Data Distributed Storage Scheme in Internet of Things Based on Blockchain

Yin Zhang and Jun Ye(✉)

School of Computer Science and Cyberspace Security, Hainan University,
Haikou, China
yejun@hainanu.edu.cn

Abstract. With the explosive growth of Internet of Things data, the storage of large amounts of data has become a problem that plagues people. These IoT devices generally do not have powerful computing and storage capabilities. Most of them assume the role of data collection and information transfer, which also brings challenges to data storage. And a large proportion of the large amount of data is the user's privacy data, so under the premise of storing a large amount of data, we must also ensure the security of the data. In view of the above problems, this article proposes an effective IoT data storage model based on secret sharing, which effectively guarantees the security of IoT data, and proposes a secure distributed storage of IoT data based on blockchain. This solution has the advantages of a complete blockchain. Under the supervision of the edge computing nodes of the entire network, it can ensure that each transaction data is not tampered with, and it can also obtain complete original data.

Keywords: Internet of Things · Secret sharing · Blockchain · Edge computing · Distributed storage

1 Introduction

The Internet of Things, the Internet of Everything, is to allow all human-related devices to access the Internet, so as to classify information, classify sharing, and information aggregation, so that people-centric devices have more functions, give them life, and create people-centric Ecosphere. According to Gartner's prediction, the number of IoT device will grow to around 30 billion by 2020, and will more than double by 2025 [1]. The explosive growth of Internet of Things data needs to be stored in a secure manner, which will undoubtedly be a huge challenge in the future.

Since its establishment, blockchain technology has broad application prospects [2]. This article combines blockchain with IoT to provide privacy protection for IoT data to achieve secure storage of data. Blockchain has many advantages to satisfy the above functions:

- Distributed layout: It replacing the previous centralized service center, transactions in this blockchain network are initiated, confirmed and maintained by full blockchain network nodes, without worrying about the problem of low fault tolerance brought by the centralized server;

J. MacIntyre et al. (Eds.): SPIoT 2020, AISC 1282, pp. 670–675, 2021.
https://doi.org/10.1007/978-3-030-62743-0_96

- Unique block structure: The newly generated blockchain will store the hash value of the previous block. This hash value is generated by the transaction data stored in the block. Once the data in a block is changed, it will discovered by nodes across the network, so the blockchain has a unique block structure and has the function of preventing malicious tampering;
- Proof of work: The process for generating new blocks is implemented in accordance with the mechanism of proof of work. The miner node gain the right to write new blocks into the blockchain only when a specific required hash value is generated and the proof of work is completed.

The initial design of IoT devices is to collect data and transfer information, but at the same time is limited by power consumption, it is impossible to have strong computing power. Edge computing is a good solution. The edge device can be any computing resource between the data source and the cloud. They have powerful computing capabilities and at the same time shorten the distance from the Internet of Things devices and respond quickly. The edge computing node can take over the data storage request from the IoT device, enter the blockchain network for the IoT device, and complete a series of work such as transaction propagation/transaction verification/workload proof/block writing, etc.

There are many problems with the traditional centralized data storage method. When a database is compromised, it means that all users' information is leaked. Distributed cloud storage servers can solve this problem well. This article will combine distributed storage and secret sharing plans. The core idea of both is decentralization, which is where they fit best. The server of the distributed storage cluster is a participant in the secret sharing plan. The feature of secret sharing is that even if any participant gets a sub-secret, he cannot guess the complete secret, this form guarantees the security of the data. The distributed storage also provides a guarantee for the secret sharing scheme. Even if some servers (as long as not all) are attacked and some sub-secrets are lost, it will not affect the recovery of the final secret.

2 Related Work

At present, there are some researches on data security storage of Internet of things. This article analyzes these schemes.

Based on the existing research, Mo [3] proposed a distributed file system architecture. This solution is very effective in optimizing task scheduling, but it does not consider the issues of data access control in the Internet of Things, which is a job we have to do. For the integrity of data, Tian et al. [4] presents a tailor-made public auditing scheme for data storage in fog-to-cloud based IoT scenarios. This solution can meet the performance and security requirements, but there are still some significant problems for ensuring data integrity and usability in the fog-to-cloud scenario. In order to store and protect a large amount of Internet of Things data, Li et al. [5] introduced the concept of distributed data storage, combined with blockchain to achieve data protection, but the process of blockchain transactions was not carefully described. Fu et al. [6] studied data processing and secure data storage. Based on fog computing and cloud computing, they designed a secure, flexible, and efficient data storage and retrieval system, but only supported two data retrieval methods. In view of the new security challenges faced by the Internet of Things between the two parties, Wang et al. [7] proposed a method that management secure cloud-assisted Internet of Things data that uses the cloud to collect, store, and access data, but the solution uses the encryption method is too complicated. Xiong et al. [8] constructed a new storage model based on CP-ABE for data storage and secure access in IoT applications, but the solution is too complicated. Yang et al. [9] proposed a medical big data storage system with adaptive access control based on the intelligent Internet of Things, which can save storage space under the premise of ensuring data security and access control. Xia et al. [10] proposed a secure, trust-oriented edge storage model that can ensure data security, but this scheme increases the energy consumption of IoT devices.

3 IoT Data Storage Model Based on Secret Sharing

In this article, Shamir's secret sharing scheme is used in the IoT data storage model based on secret sharing [11]. The process of constructing (t, n) Shamir secret sharing scheme will be introduced below.

3.1 Initialization

Firstly, selecting n elements $x_i(i = 1, 2, \ldots, n)$ in the finite field q, and distribute x_i to n different participants $Z_i(i = 1, 2, \ldots, n)$, the value of x_i is open, not secret.

3.2 Sub-secret Generation

Constructing n polynomials of degree $t - 1$, the following is the general formula

$$f(x_i) = s + a_1 x_i + a_2 x_i^2 + \cdots + a_{t-1} x_i^{t-1},\ a_i \in q,\ i = 1, 2, \ldots, t - 1 \tag{1}$$

The sub-secret $f(x_n)$ obtained by the nth participant is constructed as

$$f(x_n) = s + a_1 x_n + a_2 x_n^2 + \cdots + a_{t-1} x_n^{t-1} \tag{2}$$

Among them, s is the secret that we will share, P is a large prime number, and $s < P$, the n unequal x obtained in the previous step of initialization are taken into $f(x)$ to get n groups $(x_i, f(x_i))$, Z_i will be assigned to n different participants, and open P, destroy the polynomial, each participant is responsible for keeping his own $(x_i, f(x_i))$.

3.3 Secret Recovery

Let $Z_1, Z_2, \ldots, Z_t$ among the participants participate in the secret recovery, collect the sub-secrets $f(x_i)$ of these participants involved in the secret recovery, and get

$(x_1,f(x_1)), (x_2,f(x_2)), \ldots, (x_t,f(x_t))$ such t points. Finally, the Lagrange interpolation method can be used to restore the polynomial

$$f(x) = \sum\nolimits_{i=1}^{t} f(x_i) \prod\nolimits_{j=1, j \neq i}^{t} \frac{x - x_j}{x_i - x_j} \tag{3}$$

Then restore the secret s

$$s = f(0) = \sum\nolimits_{i=1}^{t} f(x_i) \prod\nolimits_{j=1, j \neq i}^{t} \frac{-x_j}{x_i - x_j} mod(P) \tag{4}$$

When $x = 0, f(0) = s$, we can recover s. It is worth noting that when there are less than t participants in the recovery of the secret, the polynomial cannot be recovered, but only a single sub-secret cannot infer s. So any information about the secret is not available.

4 IoT Data Storage Scheme Based on Blockchain

Combining the blockchain with the Internet of Things to provide privacy protection of the Internet of Things data, so as to realize the safe storage of data. However, there will be some changes in the operation mode of the blockchain. This article simplifies the blockchain transaction process to ensure security while reducing resource consumption. The following will introduce the blockchain-based IoT data storage solution.

4.1 Transaction Generation and Broadcasting

Taking the fast-growing body area network as an example, Internet of Things devices, such as smart watch W, are ready to store a certain amount of data in distributed cloud storage when they are idle to relieve local storage pressure. Firstly, connecting the paired mobile phone M (here is a Bluetooth connection, in other IoT types may also be other connection methods), after M receives the original data S, perform a simple string segmentation of S to generate $s_1^{'}, s_2^{'}, \ldots, s_i^{'}$, $i = 1, 2, \ldots, m$, the size of m depends on the data size, and then let $s_i^{'}$ decimalized to get $s_i, i = 1, 2, \ldots, m$.

After making the above preparations, M will send data storage requests to the distributed cloud storage server separately. The cloud server participating in the storage is equivalent to the participant $Z_i(i = 1, 2, \ldots, n)$, after the storage is successful, Z_i returns the unique address *Addr* stored in the data block, M is ready to generate a new transaction to broadcast to the blockchain network, M writes the address *Addr* into this new transaction, stores the data and generates.

The input parameter *ID* represents the unique identifier of W, where the data comes from, and *Addr* where the data goes. The *version* specifies the version rule referenced by the transaction, and *lockTime* represents the transaction lock time. There is also an important step, M will hash the transaction data of *Tran* to get H_1, and then use $W's$ private key SK_W to encrypt H_1, get an encrypted document D, D is added to the transaction as a digital signature of W, the resulting $Tran_1$ will be broadcast by M. The above process is visible in process 1.

4.2 Transaction Verification and Writing

In the previous step, after M broadcasts the new transaction $Tran_1$ to the full blockchain network, the nodes in the network will verify $Tran_1$, and after successful verification, it will be written to a candidate block B. Then, the node that wins the proof of work will connect the candidate block to the longest chain in the blockchain. The following steps need to be performed during the transaction verification process: (1) Create the Hash value H_2 of the transaction $Tran_1^{\prime}s$ data to be verified, except for the digital signature itself; (2) Use the public key of the account agreed to store PK_W to decrypt the encrypted document D (digital signature) in the transaction $Tran_1^{\prime}s$ data, and get H_1; (3) Contrast H_1 and H_2, if the two values are the same, the transaction is authorized by the owner of the private key SK_W corresponding to the account W that agreed to store, the verification is successful, otherwise it is not, the verification fails. The verification and writing of the transaction are shown in process 1.

```
Process 1 Transaction Verification And Writing
Input $Tran_1$,$PK_W$,$D$,$H_1$ Output null
start process Hash($Tran_1$) -> Obtain $H_2$
  Decrypt($PK_W$,$D$) -> Obtain $H_1$
  Judgment($H_1$,$H_2$)
  if $H_1 = H_2$
    Write($Tran_1$)
  else Abort
  return -> end process
```

4.3 Data Acquisition and Recovery

Regarding data acquisition, it is necessary to add that an access control list needs to be added to the distributed cloud storage server to limit external access to data. Due to space limitations, this article will not make too many explanations. Then there is data recovery. According to the secret recovery process in Sect. 3, the block data of the Internet of Things can be recovered, that is, the secret $s_i^{\prime}$, and then M is spliced to form the final data S. Data acquisition and recovery are as in the process 2 shows.

```
Process 2 Data Acquisition And Recovery
Input $ID$ Output $S$
start process Request($ID$) -> Obtain $Addr$
  Request($ID$,$Addr$)
  if Allow access -> Obtain $s_i$
    Restore($s_i$) -> Obtain $S$
  else Abort
  return $S$ -> end process
```

5 Conclusion

Aiming at the problem of massive data security storage caused by the surge of IoT data, this article proposes a new distributed storage solution for IoT data based on secret sharing and blockchain, which has theoretical reliability, efficiency and feasibility. And data access is under regulatory control, and only authorized users can obtain target data.

Acknowledgments. This work was partially supported by the Science Project of Hainan University (KYQD(ZR)20021).

References

1. Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025 (2016). https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/
2. Li, X., Jiang, P., Chen, T., et al.: A survey on the security of blockchain systems. Future Gener. Comput. Syst. **107**, 841–853 (2020)
3. Mo, Y.: A data security storage method for IoT under Hadoop cloud computing platform. Int. J. Wirel. Inf. Netw. **26**(3), 152–157 (2019)
4. Tian, H., Nan, F., Chang, C.C., et al.: Privacy-preserving public auditing for secure data storage in fog-to-cloud computing. J. Netw. Comput. Appl. **127**, 59–69 (2019)
5. Li, R., Song, T., Mei, B., et al.: Blockchain for large-scale Internet of Things data storage and protection. IEEE Trans. Serv. Comput. **12**(5), 762–771 (2018)
6. Fu, J.S., Liu, Y., Chao, H.C., et al.: Secure data storage and searching for industrial IoT by integrating fog computing and cloud computing. IEEE Trans. Industr. Inform. **14**(10), 4519–4528 (2018)
7. Wang, W., Xu, P., Yang, L.T.: Secure data collection, storage and access in cloud-assisted IoT. IEEE Cloud Comput. **5**(4), 77–88 (2018)
8. Xiong, S., Ni, Q., Wang, L., et al.: SEM-ACSIT: secure and efficient multiauthority access control for IoT cloud storage. IEEE Internet Things J. **7**(4), 2914–2927 (2020)
9. Yang, Y., Zheng, X., Guo, W., et al.: Privacy-preserving smart IoT-based healthcare big data storage and self-adaptive access control system. Inf. Sci. **479**, 567–592 (2019)
10. Xia, J., Cheng, G., Gu, S., et al.: Secure and trust-oriented edge storage for Internet of Things. IEEE Internet Things J. **7**(5), 4049–4060 (2019)
11. Yannuzzi, M., Milito, R., Serral-Gracià, R., et al.: Key ingredients in an IoT recipe: Fog Computing, Cloud computing, and more Fog Computing. In: 2014 IEEE 19th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD), pp. 325–329. IEEE (2014)
12. Shamir, A.: How to share a secret. Commun. ACM **22**(11), 612–613 (1979)