Rely on Negotiable Hash Algorithm of Data Exchange in the Relational Data Privacy Protection

YuHong Zhang*
City University of Macau, China, 100020

*Corresponding author e-mail: yuyong@163.cm

Abstract. In order to ensure that data privacy in the deal, people need to deal with the data privacy fully autonomous, it was a man of data made by a reflection, is also an effective method for individuals to take the initiative to protect privacy. At the same time in the data business, real-time data is a kind of valuable data, how to prevent the real-time data privacy, and ensure the data transaction security implementation is also an important part of the problem. So far, researchers have put forward a lot of privacy protection method; however, these models cannot effectively protect after fully autonomous data privacy, and for the privacy of the real-time release wasn't reasonable privacy protection model is put forward.

Keywords: Data transaction, Privacy security, Privacy protection, Model building

1. Introduction
With the rapid development of information industry, the renewal speed of data is becoming more and more quickly, in order to meet the demand of information to the people, the real-time data become indispensable and release form of great commercial value. In the data business, especially for the demand of real-time data, therefore to ensure the safety of the real-time data privacy in data transaction is of great significance.

Privacy leak has been in every aspect of human social life had a profound impact. If the attacker access to the user's credit card information and caused direct economic losses, or family contact information is used by others, harassed and even deceived and caused personal injury and so on[1]. At present, the issue of privacy security has been paid more and more attention. In practice, the demand for privacy protection from individuals, businesses and governments is increasing. However, each individual has a different understanding of privacy, which means that the privacy information belonging to each person is not exactly the same, and each person's security requirements for privacy protection are not the same.
In data transactions, to reflect the fairness, impartiality of trading, personal data should have complete autonomy to oneself, that is to say, people can choose what type of data, set up transaction data needs to be reached when the degree of privacy, and so on constraints[2]. Researchers at home and abroad have proposed a large number of privacy protection anonymous models to protect personal privacy, but most of the models are by the data publisher’s of a certain type of information as private information and focus on a common method to apply the same protection to all users[3]. This no comprehensive consider personal privacy information, and ignore the real personal privacy protection of different security requirements of privacy policies, will appear that personal privacy protected excessive or insufficient to be protected, so the privacy protection model is unrealistic and unable to achieve personal data safely.

For privacy after complete autonomy of data security, this paper puts forward the (alpha, Ω) anonymous models, this model allows the user to the owner (data) of relational data quasi identifier (QI), and the sensitive attribute value (S) for privacy is set at the same time, to reflect the user privacy has full autonomy, alpha is QI privacy constraints in the model, constraint Ω is S privacy. (alpha, Ω) anonymous model using granular computing and top-down local encoding differ to anonymous processing of data, this model can meet the demand of all the users' privacy, and the efficiency of the model is higher than the other anonymous model, at the same time also can ensure the lower degree of information loss.

2. Completely autonomous privacy protection research

In privacy protection, privacy is an important principle of autonomy, so far, most of the personal privacy protection model based on data privacy autonomy study is aimed at sensitive attributes, and constraints to sensitive attribute is relatively small. However, for the data provider, providing the data that can be published generally consists of two parts: the data on the quasi-identity attribute and the sensitive attribute and both parts of the data can affect the personal privacy security.

For the sensitive attribute data, the data provider in providing data, is not willing to unconditionally or without reservation to personal information provided to data collection, so the provider can be provided by the data preprocessing, to determine the data provided is complete or incomplete, or hidden (blank) directly. This approach fits well into a practical application scenario and better reflects the complete autonomy of the provider over its privacy[4]. Using Ω values in this paper, the anonymous model to describe the sensitivity to the attributes of the constraints, the implication of the sensitivity properties of minimum generalized degrees. Alpha parameter is expressed in personal hopes to achieve security, privacy belongs to passive protection of privacy, and Ω parameter is the provider to protect personal privacy information to make processing constraints, such privacy Settings can reflect completely autonomous.

Because most of the personalized anonymous modes related to privacy autonomy are for sensitive attributes, there is no reasonable protection mechanism for privacy protection of completely autonomous data publishing[5]. To this end, this paper puts forward the personalized (alpha, Ω) anonymous mode; the model can satisfy the users' in sensitive attributes and the sensitive attribute set constraints. In the model, granular computing and top-down local coding are used to process the data to be published in order to realize the privacy protection with complete privacy autonomy[6]. Experiments show that: (alpha, anonymous Ω) algorithm can very good satisfy the demands of each
user anonymity, and produce less information loss, is a realistic and reasonable personal privacy protection.

Table 1. The original data

| ID | Zip code | Birth-date | Disease | α   |
|----|----------|------------|---------|-----|
| T1 | 321811   | 2009/04/20 | AIDS    | 0.6 |
| T2 | 321875   | 2009/11/17 | Flu     | 0.4 |
| T3 | 321846   | 2009/10/25 | Cancer   | 0.3 |
| T4 | 321899   | 2010/04/05 | HIV     | 0.3 |
| T5 | 321802   | 2010/06/22 | Cancer   | 0.6 |
| T6 | 321847   | 2009/07/16 | Hepatitis| 0.3 |

Traditional personalized (alpha, Ω) anonymous model, publishers need to set up parameter k, and the top-down specialization, specialized processing is based on the data table object. However, such operation will make the equivalence class division too limited due to the constraint of k parameter, resulting in more information loss. We can use an example to describe this situation in detail. Table 1 is an original data table, and table 2 is an anonymous result that is specialized from top to bottom based on the whole table[7]. Since the frequency of Cancer in the table is 0.5, greater than 0.3, t3 needs to be hidden (deleted) to meet the requirements? However, if the data table is first granulated, no tuples need to be deleted and the data table with less information loss can be obtained.

Table 2. After dealing with the global specialization of data

| ID | Zip code | Birth-date | Disease | α   |
|----|----------|------------|---------|-----|
| T1 | 3218**   | 2009/04/20 | AIDS    | 0.6 |
| T2 | 3218**   | 2009/11/17 | Flu     | 0.4 |
| T3 | 3218**   | 2009/10/25 | Cancer   | 0.3 |
| T4 | 3218**   | 2010/04/05 | HIV     | 0.3 |
| T5 | 3218**   | 2010/06/22 | Cancer   | 0.6 |
| T6 | 3218**   | 2009/07/16 | Hepatitis| 0.3 |

Now effective algorithms are based on k - anonymous model assumes that the analysis of experimental data set is fixed, which only considers the data of a release. But this assumption is seriously flawed because in reality many applications have dynamic data sets that need to be published on a regular basis. In data trading, dynamic data release cannot meet people's strong demand for the latest information, so it is of great significance to realize real-time release. However, in the existing research, there is no model that can protect the privacy security in real-time data publishing. In this paper, we propose a real-time publishing privacy protection model based on the dynamic publishing privacy protection model. Among the dynamic privacy protection models, the two most common models are incremental privacy protection model and m-invariance privacy protection model. Based on the analysis of these two models, we can find the following defects in these two models. For example, table 1 is the original data table, table 2 meet (alpha, Ω) on condition of anonymity, the incremental information as follows:

\( t_1 = \{\text{Lily, 34, 463100, headache}\} \quad (1) \)

\( t_2 = \{\text{Lucy, 45, 463111, fat}\} \quad (2) \)

Under the background of full participation, the two incremental tuple needs to be saved in the candidate focus in order to prevent the h - difference attack. If the candidate set is not published, table
1.5 is the updated publication table. In summary, this model has a situation where information is published late and can only handle the dynamic publishing of data.

3. Conclusion

In this paper, in order to solve the above problem, use the thought of candidate sets and m-invariance, we put forward a new model to deal with dynamic data, and can make the model suitable for real time release. Sensitive attribute value in this model needs to meet the m-signature and will not be suitable for direct data released temporarily stored in the candidate set, then to blur the candidate after release, the advantage of the proposed model is: first, the model can deal with update operations such as add, delete and modify the dynamic data; Secondly, all information can be published, and there are no forged tuples in the model. Then, this model is suitable for real-time release, and compared with other sequences of high practicability; In addition, when released in real-time, the model is practical, operational efficiency, information loss and security were superior to other suitable sequence release model.

References

[1] Ta C D C, Thi T P. Improving the Algorithm for Mapping of OWL to Relational Database Schema[J]. 2015.
[2] Y.P. Wang, J.H. Ge, J.P. Shao. Research For Data Exchange Technology Of Heterogeneous Database Based On Xml[J]. Key Engineering Materials, 392-394:p.903-907.
[3] Mr Rakesh Ramesh Tannu, Prof. Sandip A. Kahate. The Constrained Method of Accessibility and Privacy Preserving Of Relational Database[J]. International Journal of Engineering Research and Applications, 2016.
[4] Lv J, Yang Q, Lv J, et al. Research on Algorithm of Partitioning Table Set in Data Exchange Process[J]. 2013, 13(21):4442-4448.
[5] Andre Hernich. Answering Non-Monotonic Queries in Relational Data Exchange[J]. Logical Methods in Computer Science, 2011, 7(3).
[6] Li, Zhi Ping, Jiang, Chuan Xian, Li, Zhi. A New Watermarking Algorithm for Relational Database Copyright Protection[J]. Applied Mechanics & Materials, 2013, 385-386:1713-1717.
[7] ZHANG Li-zhong, JIANG Nan. Research of fragile watermarking algorithm based on relational database[J]. Computer Engineering & Applications, 2008, 44(29):157-160.