Safe Storage Method of Energy Big Data Center Based on Blockchain Technology

Jiangbo Sha[1a*], Jia Liu[1,b], Jianhui Cai[1,c], Dongge Zhu[1,d], Rui Ma[1,e]

[*] E-mail: 20102844@163.com[a]; fanjian123391@163.com[b]; dusijiaof@163.com[c]; xinghui06348997495@163.com[d]; 33697525@qq.com[e]

Electric Power Research Institute of State Grid Ningxia Electric Power Co.,Ltd, Yinchuan, China[1]

Abstract—In order to solve the problems of long data storage and weak data storage integrity in traditional methods, this paper proposes a secure storage method for energy big data centers based on blockchain technology. First, the big data of energy is statistically analyzed. Then, based on the security threats faced by the energy big data center in the storage process, assumptions were made, and a secure storage model of energy big data was constructed. Finally, combined with the process analysis of blockchain technology and the role of the system, a secure storage solution for the energy big data center is designed to realize the secure storage of the energy big data center. The experimental results show that the data storage efficiency of the design method in this paper is higher, and the data integrity coefficient is higher, indicating that the data storage performance of this method is better.

Keywords-blockchain technology; energy big data; secure storage; statistical analysis; functional requirements

1. INTRODUCTION

The energy industry is related to the prosperity and development of the country and social stability. If energy big data is illegally exported or tampered with, national security will be seriously threatened [1]. Therefore, focusing on the security of energy big data centers and analyzing the current status and existing problems of energy data management will help prevent the leakage and tampering of energy and power data. At this stage, relevant scholars and experts have conducted certain research on this issue and have obtained many research results [2].

Li et al. proposed an intelligent storage algorithm for distributed big data under cloud computing. By analyzing the distributed big data sequence, the normalized RGB histogram is obtained, the absolute difference of the big data histogram is calculated, and the change of the data sequence is mapped, so as to classify the distributed big data. Then we used the k-means algorithm to select the cluster centers, and determined the number of clusters and the standard

ICEMME 2022, November 18-20, Nanjing, People's Republic of China

DOI 10.4108/eai.18-11-2022.2327144

evaluation threshold. On this basis, we clustered the data, and the corresponding data in all clusters is supplemented by long and short buffers. Then, it is stored in the tag in the form of data stream to form a complete distributed big data file, so as to achieve the purpose of intelligent storage of distributed big data. The simulation results show that this method is not easy to be disturbed by the external abnormal, and effectively enhances the storage effect of big data, so as to improve the intelligent storage performance of distributed big data, but there is the problem of poor data integrity[3]. Xia et al. proposed an improved medical record map storage scheme, according to the original medical record data has the characteristics of multiple relations, designed the transformation scheme from multiple relations to RDF triples. On this basis, entity type table is designed, and SPARQL to SQL query conversion algorithm is used for medical record atlas. The experimental results show that the scheme has high query efficiency, but the integrity of data storage is not high [4]. Huang et al. proposed an optimization method of redundant optical fiber data storage based on traditional genetic algorithm and data compression algorithm is proposed. Combining with Dopplerlet transform, the global optimization of the best basis function is found, and the redundant characteristics of optical fiber data are analyzed and filtered. Based on this result, the traditional genetic algorithm is used to compress the redundant optical fiber data. On this basis, the K-L feature is used to reduce the load of optical fiber data storage, and the compression optimization of redundant optical fiber data is completed to realize the optimal storage of redundant optical fiber data. The experimental results show that this method can effectively compress the redundant off fiber data, but there is the problem of long time-consuming data storage[5]. Aiming at the problems of long time consuming and incomplete data storage in traditional methods, this paper proposes a secure storage method of energy big data center based on blockchain technology.

2. SECURE STORAGE METHOD OF ENERGY BIG DATA CENTER BASED ON BLOCKCHAIN TECHNOLOGY

2.1 Statistical analysis of energy big data

Due to the large number and types of energy big data, in order to improve data storage efficiency, we first conduct statistical analysis on energy big data [6]. Suppose the original sequence of the energy big data center is $S=(s_1,s_2,...,s_n)$, use D to denote the sequence operator of the energy big data center, then there is $SD=(s_1d,s_2d,...,s_nd)$, then the following expression:

$$s_i(n)d=\frac{d(k)}{s_i}\times\varepsilon_i \tag{1}$$

Among them, $i=1,2,...,n$, s_i represent the mean sequence of the energy big data center. Assuming that the sequence length is the same, and the initial value is not zero, the expression of the comprehensive correlation degree of each sequence in the energy big data center is:

$$h(s_k)=\phi(s_{k+1},s_k)B(\tau) \tag{2}$$

Among them, $B(\tau)$ represents the relative degree of relevance; s_k represents the data factor.

Use the factor analysis method to analyze the latent variables of the energy big data center, and the results are expressed in the form of a matrix:

$$S_k^{'} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & s_{22} & \cdots & s_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ s_{n1} & s_{n2} & \cdots & s_{nm} \end{bmatrix} \tag{3}$$

Among them, m represents the original variable of the energy big data center. Combined with the above analysis, the statistical analysis of the data of the energy big data center is realized.

2.2 Modeling of energy big data storage security issues

According to the security threats faced by the energy big data center during the storage process, the following assumptions are made:

(1) The source of the attack is unlimited, and each attacker arrives individually and independently of each other;

(2) The arrival number of attackers conforms to the Poisson distribution with parameter μ, where μ is the average number of attacks per unit time;

(3) The tampering of dynamic data information caused by each attack obeys the negative exponential distribution with parameter θ;

(4) The time interval between the arrival of each attack and the damage caused are independent of each other [7].

Suppose the number of terminals in the example system is N, Y is the set of dynamic data information codes in the system, $Y = (y_1, y_2, ..., y_n)$, $y_i = [y_i, code, y_i, state]$, among them, $y_i, code$ includes the corresponding dynamic data information code, $y_i, state$ represents the status of the data file, $y_i, state \in [1, -1]$, among them, $i \in N$, the three values correspond to the three situations of tampering, no change, and forgery of the data file; Suppose P is the condition for the safe operation of the system, and $P_s(T_1, T_2)$ describes the extent of the system being attacked by various attacks within the time range of (T_1, T_2) under the safe operating condition P; $P_s(T_1, T_2, j)$ is the probability that the data file will be tampered or forged r times within the time range of (T_1, T_2) under the safe operation condition P; RP_s

is the risk factor of the system running under the security algorithm, or the robustness of the system; Δt is the probability that the system has been attacked by multiple attackers at time t [8].

It can be known from the hypothesis that when Δt is small enough, the probability of an attacker arriving in the $[t, \Delta t]$ time interval is $\Delta t \times \sigma$. Therefore, at time $\Delta t + 1$, the probability of the system being attacked by multiple attackers is $P_a(t + \Delta t)$:

$$P_a(t + \Delta t) = \frac{1}{\left[m_i(t+1) \right] \Delta t} \tag{4}$$

Among them, m_i represents the correlation coefficient between attackers.

Consider the special situation, that is, when $m_i = 0$, the possibility of the system being attacked in time interval $[t, \Delta t]$ is the following three mutually independent situations:

(1) At time t, the system is not attacked, and there is no new attack in $[t, \Delta t]$, the probability is $(1 - \varpi \Delta t)$;

(2) At time t, the system is not attacked, and a new attack occurs in $[t, \Delta t]$, the probability is $(1 - \varpi \Delta t) P_a(t + \Delta t)$;

(3) At time t, the system is attacked, and there is no new attack in $[t, \Delta t]$ the probability is $(1 - \varpi \Delta t) P_a(\mu)$.

From this, the security problem of energy big data storage can be modeled to obtain:

$$\frac{\partial x}{\partial X_1} = -\frac{z^2}{Bf_1} \frac{(\omega_2 + \alpha_2)}{(\omega_1 + \alpha_1)} \omega_1 \tag{5}$$

$$\frac{\partial x}{\partial X_2} = -\frac{z^2}{Bf_2} \frac{(\omega_1 + \alpha_1)}{(\omega_2 + \alpha_2)} K \omega_2 \tag{6}$$

Formula (6) is a set of probability equations that the system has been attacked by multiple attackers at time t. Formula (7) is a constraint function of the degree of system infringement; Formula (6) is the constraint function of the system, and at least one data file is required.

Formula (7) is the probability of the data file being destroyed K times in time $\left(T_1, T_2\right)$. The formula (7) is used to measure the average situation of the system under attack in a period of time [9]. The smaller the K value, the higher the robustness of the system, which means that the safe storage effect of the energy big data center is better.

2.3 Safe storage method of energy big data center based on blockchain

Based on the modeling results of energy big data storage security issues, combined with the analysis of the blockchain technology process and system roles, design an energy big data center security storage solution. The main requirements include the functional requirements of the blockchain and the energy big data scenario. The main functional requirements of the energy big data security storage solution are shown in Figure 1.

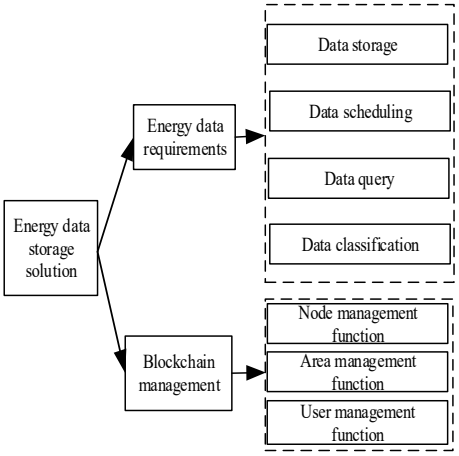

Figure 1 Security storage requirements for energy big data

Due to the rapid growth of energy business data, centralized data centers cannot meet the security and scalability required for energy data storage. In order to solve these problems, this article proposes a blockchain-based method to store power data. Analyzing Figure 1 shows that the schematic diagram of the secure storage process of energy big data is shown in Figure 2.

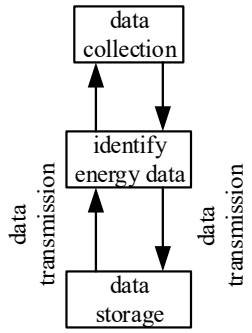

Figure 2 Schematic diagram of the secure storage process of energy big data

In the process of secure storage of energy big data, first is the collection of energy data, and then all the collected data is sent to the corresponding cloud server to identify the energy data [10]. In the last stage, after the energy data is accurately identified in the corresponding cloud, it is transmitted to each node in the blockchain through the transmission channel. The blockchain network composed of these nodes can ensure the authenticity of energy data and realize the safe storage and scheduling of energy data [11]. The decentralized, tamper resistant and traceable features of blockchain technology are just conducive to the safe storage of data, and each node in the blockchain spontaneously maintains the data in the node through consensus mechanism and consensus algorithm, and the data and version information of each node are highly consistent.

3. EXPERIMENT AND RESULT ANALYSIS

In order to verify the effectiveness of the secure storage method of energy big data center based on blockchain technology, simulation experiments are carried out. In the experiment, the distributed big data intelligent storage algorithm based on cloud computing and the redundant optical fiber data storage optimization method based on traditional genetic algorithm and data compression algorithm are selected as the comparison method of this method to compare the data storage performance.

3.1 Experimental environment

The programming software in this article mainly uses Pycharm, uses Linux as the platform, simulates the construction of a distributed environment through Docker containers, uses Python as the programming language, and stores data through the Redis database. In the above experimental environment, the different methods are compared, and the results are analyzed as follows. Traditional method 1（Distributed big data intelligent storage algorithm under cloud computing）. Traditional method 2（Data storage method based on traditional genetic and data compression algorithm）.

3.2 Analysis of experimental results

(1) Data integrity verification experiment

In the experiment of data integrity verification, multiple nodes are selected to verify the integrity of multiple data, and the results of integrity verification are statistically analyzed. Figure 2 shows the results of integrity verification of randomly selected data numbers by different methods. The integrity is represented by the integrity coefficient, which is 0.1-1. The higher the value is, the higher the data integrity is. The specific results are shown in Figure 3.

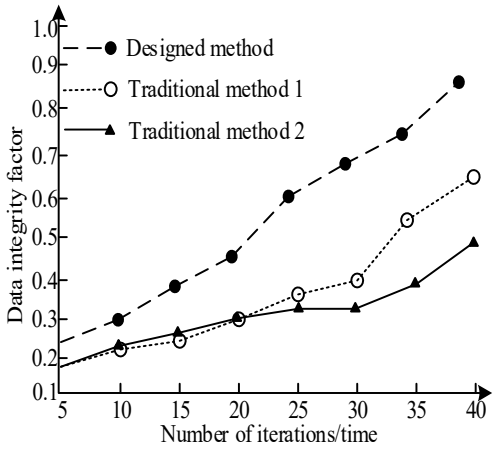

Figure 3 Data integrity comparison results

It can be seen from the analysis of Figure 3 that with the increase of iterations, the data storage integrity coefficients of different methods show a continuous growth trend. Compared with the design method in this paper, the gap is still obvious, and the data integrity coefficient of the design method is close to 0.9. The experimental results show that the integrity of the design method is high, which indicates that the data security storage range of the energy center is large, and it is not easy to lose data.

(2) Data storage efficiency verification experiment

Taking data storage efficiency as the experimental index, a comparative experiment was carried out, and the result is shown in Figure 4.

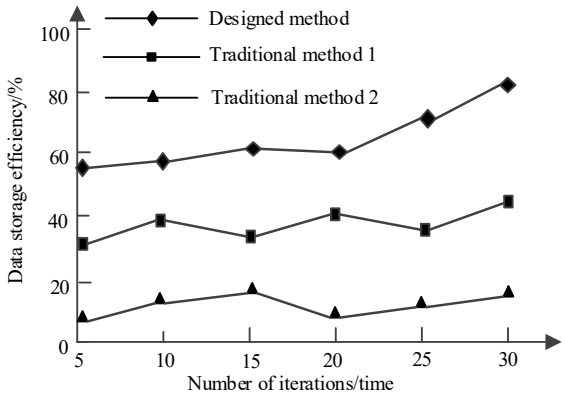

Figure 4 Comparison results of data storage efficiency

According to the analysis of Figure 4, under different iterations, the energy data storage efficiency of this design method is significantly higher than the distributed big data intelligent storage algorithm under cloud computing and the redundant optical fiber data storage optimization method based on traditional genetic algorithm and data compression algorithm. It shows that the design method in this paper can realize data storage in time when facing large-scale energy data.

4. CONCLUSION

In order to solve the problems of long data storage time and poor data storage integrity existing in traditional methods, a secure storage method for energy big data centers based on blockchain technology is proposed. The experimental results show that the data storage efficiency of the design method in this paper is high, and the data integrity coefficient is high, and the highest value of the data integrity coefficient is close to 0.9, indicating that the data storage performance of this method is better.

REFERENCES

[1] Zhou T, Tian C. Fast Erasure Coding for Data Storage: A Comprehensive Study of the Acceleration Techniques[J]. ACM Transactions on Storage, 2020, 16(1):1-24.

[2] Jagdish M, Viloria A, Vargas J, et al. Modeling software architecture design on data storage security in cloud computing environments[J]. Journal of Intelligent and Fuzzy Systems, 2020, 39(6):1-8.

[3] Li S, Huang C. Simulation of Distributed Big Data Intelligent Storage Algorithm under Cloud Computing[J]. Computer Simulation, 2020, 37(05):448-452.

[4] Xia Y H, Gao D Q, Ruan T, et al. Research on Data Storage of Medical Record Based on Knowledge Graph[J]. Computer Engineering, 2019, 45(01):9-16,22.

[5] Huang Z P, Wang L, Zhang S X, et al. Redundant optical fiber data storage optimization based on traditional genetic algorithm and data compression algorithm[J]. Laser Journal, 2019, 40(03):135-139.
[6] Sinha K, Priya A, Paul P. K-RSA: Secure data storage technique for multimedia in cloud data server[J]. Journal of Intelligent and Fuzzy Systems, 2020, 39(2):1-18.
[7] Düben P D, Leutbecher M, Bauer P. New Methods for Data Storage of Model Output from Ensemble Simulations[J]. Monthly Weather Review, 2019, 147(2):677-689.
[8] Kishani M, Tahoori M, Asadi H. Dependability Analysis of Data Storage Systems in Presence of Soft Errors[J]. IEEE Transactions on Reliability, 2019, 68(1):201-215.
[9] Marsh M, Chaput T, Smith D. Unified electronic traceability and data storage system[J]. Cytotherapy, 2019, 21(5):43.
[10] Xu G, Han S, Bai Y, et al. Data Tag Replacement Algorithm for Data Integrity Verification in Cloud Storage[J]. Computers & Security, 2021, 103(3):102205.
[11] Lu L H, Du C L. Multi-Layer Classification Storage Simulation of Complex Network User Privacy Data[J]. Computer Simulation, 2020, 37(3):410-413+444.