DOI: 10.21276/sjet

Scholars Journal of Engineering and Technology (SJET) **ISSN 2321-435X (Online)**
Sch. J. Eng. Tech., 2017; 5(6):232-237 **ISSN 2347-9523 (Print)**

(An International Publisher for Academic and Scientific Resources)
www.saspublisher.com

Review Article

K-Anonymization Techniques for Hiding Multi-Sensitive Information

Anisha Tiwari[1], Minu Choudhary[2]

[1] Dept. of Computer Science and Engineering, Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India

[2] Assistant Prof., Dept. of Information Technology, Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India

***Corresponding author**
Anisha Tiwari
Email: anishatiwari20@gmail.com

Abstract: Securing information protection is an imperative issue in microdata distribution. Anonymity strategies regularly mean to ensure singular security, with insignificant effect on the nature of the discharged information. As of late, a couple of models are acquainted with guarantee the security ensuring or potentially to diminish the data misfortune to such an extent as could be allowed. That is, they additionally enhance the adaptability of the anonymous system to make it all the more near reality, and after that to meet the various needs of the general population. Different proposition and calculations have been intended for them in the meantime. In this paper, we present two anonymization techniques for adding vertices and edges for hiding useful information. The algorithm for adding edges executes faster as compared to algorithm for adding vertices but hiding information capability is more in vertices adding algorithm because it add vertices implied that it also add edges so it becomes more robust.
Keywords: Anonymity techniques, anonymity models, privacy preserving algorithm.

INTRODUCTION

Today's databases contain a considerable measure of delicate individual information. So it's significant to outline data frameworks which may confine the noteworthy of individual information. For instance, consider a healing center that keeps up patient records [1,2].

The healing facility longings to unveil information to an organization in such some way that the organization can't deduce that patients have that diseases. One system to formally indicate protection arrangements is to particular sensitive information as inquiries and implements excellent security, a terribly durable thought of security that ensures that the other question replied by the data won't uncover any information with respect to the delicate information [3,4].

Security Preserving Data Publishing

Personal records of individuals are logically being gathered by various government and organization foundations for the requirements of information examination [7,8]. The information examination is encouraged by these associations to distribute "adequately private" thoughts over this data that are collected. Privacy could be a twofold edged brand - there should be sufficient protection to ensure that touchy information concerning the general population isn't revealed by the perspectives and at a comparative time there should be sufficient data to play out the investigation. Besides, an enemy who needs to gather delicate information from the uncovered perspectives in some cases has some data concerning the general population inside the data. The principle goal is to change over the first data into some mysterious sort to prevent from inducing its record owner's sensitive information as examined in [9].

Information Anonymization

Information anonymization is the way toward expelling by and by identifiable data from informational indexes, to make the general population unknown about whom the information describe. It allows the exchange of information over a limit, as between two offices inside focus or between two offices, though lessening the peril of accidental uncovering, and in bound conditions in an exceedingly way that grants investigation and examination post-anonymization [10,11]. This system is utilized as a part of undertakings

to expand the security of the information while enabling the information to be broke down and utilized. It changes the information that will be utilized or distributed to keep the distinguishing proof of key data. Information anonymization methods, for example, k-anonymity, l-diversity qualities what's more, t-closeness are broad.

k-Anonymity: The essential arrangement of k-anonymity is to shield a dataset against re-identified by summing up the characteristics that may be used in a linkage attacks (semi identifiers). A data set is considered k-anonymous if each data thing can't be recognized from at least k-1 elective information things [12].

l-Diversity: l-diversity qualities could be an assortment of group based for the most part anonymization that is wont to safeguard security in learning sets by decreasing the coarseness of a learning portrayal. This lessening might be an exchange off that winds up in some loss of adequacy of information administration or mining algorithm in order to accomplish some security. The l-differing qualities model is related degree expansion of the k-secrecy demonstrate that diminishes the harshness of data representation exploitation procedures and in addition speculation and concealment indicated any given record maps onto at least k elective records inside the information [13].

t-Closeness: t-closeness could be an extra refinement of l-assorted qualities bunch based generally anonymization that is acclimated safeguard security in learning sets by lessening the coarseness of a data portrayal. t-closeness could be an additional refinement of l-assorted qualities group essentially based anonymization that is wont to save protection in learning sets by decreasing the coarseness of a data delineation. This decrease could be an exchange off that winds up in some loss of viability of information administration or mining algorithm in order to understand a few protection [14].

K-ANONYMITY

k-Anonymity could be a formal model of protection [16]. The objective is to frame each record unclear from an illustrated variety (k) records if tries region unit made to detect the data. An arrangement of data is k-anonymized if, for any record with a given arrangement of characteristics, there square measure in any event k-1 elective records that match these traits. The properties can be any of the accompanying sorts.

The usage of k-anonymity needs the preparatory ID of the quasi identifier. The quasi identifier depends on the outer information accessible to the beneficiary, since it decides the connecting capacity (not all conceivable outside tables range unit open to every potential learning beneficiary); and diverse quasi identifiers will without a doubt exist for a given table [15].

Example

In the event that the previously mentioned table is to be anonymized with Anonymization Level (AL) set to 2 and the arrangement of Quasi identifiers as QI = {AGE, SEX, ZIP, PHONE}.Sensitive trait = {SALARY}. The quasi identifiers and touchy qualities are distinguished by the association as indicated by their rules and regulation.

Table-1: Table to be Anonymized

ID	Age	Sex	Zip	Phone	Salary (in Rs.)
1	24	M	641015	9994258665	78000
2	23	F	641254	9994158624	45000
3	45	M	610002	8975864121	85000
4	34	M	623410	7456812312	20000

Table-2: Anonymized Table

ID	Age	Sex	Zip	Phone	Salary (in Rs.)
*	20-50	ANY	641***	999*******	78000
*	20-50	ANY	641***	999*******	45000
*	20-50	ANY	612***	897*******	85000
*	20-50	ANY	623***	745*******	20000

Generalization

Generalization is the way toward changing over an incentive into a less particular general term. For ex, "Male" and "Female" can be generalized to "Any". At the accompanying levels generalization procedures can be connected.

- Attribute (AG): Generalization is performed at the segment level; all the qualities in the section are generalized at a speculation step.
- Cell (CG): Generalization can likewise be performed on a solitary cell; at long last a summed up table may contain, for a particular

section and values at various levels of generalization.

Suppression

Suppression comprises in averting delicate information by evacuating it. Suppression can be connected at the level of single cell, whole tuple, or whole segment, permits diminishing the measure of speculation to be forced to accomplish k-anonymity.

- Tuple (TS): Suppression is performed at column level; suppression operation evacuates entire tuple
- Attribute (AS): Suppression is performed at segment level; suppression operation shrouds every one of the estimations of a segment.

LITERATURE SURVEY

Xuyun Zhang *et al*. [16] proposes giving security and protection over the intermediate data sets become dispute problem since adversaries may retain micro data by identifying multiple data records. Encryption of all datasets in general society stage called cloud take in past systems may extremely tedious and exorbitant.

Mohammad Reza Zare Mirakabad *et al*. [17] points giving protection over the information production. Under security information usage and aversion of divulgence of individual personality is more critical. One of the information anonymization methods called K-secrecy keeps the divulgence of individual character however it is for the most part neglected to accomplish.

Min Wu *et al*. [18] proposes saving security is most basic however a similar time it is inconvenience in arrival of small scale information discharge. In the perspective of trait disclosure K-namelessness is not well. So we propose new system called an ordinal separation based affectability mind full differing qualities metric model.

Yunli Wang *et al*. [19] proposes k- anonymity neglects to accomplish qualities revelation however in l-assorted qualities plans to accomplish characteristic exposure. Second information anonymization procedure focus on cutting the illation from liberated miniaturized scale traits.

Jordi Soria Comas *et al*. [20] points information anonymization strategies save protection, k- anonymity and €-differential security are two principle protection display. The t-closeness is the augmentation of k-obscurity, the development of private sensitive data depends on Bucketization algorithm.

METHODOLOGY

In this section we present the two Anonymization techniques

- With Vertices Anonymization
- With Edges Anonymization

With Vertices Anonymization

In this the vertices are anonymized. The additional vertices are deliberately added to the network or graph to hide the degree of information present in it.

With Edges Anonymization

In this the edges are anonymized. The additional edges are deliberately added to the network or graph to hide the degree of information present in it.

Adding vertices and edges are depended upon the k-anonymization algorithm.

For Edges Anonymization – Edges Anonymization algorithm is used. The algorithm effectively anonymized the data. The algorithm is presented in fig. 1.

Algorithm: Edges Anonymization
Input: An initial multi sensitive graph G(V, E)
Output: Graph G'(V', E') – with added edges
1. Get degrees in descending order **2. Anonymize the degree** **3. Using anonymize vector, add additional degree** **4. Create subgraph for the degree vector** **5. Return Anonymized graph**

Fig-1: Shows the algorithm of Anonymization using Edges

For Vertex Anonymization – Vertex Anonymization algorithm is used. The algorithm effectively anonymized the data. The algorithm is presented in fig. 1.

Algorithm: Vertices Anonymization
Input: An initial multi sensitive graph G(V, E)
Output: Graph G'(V', E') – with added vertices
1. fetch orbits from the graph using stab graph Algorithm 2. Iterate through the orbits of the graph a. Introduce new vertex and add to the graph and include it into orbit b. Get ID of the vertex c. Connect new edges according to the orbit d. In same orbit connect them by tag and in different connect them by the regular graph connection 3. Return Anonymized graph

For adding least number of edges in the graph the following equation is used:
If

$$L_1\left(\hat{\mathbf{d}}-\mathbf{d}\right)=\sum_i\left|\hat{\mathbf{d}}(i)-\mathbf{d}(i)\right|,$$

Then the minization of egdes can be converted into the problem of minimization of L_1 distance of seuqences of degree of G and G'. Based on quation:

$$G_A(\hat{G},G)=\left|\hat{E}\right|-|E| \;=\; \frac{1}{2}L_1\left(\hat{\mathbf{d}}-\mathbf{d}\right).$$

Where G is the Graph with E edges and V vertices.

d is the distance minimum from the source. And G' is the sugraph with E' and V' with edges and vertices respectively.

RESULT

The experiment are conducted using Eclipse framework on java language. We have demonstrated two algorithm for anonymization. Firstly by adding edges and secondly by adding vertices. As adding vertices is very complex process because we cannot simply add vertices into the graph, we need also to add edges by default. Hence time taken by the adding vertices is long as compared to adding edges. Fig. 2. Shows the details of Facebook network dataset.

Fig-2: Snapshot of Facebook circle dataset

Table-3: 5-Anonlymized – Adding Edges Output

Total Vertices	**4039**
Vertices Added	0
Total Edges	95420
Edges Added	7186 (8.14%)
Time Taken	3.59 sec

Table-4: 5-Anonlymized – Adding Vertices Output

Total Vertices	**4386**
Vertices Added	347 (8.59%)
Total Edges	181487
Edges Added	93253 (105.69%)
Time Taken	16.65 sec

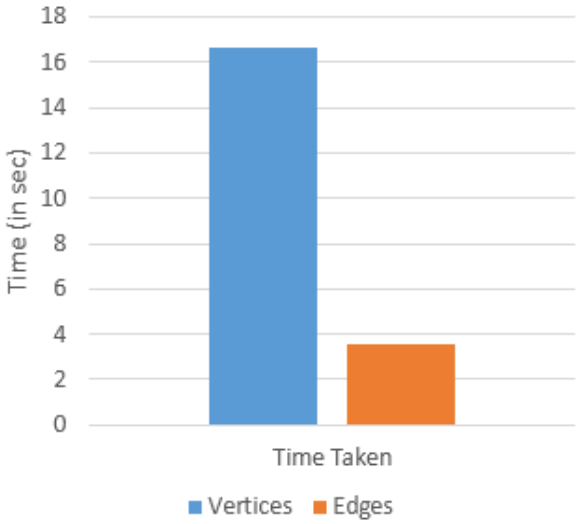

Fig-3: Shows the time taken for execution of two algorithms

CONCLUSION

The degree of a hub in a graph network, among other auxiliary attributes, can to a huge degree recognize the hub from different hubs. In this paper, we concentrated a specific graphanonymity thought that keeps the re-identification of people by an assailant with sure earlier information of the degrees. We formally defined the Graph Anonymization issue that, given an information graph requests the base number of edge increments (or erasures) that enable the change of the contribution to a degree-anonymous graph i.e., a diagram in which each hub has a similar degree with k-1 different hubs. If by change hacker knows the particular information, it cannot traverse the whole set of information. This is the concept of anonymization. Hence it is used by many organization such as Hospitals, Colleges, Universities, Secret Data holder to hide critical information.

REFERENCES

1. Kabir ME, Wang H, Bertino E. Efficient systematic clustering method for k-anonymization. Acta Informatica. 2011 Feb 1;48(1):51-66.
2. Byun JW, Kamra A, Bertino E, Li N. Efficient k-anonymization using clustering techniques. InInternational Conference on Database Systems for Advanced Applications 2007 Apr 9 (pp. 188-200). Springer, Berlin, Heidelberg.
3. Xiao X, Tao Y. Anatomy: Simple and effective privacy preservation. InProceedings of the 32nd international conference on Very large data bases 2006 Sep 1 (pp. 139-150). VLDB Endowment.
4. Zhang X, Liu C, Nepal S, Yang C, Dou W, Chen J. Combining top-down and bottom-up: scalable sub-tree anonymization over big data using MapReduce on cloud. InTrust, Security and Privacy in Computing and Communications (TrustCom), 2013 12th IEEE International Conference on 2013 Jul 16 (pp. 501-508). IEEE.
5. Goldberger J, Tassa T. Efficient anonymizations with enhanced utility. InData Mining Workshops, 2009. ICDMW'09. IEEE International Conference on 2009 Dec 6 (pp. 106-113). IEEE.
6. Terrovitis M, Mamoulis N, Kalnis P. Privacy-preserving anonymization of set-valued data. Proceedings of the VLDB Endowment. 2008 Aug 1;1(1):115-25.
7. Huda MN, Yamada S, Sonehara N. On Enhancing Utility in k-anonymization. International Journal of Computer Theory and Engineering. 2012 Aug 1;4(4):527.
8. Bhaladhare PR, Jinwala DC. Novel Approaches for Privacy Preserving Data Mining in k-Anonymity Model. J. Inf. Sci. Eng.. 2016 Jan 1;32(1):63-78.
9. Mohammed N, Fung B, Hung PC, Lee CK. Centralized and distributed anonymization for high-dimensional healthcare data. ACM Transactions on Knowledge Discovery from Data (TKDD). 2010 Oct 1;4(4):18.
10. Fienberg SE, Slavkovic A, Uhler C. Privacy preserving GWAS data sharing. InData Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on 2011 Dec 11 (pp. 628-635). IEEE.
11. Gkoulalas-Divanis A, Loukides G. PCTA: privacy-constrained clustering-based transaction data anonymization. InProceedings of the 4th International Workshop on Privacy and Anonymity in the Information Society 2011 Mar 25 (p. 5). ACM.
12. Data Anonymization", ACM 2011
13. Kisilevich S, Rokach L, Elovici Y, Shapira B. Efficient multidimensional suppression for k-anonymity. IEEE Transactions on Knowledge and Data Engineering. 2010 Mar 1;22(3):334-47.
14. Loukides G, Gkoulalas-Divanis A, Malin B. Anonymization of electronic medical records for validating genome-wide association studies.

Proceedings of the National Academy of Sciences. 2010 Apr 27;107(17):7898-903.
15. Cao J, Karras P, Raïssi C, Tan KL. ρ-uncertainty: inference-proof transaction anonymization. Proceedings of the VLDB Endowment. 2010 Sep 1;3(1-2):1033-44.
16. Loukides G, Shao J. Capturing data usefulness and privacy protection in k-anonymisation. InProceedings of the 2007 ACM symposium on Applied computing 2007 Mar 11 (pp. 370-374). ACM.
17. Zhang X, Liu C, Nepal S, Pandey S, Chen J. A privacy leakage upper bound constraint-based approach for cost-effective privacy preserving of intermediate data sets in cloud. IEEE Transactions on Parallel and Distributed Systems. 2013 Jun;24(6):1192-202.
18. Mirakabad MR, Jantan A. Diversity versus anonymity for privacy preservation. InInformation Technology, 2008. ITSim 2008. International Symposium on 2008 Aug 26 (Vol. 3, pp. 1-7). IEEE.
19. Wu M, Ye X. Towards the diversity of sensitive attributes in k-anonymity. InProceedings of the 2006 IEEE/WIC/ACM international conference on Web Intelligence and Intelligent Agent Technology 2006 Dec 18 (pp. 98-104). IEEE Computer Society.
20. Wang Y, Cui Y, Geng L, Liu H. A new perspective of privacy protection: Unique distinct l-SR diversity. InPrivacy Security and Trust (PST), 2010 Eighth Annual International Conference on 2010 Aug 17 (pp. 110-117). IEEE.