WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Hiding Multi-Sensitive Information Using Anonymization with Privacy Preserving Algorithm

Seema Sahu

Kalinga University,Raipur, Chhattisgarh, India
seemasahu1304as@gmail.com

ABSTRACT

In the dissemination of microdata, information protection is a crucial concern. With little impact on the type of information released, anonymity solutions frequently aim to provide solitary security. Recently, a few approaches have become commonplace for guaranteeing security or maybe minimizing data loss to the amount that is allowable. In other words, they also make the anonymous system more flexible to bring it closer to reality and then to accommodate the diverse demands of the populace. For these, other computations and propositions have since been made. In this article, we describe two anonymization methods that add vertices and edges to conceal relevant data. The method for adding vertices runs more quickly than the algorithm for adding edges, but the latter has a greater capacity for hiding information since adding vertices implies adding edges, making the latter technique more resilient.

Keywords— Techniques for anonymity, anonymity models, and a privacy-preserving algorithm.

Introduction

There is a significant amount of sensitive personal data in databases today. Therefore, it's important to develop data structures that can limit the significance of personal information. Consider, for instance, a medical facility that maintains patient records [1] [2].

The medical facility wants to provide information to a company in a way that prevents the company from assuming that patients have a certain ailment. One method of explicitly indicating security measures is to designate sensitive information as inquiries and use great security, a very robust idea of security that assures that the other inquiry answered by the data won't reveal any information regarding the sensitive information. [3] [4].

Security Preserving Data Publishing

Different government and organisation foundations are naturally compiling personal information about people for the sake of information examination [7] [8]. These relationships promote the distribution of "adequately private" ideas over these gathered data via the information examination. Privacy might be a double-edged sword: there should be enough security to prevent someone from disclosing sensitive information about the general public, and at the same time, there should be enough information to carry out the inquiry. Additionally, an adversary who wishes to obtain sensitive information from the uncovered views occasionally hides information about the populace [9].

Information Anonymization

Information anonymization is the process of removing consistently identifying material from informational indexes such that the general public cannot be identified as the subject of the information. It permits the sharing of information across a limited distance, such as between two offices inside a focus or between two offices, while reducing the risk of accidental discovery and under limited circumstances in a fashion that facilitates post-anonymization research and analysis [10] [11]. This system is used as part of projects to increase the security of the information while allowing it to be analysed and used. In order to maintain the distinctive evidence of the crucial data, it modifies the information that will be used or sent. Information anonymization techniques, such as k-anonymity and l-diversity characteristics Additionally, t-closeness are wide.

K-Anonymity: The fundamental idea behind k-anonymity is to protect a dataset from being re-identified by aggregating the traits that may be exploited in linkage attacks (semi identifiers). If each data item in a data collection cannot be distinguished from at least k-1 optional information items, the data set is said to be k-anonymous. [12].

L-Diversity: L-Diversity characteristics can be a variety of group-based, often anonymousization that is used to protect learning sets' security by reducing the coarseness of a learning depiction. This diminution may be a trade-off that results in a reduction in the effectiveness of information management or mining algorithms in order to achieve some security. The l-differing characteristics model is an extension of the k-secrecy demonstrate that reduces the rigour of data representation exploitation processes as well as speculation and concealment by showing that each given record maps onto at least k elective records within the data [13].

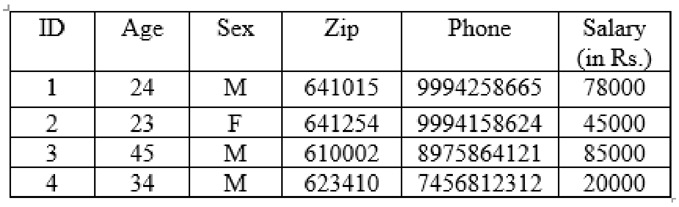
T-Closeness: By reducing the coarseness of a data depiction, t-closeness may be used to further enhance the l-assorted characteristics bunch-based generally anonymization technique used to protect learning sets' security. By reducing the coarseness of a data delineation, t-closeness might be another modification of the l-assorted characteristics group-based anonymization technique that is used to save protection in learning sets. This reduction might be a trade-off that results in some loss of viability for information management or mining algorithms in order to comprehend some security. [14].

K-ANONYMITY

A formal paradigm of protection might be k-Anonymity [16]. If attempts are made to identify the data, the goal is to isolate each record from an illustrated variety of (k) records. If there are always k-1 elective records with the same attributes as any given record with a given set of characteristics, then the set of data is k-anonymized. The characteristics may be of any of the following kinds.

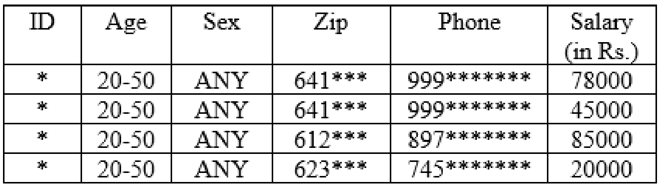
The preliminary ID of the quasi identifier is required for the use of k-anonymity. Since it determines the connecting ability (not all conceivable external tables are open to every potential learning beneficiary), the quasi identifier depends on the external information that the beneficiary has access to. Additionally, different quasi identifiers will undoubtedly exist for a particular table [15].

TABLE I: 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 II: 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

The process of generalising entails transforming an incentive into a more vague general phrase. Examples include the generalisation of "Male" and "Female" to "Any". The next layers allow for the connection of generalisation techniques.

1) Attribute (AG): Generalization is done at the segment level; at a step of speculation, all the characteristics in the section are all generalised.

2) Cell (CG): Generalization can also be done on a single cell; finally, a summarised table may include data for a specific region at several generalisation levels.

Suppression

Suppression involves avoiding sensitive information by removing it. Suppression can be connected at the level of a single cell, an entire tuple, or an entire segment, which reduces the amount of forced conjecture required to achieve k-anonymity.

1) Tuple (TS): The complete tuple is evacuated during the suppression process, which is carried out at column level.

2) Attribute (AS): Suppression is done at the segment level, and the operation hides all of the segment's estimates.

Literature Survey

In order to avoid disputes arising from attackers' potential retention of microdata through the identification of numerous data records, Xuyun Zhang et al. [16] suggest providing security and protection over intermediate data sets. It may be exceedingly time-consuming and expensive to encrypt all datasets in the general society stage of the cloud take in earlier systems.

Points made by Mohammad Reza Zare Mirakabad et al. [17] provide protection for the creation of information. Usage of information and reluctance to provide personal information are more important under security. K-secrecy, one of the information anonymization techniques, prevents the disclosure of personal information, although it is frequently ignored to carry out.

Saving security, according to Min Wu et al. [18], is essential yet at the same time slows down the release of tiny amounts of information. When it comes to trait disclosure, K-namelessness is not doing well. An ordinal separation based affectability mind full different attributes metric model is the new system that we suggest as a result.

According to Yunli Wang et al.[19], k-anonymity fails to achieve quality revelation whereas l-assorted quality plans achieve quality exposure. Cutting the illation from released miniature scale features is the second step in the information anonymization process.

Information anonymization solutions that save protection are pointed out by Jordi Soria Comas et al. [20]; the two main protection displays are k-anonymity and €-differential security. The development of private sensitive data depends on the bucketization technique, and t-closeness increases k-obscurity.

Methodology

The two anonymization strategies are presented in this section.

Vertices are anonymized, and edges are anonymized.

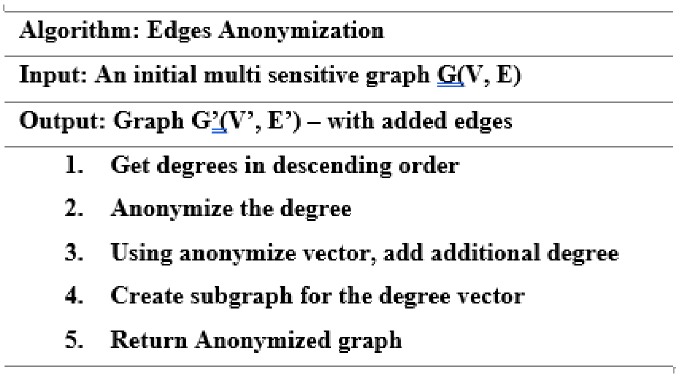
A. When Vertices Are Anonymized

Vertices in this have been made anonymous. In order to conceal the amount of information existing in the network or graph, more vertices are purposefully added to it.

Edges Anonymization B

The edges are made anonymous in this. In order to conceal the amount of information existing in the network or graph, new edges are purposefully added to it.
The k-anonymization procedure is required for the addition of vertices and edges.

The Edges Anonymization algorithm is utilised for this process. The algorithm successfully made the data anonymous. Fig. 1 presents the algorithm.



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. 2.

```
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
```

The following equation is used to the graph to add the fewest amount of edges:

If

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

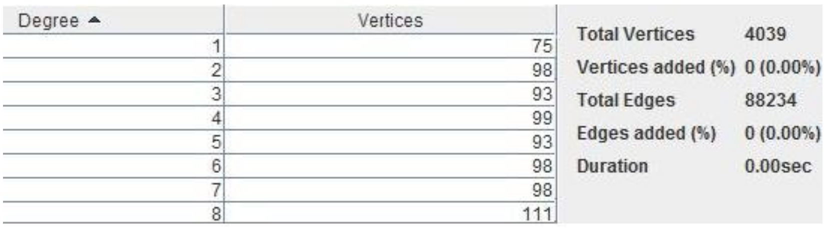
The problem of minimising of L1 distance of sequences of degree G and G' may thus be derived from the minification of edges. Using the equation:

$$\mathrm{G}_{\mathrm{A}}(\widehat{G},G)=\left|\widehat{E}\right|-|E| \quad = \quad \frac{1}{2}L_1\left(\widehat{\mathbf{d}}-\mathbf{d}\right).$$

Graph G has E edges and V vertices, and its name is G.
D is the closest point to the source. Additionally, the paragraph G' has the edges and vertices E' and V', respectively.

Results

The experiment is carried out on Java using the Eclipse framework. Two algorithms for anonymization have been shown. both by adding vertices and edges, in that order. We need to automatically create edges since putting vertices into the graph is a highly complicated procedure because we cannot just add vertices. As a result, increasing vertices takes more time than adding edges. Fig .3 displays information from the Facebook network dataset.

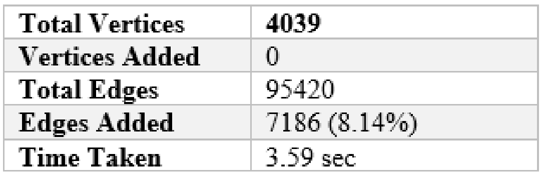


Degree ▲	Vertices
1	75
2	98
3	93
4	99
5	93
6	98
7	98
8	111

Total Vertices	4039
Vertices added (%)	0 (0.00%)
Total Edges	88234
Edges added (%)	0 (0.00%)
Duration	0.00sec

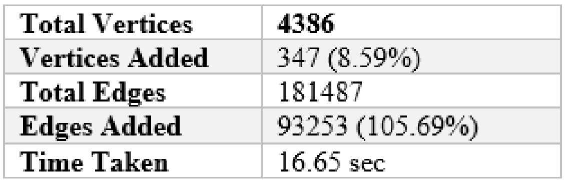
Fig. 3. Facebook circle dataset snapshot

Table: IV. Output 5-Anonlymized - Adding Edges



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

Table: IV. Output 5-Anonlymized - Adding Vertices



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

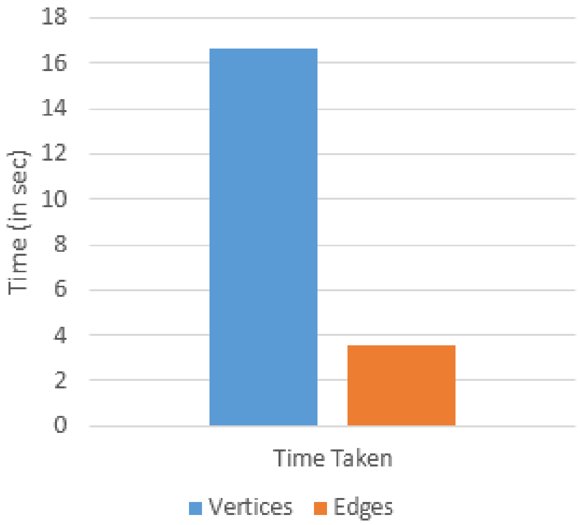


Fig. 5. Displays the duration of two algorithms' execution

Conclusion

In addition to other auxiliary properties, a hub's degree in a graph network can significantly distinguish it from other hubs. In this research, we focused on a special graphanonymity idea that prevents re-identification of persons by an attacker possessing specified previous degree information. We explicitly characterised the Graph Anonymization problem as follows: given an information graph, what is the base number of edge increments (or erasures) that permit the transformation of the contribution to a degree-anonymous graph, i.e., a diagram with k-1 distinct hubs and each hub having a comparable degree.

References

[1] M. E. Kabir, H. Wang and E. Bertino, “Efficient systematic clustering method for k-anonymization,” Acta Informatica, Springer, Vol. 48, 2011, pp. 51-66.

[2] J. W. Byun, A. Kamra, E. Bertino, and N. Li, “Efficient k-anonymization using clustering techniques,” in Proceedings of International Conference on Database Systems for Advanced Applications, 2007, pp. 188-200.

[3] X. Xiao and Y. Tao, “Anatomy: simple and effective privacy preservation,” in Proceedings of the 32nd International Conference on Very Large Data Bases, 2006, pp.139-150.

[4] Xuyun Zhang, Chang Liu, Surya Nepal, Chi Yang, Wanchun Dou, Jinjun Chen” Combining Top-Down and Bottom-Up: Scalable Sub-Tree anonymization over Big data using MapReduce on Cloud”.

[5] J. Goldberger and T. Tassa, “Efficient anonymization with enhanced utility,” Transactions on Data Privacy, Vol. 3, 2010, pp. 149-175.

[6] M. Terrovitis, N. Mamoulis, and P. Kalnis. “Privacy-preserving anonymization of set-valued data.” PVLDB, 1(1):115–125, 2008.

[7] Md Nurul Huda, Shigeki Yamada, and Noboru Sonehara, “On Enhancing Utility in k-Anonymization”, International Journal of Computer Theory and Engineering, Vol. 4, No. 4, August 2012.

[8] Pawan R. Bhaladhare and Devesh C. Jinwala, “Novel Approaches for Privacy Preserving Data Mining in k-Anonymity Model” , JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 32, 63-78 (2016).

[9] Mohammed, N. and Fung, B. C. M, “Centralized and distributed anonymization for high-dimensional healthcare data”, ACM Trans. Knowl. Discov. Data. 4, 4, Article 18 (October 2010), 33 pages.

[10] S. E. Fienberg, A. Slavkovic and C. Uhler,” Privacy Preserving GWAS Data Sharing”, 2011 11th IEEE International Conference on Data Mining Workshops.

[11] A. G. Divanis and G. Loukides,” PCTA: Privacy-constrained Clustering-based Transaction

[12] Data Anonymization”, ACM 2011.

[13] S. Kisilevich, L. Rokach, Y. Elovici, and B. Shapira. “Efficient multidimensional suppression for k-anonymity.” TKDE, 22:334–347, 2010.

[14] G. Loukides, A. Gkoulalas-Divanis, and B. Malin. Anonymization of electronic medical records for validating genome-wide association studies. PNAS, 17:7898–7903, 2010.

[15] J. Cao, P. Karras, C. Ra¨ıssi, and K. Tan. rho-uncertainty: Inference-proof transaction anonymization. PVLDB, 3(1):1033–1044, 2010.

[16] Loukides, G., Shao, J.: Capturing data usefulness and privacy protection in k-anonymisation. In: Proceedings of the 2007 ACM Symposium on Applied Computing (2007)

[17] X. Zhang, C. Liu, S. Nepal, S. Pandey and J. Chen, “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, vol. 24, no. 6, pp. 1192-1202, 2013.

[18] Mohammad Reza Zare Mirakabad School of Computer Sciences, USM, Malaysia Intern at School of Computing, NUS, Singapore reza@cs.usm.my, reza.z@comp.nus.edu.sg “Diversity versus Anonymity for Privacy Preservation”.

[19] Min Wu, Xiaojun Ye Institute of Information System and Engineering School of Software, Tsinghua University, Beijing, 100084, China “Towards the Diversity of Sensitive Attributes in k-Anonymity”.

[20] Yunli Wang, Yan Cui, Liqiang Geng and Hongyu Liu, "A new perspective of privacy protection: Unique distinct l-SR diversity," 2010 Eighth International Conference on Privacy, Security and Trust, Ottawa, ON, 2010.