European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 7, Issue 4, 2020

Sensitive Label Security Preservation with Anatomization for Data Publishing

Subapriya V, Jaichandran R, Shunmuganathan K.L, Puja Kumari, Anjali Kumari, Shivani P
Department of Computer Science and engineering, Aarupadai Veedu Institute of Technology, Vinayaka Missions Research Foundation(Deemed to be University), Paiyanoor, Tamil Nadu, India
subapriya.cse@avit.ac.in, rjaichandran@gmail.com, klsnathan@avit.ac.in
s.pujakumari@gmail.com, ba.anjalikumari@gmail.com, p.shivanikalpna@gmail.com

Abstract

Maintaining privacy in data publishing is a major challenge. In complex world sensitive information privacy is the main issue. Many algorithms are used to protect sensitive information in mined data which is not efficient because resulted output can be easily linked with public data so it shows user identity. Many techniques are used to protect privacy in data mining. Anatomization approaches aim to avoid directly use of sensitive data. The growing popularity and development of anatomization technologies bring sensitive data and protect the security of sensitive information Anatomization. The anatomization approach dissociates the correlation observed between the quasi identifier attributes and sensitive attributes and yields two separate tables with non-overlapping attributes. In the slicing algorithm, vertical partitioning does the grouping of the correlated sensitive attributes in sensitive table together and thereby minimizes the dimensionality. Consequently, it becomes increasingly important to preserve the privacy of published data. An attacker is apt to identify an individual from the published tables, with attacks through the record linkage, attribute linkage, table linkage or probabilistic attack. Two comprehensive sets of real-world relationship data are applied to evaluate the performance of our anonymization approach. Simulations and privacy analysis show our scheme possesses better privacy while ensuring higher utility.

Keywords -Anatomization, Attack, Privacy, Dimensionality, Probabilistic, Simulation.

1. Introduction

Our main research aim is to protect the privacy of the published data from the background knowledge attacks while not compromising the utility of published data. In this section, we first introduce our basic framework and then elaborate on the details of our analyzing the utility of published data with privacy to protect the privacy of the published data scheme. Due to the rapid growth of information, the demands for data collection and sharing increase sharply [1]. A great quantity of data is used for analysis, statistics and computation to find out general pattern or application which is beneficial to social development and human progress. Meanwhile, threats appear when tremendous data available for the public. For example, people can dig privacy information by getting together safe-seeming data, consequently, there is a great possibility of exposing an individual's privacy. The collection of data from published data On the one hand, it is beneficial to release data publicly and analysis activities [2]. Sensitive data contain the private or specific information on each individual. Non sensitive data can be known to the public without any concern. In this scenario, security is a major problem, especially maintaining the privacy of an individual user is a major concern. By privacy is meant that the sense of private and sensitive information of an individual is not disclosed to anyone if it is available in public [3][16]. There are many solutions which exist to protect data privacy. A detailed survey of the existing methodologies has been used to protect data privacy and the pros and cons of these approaches are discussed. Based on the background knowledge of Sensitive data an attacker with a record linkage attack, a target user can be identified from a specific record in the published tables. Sensitive data contain the private or specific information on each individual. Non sensitive data can be known to the public without any concern [4].

2. Existing System

Existing approaches to prevent the privacy leakage of the published data are categorized into the following set of analyzing operations generalization and suppression, permutation, and perturbation. The use of larger groups will help better conserve privacy, but also take longer time to collect information from the published tables [5] [6]. If the attacker does not interfere the group which the victim belongs to, the query space will be very large, which makes the probability of inferring other information on the victim be much smaller. In this section, we formally quantify the privacy guarantees following a successful execution of data based attacks by an adversary. Preserving data privacy is an essential task in order to allow such data to be published for different research and analysis purposes. As mentioned previously, privacy is defined as the reduction in entropy of sensitive information given that an adversary has access to some correlated public information [7] [8].

3. Proposed System

In many data studies, there is a need to obtain the data statistics. We define the utility based on how well one can estimate count queries, i.e. the number of records that meet a query condition needs to be found. To evaluate the anonymous data utility of our method, we first explain how to respond to a counting query to prove that our analyzing the utility of published data with privacy to protect the privacy of published data can satisfy the four privacy requirements [9] [10]. The privacy data of privacy requirements are defined to resist the record linkage attack. In this situation, the attacker aims to infer the right of the target victim with a part of Sensitive data knowledge. Privacy requirement is defined to resist the attribute linkage attack. In this case, we assume that an attacker may know a part of the victim. The attacker aims to infer the right sensitive data of the target victim. Privacy requirement is defined to resist the table linkage attack. We assume that an attacker may know a part of sensitive data of the target victim. The attacker attempts to infer whether the victim appears in the published tables. Privacy requirement is defined to resist the probabilistic attack. In this case, we assume that an attacker may know the victim's sensitive data or the probability that the victim possesses[11]. The attacker tries to infer some useful information by comparing the difference in the probability of knowing the sensitive data value before and after the analyzing.

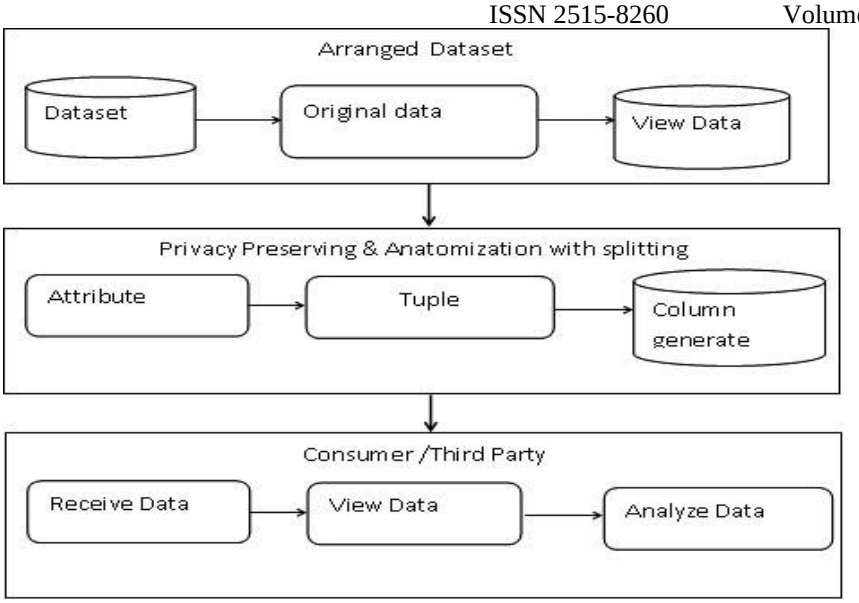

Figure 1. Anatomization with Split Dataset

4. Module Description

4.1. Admin

The process of admin is to monitor each and every action in an organization and to proceed with the action in admin. It allows system administrator to set up back-end of the system and perform basic system configuration, mainly definition of predefined drop-down fields, definition of classes time schedule, maintainability, authentication, etc. Admin allow user management which help to set up access to a single or multiple branches. Also admin can manage overall security of system like password strength, session time out, inactive accounts lockout, password reset period, etc. Security system logged all the process that are being done in the system. So it's easy to monitor overall activities of users and the values changed in the system[12].

4.2. Manager

The manager module allows users to register, log in, and log out. They benefit from being able to sign on because these associates content they create with their account and allow various permissions to be set for their roles. The manager module supports user roles, which can be set up with fine-grained permissions allowing each role to do only what the administrator permits. Each user is assigned one or more roles. By default, there are three roles: *anonymous* (a user who has not logged in) and *authenticated* (a user who is registered), and *administrator* (a signed-in user who will be assigned site administrator permissions)[13].

4.3. Employee

The employee is the actor added by the manager with certain terms. Once the employee has been added those details will be shown to the manager and the admin. They will be having complete information of the employee. The manager can able view only the persons who have been added and they will be given some task to perform by the manager. A complete information about the employees and managers will be shown only to the manager in their database [14].

4.4. Third Party

The third party is the one who will be getting some details of those employees and they will perform some actions in the local host with the collection of those data. Without the knowledge of employees, they are going to make some changes in the application they are going to theft the data and going to replace it with some other data without the knowledge of the employees. And those modified data will be shown only to the manager and the admin. Admin is the one has given the authority to handle those changed data in the database.so retrieving of data can be done only by the admin in this application [15].

4.5. Authorization

In authorization the admin will be authorizing the manager in form of approval and decline the manager will be having registration once the manager registered the details will be shown to admin by this the admin will be authorizing the manager. once the admin declines the manager will not be able to login if admin approves the manager can able to login.

4.6. Profile

In profile the manager can able to view the particular applicants who login in the application they can able view the id, designation and some other details. Where the admin can able to view the complete details.

5. Algorithm

```
        Input: Dataset in QIT, the parameter n
   Output: Split  QIT
   Procedure:
   begin
   Data=QIT, Split Box SB=∅
   While Data is not empty
{
     Remove the first box from Data
     Data=Data-{B}
      Split the box into completely two different box  using firefly algorithm
         Check the tuple from Data
         Provide  the objective function for Data
         Complete the Intensity function by the objective function
          Attractiveness == the Minimum distance between the tuples
          Using the intensity and Attractiveness from the Box Data
           Check  n  anonym it
           Data=Data  U {BI, B2}
              Else SB=SBU{B}
           Return SB
   }
   End while
   End
```

6. Implementation:

For system implementing they need to requires a suitable environment for overall system development and proper resource for completion. Following algorithm is use to split the sensitive level of data.

6.1. Flow of system Development:

First open the Register page and user has to register. After the registration complete, use have login that page and fill all information. Then anatomization technique is use to split the data into normal dataset and sensitive dataset. After using hide method sensitive dataset will be hide and secure. And if consumer want to see the data only normal dataset will be display. If consumer want to see whole information about user then consumer have to take permission from user.

Table 1. Original Dataset

Sr no	Name	Education	Gender	Salary	Disease	Medicine	Address	Pin code
1	Anu	MBBS	Female	1,00,000	Diabetes	Diafix	Patna	620026
2	Rahul	BSE	Male	50,000	BP	Lisinopril	Delhi	765432
3	Jhon	B.E	Male	70,000	Cancer	Chemotherapy	Mumbai	453279
4	Rocky	B.E	Male	80,000	Cancer	Chemotherapy	Chennai	601104
5	Riya	BSC	Female	40,000	Cold	Synus77	Chennai	601105

After applying Anatomizations technique Salary, Disease, Medicine, Address and Pin code are consider in sensitive dataset, so these all are considered as a single column and for consumer, the dataset is display like this Table.

Table 2. After Applying Anatomization, Split and Hide Rule

Sr no.	Name	Education	Gender	Column
1	Anu	MBBS	Female	**********
2	Rahul	BSE	Male	**********
3	Jhon	B.E	Male	**********
4	Rocky	B.E	Male	**********
5	Riya	BSE	Female	**********

7. Result and Analysis the dataset:

In this section, result got from using various rules such as anatomization rule, hide rule and data split rule. How much data are sensitive that is showing in this graph.

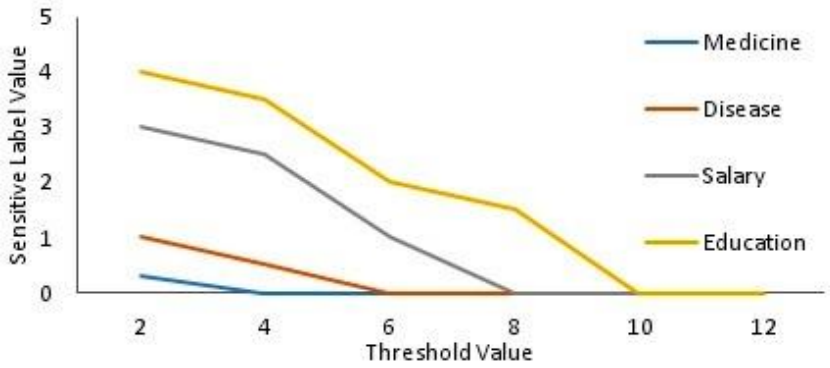

Figure 2. Sensitive Data Matrices

Anatomization technique and split technique is use to reduce the time of execution in each iteration. Here those data are more sensitive or privacy that all are convert a QI attribute and tuple in a single

column. Hidden rule hides the all sensitive data which is given by manager to admin. The first process is anatomization technique .it is use to reduce the execution time by QIT and SA. The second process is split approach, it measures how much time to require to generate the split data by attribute and tuple. And the third process hide the data after execution. According to Anatomization, split, and hiding rule graph is generate related to time execution and number of iterations.

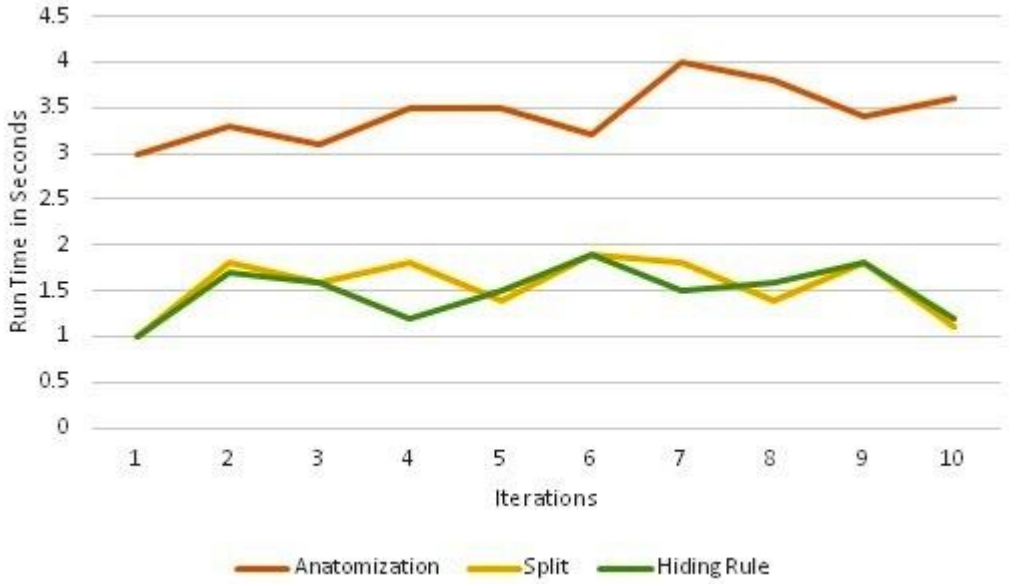

Figure 3. Anatomization using split and hide

8. Conclusion

Privacy preserving data publishing is a promising approach to information sharing, while preserving individual privacy and protecting sensitive information. To the resist attacks resulted from record linkage, table linkage and attribute linkage as well as probabilistic attacks, we propose a privacy model. Privacy protection which having complex social issues that policy make the technology and system protection. It can help to provide technical solution to the problem. This will helps in preserving co-relation, utility and anatomization minimizes the information loss. In advanced clustering algorithms exhibited its efficiency by minimizing the time and complexity.

9. References

[1] Z. H. Zou, Y. Yi, and J. N. Sun, "Entropy method for determination of weight of evaluating indicators in fuzzy synthetic evaluation for water quality assessment," Journal of Environmental Sciences, vol. 18, no. 5, pp. 1020–1023, (**2006)**.

[2] A. Gkoulalas-Divanis, G. Loukides, and J. Sun, "Publishing data from electronic health records while preserving privacy: A survey of algorithms," Journal of Biomedical Informatics, vol. 50, no. 8, pp. 4–19, **(2014)**.

[3] P. Samarati, "Protecting respondents identities in microdata re-lease," IEEE transactions on Knowledge and Data Engineering, vol. 13, no. 6, pp. 1010–1027**,(2001)**.

[4] L. Sweeney, "Achieving k-anonymity privacy protection using generalization and suppression," International Journal of Uncertain-ty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 05, pp. 571– 588, **(2002)**.

[5] S. Kiyomoto and T. Tanaka, "A user-oriented anonymization mechanism for public data," in International Conference on Inter-national Workshop on Data Privacy Management. ACM, **(2010)**, pp. 22–35.

[6] V. S. Iyengar, "Transforming data to satisfy privacy constraints," Kdd, pp. 279–288**,(2002).**

[7] X. Zhang, C. Liu, S. Nepal, and J. Chen, "An efficient quasi-identifier index based approach for privacy preservation over incremental data sets on cloud," Journal of Computer and System Sciences, vol. 79, no. 5, pp. 542–555, **(2013).**
[8] K. LeFevre, D. J. DeWitt, and R. Ramakrishnan, "Incognito: Ef-ficient full-domain k-anonymity," in Proceedings of the 2005 ACM SIGMOD international conference on Management of data. ACM, **(2005)**, pp. 49–60.
[9] R. Chen, B. C. M. Fung, N. Mohammed, B. C. Desai, and K. Wang, "Privacy-preserving trajectory data publishing by local suppres-sion," Information Sciences An International Journal, vol. 231, no. 1,83–97, (**2013)**.
[10] M. Serpell, J. Smith, A. Clark, and A. Staggemeier, "A prepro-cessing optimization applied to the cell suppression problem in statistical disclosure control," Information Sciences, vol. 238, no. 7,22–32, **(2013)**.
[11] B. C. M. Fung, K. Wang, and P. S. Yu, "Anonymizing classification data for privacy preservation," IEEE Transactions on Knowledge & Data Engineering, vol. 19, no. 5, pp. 711–725, **(2015)**.
[12] B. C. Fung, K. Wang, and P. S. Yu, "Top-down specialization for information and privacy preservation," in 21st International Conference on Data Engineering (ICDE'05). IEEE, (**2005**), pp. 205– 216.
[13] X. Sun, L. Sun, and H. Wang, "Extended k -anonymity models against sensitive attribute disclosure," Computer Communications, vol. 34, no. 4, pp. 526–535,**(2011)**.
[14] M. Ye, X. Wu, X. Hu, and D. Hu, "Anonymizing classification data using rough set theory," Knowledge-Based Systems, vol. 43, no. 2,82–94, (**2013)**.
[15] R. Mahesh and T. Meyyappan, "Anonymization technique through record elimination to preserve privacy of published data," in International Conference on Pattern Recognition, Informatics and Mobile Engineering, **(2013)**, pp. 328–332.
[16] Raja, S. Kanaga Suba, and T. Jebarajan. "Level based fault monitoring and security for long range transmission in wban." International Journal of Computer Applications 64, no. 1 (2013).