An Efficient Role-based Privacy Data Dynamic Desensitization Method

Chen Jianfei[1], Tan Hu[2], Yu Hao[3], Wang Wenting[3], Zhang Hao[3]

1.State Grid Shandong Electric Power Company Jinan City,China
2.State Grid Weifang Power Supply Company Weifang City, China
3.State Grid Shandong Electric Power Company Electric Power Research Institute Jinan City, China
164656637@qq.com;445363918@qq.com;yuhao_dky@163.com;
13853115319@163.com;yuhuaxuan0677@163.com

Keywords: Privacy protection; Data integrity; Data dynamic desensitization;

Abstract: Privacy protection is very important in traditional fields such as finance and medical treatment, and it has been paid more and more attention in e-commerce and social network. Privacy protection technology needs to ensure both data use and privacy data security. In order to make the protected data still worth reading, previous scholars used data desensitization to maintain data integrity. In real life, multiple roles can be set for application systems, and different system access rights can be set for different roles. The role-based privacy data dynamic desensitization method and system proposed in this paper is designed to provide users with desensitized and protected security data sets while maintaining plaintext real data in the production database.

I. INTRODUCTION

At present, the Internet has produced quite large data, which contains privacy information. However, the incident of privacy leakage caused by insufficient protection of personal data is continuously increasing, which seriously affects people's life. The commercial value of user data has gradually formed a chain that collects, stores, sells and uses privacy data. In order to seek benefits, hackers steal user data for sale or extortion. On June 2, 2017, hackers attacked a plastic surgery hospital in Lithuania, stealing more than 25,000 private photos and private information, and asking the hospital and individuals for a ransom. Some insiders also use their positions to sell user data for high profits.

In an enterprise, different users or applications could access the database. The data in database may contain a large amount of personal privacy. If the manager cannot manage these data effectively and securely, it will cause a privacy disclosure. In real life, we can set multiple roles for an application system such as administrators, common users, and VIPs, then assign different system access permissions for those roles, and finally assign roles to users who will access the system, to implement role-based system access control.

In order to make the protected data still have reading value, many scholars use data desensitization [1-3] to transform and modify sensitive data under given rules and strategies to achieve protection of private data.

Because different roles have different protection requirements for the same or different data, this paper studies the privacy protection method from the perspective of roles based on the data desensitization. This paper studies the role-based privacy protection method by dynamic data desensitization technology.

II. RELATED WORK

Privacy protection was significant in traditional fields such as finance and medical care, and had become more and more popular in e-commerce and social networks [4]. Privacy protection technology requires both data usage and privacy data to be secure. According to different technologies, data distortion [5], data encryption [6], restricted release [7], data anonymity[8-10] and other privacy protection technologies[11] have emerged.

Currently, the research directions of privacy data protection mainly include general privacy data protection technology, privacy mining protection technology, privacy data release protection technology and privacy data protection algorithm. The specific content is shown in Table 1:

  DOI: 10.25236/icemit.2019.053

In 2011, Pomroy S P et al. [32] proposed data desensitization methods and systems, and designed a system to support data desensitization technology, which further pointed out the feasibility of data desensitization technology. In 2013, Barbas P et al. [33] used data desensitization method to achieve the protection of sensitive data. Their method desensitizes the privacy in the QUERY results to ensure the security of private data.

In 2015, Noel HE D'Costa et al. [34] and others proposed a consistent data desensitization method, which advocates each service provider to provide different types of data desensitization methods for data objects, so that desensitized data remains relevant integrity. In the same year, Shukla M et al. [35] applied dynamic data desensitization technology to protect private data, and dynamically identify and desensitize private data when using data in a production environment.

TABLE I. Research Direction Of Privacy Data Protection Technology

research direction	Related work
general privacy data protection technology	Data Anonymity[13-14], Data Disturbance [5-6], Data Randomization[7], Data Substitution [8], Data Encryption [1][9-10]
Privacy data mining protection technology	Association rules[2-3][14], data classification, data clustering [15]
Privacy data release protection technology	k-anonymity[16-19]、t-closeness[20-24]、l-diversity[24-25]、m-invariance[26]
Privacy data protection algorithm	Data Anonymity[27], Differential Privacy Algorithm [28-31], Mondrian

III. OVERALL STRUCTURE

The goal of the role-based privacy data dynamic desensitization method we proposed in this paper is to provide users with desensitized and protected data sets while storing raw data in the production database. At the same time, users with different roles require have different privacy data protection levels. Figure 1 depicts the system structure of the role-based sensitive data dynamic desensitization protection technique. The application submits the SQL request to the server. The server obtains all the query field names in the SQL query statement, and identifies the sensitive data fields according to the role of the user. Then, according to the corresponding desensitization function name and parameters in the lookup table, the server will modify the sensitive data, and then submit the SQL statement to the database. The database in the system stores the plaintext raw data.

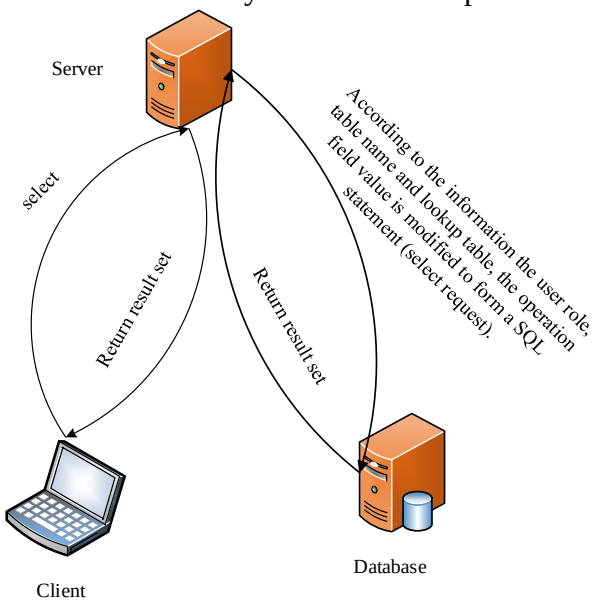

Figure 1. Overall structure

IV. IMPLEMENTATION PROCESS

First, for the relational database, the system administrator needs to set the user roles, tables, sensitive fields, desensitization function names, and desensitization parameters, and save them in the lookup table; for non-relational databases, the system administrator needs to set the user roles, sensitive data definitions, desensitization function names, and desensitization function parameters, and save them in the lookup table. Then, we can hash the contents of the lookup table according to the hash algorithm, and save the hash value in the database.

This section only introduces the privacy data protection in the relational database in detail. It consists of the following four specific processing steps, as shown in Figure 2.

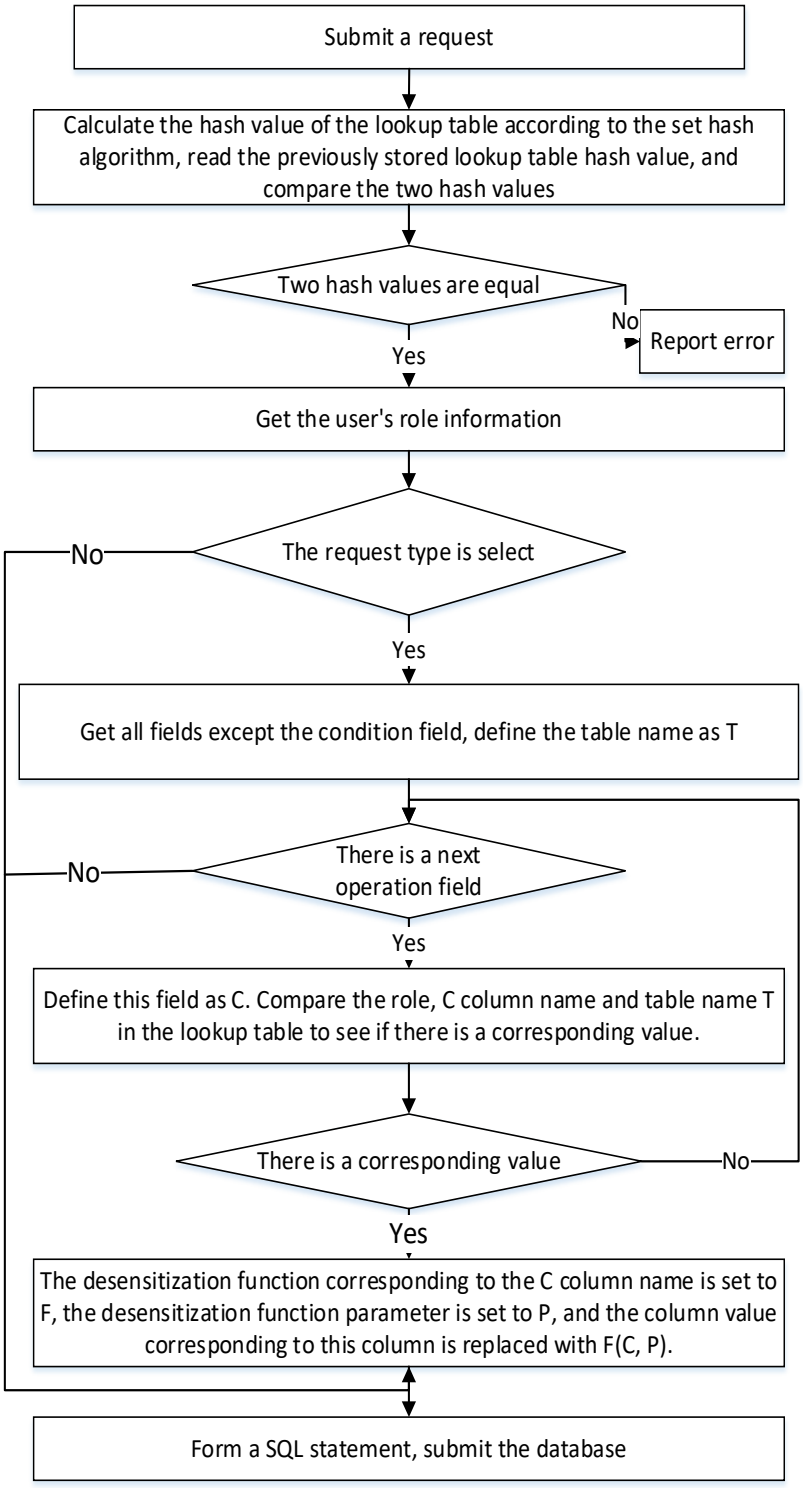

Figure 2. Processing flow chart

Specific steps are as follows:

1. Submit a query request: the client program initiates a query request;
2. Verify lookup table integrity:

The data processing unit in the server reads the lookup table, and calculates the hash value of the content of the lookup table according to the hash algorithm. Then the unit retrieves the hash value of lookup table stored previously from the database, and compares the two to verify the integrity of the contents of the table. If the two values are not consistent, the verification fails and an error appears. The lookup table is shown in Table 2.

TABLE II. Lookup Table

Role	Table Name	Column Name	Function	Parameter
customer	user	name	shield	
customer	user	age	generalize	2
VIP	user	name	shield	
VIP	user	age	generalize	2

3. Process the SQL statement and identify and modify the private data in the action field in the SQL statement according to the user's role:

If the SQL statement is a SELECT request, it will obtain the user's role information according to the login information of the user, and then obtain all the query fields and tables in the SELECT request. Next, the server searches query field value respectively in the lookup table according to the role information and the table information. Then the server obtains the corresponding desensitization functions and desensitization function parameters. The server replaces the value corresponding to the field name with desensitization function (desensitization function parameters , value), and finally assembles them into a SELECT statement.

If the SQL query request type is not a SELECT query request, the server will not process it.

For example, the server receives a SELECT query request, "select name, age from user where id = 4".

The server obtains the user's role as the customer according to the user's login information. Firstly, according to (customer, user, name), the desensitization function "shield" and the parameter are obtained in the lookup table. Secondly, it uses the function "shield" to replace the name as shield (name, null). According to (customer, user, age), the server get the desensitization function generalize and parameter 2. It uses the generalize function to replace the age as generalize (age, 2), and finally it changes the SELECT request to "select shield (name, null), generalize (age, 2) from user where id = 4.

4. Submit the modified SQL statement to the database. The desensitization functions are stored in the database. The database will use the desensitization function to desensitize the select result and then send the desensitized data the server. Then the server sends the select result to the client.

V. ATTACK MODEL AND SECURITY ANALYSIS

In this paper, we study the dynamic desensitization protection method based on the role. This method has the characteristics of maintaining the readability after the desensitization protection, and can provide different protection degrees for different roles.

This section introduces the attack model that this method may suffer from, and analyzes the security of the corresponding attack model to prove the security of the method.

1. Attack model: Crack the desensitization function

Figure 3 is an attack flow chart for cracking the desensitization function. The specific process is as follows:

1) Malicious hackers crack the desensitization function
2) Malicious hackers intercept network traffic by illegal means
3) Malicious hackers restore the acquired data to real data

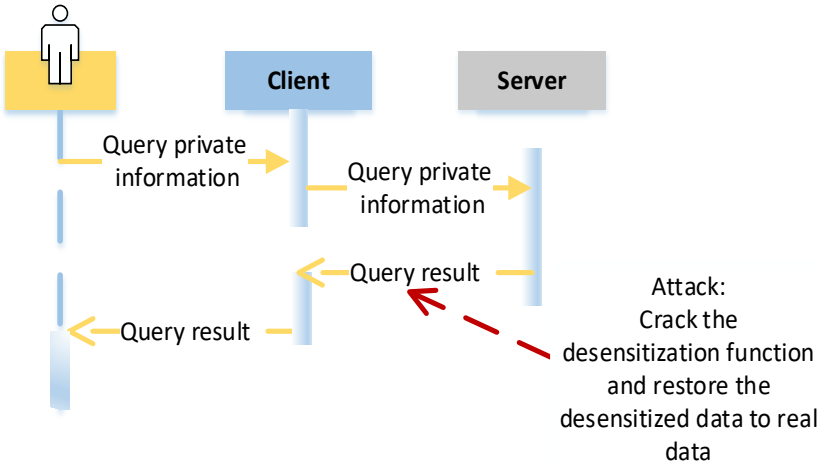

Figure 3. Crack the desensitization function

2. Security analysis

This method assigns different desensitization functions and desensitization parameters for the same sensitive fields according to different roles. When the malicious hackers crack a desensitization function, since we use various desensitization functions to protect sensitive data, it is necessary for the hacker to crack all the desensitization functions to achieve a successful attack, otherwise the attack is invalid. If we use many desensitization functions, it almost be impossible for a hacker to crack all the desensitization functions without being found.

VI. RESEARCH CHARACTERISTICS

This paper studies the role-based dynamic desensitization method for sensitive data. The database saves the raw data, so it will not affect the insert processing efficiency in an actual working environment. Only when a user reads the data, the system judges whether the user's role, the operation table and the field are sensitive, and processes the sensitive field values. The desensitization functions are stored in the database. The database will use the desensitization function to desensitize the select result and then send the desensitized data the server. The desensitization functions can get protected by the database security strategy.

The method proposed in this paper can set different desensitization methods for different fields according to different roles to meets different degrees of requirements. If a hacker wants to restore network traffic to the real data, he has to crack all the desensitization functions, so the security of the method increases if we use many desensitization functions.

REFERENCES

[1] Sivakumar T K, Sheela T, Kumar R, et al. Enhanced Secure Data Encryption Standard (ES-DES) Algorithm Using Extended Substitution Box (S-Box)[J]. International Journal of Applied Engineering Research, 2017, 12(21): 11365-11373.

[2] Zhang Peng, Tong Yunhai, Tang Shizhen, et al. An effective privacy protection association rule mining method [J]. Journal of Software, 2006, 17(8):1764-1774.

[3] Dong Bo, Wang Xue. Research on Optimization of Calculation Efficiency of Association Rules Algorithm [J]. Computer Simulation, 2017, 34(9):247-253.
[4] The Dictionary of Law, China University of Political Science and Law Press , 1990 年:416-419.
[5] Fang Yuejian, Zhu Jinzhong, Zhou Wen, et al. A review of data mining privacy protection algorithms [J]. Information Network Security, 2017(2):6-11.
[6] Tao Weiping. Classification mining method based on data perturbation privacy preservation [J]. Digital Technology and Application, 2010(9):53-53.
[7] LIen N O, Patapoutian A, Pream J J, et al. Multi-dimentional data randomization: U.S. Patent 9,576,624[P]. 2017-2-21.
[8] Mivule K. Data Swapping for Private Information Sharing of Web Search Logs[J]. Procedia Computer Science, 2017, 114: 149-158.
[9] Gai K, Qiu M, Zhao H. Privacy-Preserving Data Encryption Strategy for Big Data in Mobile Cloud Computing[J]. IEEE Transactions on Big Data, 2017.
[10] Chen Zhuang, Ye Chengyin. Cloud Audit Data Encryption Scheme Based on AES and ECC [J]. Computer Science, 2017 (S1): 333-335.
[11] Anonymous. Personal information security and privacy protection report released eight adults have encountered "unfamiliar calls" [J]. Journalist, 2016(12):94-94.
[12] Chen S Y C, Cotner C L, Kiernan G G, et al. Masking sensitive data of table columns retrieved from a database: US, US8983985[P]. 2015.
[13] Raghunathan B, Saxena V K, Subbarao V, et al. Methods and systems for runtime data anonymization: US, US 8930381 B2[P]. 2015.
[14] Pradhan G N, Prabhakaran B. Association Rule Mining in Multiple, Multidimensional Time Series Medical Data[J]. Journal of Healthcare Informatics Research, 2017, 1(1):92-118.
[15] Han X H, Quan L, Xiong X Y, et al. A novel data clustering algorithm based on modified gravitational search algorithm[J]. Engineering Applications of Artificial Intelligence, 2017, 61: 1-7.
[16] Pramanik M I, Lau R Y K, Zhang W. K-Anonymity through the Enhanced Clustering Method[C]// IEEE, International Conference on E-Business Engineering. IEEE, 2017:85-91.
[17] Gao Y, Luo T, Li J, et al. Research on K Anonymity Algorithm based on Association Analysis of Data Utility[J].
[18] LATANYA SWEENEY. ACHIEVING k-ANONYMITY PRIVACY PROTECTION USING GENERALIZATION, AND SUPPRESSION[J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002, 10(05):571-588.
[19] Sheshikala M, Prakash R V, Rao D R. Implementation of K-Anonymity Using Android SDK[C]//Advance Computing Conference (IACC), 2017 IEEE 7th International. IEEE, 2017: 866-869.
[20] Ninghui Li, Tiancheng Li, Suresh Venkatasubramanian. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity[C]// IEEE, International Conference on Data Engineering. IEEE, 2007:106-115.
[21] Soria-Comas J, Domingo-Ferrert J. Differential privacy via t-closeness in data publishing[C]// Eleventh International Conference on Privacy, Security and Trust. IEEE, 2013:27-35.
[22] Jordi Soria-Comas, Josep Domingo-Ferrer, David Sanchez, et al. t-Closeness through Microaggregation: Strict Privacy with Enhanced Utility Preservation[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(11):3098-3110.
[23] Soria-Comas J, Domingo-Ferrer J, Sánchez D, et al. t-closeness through microaggregation: Strict privacy with enhanced utility preservation[J]. IEEE Transactions on Knowledge and Data Engineering, 2015, 27(11): 3098-3110.
[24] Sei Y, Okumura H, Takenouchi T, et al. Anonymization of Sensitive Quasi-Identifiers for l-diversity and t-closeness[J]. IEEE Transactions on Dependable and Secure Computing, 2017.

[25] Yang G, Li J, Zhang S, et al. An enhanced l-diversity privacy preservation[C]//Fuzzy Systems and Knowledge Discovery (FSKD), 2013 10th International Conference on. IEEE, 2013: 1115-1120.
[26] Liao J, Jiang C, Guo C. Data privacy protection based on sensitive attributes dynamic update[C]//Cloud Computing and Intelligence Systems (CCIS), 2016 4th International Conference on. IEEE, 2016: 377-381.
[27] Qiao Hongming, Liang Wei. Discussion on Data Desensitization Methods for Operators for Big Data Applications [J]. Mobile Communication, 2015(13):17-20.
[28] Xiong Ping, Zhu Tianqing, Wang Xiaofeng. Differential Privacy Protection and Its Application [J]. Journal of Computer,2014,37(1): 101-122.
[29] Yang Y, Zhang Z, Miklau G, et al. Differential privacy in data publication and analysis[C]// 2012:601-606.
[30] Wang Junli, Guan Min, Wei Shaochen. A Survey of Differential Privacy Protection for Social Network Analysis [J]. High-tech Communication, 2015, 25(3):239-248.
[31] Piao C, Shi Y, Zhang Y, et al. Research on Government Data Publishing Based on Differential Privacy Model[C]//2017 IEEE 14th International Conference on e-Business Engineering (ICEBE). IEEE, 2017: 76-83.
[32] Pomroy S P, Lake R R, Dunn T A. Data masking system and method: US, US7974942[P]. 2011.
[33] Barbas P, Clifford A, Jenkins G, et al. Data masking:, US 20130282697 A1[P]. 2013.
[34] D'Costa N H E, Hagelund P, Henderson D J, et al. CONSISTENT DATA MASKING:, US20150113659[P]. 2015.
[35] Shukla M, Joseph J, Vidhani K, et al. Dynamic data masking:, US9171182[P]. 2015.