Big Data Secure Storage Privacy Protection Scheme Based on Homomorphic Encryption Mechanism

Kun Qi

Shenzhen Polytechnic, Shenzhen, Guangdong 518055, China

Keywords: Homomorphic encryption, Privacy protection, Big data, Security protection

Abstract: With the rapid development of Internet technology, massive amounts of data are generated all the time on the network. Data storage security and privacy protection are major security challenges for big data. In the era of big data, most of the data is distributed storage and stored in multiple independent sites. The typical data encryption algorithm is not suitable for the security data related to big data background, so the homomorphic encryption technology can be used for data. Storage security and privacy protection. However, the existing homomorphic encryption technology based on the public key cryptosystem has high computational complexity, and is not suitable for the case where the number of users is large, the amount of calculation is large, and there is no trusted cloud center. This paper first expounds the definition of homomorphic encryption algorithm and the theory of big data privacy protection, and introduces the research status of homomorphic encryption algorithm and the protection of big data security storage privacy. Aiming at this problem, the homomorphic encryption technology is improved. By using the advantages of good security performance and low computational complexity, a big data security storage privacy protection scheme based on homomorphic encryption algorithm is proposed. The algorithm is theoretically analyzed and experimentally prove. Compared with the prior art, the method has its advantages in terms of privacy protection and computing performance.

1. Introduction

With the rapid development of computer and Internet technologies, massive data is flooding all corners, and big data has been applied and played a large role in many fields. Big data environments have created enormous challenges for network users' data privacy protection[1]. Due to the cross-domain connectivity of big data, once the server of the network service provider is attacked by an intrusion, a large amount of private information may be leaked[2]. How to ensure the safe storage and storage of user data, timely and effective detection and prevention of accidental loss and damage of data has become a hot issue of common concern in academia and industry. It is of great practical significance to study and solve the security problems caused by the protection of big data storage privacy[3]. When the server of the network service provider is invaded, the previously applied user privacy information security protection method is a security protection method for user privacy information in big data based on information decomposition and matrix and vector operation[4]. The method decomposes the user's private information, stores it in the outsourced database through encryption, first encrypts the user's private information in the big data by using various operations of the vector and the matrix, and supports fuzzy retrieval and matching of the encrypted string[5]. The four arithmetic operations of adding, subtracting, multiplying, and dividing the encrypted numerical information are performed, and the encrypted secret key is retained on the client, and only part of the information including the identification information is deleted when the information is deleted, thereby completing the security protection of the user's private information in the big data[6]. The method can improve the user privacy information while maintaining the practicability of the user's private information to a certain extent, but the encryption process is cumbersome[7]. Based on homomorphic encryption, the big data security storage privacy information security protection method, the experimental results show that the proposed method can effectively improve user information security, and the encryption and decryption speed is faster.

  DOI: 10.25236/ISMHI.2019.011

2. An Overview of the Theory of Homomorphic Encryption Algorithms and Big Data Privacy Protection

2.1 Definition of Homomorphic Encryption

Homomorphic encryption is an encryption technique that performs algebraic operations on ciphertext data that can be performed on plain-text without decrypting the ciphertext[8]. That is to say, the homomorphic encryption technique can realize that a specific algebraic algorithm for plain-text is equivalent to another (possibly different from the former) algebraic algorithm for ciphertext.

The formal definition of homomorphic encryption is as follows:

Let m_0 and m_1 be the elements in the plain-text set M, f be the operation on M, and E be the encryption algorithm on M, if there is a valid operation function F, so that

$$F\left(E\left(m_0\right), E\left(m_1\right)\right) = E\left(f\left(m_0, m_1\right)\right) \tag{1}$$

It is said that the encryption algorithm E is homomorphic to the operation f.

Full homomorphic encryption allows anyone, including those who do not decrypt the dense copper, to perform any algebraic operations on the ciphertext data, and the result obtained by decrypting the obtained value is the same as the corresponding algebraic operation of the corresponding plain-text.

2.2 Homomorphic Encryption

The study of homomorphic encryption technology can be traced back to the 1970s. In 1978, Rivest et al. proposed secret homomorphic techniques, which was the first time that the concept of “homomorphism” was proposed. The homomorphic encryption algorithm, which has a multiplicative homomorphism, is a single homomorphism, and is a form of encryption. Its biggest feature is that it can perform specific algebraic operations on the ciphertext directly, and the obtained operation result is still encrypted. The result is the same as the result of the same operation on the plain-text after decryption.

2.3 Information Homomorphic Encryption Security Storage Privacy Protection Principle in Big Data

In information technology, big data refers to large and complex data sets that are difficult to process using existing database management tools or traditional data processing applications. As the name implies, big data can also be called huge amount of data or massive data[9]. Big data is significantly different from traditional data in terms of structure type, processing speed, data volume, value density, etc., in data storage, query mode, and analysis applications. In order to solve the security and privacy of big data[10]. Redesigning and building big data security infrastructure and open data services, we need to deploy the overall security and privacy solution from all aspects, so as to ensure big data in data collection, data computing, data integration, data storage, data Analysis, data mining and then data security on the entire chain. According to the application process of big data, big data security technologies mainly include data collection security technology, data storage security technology, data mining security technology and data release security technology.

2.4 The Significance of Piano Accompaniment in Vocal Music Teaching

In order to effectively improve the confidentiality protection of user privacy information during transmission, it is necessary to design privacy encryption protection for big data information. The process of security protection transmission needs to balance the energy of each node to protect information integrity. . The effective protection of secure storage of private information in big data, by using the homomorphic function, compresses the private data of the big data user into an integrated structure, specifically by using the group encryption method to compress the information into an encryption source-optimizing the evolution structure, and then Based on the completion of

encryption derivation, the seed set is generated on the basis of this, and the binary discrete chaos optimization process is implemented by selecting the competition mechanism of the encrypted user privacy information through the seed set competition mechanism, and the coupling control game is selected, and the output after the coupled control game is selected. As a result, the selected encrypted user privacy information ciphertext needs to meet the requirements of the optimal game performance and the lowest chaotic dispersion coefficient, so as to realize the adaptive encryption protection process of the privacy information of the big data network users.

According to the orthogonal conversion principle, the data space set encrypted by the user privacy information is represented by F, the matrix corresponding to the privacy information source signal space is represented by A, the complete set space threshold is represented by Ω, and the error factor is represented by β, then the expression of F as follows:

$$F = \arg\left\{\left\| B - AX \right\| \right\} + \alpha \left\| x_i \right\| \tag{2}$$

On the basis of the corresponding coupling processing of F, in order to realize the orthogonality of the privacy information data space F and x, we need to further control the F (according to the positive price theory), when Fand B are coupled to the game control. After that, the coupling factor is defined as follows:

$$Idex(F_m, B_n) = \left[l(F_m, B_n) \right]^{\alpha} \left[k(F_m, B_n) \right]^{\beta} \left[h(F_m, B_n) \right]^{\mu} \tag{3}$$

Finally, the encrypted private information data space that will be finally obtained is realized, thereby implementing encryption protection of the secure storage household privacy information data in the big data.

3. Research Status of Homomorphic Encryption Algorithm and Big Data Security Storage Privacy Protection

3.1 Research Status of Homomorphic Encryption Mechanism

After the database encryption by the traditional method, the size and order of the data have changed greatly. If you want to retrieve specific data, you first need to download the ciphertext data from the database server to the local, then decrypt it, and then retrieve the desired result. Throughout the process, the encryption and decryption process of massive data requires high performance on the client side, and the data also needs to be transmitted back and forth between the server and the client, consuming bandwidth resources. The homomorphic encryption is based on the premise of not destroying the confidentiality of the data, and realizes the operation and processing of the ciphertext data, and can ensure that the ciphertext after the operation is still valid. The process of encrypting a database with homomorphism is equivalent to a database security upgrade that is transparent to the user. The secret homomorphic technique proposed by Rivest et al. is based on the problem of large number decomposition. From the perspective of large number decomposition, it is indeed safe. IBM researcher Gentry is proposing a new, homomorphic encryption solution that is an epoch-making effort. “Full homomorphism” means that any operation on plaintext can be implemented on ciphertext, and the ciphertext results can be correctly decrypted. However, if the noise introduced in the data encryption process is higher than the threshold, the decryption operation cannot be completed smoothly. Usually, the noise is getting bigger and bigger when processing the ciphertext, which makes the decryption impossible. Furthermore, the full homomorphism scheme is optimized, and in theory, the ciphertext can be subjected to any number of arbitrary and arbitrary forms of operations. At this time, the security parameter λ needs 72 bits, the public key size reaches 3.2 GB, and the key update time is long. Based on the research results of Gentry, the homomorphic encryption scheme on integer is proposed, which is based on modulo operation, but the computational efficiency is not high. With the improvement of the homomorphic scheme, the integer-based homomorphic encryption scheme and the security of the algorithm are ensured through difficult problems, so that the operation of encrypting 1 bit plain text once in the

original scheme becomes one encryption at a time. The comparison of different types of homomorphic encryption algorithms is shown in Table 1.

Table 1 Comparison of Homomorphic Encryption Algorithms

Homomorphic algorithm	Key	Safety	Algorithm complexity	Characteristics
RSA algorithm	Large prime function	Relying on the problem of large number decomposition	O(nlogn)	Only satisfy the multiplicative homomorphism
Homomorphic Encryption Algorithm Based on Ideal Lattice	Matrix	Dependent on discrete subset summation problem	O(n6)	Theoretically satisfying the homomorphism
Homomorphic Encryption Algorithm Based on Integer Ring	Polynomial	Dependent on discrete subset summation problem	O(n5)	Theoretically satisfying the homomorphism
Homomorphic encryption algorithm on integer	Large integer	Relying on the approximate maximum common factor problem	O(n3)	Does not satisfy the full homomorphism after exceeding the limit of the number of calculations

It can be seen from the table that the homomorphic encryption algorithm on the integer has high security and low algorithm complexity, and has a good application prospect. While conducting a deeper study of homomorphic encryption, the existing results of homomorphic encryption are also widely used in all aspects of life. However, in practical applications, homomorphic encryption still has many problems such as large computational complexity and high complexity, and most of the homomorphic encryption schemes remain at the theoretical level, and the practicality needs to be improved.

3.2 Big Data Security Privacy Issues

In the context of big data, the security needs of various industries are changing. From data collection, data integration, data storage, data analysis, data mining to data release, this process has become a new complete chain. As the amount of data increases and concentrates, the security threats of data in the entire chain are also increasing, and the security and privacy protection of data is becoming more and more difficult. The security and privacy issues of big data have become the focus of attention of enterprises. In addition, the development of big data creates conflicts between users' privacy and convenience. Consumers benefit from big data technology and will buy more products that meet their needs at a lower price, but at the same time, with personal purchasing preferences, Massive data on health and financial conditions are collected and people's privacy is destroyed. Overall, the current big data mainly includes security issues such as infrastructure security, storage security, network security, and privacy security.

4. Design and Implementation of Big Data Security Privacy Protection Scheme Based on Homomorphic Encryption Algorithm

4.1 Toverall Design of Big Data Security Privacy Protection Scheme Based on Homomorphic Encryption Algorithm

The homomorphic hash function $H(m): Z \to G$ where G is a multiplicative group, refers to a hash function with homomorphic properties, which satisfies the following properties:

(1)Homomorphism: for any two messages m_1, m_2 and positive integers w_1, w_2 ,then

$$\mathrm{H}(\mathrm{w}_1 m_1 + w_2 m_2) = H(m_1)^{w_1} H(m_2)^{w_2} \tag{4}$$

(2)Collision-free: the attacker does not exist. The probability polynomial algorithm can forge $(m_1, m_2, m_3, w_1, w_2)$ and satisfy $m_3 \neq w_1 m_1 + w_2 m_2$, so that

$$\mathrm{H}(\mathrm{m}_3) = H(m_1)^{w_1} H(m_2)^{w_2} \tag{5}$$

4.2 Establish a Big Data Security Privacy Information Protection Model

In the process of protecting user privacy information in big data, firstly, the clustering route is used to balance the energy of each node in the big data, the homomorphic hash function is used to protect the integrity of the user's private information, and the cluster key is introduced to reduce the privacy of the node user. The probability of information being deciphered, the specific process is as follows:

It is assumed that the criterion for measuring whether a node in a big data can be selected as a cluster head node is mainly composed of the current remaining energy and the number of loops of the point. Therefore, the probability i of the node i in the big data is selected as the cluster head node in the t round $pi(t)$ calculated using equation (7):

$$\mathrm{P}_i(t) = \max\left\{ p_{prob} \bullet n / x \bullet (E_{i-init}, P_{\min}) \right\} \tag{7}$$

After the key vector of the node i in the big data, the homomorphic Hash verification code and the custom user privacy information are generated, the user privacy information data packet is encrypted as a whole and transmitted to the cluster head node. The key vector and the custom user privacy information are separately added and combined to obtain the cluster key vector and the cluster user customized user privacy information in the big data, respectively

$$\mathrm{A}_{Hi-\min} = \sum_{i=1}^{Z} A_i = \left(\sum_{i=1}^{Z} a_{i-1}, \sum_{i=1}^{Z} a_{i-2}, \ldots \sum_{i=1}^{Z} a_{i-p} \right) \tag{8}$$

$$R_{Hi-\min} = \sum_{i=1}^{Z} R_i = \sum_{i=1}^{Z} d_i + \sum_{i=1}^{Z} K_{i-\min} \tag{9}$$

$$\mathrm{R}_{Hi-pro} = \prod_{i=1}^{Z} H(d_i) = \sum_{i=1}^{Z} g^{d_i} \bmod m \tag{10}$$

After the cluster head node in the big data passes the above calculation, the cluster key vector, the cluster user customized user privacy information and the cluster head verification code product are encrypted, and the multi-hop method is selected to be transmitted to the base station node. In the process of security protection of user privacy information in big data, based on the established security model of user privacy information, combined with greedy theory to quantify the loss of different types of user privacy information brought about by the encryption process of user privacy information in big data.

4.3 Experimental Results and Analysis

In order to prove the effectiveness of the proposed method for protecting user privacy information in big data based on homomorphic encryption, an experiment is needed. The experimental platform adopts Windows10, and the experimental data comes from the information data set in China's mass online shopping. The original data set contains 52,000 customer consumption information data. After deleting the null value records in the data set, 2150 pieces of data are randomly selected as experiments. data. The homomorphic encryption method and DES encryption method are used to conduct user privacy information security protection experiments in big data. Under the same key length, the user privacy information encryption time (s) and decryption time (s) are compared by two different methods. The comparison results are used to measure the computing performance of the user privacy information security protection of two different methods(see in Figure 1). It is proved that the adaptive encryption method of user privacy information designed in this paper can effectively improve the security of private information, and

the encryption and decryption speed can meet the actual needs of users, which has certain practicability.

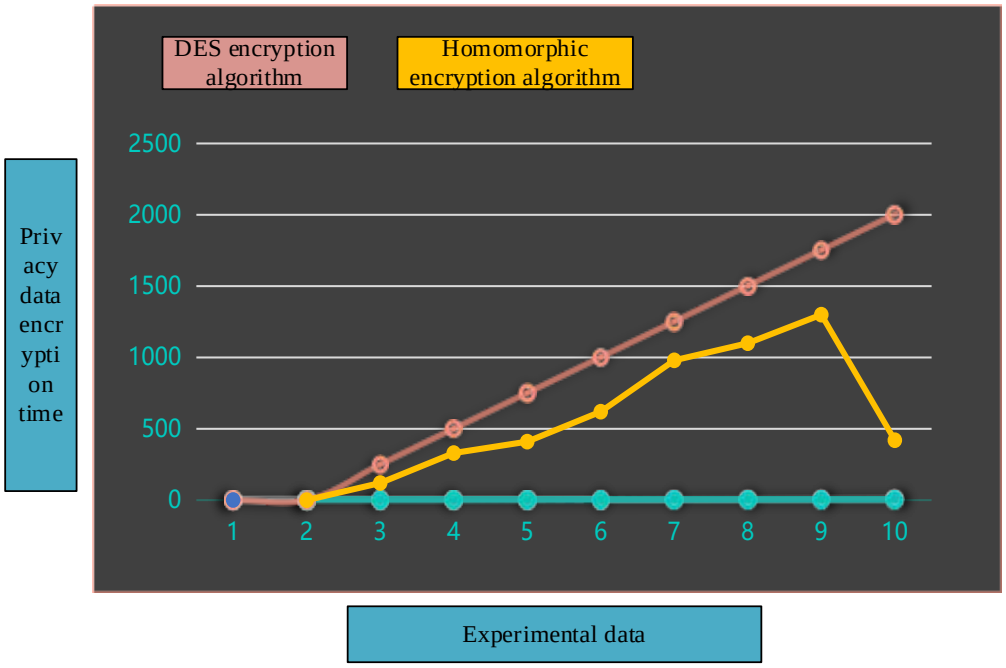

Fig.1 Ure 1 Comparison of User Privacy Encryption Time in Different Methods

Analysis of Figure 1 shows that the computational performance of user privacy information security protection in big data is better than DES encryption method by homomorphic encryption. This is mainly because the use of homomorphic encryption method for user privacy information security protection. Firstly, the homomorphic encryption function is used to protect the integrity of the user's private information, and the greedy theory is used to quantify the loss of different types of user privacy information brought by the encryption process of the user's private information in the big data, thereby completing the user's private information in the big data. Security protection makes the computing performance of user privacy information protection under big data network using homomorphic encryption method superior.

5. Conclusions

In the context of big data network environment, the existing user privacy information protection methods have poor privacy protection effect and poor security of encrypted private information. The main research is on the implementation path of private data encryption of big data network users. The different types of user privacy information generated by the process are different. The method of security protection of private information based on homomorphic encryption is proposed. Firstly, the energy of each node in the big data is balanced by using clustered routes, so as to ensure the integrity of the user's private information. The homomorphic Hash function is used, and the probability of node information being deciphered is effectively reduced by the introduction of the cluster key, thus realizing the security protection of the privacy information of the big data network users. The test results show that the method can effectively improve the security of private information, and the encryption and decryption speed can meet the actual needs of users. The existing privacy protection method has the problems of poor privacy protection for user privacy information in big data, low processing performance, and low security of encrypted user privacy information. This paper proposes a method for protecting user privacy information in big data based on homomorphic encryption. The experimental results show that the proposed method can effectively improve the security of user privacy information, and the encryption and decryption speed is faster.

References

[1] D. Zhuravlev.(2015).Towards practical private information retrieval from homomorphic encryption. 19(2):302–312.

[2] Chen Z, Song X, Zhang Y.(2015).A Fully Homomorphic Encryption Scheme Based on Binary-LWE and Analysis of Security Parameters. 47(2):75-81.

[3] Liping Zhang, Shaohui Zhu, Shanyu Tang.(2017).Privacy Protection for Telecare Medicine Information Systems Using a Chaotic Map-Based Three-Factor Authenticated Key Agreement Scheme. IEEE Journal of Biomedical & Health Informatics, 21(2):465-475.

[4] Tonghao Yang, Junquan Li, Bin Yu.(2016).A Cryptographic Access Control Scheme Providing with Fully Homomorphic Encryption. Journal of Computational & Theoretical Nanoscience, 13(4):2433-2438.

[5] Veugen P J M, Blom F, Hoogh S J A D, et al.(2015).Secure comparison protocols in the semi-honest model. 9(7):1217-1228.

[6] Khedr A, Gulak G.(2018).SecureMed: Secure Medical Computation using GPU-Accelerated Homomorphic Encryption Scheme. 22(2):597-606.

[7] Bai L Q, Li L, Qian S G, et al.(2017).Source-location privacy protection algorithm in WSNs based on ellipse model.32(2):255-261.

[8] Ashok George, A.(2019).Sumathi. Dyadic product and crow lion algorithm based coefficient generation for privacy protection on cloud. Cluster Computing, 22(4):1-12.

[9] Xinyan Li, Huajian Mou, Dianjun Lu.(2019).An Improved Ciphertext Retrieval Scheme Based on Fully Homomorphic Encryption.Wuhan University Journal of Natural Sciences, 24(3):218-222.

[10] Peng Zhang, Tony Thomas, Tao Zhuo, et al.(2017).Object coding based video authentication for privacy protection in immersive communication. 8(6):871-884.