impossible to achieve, because sensitive attribute is added in multiple columns.

So, check for security, anonymity and diversity is applied but not surely prevent security. So, in this step we add a extra QI to set of attributes to prevent against various attacks and to increase diversity. From table 7 Sex, Zip code is considered as a EQI (Extra Quasi-identifier).

Create FD: After selection of KA, QI, SA, OA from SA, EQI from QI data publisher can decide the FD attribute set for collecting correlated attributes in the same cell. So, in this section dimensionality reduction is also performed. Table 8 shows the result of creation of FD. As result shown in table 8 eight attributes are divided or arranged in three attribute set. The decision of attribute set is taken by data publisher. In table 8 three attribute sets are {Age, Sex, and Disease}, {ZIP, Disease}, {Qualification, Salary}, in set of Age, Sex and Disease we need to maintain correlations for age and disease so we add sex as EQI. Zip is EQI for set of Zip and Disease because in linking attack is performed then single sensitive attribute and set 1 may be causes the discloser of SA and details of individual. Qualification and salary does not needed a EQI as data publisher not considered it as sensitive set and even it is not contain overlapped attribute. Check for Diversity: After completion of all steps the last step of KAMOA is diversity checking. It is most important step because this step specially prevents against homogeneity and background knowledge attack. Diversity check is applied for each anonymous set and ensures that no two sensitive attributes are same in each set if anonymous set size is 2 or 3. If K value is min.5 then this rule checks that more than 2 SA should not shows same SA. If this happens, generalization level will be automatically increase for EQI and this step will be continues till resultant data set will satisfy diversity factor. Table 8 shows the outcome of KAMOA model which satisfies diversity also.

5. CONCLUSION

KAMOA is a model for Privacy Preserving data publishing with presence of overlapped attributes. K-anonymity model [8] was proposed to protect sensitive data when these type of data is publically released. N-SA K-anonymity Model [2] was proposed to protect multiple sensitive data because when multiple sensitive attributes are present in dataset the possibility of attack and membership disclosure is increases. Both models use generalization to prevent disclosure against several attacks. Slicing [1] is a new and effective technique to protect data and also shows drawbacks of generalization which emphasis on a new concept of maintaining corelations, but to maintain corelations dataset contains overlapped attributes which are present in more than one column then data publisher needs to use a technique which is capable to protect data twice because same attributes are present two or more columns and previously proposed all models are not capable to prevent information disclosure effectively and QI are responsible for attack is true but without QI we are not able to protect or prevent against attack is also true.

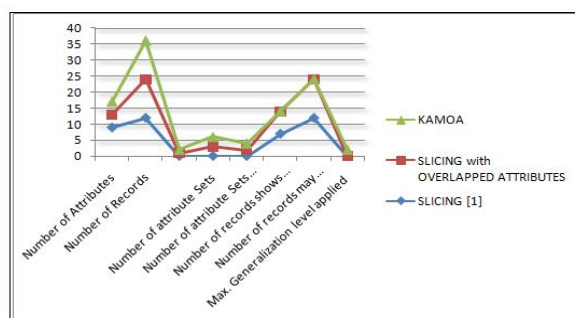


Figure 3: Graphical Comparison between KAMOA and Previous Model [1]

KAMOA is a effective model based on K-anonymity model which uses generalization technique to prevent data loss against all severe attacks also maintain corelations even using generalization technique. Results and analysis shows in section 4 represents effectiveness of model. This work can be extended for Design data mining tasks using the Anonymized data Computed by various Anonymization techniques.

REFERENCES

- [1] Tiancheng Li, Ninghui Li, Jian Zhang, Member, Ian Molloy, “” Slicing: A New Approach for Privacy Preserving Data Publishing”, IEEE Transactions on knowledge and data engineering, vol. 24, no. 3, march 2012”
- [2] Nidhi Maheshwarkar, Kshitij Pathak, Vivekanand Chourey “N-SA K-anonymity Model: A Model Exclusive of Tuple Suppression Technique”, IEEE, 2012 Third Global Congress on Intelligent Systems, Wuhan, China. November 6-8 2012. pp 229-232. 987-0-7695-4860-9/12 DOI 10.1109/GCIS.2012.77.
- [3] Nidhi Maheshwarkar, Kshitij Pathak and Narendra S Choudhari. Article: K-anonymity Model for Multiple Sensitive Attributes. IJCA Special Issue on Optimization and On-chip Communication ooc (1):51-56, February 2012. Published by Foundation of Computer Science, New York, USA.
- [4] Nidhi Maheshwarkar Kshitij Pathak and Vivekanand Chourey “Performance Evaluation of various K-anonymity Techniques”, Proc. SPIE 8350, Fourth International Conference on Machine Vision (ICMV 2011): Computer Vision and Image Analysis; Pattern Recognition and Basic Technology 83501Y (); doi:10.1117/12.921002; <http://dx.doi.org/10.1117/12.921002>.
- [5] Nidhi Maheshwarkar, Kshitij Pathak, Vivekanand Chourey “Privacy Issues for K-anonymity Model”, International Journal of Engineering Research and Application, 2011, Vol. 1, Issue 4, pp.1857-1861 ISSN No. 2248-9622.
- [6] Nidhi Maheshwarkar, Kshitij Pathak, Vivekanand Chourey “Performance Issues of Various K-anonymity Strategies”, International Journal of Computer Technology and Electronics Engineering (IJCTEE), 2011, ISSN No. 2249-6343.
- [7] Yingjie Wu, Xiaowen Ruan, Shangbin Liao, Xiaodong Wang,” P-Cover K-anonymity model for Protecting Multiple Sensitive Attributes”, IEEE, The 5th International Conference on Computer Science & Education Hefei, China. August 24–27, 2010. 978-1-4244-6005-2/10/2010 IEEE.
- [8] V. Ciriani, S. De Capitani di Vimercati, S. Foresti, P. Samarati, “K-Anonymity”, Springer US, Advances In Information Security (2007).
- [9] Tiancheng Li, Ninghui Li “. Injector: Mining Background Knowledge for Data Anonymization”, IEEE ICDE 2008.
- [10] Aristides Gionis Amon Mazza*, Tamir Tassa”. K-Anonymization Revisited”, IEEE ICDE 2008.
- [11] Xiaokui Xiao Yufei Tao” Dynamic Anonymization: Accurate Statistical Analysis with Privacy Preservation” ACM Digital Library 2008.
- [12] Wenliang Du, Zhouxuan Teng, and Zutao Zhu” Privacy-MaxEnt: Integrating Background Knowledge in Privacy Quantification”, 2008.
- [13] Manolis Terrovitis, Nikos Mamoulis, Panos Kalnis “Privacy preserving Anonymization of Setvalued Data” VLDB '08, August 24-30. ACM Digital Library.
- [14] Jiuyong Li, Raymond Chi-Wing Wong, Ada Wai-Chee Fu, Jian Pe,” Anonymization by Local Recoding in Data with Attribute Hierarchical Taxonomies” IEEE Transactions on Knowledge and Data Engineering, VOL. 20, NO. 9, Sept. 2008.
- [15] Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian “t-Closeness: Privacy Beyond k-Anonymity and t-Diversity”, ICDE, 2007.
- [16] M. Ercan Nergiz, Maurizio Atzori, Christopher W. Clifton “Hiding the Presence of Individuals from Shared Databases” SIGMOD'07, June 12–14, 2007.
- [17] Arik Friedman, Ran Wolff, Assaf Schuster” Providing k-Anonymity in Data Mining”, VLDB Journal.
- [18] Justin Brickell and Vitaly Shmatikov” Efficient Anonymity-Preserving Data Collection”, KDD'06, August 20–23, 2006.
- [19] Kristen LeFevre David J. DeWitt, Raghuram Ramakrishnan” Mondrian Multidimensional K-Anonymity” 2006.
- [20] Benjamin C.M. Fung, Ke Wang, and Philip S. Yu” Anonymizing Classification Data for Privacy Preservation”, IEEE Transactions on Knowledge and Data Engineering, VOL. 19, NO. 5, MAY 2007.

Table 1: Private Table

S. No.	Sex	Age	Marital Status	Zip Code	Disease
1.	M	23	Single	456001	Hiv
2.	M	25	Married	456010	Cancer
3.	M	26	Single	456001	Flu
4.	F	23	Divorced	456010	Cold
5.	M	23	Married	456001	Headach

Table 2: Public Table

S. No.	Name	Sex	Age	Zip Code
1.	Jasmin	F	23	456010
2.	Bob	M	26	456001
3.	John	M	29	456010
4.	Isha	F	30	456001
5.	Sam	M	23	456001
6.	Rahul	M	23	456001

Table 3: Example of linking attack

S. No.	Name	Sex	Age	Marital Status	Zip Code	Disease
1.	Rahul	M	23	Single	456001	Hiv
2.	Maddy	M	25	Married	456010	Cancer
3.	John	M	26	Single	456001	Flu
4.	Jasmin	F	23	Divorced	456010	Cold
5.	Sam	M	23	Married	456001	Headach

Table 4: Dataset

S.No.	Age	Sex	Zip Code	Disease
1.	22	M	456001	Flu
2.	23	F	456001	Headache
3.	33	F	456010	Headache
4.	54	F	456010	Cold
5.	57	M	456011	Headache
6.	70	M	456011	Flu
7.	72	M	456012	Flu
8.	63	F	456012	Gastrictis

Table 5: Sliced Dataset

S. No.	(Age	Sex)	(Zip Code	Disease)
1.	(22,	M)	(456001,	Flu)
2.	(23,	F)	(456001,	Headache)
3.	(33,	F)	(456010,	Headache)
4.	(54,	F)	(456010,	Cold)
5.	(57,	M)	(456011,	Headache)
6.	(70,	M)	(456011,	Flu)
7.	(72,	M)	(456012,	Flu)
8.	(63,	F)	(456012,	Gastrictis)

Table 6: Master Table

S.No	Name	SEX	Age	ZIP	DISEASE	WEIGHT	OCCUPATION	QUALIFICATION	SALAR Y
1	Sunil	M	21	432157	CANCER	78	MANAGER	MBA	10000
2	Ram	M	22	123456	HEART ATTACK	44	CLARK	BA	3000
3	Koshni	F	34	345678	HIV	57	STUDENT	UG	5400
4	Jasmin	F	45	678920	BRAIN TUMOR	68	STUDENT	UG	56000
5	Mishel	F	67	234567	FLU	66	BE	BCA	42000
6	Seema	F	63	234589	HEADACHE	66	CA	BCA	48000
7	Abshih	M	45	234567	CANCER	66	DEVELOPER	BCA	36000
8	Beena	F	23	678901	GASTRIC PROBLEM	68	CA	CA	56000
9	Sam	F	41	345678	CANCER	57	MANAGER	ME	54000
10	Rashmi	F	45	234567	KIDNEY FAILURE	46	DEVELOPER	BE	45000
11	Ramesh	M	46	123456	HEADACHE	40	CLARK	BA	30000
12	Rohit	M	21	432156	FLU	78	MANAGER	MBA	100000

Table 7: Revised Master Table

Table 8: Resultant Table of KAMOA

S.No.	SEX	Age	ZIP	DISEASE	WEIGHT	OCCUPATION	QUALIFICATION	SALARY
1	M	21	432187	CANCER	78	MANAGER	MBA	10000
2	M	22	123456	HEART ATTACK	44	CLARK	BA	3000
3	F	34	345678	HIV	57	STUDENT	UG	5400
4	F	45	678920	BRAIN TUMOR	68	STUDENT	UG	56000
5	F	67	234567	FLU	66	BE	BCA	42000
6	F	63	234589	HEADACHE	66	CA	BCA	48000
7	M	45	234567	CANCER	66	DEVELOPER	BCA	36000
8	F	23	678901	GASTRIC PROBLEM	68	CA	CA	56000
9	F	41	345678	CANCER	57	MANAGER	ME	54000
10	F	45	234567	KIDNEY FAILURE	46	DEVELOPER	BE	45000
11	M	46	123456	HEADACHE	40	CLARK	BA	30000
12	M	21	432186	FLU	78	MANAGER	MBA	100000

Table 9: Comparison between KAMOA and Previous Model

Slicing Algorithm	KAMOA Algorithm
Input: PT Private Table, Master Table MT, quasi-identifier QI, KA key Attributes, AS attribute Set, DF diversity factor. Output: satisfy DF and the lowest information loss in Outcome table.	Input: PT Private Table, Master Table MT, quasi-identifier QI, KA key Attributes, EQI extra quasi-identifier, AS attribute Set, OA overlapped attribute, GL generalization level, DF diversity factor. Output: Satisfy DF and no information Loss and increases diversity factor.
1. Select QI and SA from M.	1. Select KA, QI and SA from M.
2. Copy all records of Master to Temporary.	2. Remove KA and then Copy all records of Master to Temporary.
3. Create AS.	3. Select EQI from QI.
4. Apply Bucketization on AS.	4. Create AS.
5. Check for diversity.	5. Apply Generalization on EQI.
6. If table of group shows lake of diversity change the swapping manner and buckets and again check for diversity till diversity not achieved.	6. Apply Bucketization on AS.
7. Finally the result of this approach shows a table which satisfies diversity for data set contains AS without overlapped attributes and generalization.	7. Check for diversity
	8. If table of group shows lake of diversity, increase the level of generalization and again check for diversity till diversity not achieved.
	9. Finally the result of this approach shows a table which satisfies diversity for data set contains overlapped attributes.

S.No.	Age,Sex,Disease	Zip Code ,Disease	Qualification, salary
3	45*,Brain Tumor	1234**,Heart Attack	BE,45000
9	45*,Kidney Failure	1234**,Headache	UG,5400
11	45*Cancer	2345** Headache	CA,56000
12	23*, Gastric Problem	2345** Cancer	BCA,36000
4	21*Cancer	2345**,Flu	BA,30000
7	46*Headache	2345**,Kidney Failure	MBA,10000
1	21*Flu	3456**,HIV	UG,56000
10	63*,Headache	3456**,Cancer	BCA,42000
5	34*,Flu	4321**,Cancer	MBA,100000
2	41*,Cancer	4321**,Flu	BCA,48000
8	22* Heart Attack	6789**,Brain Tumor	BA,3000
6	67*,HIV	6789**,Gastric Problem	ME,54000

Table 10: Measures for Comparison between KAMOA and Previous Model [1]

S.N O.	MEASURES	SLICING [1]	SLICING with OVERLAPPED ATTRIBUTES	KAMOA
1	Number of Attributes	9	4	4
2	Number of Records	12	12	12
3	Number of Overlapped Attributes	0	1	1
4	Number of attribute Sets	0	3	3
5	Number of attribute Sets contains overlapped Attributes	0	2	2
6	Number of records shows information disclosure for sensitive Disease	7	7	0
7	Number of records may involved for linking attack	12	12	0
8	Max. Generalization level applied	0	0	2