Research on Real-time Crawling and Marking Technology of Sensitive Access SQL Statements Based on Information Network Security Isolation Device

Dong Wang1,*, Huanyu Fan2,*, Guomin Li3, Xinnan Wang3, Jia Yu4
1Network Information Centre, State Grid Blockchain Technology (Beijing) Co., Ltd., Beijing, China; State Grid E-Commerce Co., Ltd., Beijing, China; State Grid Blockchain Technology Laboratory, Beijing, China
2Customer Service Department, State Grid JiangSu Marketing Service Centre, Jiangsu, China
3Technology Department, State Grid Blockchain Technology (Beijing) Co., Ltd., Beijing, China; State Grid E-Commerce Co., Ltd., Beijing, China; State Grid Blockchain Technology Laboratory, Beijing, China
4Management Department, State Grid Changzhou Electric Power Supply Company, JiangSu, China
*a-mail wangdong@sgec.sgcc.com.cn
*b-mail lilili@sgec.sgcc.com.cn
*c-mail li-guo-min@126.com
*d-mail 2283569717@qq.com
*e-mail 1145569491@qq.com

Abstract. Based on the information network security isolation device, this paper mainly analyzed the application status of dynamic desensitization technology in the data internal and external network penetration network. And on that basis, combining with the actual power system network, a real-time capture and marking technology of sensitive access SQL statements is proposed.

1. Introduction
The current three-line defense architecture of the power grid, the information internal and external network through the network security isolation device to achieve strong logical isolation, according to functional positioning, the isolation device can only transparently transmit legal SQL statements, the traditional dynamic desensitization technology is not suitable for the application environment of strong logical isolation, it can not achieve user-level fine-grained desensitization [1].

Based on the SQL parsing engine of the information network security isolation device, this paper researched on the in-depth analysis technology of SQL statement based on syntax tree and the secure transmission protocol based on information network security isolation device, and proposed the fact-grabbing and marking technology of sensitive access SQL statements based on information network...
security isolation device, in order to collaborate with external network business systems and isolation devices to achieve application-level data differentiation and fine-grained dynamic desensitization.

2. Situation analysis

2.1. Status of information network security isolation

In order to build a three-to-depth line of defense for power grid information security, based on the information network security isolation device, high-strength logical isolation of information internal and external network systems and equipment is realized, preventing important data in the company's information network from being stolen and tampered with, and effectively ensuring the security of the company's information construction Stable operation, allowing only necessary business data between internal and external networks to be exchanged in a controllable database communication mode, the data access process is controllable, and the interactive data is true and reliable[2].

Figure 1. Schematic diagram of the architecture of data internal and external network penetration.

2.2. Dynamic data desensitization technology

Dynamic data desensitization is often used to desensitize sensitive data immediately when accessing sensitive data. It is generally used to solve scenarios where different levels of desensitization need to be performed on the same sensitive data according to different situations in the production environment[3].

Dynamic data desensitization is a process of uniquely shielding, encrypting, concealing, auditing, or blocking access to data at the user level. When applications, maintenance, and development tools request through dynamic data desensitization (DDM), the requested SQL is filtered in real time Statements, shield sensitive data based on user roles, permissions, and other desensitization rules, and can use horizontal or vertical security levels, while limiting the number of rows returned in response to a query [4][5]. According to the deployment method of the dynamic data masking system, there are three types as follow:

2.2.1. Proxy gateway type. A common deployment mode of dynamic desensitization systems is a bypass method logically and a serial method physically. The original application system establishes a connection with the database. In order to realize data desensitization, the SQL data connection request of the application system is first forwarded to the desensitization proxy system, and then the dynamic desensitization system parses the request and forwards the SQL statement to the database server. The data returned by the database server is also returned to the application server by the desensitization system after passing through the dynamic desensitization system.

This deployment method does not need to install software on the database server and application server to achieve desensitization, but it also needs to change the calling address of the application server to the database. This deployment method requires changing the application server connection from the original database to dynamically shielding the proxy gateway. This deployment mode can achieve coarse-grained desensitization for application users, and can also achieve desensitization for
operation and maintenance. The problem is that application users cannot achieve different user-level desensitization algorithms and effects. At the same time, there is a danger of bypassing operation and maintenance desensitization. DBAs may bypass the dynamic desensitization system and directly access the database address.

2.2.2. Transparent gateway type. This deployment mode is to connect the dynamic desensitization system in series between the application server and the database. Since the dynamic desensitization system can work on the second layer of OSI, no IP address is needed. To the application server and the database server, it looks like the original the same access to their real IP addresses, the dynamic desensitization system analysed the SQL statements in the traffic through protocol analysis to achieve desensitization. This deployment method does not need to change the connection settings between the application server and the database server, but it will form a single point of failure in the network. Although BYPASS technology is often used as support, all traffic will pass through the gateway, which will cause the gateway performance bottleneck.

2.2.3. Software Agent mode. The proxy service software is installed on the database server to monitor data access requests. When the requested data is sensitive data, the proxy service software will use a desensitization algorithm to desensitize the data. This deployment method requires software to be installed on the database server, and its advantage is that operations and maintenance personnel cannot bypass it. The problem is that the proxy service software will increase the burden on the database server, and there are compatibility issues. After upgrading the database server, you need to re-evaluate the compatibility of the agent [6].

3. Demand and technology
With the continuous advancement of the company's power information system construction and the continuous development of the digital economy, data interactions between the internal and external networks of the information under the new form of the Internet have become more frequent, and a large number of sensitive data on the internal network has been accessed by the information external network system[7]. Because the information network security isolation device only allows the necessary business data between the internal and external networks to be exchanged under a controllable database communication method, and prohibits any access to the Internet by the information internal network host, resulting in information internal network data and information external network business system users Permission isolation, traditional dynamic data desensitization technology cannot achieve fine-grained and differentiated data access to information extranet systems[8].

In order to reduce the sensitive attributes of sensitive data in the information intranet when used in the information extranet, strengthen the security protection of sensitive data, and realize the dynamic management of sensitive data, this paper adopts real-time capture and marking technology based on sensitive access SQL statements, to achieve data collaboration between internal and external networks.

![Diagram](image.png)

Figure 2. Real-time capture and marking technology of sensitive access SQL statements based on information network security isolation device.
The real-time capture and marking technology of sensitive access SQL statements based on the information network security isolation device includes two parts as shown in Figure 2. The first is the real-time capture technology of sensitive access SQL based on in-depth analysis, using the existing SQL analysis function of the information network security isolation device to realize the extraction of sensitive parameters of the known command mode SQL and the in-depth analysis of the unknown command mode SQL, after the extraction of sensitive parameters. The second is to develop sensitive SQL statement tagging and secure transmission protocols, and realize the secure transmission of sensitive SQL parameters by defining reserved fields in the secure transmission protocol or expanding the message header in the secure transmission protocol.

3.1. Real-time crawling of sensitive access SQL based on in-depth analysis
The real-time capture technology for sensitive access to SQL based on in-depth analysis is shown in Figure 3. First, the new instruction is matched with the SQL instruction pattern set to prevent repeated pattern construction. When the pattern matching fails, the pattern analysis of the new instruction is performed, and the newly constructed pattern is added to the SQL instruction pattern set [9]. Then extract the sensitive parameters of the SQL command.

3.1.1. Hash matching method. For each SQL statement template, two different Hash algorithms are used for calculation. The calculated hash value combination is used as the unique identifier of the sentence. At the same time, the hash value is used to generate the hash value and the index of the location of the sentence parsing content for quick retrieval. When a new instruction set arrives, it can be quickly retrieved after hash mapping.

3.1.2. Aho-Corasick automata matching method. All the pattern strings in the SQL instruction pattern set are constructed into a finite state machine. Each character or each small string corresponds to a different state, and then it will reach a different state under different subsequent inputs. Certain specific states indicate that the pattern match is successful. The new SQL command mode is used as the input of the finite state machine, and the state conversion is continuously performed according to the state conversion process of the state machine, and finally the terminal state of the input is obtained. If the termination state reaches some specific state, it means that the input pattern successfully matches the pattern in the SQL instruction pattern set; otherwise, it means that the match fails.

For unknown SQL commands that fail to match, perform lexical and grammatical analysis and generate a syntax tree, and then add the SQL command pattern set. The SQL parsing technology based on the syntax tree first performs transcoding, and converts the case, code, and comment characters of the characters in the user input data. Then, according to the lexical method of SQL, the SQL statement is decomposed to generate individual words. Syntax analysis is based on lexical analysis, and judge
whether the SQL grammar conforms to the grammatical logic according to the SQL grammar, and finally construct the SQL grammar tree.

3.2. Sensitive SQL statement marking and security protocol transmission

The information network security isolation device provides the function of accessing the internal network database by the external network business system under the conditions of the isolation of the information internal network and the information external network [10][11]. When the external network business system accesses the internal network database through the isolation device, the business system uses SG-JDBC to access the information network security isolation device through a secure custom protocol. The information network security isolation device calls OCI driver (for Oracle) or ODBC driver (for Non-Oracle database) realizes the call of information intranet database. Sensitive SQL statement marking and security protocol transmission are realized by extending the definition of the custom security protocol of the information network security isolation device, as shown in Figure 4.

![Figure 4. Schematic diagram of custom security protocol extension.](image)

In addition, for custom security protocol extensions, when the external network business system and application establish a connection with the information network security isolation device, the protocol version negotiation is completed to determine whether the sensitive identification field is defined. We can refer to the TCP three-way handshake protocol for protocol version negotiation process design.

The first handshake: When establishing a connection, the external network business system SG-JDBC driver sends a SYN packet (handshake signal) to the information network security isolation device, which contains the supported custom security protocol version, and enters the SYN_SENT state, waiting for the server to confirm.

The second handshake: When the information network security isolation device receives the SYN packet, it must confirm the SYN of the client, and at the same time send a SYN packet, that is, the SYN+ACK packet (receive the message and respond), at this time, the information network security isolation device enters the SYN_SENT state.

The third handshake: The SG-JDBC driver of the information extranet business system receives the SYN+ACK packet from the information network security isolation device and sends an acknowledgment packet ACK to the server. As this package is sent, the external network business system SG-JDBC driver and the information network security isolation device enter the Established state. That is, the connection is successful and the three-way handshake is completed.

4. Implementation method

To sum up, in the new Internet business situation where the data exchange between the internal and external networks of information is more and more frequent, a large number of sensitive data on the internal network appears to be accessed by external systems. This is aimed at the internal network data due to the strong logical isolation of the internal and external networks of the State Grid Information System. In the case of isolation from the user authority of the information extranet business system, a real-time capture and marking technology for sensitive access SQL statements is proposed.

Since the information network security isolation device is a necessary node for data interaction between the internal and external networks of the information, its reliability, stability, data throughput
and other performance indicators will affect the service performance of the information external network business application system, in order to reduce the network as much as possible. The security isolation device is modified to avoid unnecessary additional load. A data intranet and external network penetration architecture that incorporates real-time capture of sensitive access SQL statements and marking technology is proposed. As shown below.

Figure 5. Schematic diagram of the dynamic desensitization technology architecture for data internal and external network penetration.

First, as shown in Figure 5, relying on the powerful SQL analysis capabilities of the information network security isolation device, a new sensitive SQL statement capture function was added to identify and mark the sensitive SQL statements in all data requests that penetrate the information network security isolation device[12]. And then, on the basis of the existing secure transmission protocol of the information network security isolation device, a new sensitive SQL tag transmission function is added. By newly defining reserved fields in the secure transmission protocol or expanding the header of the secure transmission protocol, it will be distinguished and mark the key information of sensitive SQL statements to be transmitted to the application server of the information extranet system to realize the safe transmission of sensitive SQL parameters.

5. Conclusions

In view of the frequent data interaction between the internal and external networks of the State Grid information system and the external network system accessing sensitive data on the internal network, a real-time capture and marking technology of sensitive access SQL statements based on the information network security isolation device is proposed. This method uses in-depth analysis-based sensitive access SQL real-time capture and information network security isolation device self-defined security protocol as a means, without increasing the burden of information network security isolation device, to achieve a collaborative external network business system and isolation device Differentiated and fine-grained dynamic desensitization of data at the application level.

Acknowledgments

This work was financially supported by the State Grid Corporation of China Science and Technology Project " Research and Application of Key Technology of Data Sharing and Distribution Security for Data Centre " (5700-202090192A-0-0-00).

References

[1] Wang Cunke. The application of security isolation technology in power information network security protection[J]. Communication World, 2018(03):302-303.
[2] Liu Peng. Analysis of the status quo and strategy research of network information security under the background of the Internet [J]. China Information Technology, 2018(02): 81-84.

[3] Zhang Ningchi, Zhu Xiaojuan, Zhang Yu, Tian Ye. Discussion and research on big data desensitization methods under the Internet business model [J]. Automation Technology and Application, 2021, 40(01): 150-154.

[4] Chen Guo. Application of NoSQL database query technology in the era of big data [J]. Office Automation, 2021, 26(05): 59-60+46.

[5] Cheng Xi, Cao Xiaomei. SQL injection attack detection method based on information carrying [J/OL]. Computer Science: 1-13 [2021-04-30].

[6] Wang Zhonghua, Zhang Wenfei, Zhu Xianqing, Li Hui. Research and application of database firewall based on reverse proxy[J]. Qinghai Electric Power, 2021, 40(01): 43-47+68.

[7] Liu Xiaocui, Wang Chenchen, Yan Juan, Zhang Xiaoyu. The application of security isolation technology in power information network security protection[J]. Communication World, 2016(24): 190-191.

[8] Zhao Weiwei. The application of security isolation technical measures in the field of power information network security protection[J]. Theoretical Research on Urban Construction (Electronic Edition), 2017(23):166.

[9] Yu Bo, Peng Dunlu. SQL automatic generation model for complex query requests [J/OL]. Small microcomputer system: 1-8 [2021-04-30].

[10] Lin Qing, Wang Wei, Zhang Hongna, Zhang Xiaolin. Network information security system with isolation and information exchange[J]. Electronic Technology and Software Engineering, 2018(03):224.

[11] Liu Ling. Application of security isolation technology in power information network security protection[J]. Communication World, 2017(21):161-162.

[12] Shi Xuning, Jiang Nan, Jiang Qingshan. Database Security under SQL Injection Attacks - Effective Prevention of SQL Injection Attacks under SQL Server [J]. Computer Knowledge and Technology, 2021, 17(09): 25-26+29.