Design of encryption and decryption system of heterogeneous database based on Data Mining

BAI Yan-hui[1],MA Xi-jie[2]

{gaozhenyu101@2980.com[1], tiantai0200@tom.com[2]}

(1.School of Mathematics and Computer Science,Chifeng University,Chifeng 024000, China;

2.Nanjing Institute of Technology, Nanjing 211167 china)

Abstract: In view of the problem that the error rate is too high in the use of the original heterogeneous database encryption and decryption system, the original encryption and decryption system is optimized by data mining calculation, and the heterogeneous database encryption and decryption system based on data mining is designed. Through the data acquisition equipment, data storage equipment and system data transmission equipment to complete the hardware design of the system; according to the database user rights to design the database encryption dictionary, using data mining technology to complete the data preprocessing, using the form of homomorphic encryption to complete the encryption and decryption process of the database. At this point, the implementation of heterogeneous database encryption and decryption system design based on data mining. In the system test link comparing with the traditional database encryption and decryption system, by comparing the bit error rate, it is verified that the designed system can effectively reduce the bit error rate and improve the database security.

Key words: Data mining; Heterogeneous database; Data encryption; Data decryption;

1 Introduction

Database system is the most extensive and important carrier of data in the network era. Data security derives from database system security and becomes one of the basic supporting technologies in the future based on cloud computing. At present, all kinds of data stealing supported by groups or countries are more and more frequent. At the same time, data stealing behavior from internal personnel is also one of the main threats. Such unauthorized access to user access often leads to data information disclosure, user identity disclosure and authorized card number loss, which may cause a lot of economic losses, but may cause sensitive information or countries Serious consequences of disclosure of confidential information [1-2].

EAI MOBIMEDIA 2020, August 27-28, Harbin, People's Republic of China

DOI 10.4108/eai.27-8-2020.2295348

The security of database system is related to whether the data involved in the database system is acquired by unauthorized users. Although the process of database intrusion is quite different, the consequences are the same, all of which are to obtain the information of database data in the unauthorized state, which leads to serious consequences. At the same time, it also results from the abuse of the rights granted to users, the rights granted by misoperation, and the malicious intrusion of not granting the right to steal, tamper and delete data [3]. In order to ensure the information security of database, heterogeneous database is designed on the basis of traditional database. Heterogeneous database is a collection of related databases, which can realize data sharing and transparent access. While improving the autonomy of database system and realizing data sharing, each database system still has its own application characteristics, integrity control and security control. Through the process of encrypting and decrypting the internal data of the database, the information security of the database is guaranteed. In the process of using the original heterogeneous database encryption and decryption system, the problem of high error rate of data information often occurs. Therefore, this paper uses data mining technology to optimize the original system.

2 Hardware design of heterogeneous database encryption and decryption system based on Data Mining

In view of the problems in the use of the original system, the hardware design of the system only needs to complete the data acquisition and storage process. In order to improve the data processing ability of database encryption and decryption system, the hardware part of this paper is introduced into the original system hardware. In order to ensure the normal operation of the system designed in the application paper, the new hardware architecture is set as follows.

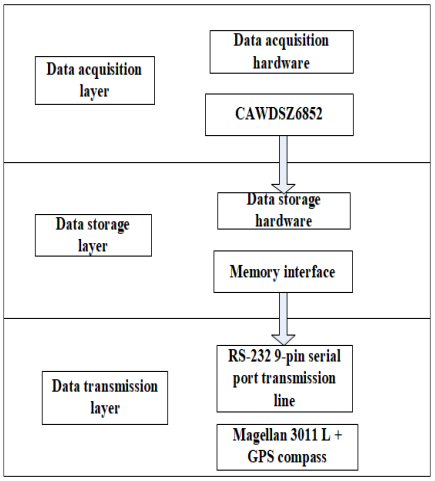

Fig. 1 System hardware design framework

According to the above hardware architecture, complete the hardware design. In view of the problems in the use of the original system, the information collector of 4-core processor is selected as the main part of the hardware design. The specific design process of the equipment is as follows.

2.1 Data acquisition equipment

In this design, high-precision data acquisition chip is used to complete the collection of data information in the database. The internal chip of data acquisition equipment is CAWDSZ68526852 [4], the chip is introduced by TI company and has the characteristics of multi-functional multimedia application. The collector has 8 parallel processors, 5 data multiplexing interfaces and 3 multi form data interfaces, which is convenient for collection, compression and transportation. The data collector is automatically connected with the hardware network of the system. After data acquisition and synthesis, it is transmitted to the hardware data terminal of the system, stored in the terminal, and recorded in the hard disk. When data acquisition equipment collects data information, SCLK is used as clock to record all data signals. The internal VPO configuration mode of the collector is rawd mode, and the input data is the basis of data encryption. In this design, the data collector is connected with PCI and HPI, the bus interface is Ethernet interface, the data path is connected with B3 and B19 by PCI bus, and the data transceiver can receive and transmit 10-100m physical layer data. In order to realize the simultaneous processing and encryption of

various database data, AP685 chip is added in the system [5], the uplink rate is 3.5 Mbit / s and the downlink rate is 6.7 Mbit / s. In order to ensure the stable operation of the above chips in the system hardware, the optimization of the computer using the chips is carried out. The optimized host parameters are set as follows.

Table 1 Computer parameter setting after optimization

Direction of use	equipment	parameter
Hardware design	chip	CAWDSZ6852
	Memory storage	16GB
		4GB RAM
		20GB ROM
	Expand storage	TF Card
Software module part	Decoding mode	EAN-8
		UPC-E
		EAN12
		UPC-A
		UCC/EAN128
		CODE BAR
		MSI PLESSEY

The optimized host is connected to other hardware devices as the hardware foundation of the system.

2.2 Data storage device

In order to improve the storage efficiency of the memory, a flash memory with large storage range and low manufacturing cost is selected in this paper. An intelligent chip is added to the memory, which greatly increases the storage capacity and reduces the occupied area. The memory design results are shown in the figure below.

Fig. 2 Data memory design results

There are 6 buses outside the memory, each bus is connected to an FPGA interface, and different interfaces connect different signals. Bus 1 connects signals in I / O mode, with a link bit width of 52, and inputs and outputs in a two-way manner to realize two-way data transmission. The remark method is I/O. The bus 2 connects signals in the OUT mode with a link bit width of 84. It outputs data in a unidirectional manner and controls the data in and out of the database. The bus 3 connects data in BSC mode with a link bit width of 16 and outputs partial data information in a unidirectional manner. Bus 4 is connected to the database by clas, the link bit width is 6, and the data is input in one-way way way to realize signal selection. The data input method of different types of database is different. Bus 5 is connected to the database in the BUSY mode, with a link bit width of 81, and a variety of data input methods are used for input. The bus 6 connects signals in the ADD mode, the link bit width is 27, and the address signal is output in a unidirectional manner [6-7].

2.3 System data transmission equipment

Serial port is the abbreviation of serial interface. In the process of data transmission, serial bit by bit transmission is adopted. The 9-pin COM port on the computer is the serial communication interface. According to the different communication modes, it can be divided into synchronous serial communication and asynchronous serial communication [8].

In asynchronous serial communication, the time interval between each bit in a single frame is fixed, while the time interval between adjacent frames is not. The

following four bits constitute a frame of asynchronous serial communication: start bit, data bit, check bit and stop bit. The maximum baud rate of asynchronous serial communication is 115200bps.

In the original serial port selection, although the data rate of the selected data serial port is relatively fast, its bit error rate is relatively high compared with other data transmission methods, and the transmission line is simple. Therefore, in this design, multi serial port is used to realize the two-way communication of the system. In this design, RS-232 9-pin serial port transmission line commonly used in embedded equipment is used to realize data transmission. The serial port is simple in structure, convenient in use, and effectively improves the security and integrity of data transmission. The specific data transmission lines and equipment settings are as follows.

(a) RS-232 9-pin serial port transmission line

(b) Magellan 3011 L + GPS compass

Fig. 3 Data transmission equipment setting

In the fig. above, the 3011 L + high-precision positioning and orientation GPS compass of Magellan navigation and positioning company of the United States selected by the system in this paper is shown. It uses serial port for data transmission, has two RS232 cdb9 interfaces, and supports the transmission baud rate of 1200115200.

The above design hardware is used in the original system hardware framework to complete the hardware design process. In the subsequent software module design, the hardware framework will be the basis of the function realization of the software module.

3 Software design of heterogeneous database encryption and decryption system based on Data Mining

In view of the problem that the error rate of the original system is too high in the process of database encryption and decryption, the encryption dictionary module and data information processing module are added to the system function module to ensure the data integrity in the process of data encryption and decryption and avoid the generation of error rate. The optimized software modules are as follows.

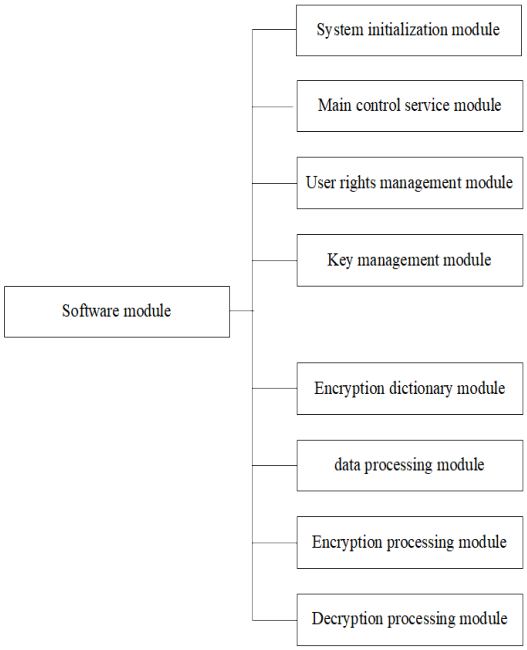

Fig. 4 System software module design framework

According to the above framework, data mining technology is used to optimize the shortcomings of the original system, and the hardware designed in this paper is used to achieve the performance of the optimized module, to ensure the realization of the encryption and decryption function of the system.

3.1 Set database encryption dictionary

In this design, heterogeneous database is used to design encryption and decryption system. Heterogeneous database is a collection system of multiple structure databases. When encrypting this kind of database, it is necessary to set the corresponding database user authority table to control the users who can encrypt and decrypt the database. The user authority table is established and maintained by the central authority of the database, which specifies the safe operation authority of each user class to each data class in detail. Table 2 is the user permission table of four user classes to three data classes, in which the field UCID is the user class ID; D is the user class's access right to the i-th data class, 3 is full control, 2 is read-write, 1 is read-only, 0 is no access; E is the number of key exports of the i-th data, and the system exports this kind of data according to the user class key and the key of a certain kind

of data Remark records some information about the user's permission. The data in Table 2 shows that user class 1 is a super user and has full control over all three types of data; user class 2 has permissions on all three types of data respectively: full control, read-write and read-only; user class 3 has only read permissions on all three types of data; user class 4 can only read the third type of data and has no permission to access other data. As shown in Table 2:

Table 2 User encryption and decryption permission settings

UCID	1	2	3	4
D1	3	3	1	0
E1	...	...		...
D2	3	2	1	0
E2	...	...	...	...
D3	3	1	1	1
E3	...	...	...	...
REMARK	information	...	...	...

Through the above-mentioned user encryption and decryption permission setting, the design process of heterogeneous database encryption dictionary is completed. The encryption dictionary is also a table, which stores the type, size, encryption and other information of each field in the data table. Through the dictionary, you can find the identity of the security data class to which the user access field belongs, and then you can find the number of key exports to export the access data class in the user permission table according to the identity. Use this encryption dictionary for effective control.

3.2 Heterogeneous database information processing based on Data Mining

Through the above processing, set the database encryption and decryption permission process. In this design, the encryption and decryption of heterogeneous database will be set. The data in heterogeneous database is too complex. In order to reduce the operation difficulty of database encryption and decryption process, data mining technology is used to preprocess the internal data of database. The specific pretreatment process is as follows.

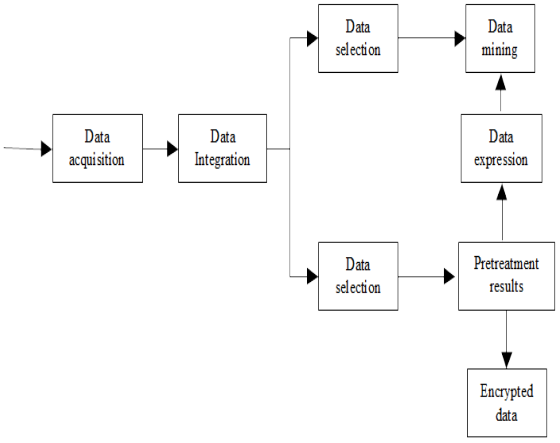

Fig. 5 Data preprocessing process

In this design, we use the classified data mining method in data mining technology to complete the data preprocessing process in heterogeneous database. Firstly, the internal data of the database is formed into training set to select the classification features. According to the selected classification features, the training data set is trained to form a classifier of heterogeneous database. Then, the classifier is used to classify the sample data of the classification database to be carried out, and finally the classification mining is completed. The whole process is divided into two stages: building the classification data mining model and using the classification data mining model to predict the target data set.

Using the data after classification mining, using Gaussian filtering and other methods to complete the preprocessing process. The specific data preprocessing process can be summarized as follows: data collection, data classification mining, data cleaning, data filtering and data composition. Through the above process, the generation of error rate in data processing is reduced, and the integrity of encrypted database content is ensured [9].

3.3 Building data encryption and decryption model

The database content after preprocessing is transformed into plaintext, which is input into the encryption and decryption module [10]. In the encryption module, there are three algorithms included in the encryption model, which will work with other modules. The SQL statement rewriting module will extract the plaintext value

involved in the database operation statement and pass it to the encryption and decryption module as input. The encryption and decryption module will determine the encryption model to be used according to the corresponding column, record location and operation in the current statement of the plaintext value in the database. If the working key needs to be generated in this process, the key management module will be called, if necessary Get the existing key, then call the metadata management module to read the key from the database. In the process of data preprocessing, data mining technology is used to complete the classification of data. Therefore, homomorphic encryption algorithm is used in this encryption module to complete the encryption and decryption of database.

Set the whole homomorphic encryption process to A, the known homomorphic encryption process consists of key generation, data encryption, decryption and data evaluation, set generation key to K, encryption process set to B, decryption process set to C, the evaluation process is D. Then the formula can be expressed as:

$$A = [K, B, C, D] \quad (1)$$

Suppose that the public key mK and the private key nK jointly generate the corresponding data security parameters, aet to o. Use mK for encrypting plaintext, nK used to decrypt ciphertext. Set plaintext $K \in n$, Where n is a positive number, then Kn is the set of integers. Homomorphic encryption expressed as $A_{cv}(w)$, then the formula can be expressed as:

$$A_{cv}(w_1 + w_2) = A_{cv}(w_1) \oplus A_{cv}(w_2) \quad (2)$$

$$A_{cv}(pw_1) = p \otimes A_{cv}(w_1) \quad (3)$$

At the same time, the data decryption process is set after the data is encrypted. Set data decryption process to C. Decrypt the ciphertext U through the private key, which can be expressed by the formula as follows:

$$U = u(C, nK) \quad (4)$$

Finally, the evaluation of the decryption results completes the calculation process of homomorphic encryption. Suppose t is the evaluation function, and the ciphertext is set to R. the evaluation algorithm evaluates the ciphertext E by evaluating the key Y evaluation function t.

$$E = C(Y, i, \ R) \quad (5)$$

Through the above formula to complete the encryption and decryption process of database, to ensure the security of heterogeneous database data.

Through the hardware part and software module part of the design, the design of heterogeneous database encryption and decryption system based on data mining is completed.

4 Experimental demonstration and analysis

Combined with the above hardware design results and software module design, the design of heterogeneous database encryption and decryption system based on data mining is completed. In order to ensure the effectiveness of the design, the performance differences between the original encryption and decryption system and the encryption and decryption system designed in this paper are compared by using the test link of the construction system.

4.1 System test platform

The hardware facilities of encryption and decryption system test of heterogeneous database based on data mining include server and other computer equipment. A system server, which is specifically configured as: CentOS version 6.5 64 bit operating system, high configuration processor, 16g memory, 4T hard disk; other computer equipment borrows computers from the existing network platform of a unit.

The system server configures the running environment and database environment required by the system, installs MySQL database, and opens the database security audit function. The development environment and language are java + mysql5.6, and the development tool is idea.

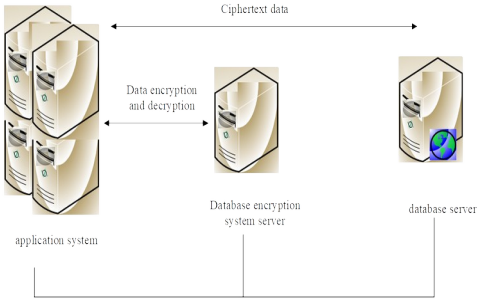

Fig. 6 System test platform architecture

Using the above platform, the system test is completed, and the performance difference between the original system and the designed system is obtained. In order to ensure the validity of the system test results. Set system test samples to provide data basis for system test. Specific system test samples are shown in the table below.

Table 3 Test sample data

Test sample No	Sample data volume	Data form
1	2000	TEXT
2	5000	TEXT
3	10000	TEXT
4	25000	TEXT
5	50000	TEXT

Using the above test sample data, using the original system and the design system to process the above data, to obtain the system test results.

4.2 Analysis of experimental results

Through the above design, the system test process is completed, and the specific system test results are as follows.

Table 4 Comparison of system test results

Test sample No	Original system bit error rate/%	Bit error rate of system

		designed in this paper/%
1	5.06	1.2
2	5.58	1.45
3	6.27	1.64
4	6.51	2.03
5	6.60	1.89

According to the above test results, the error rate of the designed system is significantly lower than the original system. Therefore, the data processing and encryption / decryption process of the system can effectively improve the data integrity in the data processing process and ensure the security and integrity of the database data content. The data processing process of the original system is relatively simple, and it is easy to create the situation of data missing, which does not meet the requirements of data security. In conclusion, the encryption and decryption performance of the design system in this paper is better than the original encryption and decryption system. Applying the design system in this paper to display life can effectively improve the internal security of the database.

5 Concluding remarks

Because the traditional heterogeneous database encryption and decryption system has a high error rate in the process of encryption and decryption, this paper designs a heterogeneous database encryption and decryption system based on data mining. Through the data acquisition equipment, data storage equipment and system data transmission equipment design system hardware; use data mining method to preprocess the data, in the form of homomorphic encryption, complete the database encryption and decryption system software design. The experimental results show that the error rate of this system is low, and the encryption and decryption performance is better than the original encryption and decryption system. We should popularize the design system in this paper, and improve the security of heterogeneous database in daily life. In this design, the description of the data processing part is relatively simple. In future applications, if there are data processing problems, they should be corrected in time to ensure the effectiveness of the data.

Reference

[1] Zhu, R. Zhou, C. Gao, R.: Research on customer relationship management system based on data mining. Modern Electronics Technique. Vol. 41, pp. 182-186 (2018)

[2] Chen, L. Huang, J. Wang R.: Overview on Security Issues and Solutions of Hadoop Big Data Platform. Computer Systems & Applications. Vol. 27, pp. 1-9 (2018)

[3]Yan, W, Z.: Simulation of Dynamic Data Security Transmission under Big Data. Computer Simulation. Vol. 35, pp. 153-156 (2018)

[4] Liu, G, R. Liu, D, X. Wang, L, F. et al.: Data encryption technology based on cloud and terminal collaboration. Telecommunications Science. Vol. 34, pp. 110-114 (2018)

[5]Zhou, Y, F.: An Improved Data Encryption Algorithm Designed for Cloud Storage Platform. Journal of Inner Mongolia Normal University(Natural Science Edition). Vol. 47, pp. 237-240 (2018)

[6] Zhang, L, P.: Research on mixed data encryption algorithm of Internet-based real-time systems. Journal of Huanggang Normal University. Vol. 38, pp. 56-60 (2018)

[7] Lei, X, F. Yang, M.: Research Progress and Trend of Educational Data Mining. Journal of Beijing University of Aeronautics and Astronautics (Social Sciences Edition). Vol. 31, pp. 108-114 (2018)

[8] Wen, Z, Y. Li, Y, D.: A big data mining algorithm based on fuzzy hierarchical clustering analysis. Journal of Hennan Institute of Engineering(Natural Science Edition). Vol. 30, pp. 70-74+80 (2018)

[9] Ye, X, Y.: Big Data Era Data Encryption Technology Application Analysis. Journal of Hubei Open Vocational College. Vol. 32, pp. 109-110+117 (2019)

[10] Tong, W. Huang, Q, P. Wang, Z, T.: Research on Location Big Data Encryption Method Based on Privacy Protection. Journal of Anhui Electrical Engineering Professional Technique College. Vol. 24, pp. 118-122 (2019)