The Frontiers of Society, Science and Technology
ISSN 2616-7433 Vol. 1, Issue 8: 94-100, DOI: 10.25236/FSST.2019.010816

Research on Key Technologies of Privacy Protection Based on Big Data Environment

Hui Zeng

Header Information Technology Co., Ltd, Guangzhou Guangdong 510000, China

ABSTRACT. *With the wide application of big data in various fields, the security of big datais also receiving more and more attention, especially the protection of user privacy. In order to solve the series of security and privacy protection problems brought by big data, domestic and foreign technology researchers have put forward many countermeasures, which laid the foundation for the research of the article. Based on the description of the key technologies of big data privacy protection widely used at present, this paper proposes a whole privacy protection scheme under the big data environment, which has certain practical guiding value.*

KEYWORDS: *Big data; Privacy; Protection; Encryption; Permissions*

The development of the Internet has changed the human society. With the deep integration of social informationization and networking, the mutual penetration of Internet of Things, smart cities, mobile devices, social networks and other fields has stimulated the explosive growth of data and promoted the arrival of the era of big data. In the era of big data, the depth and breadth of data analysis has been broadened, and all industries are conducting big data analysis and mining, such as location information, consumer behavior, network access behavior and predictive analysis. But while enjoying the value and convenience of big data, we face a series of security issues in the process of data collection, storage, distribution and use. Among them, big data security and privacy protection are one of the most important issues. In recent years, data leakage incidents have occurred frequently, and all kinds of information security of users and enterprises have been invaded. According to data from the Gemalto survey, in the first half of 2018, the world's 4.5 billion data records were infringed, stolen, lost or leaked, and the number of data records increased by 133% compared with the first half of 2017. It can be seen that data security and privacy issues have become a global problem in the era of big data, and it also highlights the importance of big data security and privacy protection. The article mainly analyzes the key technologies of privacy protection and proposes a comprehensive privacy protection scheme in the big data environment.

1. Key technologies of big data privacy protection

In addition to protecting data, privacy protection technology in big data

ISSN 2616-7433 Vol. 1, Issue 8: 94-100, DOI: 10.25236/FSST.2019.010816

environment needs to prevent data theft, loss or leakage through database encryption and network security performance.

1.1 Privacy Protection Technology

1.1.1 Raw Data Protection Technology

The original data privacy protection processes the original data, such as adding interference data or anonymizing, so as to form a new data set, so that the information provided by the user does not contain obvious personal attributes, and the protection of individual privacy data is realized.

1.1.2 Access Control Privacy Protection Technology

Data access control technology is used to determine which users, what permissions, when, where, and which data to access, in order to give visitors appropriate permissions based on appropriate policies. This method starts with role access control and determines the data that each role can access through access control to prevent unauthorized users from accessing data. Access control methods usually include passwords, UKEYs, login controls, resource authorizations, logs, audits, and more. If a visitor needs to access big data, firstly it needs to be authenticated by login, and then the system will recognize its role and give it the corresponding operation permission. Only users with sufficient permissions can access the data that stores big data, and will log with the log system. The user's access record is audited on a regular basis. These actions are performed to ensure that big data has not been accessed by unauthorized users, and that responsibility is also assigned after a problem has occurred.

1.1.3 Privacy Protection System

The big data privacy protection system mainly integrates authentication services, access control services, privacy protection services, etc., in order to protect the privacy of big data from multiple dimensions. The big data privacy protection system framework mainly includes three parts: data storage, data access and data application, and adopts protection measures of big data anonymization and role access control.

1.2 Database encryption technology

Database encryption technology is to achieve encrypted transmission and storage of data from the lowest level to ensure data availability, integrity and reliability, while ensuring database availability, security, integrity, auditing and identity authentication.

1.2.1 Homomorphic encryption technology

For massive data, if all data is directly encrypted, on the one hand, it requires a lot of system overhead, and on the other hand, the encrypted data cannot be directly

ISSN 2616-7433 Vol. 1, Issue 8: 94-100, DOI: 10.25236/FSST.2019.010816

retrieved and calculated. The homomorphic encryption operation can be used to encrypt the personal data and send it to the database server, denoted by *f*(a, b, c …), then calculate the ciphertext stored in the database and return the result. The ciphertext can be used {*Enc*(a), *Enc*(b), *Enc*(c), …} means that the result is also ciphertext which can be represented by *f*({*Enc*(a), *Enc*(b), *Enc*(c), …}). The database does not know the specific content of the request, and returns the ciphertext to the user, and the user decrypts the ciphertext content. In the above process, only the user knows what the specific data is, and the storage provider cannot analyze the encrypted information, so that the user information can be prevented from being known or leaked by the database provider.

1.2.2 Database encryption system

The database encryption system needs to adhere to the principle of not changing the structure of the database itself [3]. In order to achieve this principle, the database encryption system and the database server can be separated. The database encryption system is mainly composed of communication management, data encryption and decryption, and key management, as shown in Figure 1. The architecture consists of a client computer, an application server, a proxy server and a DBMS server. The proxy server module includes a communication management module, an encryption and decryption module and a key management module. The DBMS includes an encryption key table and encrypted data.

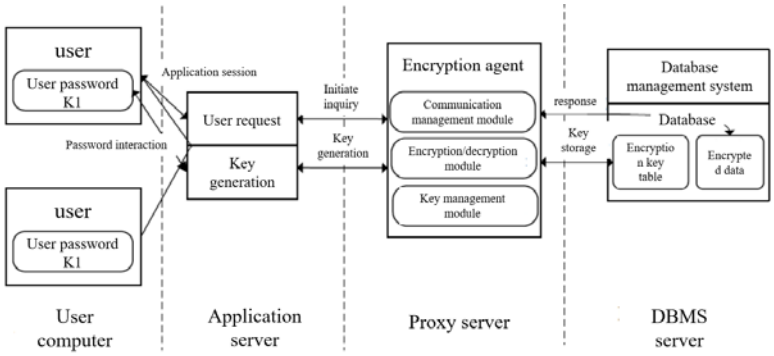

Figure.1 Database encryption system architecture

1.3 Network security technology

In the era of big data, the improvement of network security performance also protects data privacy.

1.3.1 Physical air gap

Hardware device security isolation air gap based on embedded software is divided into two architectures, namely, the dual host structure of 2 1 and the three host structure of three systems. The purpose of the air gap is to create an environment in which the internal and external networks are physically disconnected.

ISSN 2616-7433 Vol. 1, Issue 8: 94-100, DOI: 10.25236/FSST.2019.010816

The principle is to use SU-Gap security isolation technology to protect the internal network data from external network attacks.

1.3.2 Anti-DDosSystem (ADS)

DDoS attacks are one of the main forms of data infringement. DDoS attacks have the characteristics of using effective protocols, using source IP spoofing, massive distribution, and diversified attack types. At present, the anti-denial service can be used to professionally judge and identify DDOS attacks, and the ADS device can be used to pull the attack traffic in real time, so that normal access traffic can be undisturbed.

1.3.3 Next-generation Firewall (NF)

Port access control is obsolete, and APT attacks, ambiguous physical boundaries, diverse threat propagation paths, and low UTM performance have spawned the birth of next-generation firewalls. The Next-generation Firewall (NF) includes the Datacom Engine and the All-in-One Security Engine. The Datacom Engine implements the basic firewall function as the core of the entire system operation, and achieves high integration by driving the underlying data processing and high-level application security processing. Integrated security engine technology integrates multiple security policies, enabling configuration and viewing of most functions on one interface. At the same time, the user and application information obtained in the previous introduction is introduced into the integrated strategy to provide better ease of use, and the security functions of each layer are integrated into an integrated security engine. Through this engine, one decoding can be performed in parallel to process multiple times. The purpose of this function far exceeds the performance that UTM devices can provide.

2. Big data privacy protection scheme

Combined with the above-mentioned key technologies of big data privacy protection, the big data privacy protection scheme is constructed by using physical security equipment, anonymous publishing model and database encryption storage. The specific scheme is shown in Figure 2.

ISSN 2616-7433 Vol. 1, Issue 8: 94-100, DOI: 10.25236/FSST.2019.010816

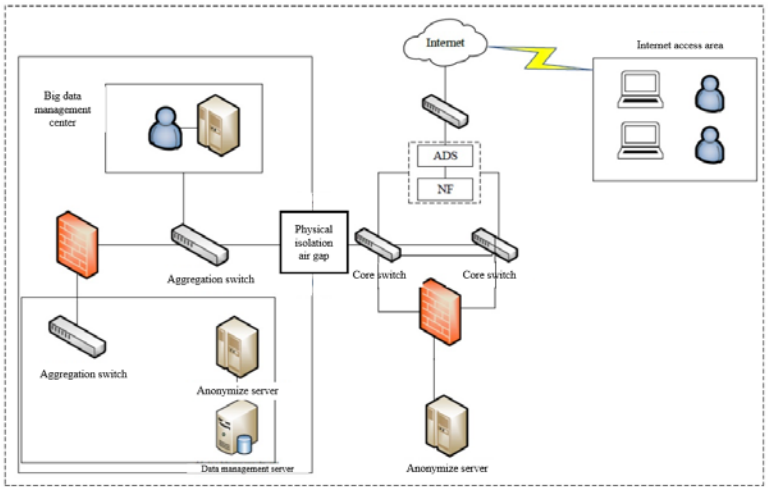

Figure.2 Big data privacy protection scheme in cloud environment

The big data privacy protection scheme proposed in the article mainly includes internal and external network isolation module, anti-malware attack module, big data anonymization release module and big data storage module. The internal and external network isolation module and the anti-malware attack module are implemented by security devices such as physical air gap, ADS, and FW. The big data anonymization publishing module is mainly implemented by an anonymization model, and the big data storage module is implemented by a big data encryption strategy. In the response to the threat between the internal network and the external network, the isolation air gap is used, and when dealing with the threat between the anonymized data and the external network, the ADS+NF scheme is used, the main reason for this difference in design is to fully consider the degree of protection and interaction required for data. The security requirements are significantly lower than the intranet. The ADS+NF solution has strong security and meets actual needs.

2.1 Internal and external network isolation module

Physical isolation adheres to the principle of physical isolation between internal and external networks and controllable information exchange between internal and external networks. The original big data in the intranet does not allow direct access by the external network, but it also needs to release the processed big data on the external network in a timely manner. Therefore, it is reasonable to use physical network gates to isolate internal and external networks, which not only can greatly improve the security of big data, but also meet the needs of application scenarios.

ISSN 2616-7433 Vol. 1, Issue 8: 94-100, DOI: 10.25236/FSST.2019.010816

2.2 Anti-Malware Attack Module

In order to prevent DDos attacks, an ADS system was introduced. Adopting ADS+NF design scheme and serial deployment mode, the two have their own functions and cooperate with each other to play a stronger protective role. ADS is responsible for filtering illegal traffic, and NF identifies and intercepts illegal attacks. The deployment method is shown in Figure 3.

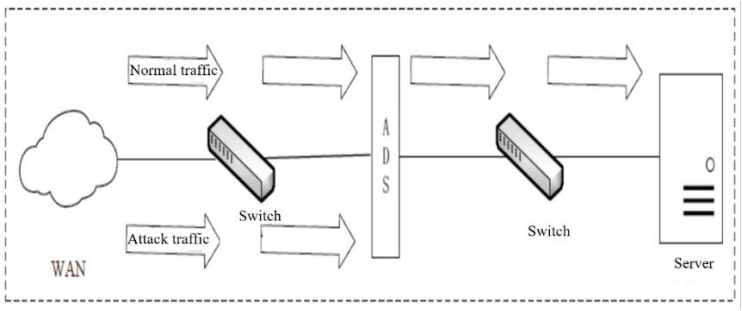

Figure. 3 ADS serial deployment.

2.3 Big Data Anonymous Publishing Module

In order to protect the original data, the anonymization model is used to anonymize the big data and then publish. Processed big data can effectively resist link attacks and consistency attacks. The principle is shown in Figure 4.

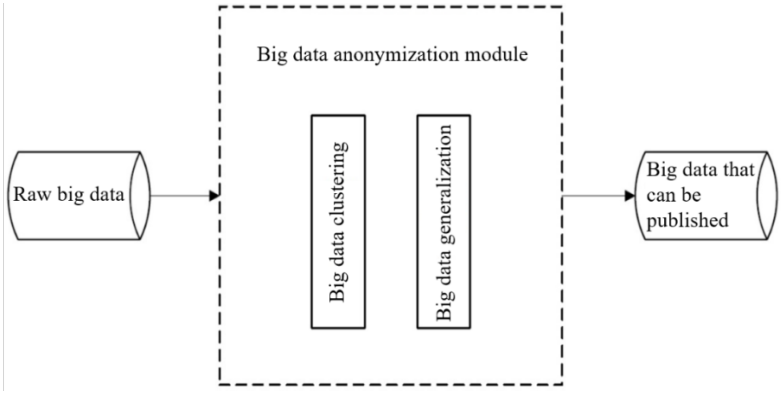

Figure. 4 Principle of big data anonymous publishing module.

2.4 Big Data Storage Module

In order to further protect the user's big data, the efficient big data encryption

ISSN 2616-7433 Vol. 1, Issue 8: 94-100, DOI: 10.25236/FSST.2019.010816

strategy is applied in the data storage layer, and the user privacy in the data is protected by the principle of encrypting the data as little as possible and consuming the system resources as little as possible. The principle is shown in Figure 5.

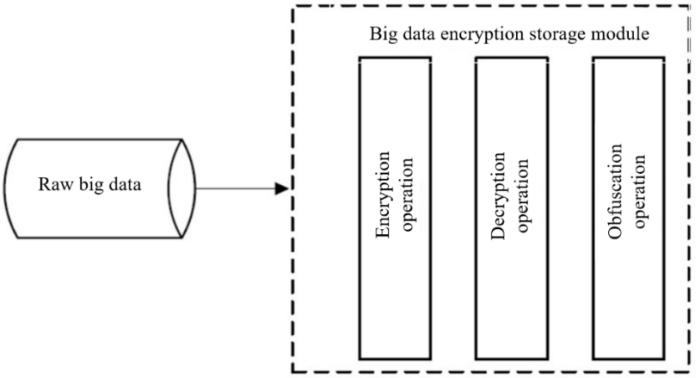

Figure. 5 The principle of the encryption module.

3. Conclusion

By applying the key technologies of big data privacy protection, the article builds a universal data privacy system that can be applied to big data privacy protection in multiple industries. However, in the actual application process, it needs to be adjusted and optimized according to the application scenario.

References

[1] Wang Songkui (2017). Discussion on Big Data Security and Privacy Protection. Network Security Technology & Application, no.9, pp.28-29.
[2] Yan Fei (2018). Research on key technologies of big data security and privacy protection. Liaoning University of Technology.
[3] Jing Xueshi (2017). Research and implementation of health big data privacy protection technology under cloud environment. University of Electronic Science and Technology of China.