Research on Internet data security and privacy protection

ZHANG Ke¹,a, LI Yongzhen²,b*
¹Yanbian University of Technology, Yanji, Jilin, China
aemail: 1742582475@qq.com
*LI Yongzhen: bemail: lyz2008@ybu.edu.cn

Abstract: The company's data leakage has always been a very serious problem. Many of the leaked data is the user's personal privacy data, which will bring a lot of trouble to users. In order to improve the security of the storage and use of personal private information in the database, two methods of lightweight private information encryption and social engineering anonymization are proposed to protect database information. We compare the proposed method with the application layer encryption and pre encryption methods. The experimental results show that the proposed method can effectively protect the user's personal privacy data, and can detect and recover the privacy data when managers need it. The storage efficiency of the three methods is close. Therefore, it has good application value in privacy protection.

1. Introduction
With the development and application of network information, how to effectively protect personal information security has become a growing concern. Encryption of database is one of the effective measures to protect personal information security. At present, in the database encryption technology, there are still two main problems to be solved: one is how to effectively protect personal privacy information, the other is to improve the security of data in the database under the condition of ensuring the storage and search efficiency as far as possible[1]. In recent years, some scholars have conducted in-depth research on the security of database, and achieved some good results. For example: Liu X.Y [2] uses attribute encryption method to identify the users accessing the database. In this method, only the users meeting the attribute characteristics can access the database and save the data. But this method only realizes the hierarchical access of users, and cannot protect the user's privacy data. Yao.H[3] proposed a method of pre encryption. Although this method can effectively protect the privacy data, it needs to modify the underlying code of the database, which is difficult to achieve in practice. Zhang.Y et al. [4] proposed an encryption platform and terminal security management module technology based on database agent. Although this method can effectively protect encrypted data and realize hierarchical access to database, the encryption process of this method is too complex. It loses the advantages of database in concurrent processing and query optimization, and the performance and stability of the system depend more on security agent. Based on the above research, this paper proposes a lightweight privacy data encryption and social engineering anonymization method to protect the security of the database, and proves the effectiveness of the method through experiments.

2. Materials and Methods

2.1 Lightweight privacy data encryption
Lightweight privacy data encryption is to protect the user's privacy data by properly processing the
privacy data without affecting the efficiency of the system. This paper takes a company as an example: When employees use the software developed by the company, they first need to enter their personal information for registration and is stored in the company database\cite{5}. The registration information contains the employee's name, ID number, mobile phone number and other information, among which the ID number and mobile phone number are unique personal privacy data. Then use the simple encryption mode to encrypt the employee's ID number, mobile phone number and other private information, we can change the content of the private information, so as to better protect the private information\cite{6}. In order to avoid the risk of decoding, it is necessary to set the encryption method and content reasonably. The lightweight privacy data encryption method proposed in this paper aims at the user's unique personal privacy data encryption.

2.1.1 Experimental design of lightweight privacy data encryption

Take a company as an example to illustrate the lightweight privacy data encryption method designed in this paper. The specific steps are as follows:

Step 1, employees will register with personal information (including mobile phone number and ID card number) when they first use the company's software or programs.

Step 2, use the company's own encrypted numeric code 1 to perform XOR operation on the mobile phone number and ID number of the employee.

Step 3, use our company's own encryption code 2 to XOR the result of the first XOR for the second time, and convert the result to hexadecimal. So as to realize the lightweight encryption of private data.

Step 4, integrate encrypted data with other input information, connect to the corresponding database by using pymysql program, and store the integrated data in the corresponding table of the database. The process of lightweight privacy data encryption is shown in Figure 1.

In order to prevent the same employee from using lightweight privacy data encryption to register in different companies after leaving, different companies should set their own two groups of company code on their software to ensure the information security of employees. Through the lightweight privacy data encryption of personal data, it effectively reduces the situation of information leakage caused by dishonest companies selling personal data of employees and corporate database stolen by hackers.

2.2 Social engineering data disclosure Technology

Social engineering refers to the process of analyzing people in the society through a series of ways,
using various ideas, and achieving the goal through means application, combination analysis and other methods\cite{7}. In this paper, social engineering technology is applied to data disclosure\cite{8}. When using social engineering for data disclosure, it is necessary to change the original user's name to a pseudonym and then make it public to protect the privacy of the user's name. In this method, the change of user's alias is not to change the user's name to "Huang" or "Huang XX", but to change the user's name to other Chinese characters at random. When selecting a Chinese character library, if the existing Chinese character library (such as GBK) is used for random name replacement, the random name after replacement may duplicate the real name of other users saved in the database. It may result in unnecessary misunderstanding. In order to avoid this situation, this experiment did not use the conventional Chinese character library, but chose to build a very useful Chinese character library. In this way, in the random generation of pseudonym, the use of rare words or Chinese characters which are not commonly used in the regular name can greatly reduce the situation of duplicate names.

2.2.1 Anonymous experiment design of social engineering

After inputting the user's personal information in the program, the specific steps of social engineering anonymization are as follows:

Step 1, extract the name field of personal information.
Step 2, the user's name is renamed randomly according to the self built name database
Step 3, the corresponding code of the new name generated by random naming in the name database is used as the key, and the original user's name is encrypted by AES and encoded by Base64.

Step 4, the random new name, encrypted user name and other user information are integrated and stored in the database. The specific implementation steps are shown in Figure 2.

![Fig.2 flow chart of database name field name change](image)

When we need to check the name information in the database and detect the authenticity of the data, we need to restore the random name corresponding to the user's real name. The steps to restore the randomly changed user names are as follows:

Step 1, decode the random name field with Base64.

Step 2, find the corresponding code of the random name in the Chinese character library through the program, take it as the key, and use AES decryption to get the user's real name.

Step 3, the real name obtained by decryption is integrated with other data and exported. The specific implementation is as shown in Figure 3.
The restore operation is only used by database administrators when they need to verify, and is not open to users. Through the anonymous random name changing program, it can realize the operation of user's random name changing, and store it in the database after integration. The stored information can be searched and applied in the database as usual, and the user's real name can be restored under the administrator's operation when necessary.

3. Results & Discussion

3.1 Experimental results of lightweight privacy data encryption

The experimental results of lightweight privacy data encryption are shown in Figure 4. From Figure 4, we can see that the lightweight privacy data encryption method proposed in this paper changes the user ID number and mobile phone number field, and can effectively protect and protect the privacy information of company employees.

![Fig.4 part of the data input diagram of lightweight privacy data encryption](image)

The personal data encrypted with lightweight privacy data is stored in the database, some of which are shown in Figure 5, and the data can be stored normally.

![Fig.5 information table of some lightweight privacy information encrypted storage](image)
3.2 Input and data restoration of anonymous experiments in social engineering

3.2.1 Input of anonymous experiment in social engineering
Before inputting batch data, the test data is saved in the Excel document. Then the data in the document is read and anonymized by the program. The experimental results are shown in Figure 6.

Fig. 6 schematic diagram of partial input batch data

After inputting batch data, the system integrates the input data cycle through random name changing, AES encryption and re-coding, and then saves the integrated data in the database. As shown in Figure 7.

Fig. 7 partial data into database

3.2.2 Data restoration of anonymous experiments in social engineering
The specific operation process of data restoration is as follows: first, extract the data from the database to the specified excel file, and then restore the data in this excel by running the program. Some experimental results are shown in Figure 8. After the data is restored, the randomly changed name corresponds to the user's real name one by one, which is convenient for the data manager to check the name.

Fig. 8 schematic diagram of partial data restoration

3.3 Experimental comparative analysis
In order to verify the effectiveness of the proposed method, the method is compared with the pre-proxy database encryption method[3] and the application layer encryption method[9], and the results are shown in Table 1. The method proposed in this paper mainly aims at the protection of privacy data, while the pre-proxy database encryption and application layer encryption methods encrypt the tables in the database as a whole, but keep the data itself clear, and there is no design for privacy data protection. In the aspect of data restoration, the method proposed in this paper can restore the data quickly, while the application layer encryption and pre-proxy database encryption methods need to decrypt the database.
Table as a whole, which is inefficient.

### Table 1 security comparison of different encryption methods

| Performance comparison | Different encryption methods                      |
|------------------------|---------------------------------------------------|
|                        | Application layer encryption and pre proxy database |
|                        | encryption                                         |
| Data encryption method | The data is saved in plaintext and the database    |
|                        | table is encrypted as a whole                      |
| Data decryption and    | Decrypt the database table as a whole              |
| restore mode           | Decrypt and restore field data                     |
| Impact on data reading | The overall decryption speed of database table is  |
| efficiency             | slow                                              |
| Whether protect        | No                                                |
| personal privacy data  | Yes                                               |

### 4. Conclusions

Citizens are an important part of the Internet, and the disclosure of personal privacy data will cause great harm. After collecting users' personal information, companies need to attach great importance to personal privacy data and strengthen protection. Therefore, it is necessary to study the protection of privacy data in database. The comparative test shows that the two methods of lightweight privacy data encryption and social engineering anonymization, which can effectively protect the privacy information and reduce the risk of privacy data leakage. They are better than the methods in literature[3] and literature[9]. As a result, this method has good application value in the field of improving privacy information protection. In the future research, we will use more kinds of databases for testing experiments, continue to improve the method of this paper, and do more in-depth research on practical application.

### References

[1] Jin, L. (2015) Research on Application of trusted computing platform for information system security. Information Engineering University, Henan, China.
[2] Liu, X.Y. (2016) Study on Attribute-based Data Access Control Methods. Lanzhou University of Technology, Lanzhou, China.
[3] Yao, H. (2015) Comparison of common database encryption technology. Computer & Network, 41:36.
[4] Zhang, Y., Huang, Y., Gao, W.S. (2020) A database encryption scheme for real estate service applications. Land and Resources Informatization, 05:60-64.
[5] Ni, X.H. (2019) Research on personal information protection in the context of big data. Zhejiang University of Technology, Zhejiang, China.
[6] Justyna Ź. (2019) Personal data protection as an element of management security of information. Multidisciplinary Aspects of Production Engineering, 2:515-522.
[7] Li, Y.L. (2020) Social engineering of network information security. Technology and Economic Guide, 28:24-25.
[8] Ye, Z.W., Guo, Y.B., Ju, A.K et al. (2020) A Risk Analysis Framework for Social Engineering Attack Based on User Profiling. Journal of Organizational and End User Computing, 32:37-49.
[9] Jiang, T. (2018) Data encryption for Cloud Computing Security. Computer & Network, 44:52-53.