Application of Blockchain in Student Information Management System

Xuechun Li, Jinsong Xu(✉), Huiling Le, Yingying Ding, Yiwen Li, and Xingning Chen

Jiangsu Normal University, Xuzhou, China
casxjs2003@jsnu.edu.cn

Abstract. Blockchain technology has the characteristics of decentralization, traceability, distributed storage, and non-tampering. It effectively solves the data security problem faced by the massive data of the information system, in order to effectively guarantee the information security of student records, avoid tampering and theft of information, guarantee information security and credibility of electronic file data, the blockchain and the equal Practical Byzantine Fault Tolerance (EPBFT) algorithm are combined. Design a student information management system to improve the consensus efficiency and solve the problem that the blockchain consensus algorithm consumes more communication overhead. And the interplanetary file system (IPFS) and alliance chain are used as data storage backup protection methods to ensure data security.

Keywords: EPBFT · consensus algorithm · IPFS · federated chain

1 Introduction

College student files are one of the main components of national personnel files, and are the main information on students' personal development. College student files include basic information, self-description, education, career goals, courses taken and work experience, etc., and are used by employers to assess and recruit important reference. Because the content of paper archives is not easy to transfer, refurbish, and retrieve, although electronic archives are popularized in most colleges and universities, there are still some problems in the management of university student archives, such as data islands, poor traceability, easy data tampering.

Security is difficult to guarantee. Therefore, it is imminent to build a safe and reliable file management system. Since the blockchain has the characteristics of decentralization, tamper-proof, distributed storage, and traceability, it has obvious advantages in terms of network-wide records, security, and traceability. Therefore, this paper proposes to combine blockchain technology and file management system to further enhance the information security of student files, avoid information leakage, and provide an effective solution to problems related to student information management systems.

D. Kumar et al. (Eds.): IEIT 2023, AHSSEH 10, pp. 748–757, 2023.
https://doi.org/10.2991/978-94-6463-230-9_89

2 Related Theories

Using distributed ledger technology and advanced encryption algorithms, the blockchain can achieve decentralization, thereby establishing a more efficient and reliable trust mechanism to ensure the security of the database [1]. The main key technologies of blockchain cryptography, including hash algorithms and asymmetric encryption. Hash algorithm, also known as hash algorithm, is used to check whether transaction data has been tampered with. The basic principle is to input information of random length, and then use hash algorithm to convert it into a fixed-length output. The basic principle of this mapping is the corresponding hash algorithm, and the binary string reflected in the original data is the hash value. The hash algorithm is a one-way algorithm, that is, it is difficult to reversely deduce the original information from the hash value, which ensures data security. Another algorithm is asymmetric encryption [2]. The receiving parties of asymmetric encryption have corresponding public and private keys respectively. The private key encrypts and the public key decrypts. After the information transmitter sends the information, the receiver can obtain the paired private key through the public key and decrypt it to obtain the information. Asymmetric encryption reduces the possibility of private keys being stolen and better protects transaction data.

3 PBFT

The PBFT algorithm's core function is to trade communications for credit as the index reduces to the polynomial level, which is more suited for actual systems. Its drawback is that due to its limited scalability and dependence on the number of participating nodes, it is unsuitable for blockchain systems with a large number of nodes. Since the system nodes are largely fixed, they are primarily suitable for environments including alliance chains or private chains.

The Byzantine problem can be addressed when the total number of nodes surpasses $3f + 1$, as required by the PBFT algorithm [3], where f denotes the number of error nodes. Nodes that experience system failure shouldn't make up more than one-third of the overall network's nodes. The fault tolerance rate is relatively low, and transaction information cannot be stored and recorded well.

Practical Byzantine Fault Tolerance Algorithm divides nodes into three categories: master nodes, check nodes and clients [4].

The master node receives a request from the client, and upon receipt, sends the request as a broadcast to the check node. The check node receives the request, checks the data information to see if it is accurate, and if it is, broadcasts once more and returns the answer.

There are five phases in PBFT: request, pre-prepare, prepare, commit, and reply. The consensus stages and the key stages of PBFT are the pre-prepare, prepare, and commit stages, among others.

Following is a general outline of the PBFT algorithm's process:

i. request, the request stage. Master node C receives request 0 from the client.

ii. pre-prepare, pre-preparation stage. The master node will give the request it receives from the client a sequence number and process the request, and then broadcast pre-prepare information to all nodes in the entire network.

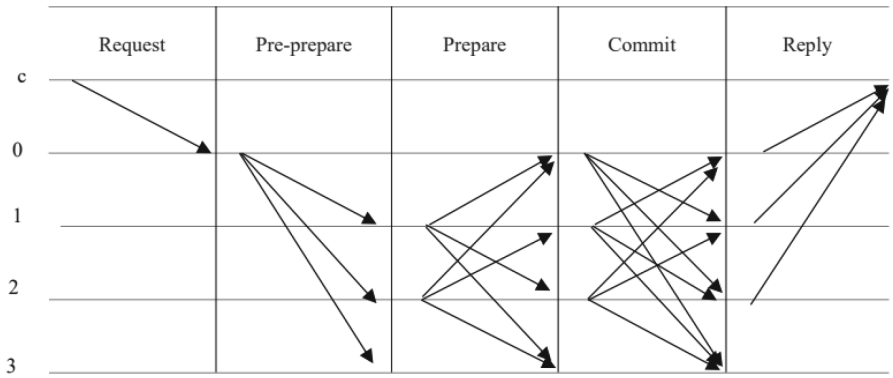

Fig. 1. PBFT consensus algorithm

iii. prepare, preparation stage. After the verification node a receives the information in the pre-preparation stage, it will verify it, and after verifying that the information is correct, it will enter the preparation stage. The first two procedures are recorded in the message log by check node A, which also sends a prepare message to every other check node in the network.

iv. Serial number confirmation (commit). A node sends an acknowledgment message to all nodes during this phase. The acknowledgment phase is complete when the node has received 2f acknowledgments identical to the pre-prepare message.

v. Response (reply). The client counts the response messages from different service nodes, and confirms that the request is successfully executed when it receives the same response from f + 1 different nodes.

A flowchart of the useful Byzantine method is shown in Fig. 1 [5].

4 Improved EPBFT Algorithm

4.1 The Development of EPBFT

The traditional POW requires massive calculations, consumes power, and has low network performance. The consensus process of the traditional PBFT algorithm includes pre-prepare, prepare, and commit stages, and multiple broadcasts are required. Every time the number of broadcasts is increased, the network bandwidth will be consumed. Since this study uses the alliance chain, the operating environment is relatively stable, and each node has the same status, which is equivalent to equalizing each node, and each node has the role of a master node. This approach eliminates the need to elect the main node to reduce the process of view changes and improve consensus efficiency. EPBFT [6] is a consensus protocol based on PBFT. It offers the same security and durability as PBFT because it inherits all of its advantages. In addition, EPBFT is appropriate for the Internet because it operates on the premise of weak synchronization. The communication complexity has been lowered from the quadratic level to the main level, and the fixed stage has been eliminated to shorten the delay as a result of the aforementioned issues. Better scalability is provided by EPBFT, which enables nodes to join or leave

the network as needed. In addition, in order to improve the reliability of the system, EPBFT introduces a reputation evaluation mechanism and a node error recovery protocol. In comparison to PBFT, EPBFT offers superior communication performance, lower latency, and is more dependable and useful.

4.2 Workflow of EPBFT

4.2.1 Workflow of EPFT Nodes

Nodes are divided into two types, basic nodes and verification nodes. The basic node is used to accept the broadcast message in the blockchain system, and the verification node is used to verify the information and verify the legitimacy of the information [7].

i. set up the system settings and backup and verify the data.

ii. Back up data to verification nodes and other nodes, and eliminate nodes that interfere with the system.

iii. To synchronize the data, node a to be synchronized sends a synchronization request to node z with the longest chain. Node a requests synchronization from node z. Node z will send a synchronization message to node a if the request's view number matches that of node z's request. If it is not consistent, it must send other nodes a synchronization request along with the relevant backup data block based on the data blocks those nodes own.

iv. The backup data block transmitted by the z node is received by the node, which is followed by a validity check. The verification pass message will be distributed throughout the network if it is legal.

4.2.2 EPBFT Consensus Process

In order to reduce communication overhead and delay, the consensus stage [8] is designed as follows:

i: The nodes of the entire network undertake data backup and verification after system activation is complete.

ii: Data from node z must be backed up by every node throughout the whole network chain, and any node requesting backup data must transmit request information to node z in a specific way.

iii: Once the backup data has been received and the verification has been completed, the backup node is required to check the data format. The backup data verification won't be successful until the nodes above $2f + 1$ pass the verification and save.

iv: The backup node is required to verify the message after receiving the backup data after the verification is successful, and sign on the verification result of s, a is the data digest of the node information, and save it after passing the verification. Only nodes above $2f + 1$ pass the verification and save the data information, and the backup data verification is successful.

v: After all nodes on the chain reach the same height, check whether there is a new area piece that has not reached a consensus. The transaction initiator node c of the new block broadcasts the data format of the signed transaction information to the entire cluster nodes.

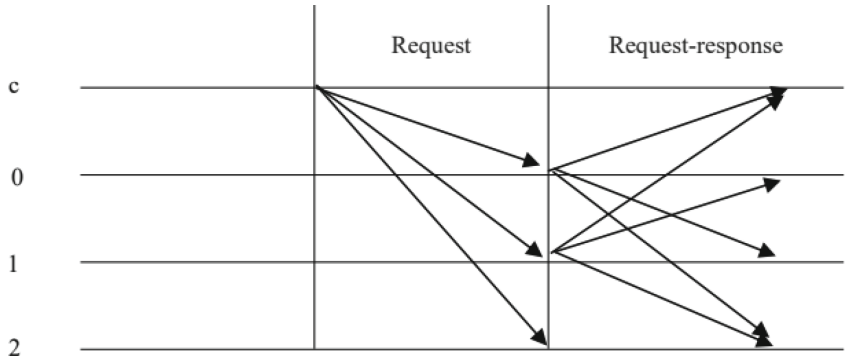

Fig. 2. EPBFT consensus algorithm

vi: After the consensus node receives the transaction information, it verifies the legitimacy of the transaction. After passing the review, return the successful verification information to broadcast the message to the whole network, and enter the REQUEST stage. When the node receives the successful verification information returned by other nodes and the amount exceeds 2f + 1, it enters the next stage.

vii: Enter the stage of prepare-response, check the transaction data for legality, and if it is, go back to node c and store the data. Store the freshly created block information into the blockchain structure after receiving a certain amount of transactions.

The EPBFT consensus algorithm can be applied by the student information management system. It has the advantages of small amount of calculation and high efficiency. The EPBFT consensus mechanism served as the foundation for this study's internal consensus algorithm construction.

The flowchart of the EPBFT consensus algorithm is shown in Fig. 2.

5 EPBFT Advantage Analysis

In terms of communication overhead, the communication overhead of the traditional PBFT algorithm requires three stages of broadcasting in PBFT. In the process of node broadcasting, they need to independently process data information to ensure security, it takes a lot of data transfer to broadcast to every other node in the network when the block verification passes. And the consumed communication resources can be calculated by the following assumptions. Assuming that the number of nodes in the entire network is m, the master node will automatically assign a serial number to the received request information after receiving each request information sent by the client, and the number of communications when broadcasting the pre-prepare message to the verification node is m − 1, the verification node enters the preparation stage after confirming that the broadcast message is correct, and will send a prepare message to all verification nodes in the entire network, and the number of communications in this stage is m × (m − 1), in the serial number confirmation phase, the node sends a confirmation message to all nodes, and the number of communications is also m × (m − 1), so the total number of communications (Mc) should be three stages The sum of communication times, $Nc = 2m^2 - m - 1$,

shows that the complexity of the network reaches the quadratic level. The following calculates the communication overhead of the EPBFT algorithm. In order to reduce the communication complexity, in EPBFT, the independence between nodes is reduced, and the communication complexity of the message is greatly reduced. Mc $= m^2 - 1$ represents the number communication times. Assuming that the ratio of communication overhead between the two is A $= (2m^2 - m - 1)/(m^2 - 1)$, then A is approximately equal to 2, that is to say, the communication overhead of EPBFT is It is half the communication overhead of traditional PBFT.

With regard to latency, the system performs better the lower the latency. The formula for calculating delays is: $T_{delay} = T_{1broadcast} + T_{consensus} + T_{block\ broadcast}$ [9].

By comparing the average delay of the PBFT algorithm and the EPBFT algorithm, we can find that under the same propagation speed, the transaction delay of EPBFT is lower, which proves that it is better.

6 System Module Design

6.1 Overview of System Module Design

Nowadays, electronic archives are developing rapidly, but problems such as data leakage, theft, and centralization have also followed. By analyzing the problems existing in electronic archives today, a system architecture design has been made for the use of blockchain technology in student archives management. Ensure data security, anti-tampering and theft.

6.1.1 System Requirements Analysis

According to system requirements, user roles are divided into student users and file administrators. The following introduces the system functions according to different user requirements.

i: student user

For student users in the file management system, the user first needs to register, log in, and fill in the verification email address and mobile phone number. If the user forgets the password, he needs to verify and modify it through the email address or mobile phone number used during registration. After successfully logging into the system, you need to complete your personal information, including education experience, grade certificates, etc. If you are interrupted during filling, you can also use the editing function until it is completed and uploaded to the system.

ii: archivist

The main functions of the administrator are: user information management, file information management, and file publicity. The user information management of the archivist is the same as that of the student user, including functions such as user registration, user password retrieval, and material modification. File information management includes functions such as modifying, updating, deleting, and transferring student files, and can follow up student files in real time. The file publicity function is mainly for students to view their file information in the system according to their user name and password after submitting materials.

6.1.2 Non-functional Requirements

In addition to meeting the above functional requirements, the student information management module also needs to meet non-functional requirements.

i: During the peak hours of system use, it is necessary to ensure the stable operation of the system and prevent the system from crashing.

ii: In addition to paying attention to the problem that data is easy to be stolen, the system needs to take anti-intrusion measures, and it is necessary to limit the access address and the number of connections to prevent malicious attacks.

iii: The system interface should be concise and clear, so that you can find the required functions more conveniently and quickly after logging in.

6.1.3 System Module Design

The student file management system is divided into an application management module, a network consensus module, and a data storage module. As shown in Fig. 3.

i: The application management module includes a user interface layer and a business logic layer. The user interface layer is the entrance for users, students and file operators, and the business logic layer can realize functions such as user registration, password modification, login verification, adding and updating file data, etc. User students need to register and log in first, and complete their own information. Then submit your own profile information through the login system, including student status materials, grade certificates, political affiliation, etc. The system saves it, and the key center will have two "keys" one-to-one for the user-public key and private key. The private key will be kept a secret from users while the public key will be broadcast throughout the blockchain network. Students can use their own private key to encrypt and sign the file information. When the file operator sends the ciphertext to the student user, the operator uses the student user's public key to encrypt, and the student user can decrypt it with his own private key. When a student user submits a file, he can encrypt it with the public key of the file operator, and when the file operator receives the information, he can decrypt it with his own private key. At the same time, the archivist can add the contents of the

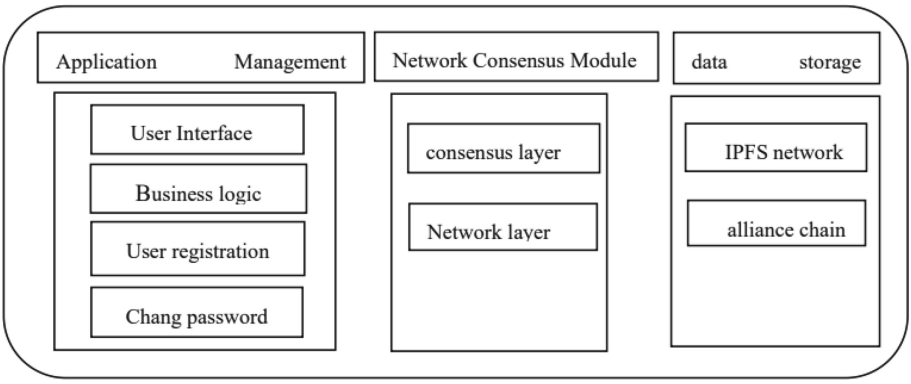

Fig. 3. Student file management system.

archives, such as the basic information of the students, and can also add or delete various attributes for each file.

ii: Both a network layer and a consensus layer are part of the network consensus module [10]. The network layer includes the P2P network of the blockchain, information dissemination mechanism, and data verification mechanism. Second, the consensus layer encapsulates various consensus mechanisms. Many nodes that have nothing to do with each other achieve unification with each other through consensus algorithms. The consensus algorithm among them is capable of applying EPBFT (Equal Practical Byzantine Fault Tolerance).

System initialization, data backup, and verification make up the first three steps of the student information management system's internal node consensus technology, which is based on blockchain. When using the EPBFT consensus technique. The particular steps involved in each stage are as follows:

The information system administrator sends a request to the system client.

1. Initialize the system, and need to backup data b to every node in the whole network.

2. Request backup data node a (administrator) sends request b to node z with the longest chain and needs to accept the verification. Only after more than $2f + 1$ nodes pass the verification and save the data information, the data backup is considered successful.

3. After all nodes have the same status, check whether there is a new block without consensus.

4. The consensus node verifies the legitimacy of the transaction after receiving the transaction information, and returns a successful verification message after the verification is passed.

5. Check the accuracy of the transaction data information during the Prepare-response step.

Finally, the consensus results will be fed back to the departments.

iii: data storage module

Since the operation of the student file management system relies on the internal network of the university, the focus should be on preventing system intrusion attacks and preventing data from being tampered with. The data storage method in this paper adopts the method of combining the alliance chain and the IPFS storage network. The alliance blockchain can effectively protect the data security and data confidentiality in the student documents, and achieve data transfer without loss. The purpose of the IPFS storage cluster is to store Data and timely backup of student files. The alliance blockchain is responsible for storing the hash value of the file data. The hash value is used to verify whether the transaction has been tampered with. IPFS access is efficient and free, which solves the low efficiency and cost of blockchain access. High question. When the archivist sends the student files to the IPFS cluster, the unique release address of the files is signed with the private key of the student user. When other users access the IPFS cluster network, it can verify the signature to ensure legitimacy.

7 Conclusions

With the development of digitization, paper files are gradually replaced by electronic files, and the rise of electronic files saves a lot of resources. However, the information of electronic archives is easy to be stolen and tampered with, and has problems such

as high centralization and insecure data storage. This paper uses the characteristics of blockchain, such as openness, transparency, traceability, safety, efficiency, and tamper-proof, to provide solutions for the security of student archives. A kind of solution thought.

The research done in this paper is as follows:

1) Through analysis, since the PBFT algorithm is mainly applied to the distributed system architecture under the static network architecture, this paper proposes a distributed system suitable for the dynamic network architecture, and the nodes can dynamically adjust the EBFT consensus fault-tolerant algorithm, which solves the problem of the traditional PBFT algorithm. The problem of consuming more resources and high communication overhead.

2) EPBFT has lower latency than PBFT, and communication is easier.

3) The system architecture is divided into three modules, the application management module, the network consensus module, and the data storage module. When a module needs to be optimized or adjusted, it can be better hierarchically managed.

4) The alliance chain is combined with the IPFS network to ensure data security.

These studies can effectively protect the security of files, provide protection means for student file information, and provide safe and reliable data information for future employment files. At the same time, the EPBFT scheme used reduces resource consumption while ensuring system throughput and consensus efficiency, and uses a combination of alliance chain and IPFS network to store data, which is efficient and safe, and can meet application requirements.

References

1. Ali SIM, Farouk H, Sharaf H (2022) A blockchain-based models for student information systems. Egypt Inform J 23(2):187-196. https://doi.org/10.1016/j.eij.2021.12.002
2. Naz M, Al-zahrani FA, Khalid R, Javaid N, Qamar AM, Afzal MK, Shafiq M (2019) A secure data sharing platform using blockchain and interplanetary file system. Sustainability 11(24). https://doi.org/10.3390/su11247054
3. LIU Y Z, LIU J W, ZHANG Z Y, XU T G, YU H. Research on consensus mechanisms of blockchain technology[J]. Journal of Cryptologic Research, 2018, 5(5): 1-2. doi: https://doi.org/10.13868/j.cnki.jcr.0000XX
4. Hui Cao, Hui He, and Jiahe Tian, A Scientific Research Information System via Intelligent Blockchain Technology for the Applications in University Management. https://doi.org/10.1155/2022/7512692
5. MA Junwei, WEN Zhi, XUE Honglin, WANG Yao, BI Sheng and DING Yixin, Blockchain-based enterprise personnel file management system. Doi: https://doi.org/10.3969/j.issn.1672-9528.2021.12.009
6. H. Tang, Y. Sun and J. Ouyang, Excellent practical byzantine fault tolerance, *Journal of Cyber Security*, vol. 2, no.4, pp. 167-182, 2020.https://doi.org/10.32604/jcs.2020.011341
7. WANG Hui, CHEN Bo and LIU Yu-xiang, Research on Personnel File Management System Based on Blockchain. doi:https://doi.org/10.11896/jsjkx.210300051
8. Khan, D., Jung, L. T., Ahmed Hashmani, M., & Wagas, A. (2020). A critical review of blockchain consensus model. In 2020 3rd international conference on computing, mathematics and engineering technologies (¡CoMET) (pp. 1-6). IEEE. doi:https://doi.org/10.1109/iCoMET48670.2020.9074107

9. Huanrong Tang, Yaojing Sun and Jianquan Ouyang, “Excellent Practical Byzantine Fault Tolerance”, School of Cyberspace Security, Xiangtan University, Xiangtan, 411105, China. https://doi.org/10.32604/jcs.2020.011341
10. S.Yu, “Application of blockchain-based sports health data collection system in the development of sports industry,” Mobile Information Systems, vol. 2021, pp. 1-6, 2021. https://doi.org/10.1155/2021/4663147