# A Hybrid Consensus Algorithm for Collaborative Protection of Multi-domain Education Data

*Xianglin Wu [1,2], Tianhao Meng [1], Haotian Huang [1], Lianhai Liu [1*], Jingwei Zhang [1]*

*[1] Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, China*
*[2] School of Artificial Intelligence, Hezhou University, China*
*wuxianglin2015@gmail.com, m1121229656@gmail.com, huanghaotian42@gmail.com, liulianhai@guet.edu.cn, gtzjw@hotmail.com*

## Abstract

Sharing education data among universities can inspire many new applications and developments. At the same time, blockchain technology can ensure the security and sharing of education data. However, the consensus algorithm cannot meet the requirements of low latency and high throughput in the current education blockchain. Therefore, we propose a hybrid consensus algorithm based on a master-slave blockchain (MSB) for multi-domain education data management (EDM) to maintain data consistency. First, we design a double-layer architecture of the MSB that can efficiently and securely handle large-scale education data from universities. Second, facing low consensus efficiency for EDM, we propose the hybrid consensus algorithm that combines the reputation-based RAFT (R-RAFT) and the multi-party optimized PBFT (M-PBFT). The experiment proves that the proposed solution can obviously improve the throughput compared with the single chain. Furthermore, it also performs well on latency, consensus speed, and Byzantine fault tolerance.

**Keywords:** Education data management, Master-slave blockchain, Consensus algorithm

## 1 Introduction

At present, we can mine more valuable information from the shared data among multiple universities, such as the course video recommendation, the learning path recommendation, etc. In addition, sharing the data provides convenience for students who transfer to other universities and for teachers who want to query students' relevant information. However, it is difficult to obtain the data, and the data are at risk of being modified. Blockchain is just suitable for data sharing. It mainly uses the consensus algorithm to ensure data consistency among multiple universities. In this way, it can prevent data modification and facilitate access to data.

Blockchain was first proposed in the paper "Bitcoin: A Peer-to-Peer Electronic Cash System" [1] by Satoshi Nakamoto in 2008. It has the characteristics of decentralization, cryptographic security, and immutability. Blockchain was originally used as a cryptocurrency, and it has been developed as a platform for various applications in different fields [2-10], such as government affairs, finance, health care, supply chain, and IoT. Compared with traditional distributed systems [11], blockchain can be regarded as a special distributed system with its own storage and query rules. Consensus among multiple nodes is reached through the consensus mechanism [12], so they can monitor the authenticity and integrity of the ledger without the support of a third party. Therefore, the innovative way of combining blockchain technology with education data will bring new opportunities for education applications. Among them, an efficient consensus algorithm of blockchain is a basis for providing credible education data.

The main characteristics of education data are diverse data types, large-scale data, multiple stakeholders (such as students, teachers, universities, government agencies, etc.), and frequently updated data. How to maintain the authenticity and reliability of the education data? The consensus algorithm plays an important role in dealing with the problem of the decentralized blockchain system among different universities. It achieves efficient and rapid data consistency through decentralized decision-making. If blockchain technology is directly applied to education data management, the following problems still need to be solved: First, the university nodes have limited resources. When the blockchain system has large-scale education data and frequently uploads diverse data from different universities, it will put storage and processing pressure on university nodes. Second, the current system involves many stakeholders. If all of them participate in the consensus algorithm, the communication times between nodes will be huge. Third, when multiple universities jointly participate in maintaining data, there may be Byzantine nodes. If there isn't a good consensus algorithm to deal with these problems, it will affect the data consistency. Furthermore, we need to build a suitable blockchain architecture for the consensus algorithm. The contributions of this paper are summarized as follows:

- We design a double-layer architecture based on the MSB. Through the architecture, universities can concurrently manage their own data while preventing other universities' data from being modified.
- We propose a hybrid consensus algorithm for EDM.

*Corresponding Author: Lianhai Liu; E-mail: liulianhai@guet.edu.cn
DOI: https://doi.org/10.70003/160792642024122507002

The RAFT algorithm based on reputation value can quickly select a stable node as the leader node to complete local consensus. On the one hand, the RAFT algorithm is suitable for quickly handling large amounts of private data from a single university. On the other hand, the improved algorithm based on reputation value can further reduce the time of leader rotations when a large amount of education data are sent to the blockchain. The multi-party optimized PBFT algorithm can eliminate Byzantine nodes when multiple universities join the blockchain system. In addition, the PBFT algorithm reduces communication times and maintains the fairness of packaging nodes from different universities. Through the combination of the two algorithms, the system can quickly complete global consensus, thereby achieving collaborative protection of multi-domain education data.

- We implement the system based on the MSB and verify the system from latency, throughput, and Byzantine fault tolerance. The results show the system is feasible and efficient. It provides a scalable and secure approach for EDM.

The remainder of this paper is organised as follows: The second part presents the education applications, the research progress of multi-chain technologies, and consensus protocols. The third part describes the architecture of MSB. The fourth part mainly introduces the hybrid consensus algorithm of the MSB. The fifth part simulates and analyzes the proposed MSB and the consensus algorithm. The sixth part summarizes our work.

# 2 Related Work

Scholars have recently begun to study the application of the combination of blockchain and education. In addition, multi-chain technologies and consensus protocols are the core parts of improving the performance of blockchain. In this section, we mainly outline three aspects: education applications, multi-chain technologies, and consensus protocols. These aspects provide references for consensus algorithms about EDM.

## 2.1 Education Applications

With the improvement of blockchain technology and its applications in all walks of life, scholars began to pay attention to the applications of blockchain in the field of education. Mishra et al. [13] proposed a tamper-proof, privacy-preserving, and easy-to-share blockchain architecture for secure sharing of students' credentials. The system used Ethereum's smart contracts to implement the privacy-preserving architecture, and experiments proved its economic feasibility. This architecture used blockchain technology to reduce the existing security-related problems among students, schools, companies, professors, and governmental authorities. Rahman et al. [14] proposed an education data management solution combining blockchain and microservices, which solved the problems of large-scale data privacy protection and data transaction security. The system realized the flexible, reliable, and secure handling of education data through blockchain technology. Li et al. [15] proposed a privacy-preserving authentication system with blockchain for ensuring multimedia resource integrity. The system exploited a hybrid storage pattern that stored multimedia content off the blockchain and the hash values on the blockchain. Although the solution reduced blockchain storage pressure, it required trusted hardware to maintain multimedia data privacy. Ali et al. [16] proposed three models for using blockchains to implement a student information system that maintains transactions such as students' and faculty members' records, course registration records, and student marks. It enabled reliable and secure storage and access to education data. Dewangan et al. [17] proposed a blockchain-based secure, students privacy-preserving certificate sharing and employment conversion system. It used signature and encryption technology to enable the safe and trustworthy management of identities and storage of students' certificates in a decentralized manner. Li et al. [18] proposed a blockchain-based secure storage and sharing scheme for electronic learning records in MOOCs learning systems. It used a combination of smart contracts, encryption and decryption technologies to complete user registration, authentication, data access, and other tasks to achieve safe data sharing. Zhao et al. [19] proposed a blockchain-based student e-portfolio platform integrating a hybrid access control approach. It realized dynamic, decentralized, student-centric, and fine-grained access control management for student data. Li et al. [20] proposed a storage and sharing mode based on blockchain. The system realized student data privacy protection through smart contracts. However, some methods usually use a single chain to maintain education data without considering the data storage pressure of nodes, especially with the increase in the number of education institutions. The other methods use the cloud or database to improve system storage capacity, but there may be risks of data leakage.

## 2.2 Multi-chain Technologies

With the current development of blockchain, multi-chain has become an important means to improve data processing capabilities and reduce isolated data islands. Each chain is relatively independent of the other chains. They can store different types of data on their own chains. Such it can realize data aggregation without affecting the performance of the respective blockchains and improve the flexibility of data management.

Some researchers have proposed the multi-chain to deal with various data in many different fields. Chang et al. [21] proposed a SynergyChain system based on a three-tier architecture to achieve multi-chain data sharing in IoT applications. The system achieved data reliability by aggregating and reorganizing the data from multiple blockchains. Guo et al. [22] proposed a master-slave chain based trusted cross-domain authentication mechanism in IoT to improve the efficiency and credibility of authentication. Feng et al. [23] proposed a cross-domain authentication scheme for drones through the combination of the private blockchain and the consortium blockchain. The private blockchain was used for storing information about registered

drones in the domain. The consortium blockchain was used for exchanging digital certificates to achieve authentication. He et al. [24] designed a multi-chain 5G network slicing service quality computing model to calculate the service quality parameters of the network slicing, which provided optimal service quality parameters for customizing virtual networks to meet the differentiated needs of customers. Xiong et al. [25] proposed a notary group-based cross-chain interaction model to achieve exchange between different blockchains. It ensured the transactional properties, security, and success rate of cross-chain transactions. Huang et al. [26] proposed an efficient energy transaction management model based on multi-chain, so the security and privacy of energy trading were guaranteed. Bai et al. [27] proposed a multi-chain structure that accommodated thousands of edge data to improve the efficiency of on-chain data and realize cross-chain edge data sharing in heterogeneous blockchain systems. Yu et al. [28] proposed a security access control method to protect the data of multi-level security systems. The system used multi-chain technology to divide resources into different domains and used smart contracts to achieve accurate access to these different domains. At the same time, the side chain stored access records. This method realized dual protection for the system. He et al. [29] proposed a scheme that used multi-chain to manage the electric vehicle shared charging platform. It stored different types of information on different blockchains to improve storage and query efficiency. Hao et al. [30] proposed an interoperable hybrid blockchain system that was maintained by different organizations. It used an interoperable consensus group to maintain the consistency of the local blockchain and the global blockchain.

For the above various multi-chain schemes, some schemes do not consider the reputation value of the node. If the nodes with a low reputation value are elected as the consensus nodes in the education field, it will affect the consensus performance. However, the other schemes do not consider the global consistency of data. Similarly, the system cannot really prevent the data from being modified without considering the global consistency of the data for EDM.

### 2.3 Consensus Protocols

The consensus algorithm is the core component of the entire blockchain system, which determines the overall performance of the blockchain. The following four consensus algorithms are widely used in blockchain:

PoW [31] introduces the workload proof for Bitcoin. It solves complex mathematical problems through mining with high-performance computers. The miner tries to use different nonce values as input and continuously performs the hash operations of SHA-256. When the calculated hash value is less than the target threshold, the miner obtains the current block accounting right. The entire process has high energy consumption, low throughput, and poor fairness. Its advantages are high security and stability.

PoS [32] is an efficient and more competitive consensus algorithm. A higher coinage will lead to a higher possibility for a node to obtain transaction packaging rights. The PoS algorithm is weaker than PoW in terms of security. Users are encouraged to hoard coins. It is at risk of long-range attacks and nothing-at-stake attacks. Compared to PoW's 7 transactions per second, the speed of PoS processing transactions can reach 30 transactions per second.

PBFT [33] doesn't consume computing power and has no forks. It can be applied to a blockchain system without tokens. The PBFT algorithm goes through 5 stages to complete a round of consensus, including request, pre-prepare, prepare, commit, and reply. By exchanging information among nodes, the system prevents interference from Byzantine nodes. Finally, the system realizes the data consistency of all nodes. When the system has 3F+1 nodes, it can tolerate the number of F fault nodes or evil nodes. The performance of the algorithm will decrease sharply with the increase in the number of nodes, such as low throughput and high latency. The bottleneck of PBFT is mainly focused on the heavy communication among nodes.

RAFT [34] can initially solve the consistency problem in the distributed system environment. It is also used in private chains. It mainly includes two stages: leader election and log replication. The leader node maintains the connection with other nodes through heartbeat information. In the log replication stage, the follower synchronizes the leader's request and returns the execution result to the leader. When the leader node fails, the system will reselect a leader node for the consensus algorithms. This algorithm cannot handle the Byzantine nodes, but it can tolerate some crashed nodes.

By analyzing the advantages and disadvantages of the above four classic consensus algorithms, we need to improve a consensus algorithm to meet the rapid consistency of the data for EDM. Furthermore, we also consider the Byzantine node problem when multiple universities participate in the blockchain system.

## 3 MSB for EDM

### 3.1 Problem Statement

Our goal is to achieve fast data consistency for EDM in the field of education. Two challenges need to be resolved. The first challenge is how to process large amounts of education data from multiple universities with a blockchain system. The second challenge is how to optimize the consensus algorithm for the presence of a large number of nodes.

The method proposed by Hao et al. [30] provides us with some references, but it does not take into account the impact of the reputation value of nodes on consensus performance. In addition, the method stores all data from local blockchains on the global blockchain to maintain global consistency, which will put storage pressure on the global blockchain. In response to the above challenges, we need to design a double-layer blockchain system to concurrently process data from multiple universities. At the same time, the master chain only stores the summary of the data from the slave chain to maintain global consistency. What's more, for the fast consistency of the education data, considering issues such as the reputation value of nodes and the communication load of PBFT, we propose a hybrid consensus algorithm combining RAFT and PBFT to enhance robustness, increase throughput, and reduce latency.

### 3.2 Overall Architecture

In order to meet the requirements of quickly processing large-scale education data, we built a system based on the MSB for EDM. The slave chain concurrently processes the data. The master chain maintains global data consistency so that it can guarantee the reliability of the data among universities. In this way, the data from multiple universities can be efficiently managed. According to the advantages of the MSB, a double-layer architecture is proposed for EDM, as shown in Figure 1.

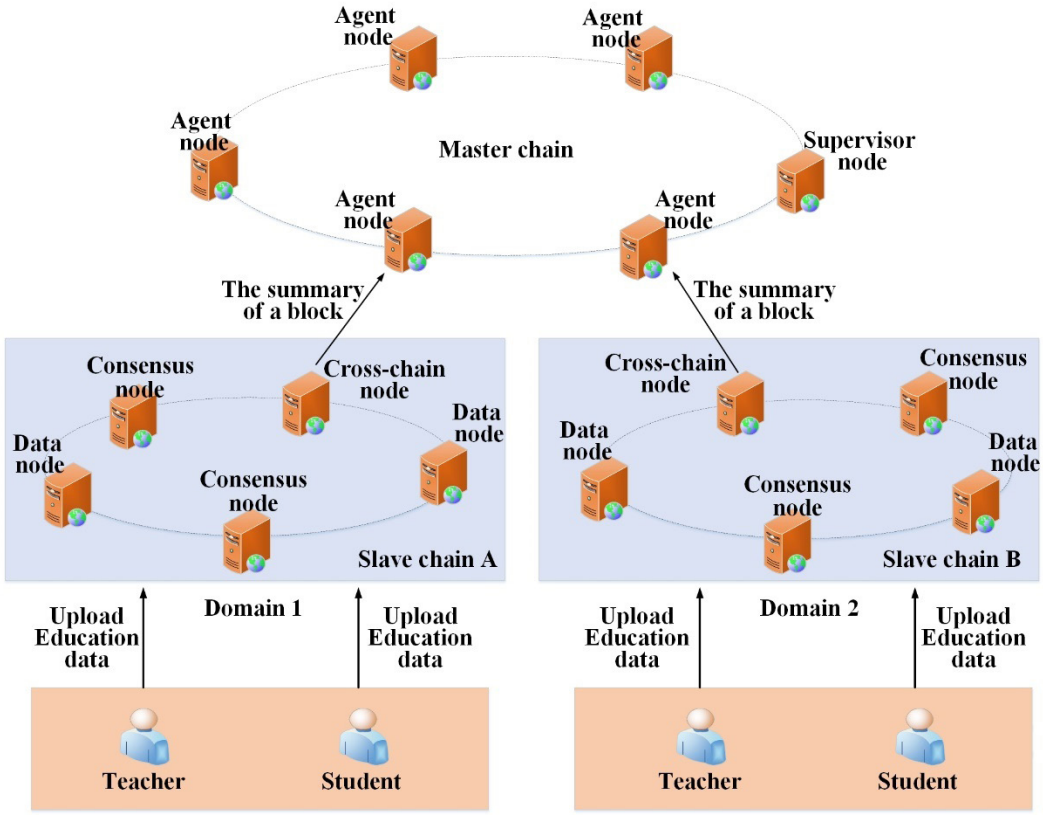



**Figure 1.** The double-layer architecture for EDM

The bottom layer consists of multiple domains. Each domain has a slave chain. It quickly forms a block referred to as an intra-university block on the slave chain. The upper layer is the master chain. It is composed of some nodes from different salve chains. The summary of the block from slave chains is stored on the master chain. The master chain block is referred to as an inter-university block for the global consistency of data. If the intra-university block is changed on the slave chain, the hash value of the block will not be found on the master chain. At the same time, the data of the blocks on the master chain are maintained by multiple different university nodes. In this way, we can guarantee that the data will not be modified among domains.

Based on the double-layer architecture, there are five types of nodes: the data node, the consensus node, the cross-chain node, the agent node, and the supervisor node. The function description of each type of node is shown as follows:

Data node: It provides verification and forwarding services for education data from the system. Various data of students and teachers are stored on the data nodes. By setting multiple data nodes, students and teachers can easily upload different types of information to data nodes. Eventually, formal education data are formed after verification by the data node.

Consensus node: It provides computing services for consensus on the slave chain. The consensus nodes need a good hardware and software environment so that they can undertake more tasks. The system forms the intra-university block through a consensus algorithm. The data nodes actively synchronize intra-university blocks from the consensus nodes, which promotes the rapid consistency of data within a university.

Cross-chain node: It connects the consensus nodes and the agent nodes between the master chain and the slave chain. It transmits the summary of the intra-university block to the agent node. Cross-chain nodes provide data transmission services for the master chain.

Agent node: The system selects the consensus nodes from the slave chain as the master chain node. Because these nodes come from different universities, they are called agent nodes. The master chain forms an inter-university block through a consensus algorithm. The agent nodes provide computing services for global consistency. By selecting some agent nodes, the system can reduce the number of each university node on the master chain. It can reduce the communication times of the consensus algorithm and improve consensus efficiency.

Supervisor node: It initiates the consensus and calculates consensus results on the master chain. As a reliable node, it will provide conditions for the optimization of the consensus algorithm. This node is generally held by the government's education authority. It not only maintains the normal operation of the consensus algorithm but also supervises the behavior of university nodes. Finally, it reduces the probability of system error.

### 3.3 Block Structure of Slave Chain

The slave chain consists of data nodes, consensus nodes, and cross-chain nodes, which mainly completes the storage of education data records and completes consensus on the slave chain nodes. A slave chain represents a domain that is maintained by multiple internal nodes of the university. Each slave chain represents a university. The intra-university block consists of various education data from students and teachers on the slave chain. The MSB connects different blockchains. As shown in Figure 2, The block header includes the version number, the hash value of the previous block, the height of the current block, the signature of the packaging node, the timestamp, and the Merkle tree root obtained by calculating the hash value of the data layer by layer, etc. The *tx* represents the data record which has detailed information about education data: student ID, student name, timestamp, etc. After the record is hashed, a specific hash value is formed. The hash values of the different data records are represented by Hash (*txi*), respectively. Through the data hashed on different layers, the Merkle tree is finally formed. Any modification of the record will change the root hash value of the Merkle tree. Thus, the system realizes the data immutability of the slave chain.

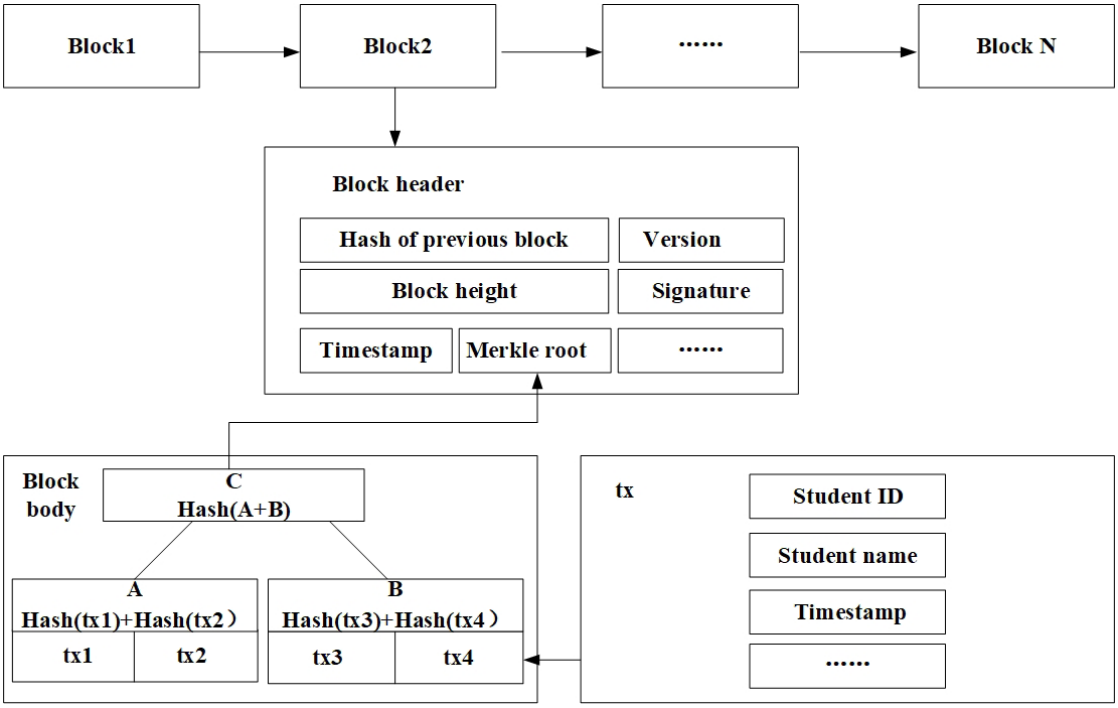



**Figure 2.** The block structure of the slave chain

### 3.4 Block Structure of Master Chain

The master chain consists of agent nodes, cross-chain nodes and supervisory nodes. It mainly stores the summary of the block from the slave chain and reaches a consensus on the master chain. The block structure of the master chain is similar to that of the slave chain. The agent nodes maintain the inter-university block on the master chain. The block structure of the master chain is shown in Figure 3. The block header includes the version number, the hash value of the previous block, the height of the current block, the signature of the packaging node, the timestamp, and the Merkle tree root obtained by calculating the hash value of the data layer by layer, etc. The MSB connects different blockchains. The *tx* represents a summary of a block from a university slave chain. The Merkle tree of the master chain is completed after some summaries are hashed from different university slave chains. The data processing steps of the master chain are as follows: When a leader node generates a block on a university slave chain, it will send the summary to the master chain, such as the hash value of the intra-university block, the slave chain ID, the timestamp, etc. The summary is broadcast to all agent nodes with the cross-chain node. After the agent nodes receive the multiple summaries from different university chains, the inter-university block is formed through the consensus algorithm of the master chain. Generally speaking, when the node initiates education data storage on the slave chain, the leader node packages the data into a block on the slave chain. At the same time, the leader node also sends the summary of the packaged data to the master chain by the cross-chain. The data are broadcast to the master chain to form blocks. Finally, the system realizes the consistency of global data. In this way, the system prevents modification of the data on both the master chain and the slave chain.

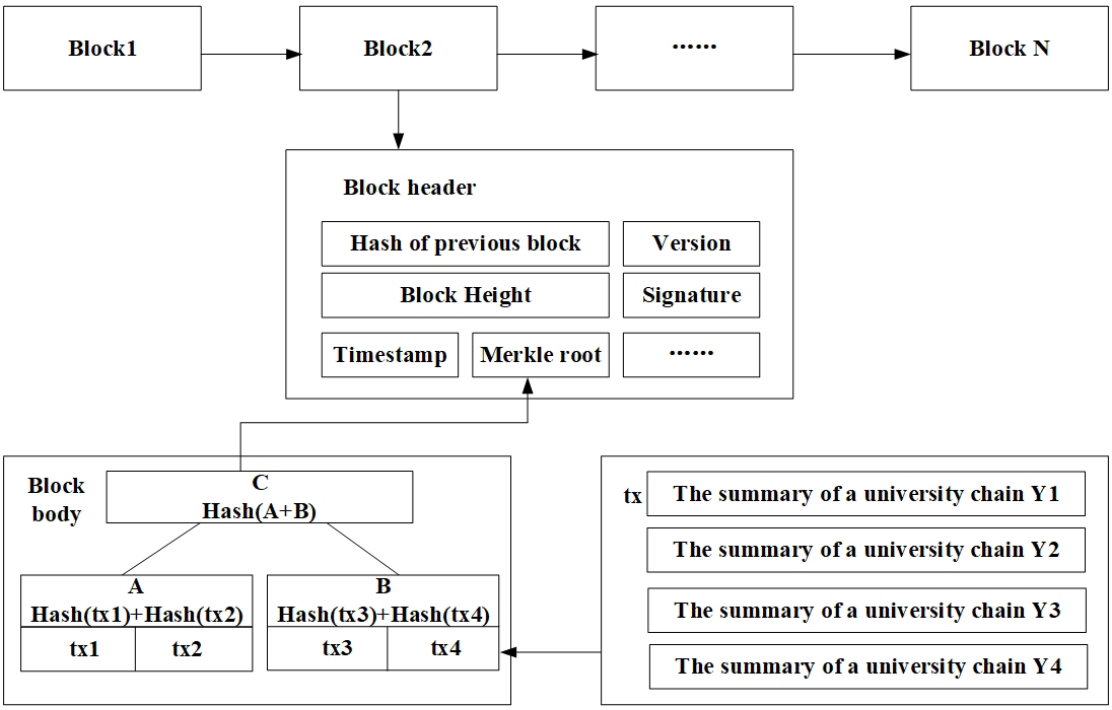



**Figure 3**. The block structure of the master chain

# 4 Consensus Algorithm for MSB

Although the MSB of the double-layer architecture can concurrently process the data of each slave chain, it requires the consensus algorithm to guarantee the consistency of the data for all nodes. For the smooth operation of the hybrid consensus algorithm, the system needs to collaborate with master chain nodes and slave chain nodes. Because some nodes have dual identities, they need to complete the consensus algorithm on both the slave chain and the master chain.

## 4.1 Main Overview

There are a large number of nodes on the slave chain, such as student nodes, teacher nodes, and other nodes. However, these nodes have different computing capabilities. In addition, some nodes are not real-time online. Meanwhile, the Byzantine node may exist on the master chain. These should be considered for the consensus algorithm of the MSB. If these problems are not reasonably solved, the efficiency of consensus will be greatly affected.

The goal of the optimized algorithm is to improve the efficiency of consensus to achieve global consistency. The slave chain adopts the RAFT algorithm based on reputation value. At the same time, the master chain uses an optimized PBFT algorithm. To achieve this goal, the idea of the consensus algorithm includes two aspects: The number of nodes participating in the consensus algorithm has a significant impact on the performance of the blockchain. Therefore, we effectively reduce the number of nodes participating in the consensus algorithm. In addition, facing consensus nodes with different performances, we propose the RAFT algorithm based on the reputation value on the slave chain. The node with the highest reputation value is selected as the leader node, which can improve the stability of the algorithm. In order to reduce the calculation pressure on consensus nodes, some nodes are designated as data nodes according to their reputation values. The nodes verify the correctness of uploading data from the student nodes and the teacher nodes on the slave chain. The cross-chain node provides data transmission service for the slave chain and the master chain. The system completes the RAFT algorithm of the slave chain through these nodes.

Byzantine nodes, the number of nodes, and other problems will affect consensus efficiency on the master chain. Therefore, we propose the PBFT algorithm based on multi-party optimization on the master chain. The main optimization includes four aspects: the agent nodes, the primary node, the Byzantine nodes, and the node communication times.

## 4.2 Reputation-Based Node Evaluation Parameter

If multiple selected leader nodes fail, it will affect the stability of the consensus algorithm. In the proposed consensus algorithm, the evaluation of consensus nodes is based on their reputation value. For the high concurrency of education data in universities, it is particularly vital to implement a stable consensus algorithm. In addition, according to the node reputation value, the system determines the number of data nodes and consensus nodes.

The reputation value of a node is mainly determined by four important indicators: the hardware performance of the nodes, the network performance of the nodes, the daily behavior of the nodes, and the abnormal behavior of the nodes. The higher the reputation value of a node is, the more likely it is to become the leader node. The specific parameters are composed of four attributes in Table 1.

$C_1$ represents the hardware performance of a node, which

determines the operation of the blockchain. CPU and RAM are used as the infrastructure of the blockchain. $PC_i$ is the evaluation of them with benchmarks [35].

$$C_1 = \log\frac{PC_i}{2} \tag{1}$$

$C_2$ is the performance index for evaluating a node's network. Due to network fluctuation and other factors, data transmission between nodes may be delayed. The threshold $L$ is used as a parameter to judge the delay. When the average delay $AL$ of the node is less than the threshold $L$, the node is determined to be reliable. When the average delay exceeds the threshold $L$, the node may cause severe consequences for the consensus algorithm.

$$C_2 = \frac{1}{2^{AL_i - L}} \tag{2}$$

$C_3$ is a quantitative indicator of the offline times of a node. $OL$ is the number of offline times of the node in a period of time. Although the RAFT algorithm can tolerate node failure due to offline, the node with fewer offline times is selected as the leader node for the robustness of the algorithm.

$$C_3 = (\frac{1}{2})^{OL_i + 1} \tag{3}$$

$C_4$ represents a measure of abnormal nodes' behavior. When the node is running, the IP packet repetition rate $P$ determines whether the node has abnormal behavior. The higher the value of $P$ is, the higher the error probability of the node is. $BC$ is a constant index.

$$C_4 = 100 - (2 - (BC)^{P_i}) \tag{4}$$

**Table 1.** Reputation value parameters

| Feature | Specific instructions | Value | Weight |
|---|---|---|---|
| Node hardware performance | RAM and CPU utilization | $C_1$ | $W_1$ |
| Node network performance | Network latency | $C_2$ | $W_2$ |
| Node daily behavior | Node offline times | $C_3$ | $W_3$ |
| Node abnormal behavior | IP packet repetition rate | $C_4$ | $W_4$ |

**Algorithm 1.** Leader node selection algorithm

**Input:** the consensus node set $A = \{X_1, X_2, \cdots, X_n\}$, the node characteristic value $\{C_1, C_2, C_3, C_4\}$, the node weight coefficient $\{W_1, W_2, W_3, W_4\}$, $NX_i$ represents the reputation value of a node $X_i$
**Output:** leader node $X_i$
  **while** $i <$ A.length **do**
  $NX_i = C_1 * W_1 + C_2 * W_2 + C_3 * W_3 + C_4 * W_4$
  $i$++
  **end while**
Descend sorting of reputation value $NX_i$
 **if** $NX_i$ is the maximum value
    $X_i$ is the leader node
  **end if**
  **if** $X_i$ loses heartbeat information
    The next node $X_j$ is selected as the leader node according to the reputation value order
  **end if**

### 4.3 R-RAFT Algorithm

The traditional RAFT consensus algorithm doesn't consider the stability of nodes, so the leader node may be constantly replaced by other nodes in the consensus process. Stable nodes are selected as leader nodes, which can improve the efficiency of the algorithm.

After each feature of the node is reasonably evaluated, it is set with a certain weight value. After sorting the reputation value of the nodes in descending order, the system selects the node with the highest reputation value as the leader node. The other nodes are consensus nodes and data nodes. The leader node selection algorithm is shown in Algorithm 1.

In the process of consensus, after the leader node is selected, the other nodes are follower nodes. The leader node periodically sends heartbeat information to all follower nodes. The leader node verifies the education data from other nodes and then packages the data into blocks. The leader node broadcasts the Append Entries RPC message containing blocks to other follower nodes. When the block information is verified by more than 1/2 of the nodes, the leader node sends the summary of the intra-university block to the master chain through the cross-chain node. At the same

time, the leader node sends an empty Append Entries RPC message to all follower nodes, which indicates that most nodes have agreed with this block. The follower nodes store this block after they receive the confirmed Append Entries RPC message. The data nodes and other nodes (student nodes and teacher nodes) send a request to the consensus node to synchronize the block. When all the nodes are synchronized successfully, the block is consistent on the slave chain. When most nodes do not receive heartbeat information from the current leader node, another node with a high reputation value will be selected as the leader node to complete the subsequent consensus process.

### 4.4 Byzantine Node Identification Algorithm on the Master Chain

Because the slave chain is a private chain, the RAFT algorithm can solve the data consistency problem well. However, the master chain is a consortium chain that is composed of multiple universities. There may be Byzantine nodes. For data reliability, the system should reasonably handle Byzantine nodes. If Byzantine nodes continue to participate in the consensus algorithm, the performance will be affected. Therefore, we propose a Byzantine node identification algorithm to reduce the number of Byzantine nodes.

The specific process of the identification algorithm is as follows: The supervisor node maintains a block voting result table $BT = \{b_1, b_2, \ldots b_i \ldots b_n\}$. The $b_i$ represents the block voting result of a node. At the end of each round of the PBFT algorithm, the supervisor node identifies Byzantine nodes according to the voting results from agent nodes. If the voting result of a node is inconsistent with 2F+1 nodes' results within the specified time, it is regarded as an abnormal behavior node. F is the maximum number of faulty replicas. Another case is that the node doesn't respond to the supervisor node within the specified time. Both types of nodes are identified as Byzantine nodes.

These Byzantine nodes will be excluded from the next round of agent nodes. It can effectively prevent nodes from continuing to participate in the consensus algorithm and affect the efficiency of consensus. In this way, we can continuously improve the stability and reliability of the master chain.

### 4.5 Selection of Agent Node and Primary Node

Each slave chain has multiple consensus nodes. If all of the nodes participate in the consensus algorithm on the master chain, the efficiency may be affected by an increase in the number of domains. In addition, how to ensure the fairness of nodes in these domains.

The multiple nodes with high reputation value are selected as agent nodes on the master chain according to the node reputation value of the nodes on the slave chain. The selected agent nodes are relatively stable. Therefore, the system reduces the number of participating nodes so as to reduce the communication times during the consensus process.

As more universities join the consortium chain, more agent nodes will participate in the consensus algorithm on the master chain. The system randomly selects a primary node to ensure the fairness of nodes among universities with a verifiable random function (VRF). Each agent node uses its private key and the height value of the current blockchain to the VRF to get a random value. If the value of a node is the minimum, the node is selected as the primary node. It is responsible for packaging, signing, and broadcasting blocks to the entire network. The method can achieve the goal of fairness for the domains.

### 4.6 M-PBFT Algorithm

The traditional PBFT algorithm can tolerate a certain number of Byzantine nodes and complete the PBFT algorithm. However, the fairness of the primary node and the elimination of Byzantine nodes from the consensus algorithm are not considered. In addition, the communication times of the PBFT algorithm will sharply increase when a large number of nodes participate in the consensus algorithm. Therefore, we optimize the PBFT algorithm based on the above aspects. The algorithm process is shown in Figure 4. The specific algorithm steps are as follows:

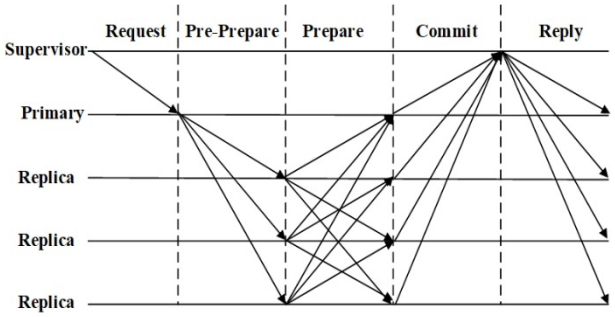



**Figure 4**. M-PBFT

1) When a node is randomly selected as the primary node according to the VRF, it obtains the packaging right.

2) The supervisor node sends block packaging requests to the primary node. It obtains the transaction list from the transaction pool, sorts the transactions, and packages the selected transaction into a block. At last, it broadcasts the block to all replica nodes (agent nodes). The system is in the pre-prepare stage.

3) When a node receives a block in the pre-prepare stage, the node enters the prepare stage. When a node broadcasts the block in the prepare stage, it also receives a block from other nodes and validates the block. After the block verification is passed, the node broadcasts the block to the supervisor node.

4) After the supervisor node receives the results of multiple nodes in the commit stage, it counts the voting results. According to the received votes of the block from other nodes within the specified time, the supervisor performs the Byzantine consensus node identification algorithm. If the number of approved votes is more than 2F+1 within the specified time, the current block is legal. F represents the maximum number of fault nodes. The nodes enter the reply phase after completing the identification algorithm.

5) After the supervisor node confirms that the current block is legal, it broadcasts the confirmation information. If agent nodes receive the final confirmation message for the block, they update the block information. Otherwise, all nodes reject the block when the supervisor node confirms

that the current block is illegal. Furthermore, if the system detects Byzantine nodes, it will reselect other nodes for the next round of consensus.

Compared with the traditional PBFT algorithm, we introduce a supervisor node to quickly process Byzantine nodes and reduce the communication times among nodes. This algorithm effectively ensures the consistency of data among domains.

# 5 Experiment and Analysis

We implement the MSB prototype based on Xuperchain [36]. Xuperchain is Baidu's self-developed underlying blockchain technology which has many internationally leading technologies such as in-chain parallel technology, a pluggable consensus mechanism, and integrated smart contracts [37]. Compared to Ethereum, Hyperledger Fabric, and Tendermint, it has stronger compatibility and more flexible scalability. Xuperchain is widely used in various consortium chain and private chain scenarios.

**Table 2.** Environment configuration parameters

| Attribute name | Parameter description |
|---|---|
| CPU | E7-4820@2.0GHz * 2 |
| Operating system | CentOS 8.0 |
| RAM | 16GB |
| Hard disk | 500G |
| Programming language | Java |
| Blockchain platform | Xuperchain |

To verify the feasibility of the MSB, we use Xuperchain and JAVA smart contracts to implement the MSB system. The system is tested in a test environment. The environment configuration parameters are set as stated in Table 2. By using the different numbers of nodes to form a MSB network, we test key indicators of the consensus algorithm. We mainly implement performance tests including latency, throughput, etc.

### 5.1 Master Chain Communication Times Analysis

According to the advantages of the M-PBFT algorithm in terms of communication overload, we conduct the following comparative analysis.

When the total number of consensus nodes is $N$, we can see that the communication process includes three stages. In the pre-prepare stage, the communication times of the M-PBFT algorithm at this stage are $N-1$. In this process, the primary node sends a message to each replica node. In the prepare stage, the consensus nodes must send a message to the other nodes. The communication times of the M-PBFT algorithm at this stage are $(N-1)*(N-1)$. In the commit stage, all nodes validate the received prepare messages. If the message is true, the replica node sends messages to the supervisor node. Each node needs to send 1 message to vote for the block, which requires $N$ communication times. The total communication times of the M-PBFT algorithm are $N*N$, while the total communication times of the traditional PBFT algorithm are $2N*(N-1)$. As can be seen from Figure 5, the M-PBFT algorithm has relatively fewer communication times than the traditional PBFT algorithm. When the number of nodes is small, the communication times of both methods are roughly the same. When the number of nodes increases, the M-PBFT can obviously reduce the communications times between the nodes. The M-PBFT algorithm will effectively improve the performance of the blockchain.

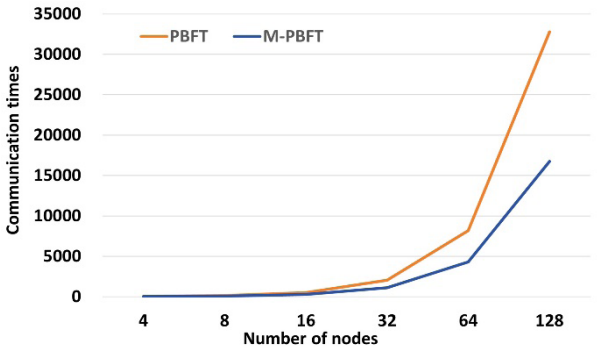



**Figure 5.** Comparison of communication times between PBFT and M-PBFT

### 5.2 Experiment Setting

In order to verify the feasibility and practicability of the proposed method, we send different amounts of data from the EduCoder dataset for this experiment. The dataset has a total of 31 attributes, 220 courses, 1580 training projects, and 2320227 pieces of user level information, which includes user ID, user name, the name of the user's school, course ID, course name, number of chapters, number of training courses, number of learners, chapter data (chapter name, description, related training projects), training project ID, training project name, level data (level ID, name, task description, number of people who have passed the task, number of people who are doing the task) and other attributes. For the performance test of the MSB, a different number of nodes are set, and the supervisor node is set separately.

#### 5.2.1 Experiment regarding the Performance of MSB

For comparison of throughput between the single chain and the MSB, the system sets 5 slave chains and 1 master chain. Each slave chain is set with 10 consensus nodes. The master chain sets 30 nodes as agent nodes. Specific parameters are shown in Table 3.

**Table 3.** Experimental configuration parameters

| Attribute name | Group1 | Group2 |
|---|---|---|
| Consensus algorithm | RAFT/M-PBFT | RAFT |
| System type | MSB | Single chain |
| Number of master chain nodes | 30 | - |
| Number of slave chains | 5 | - |
| Number of slave chain consensus nodes | 10 | - |
| Total number of nodes | 80 | 80 |

The consensus throughput is tested on both the single chain and the MSB. The experiment sends different amounts

of education data to test the processing capacity. The result is shown in Figure 6.

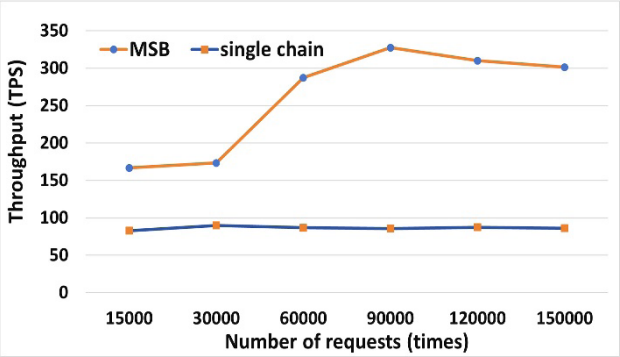



**Figure 6.** Comparison of throughput between the single chain and the MSB

Although different amounts of requests are sent to the single chain, the throughput of Group2 is kept between 80tps and 90tps. When the number of requests is 90000, Group1 reaches its maximum 327 tps. This is three times more than the single chain. The MSB can process data concurrently, so its performance can be significantly improved. In the case of the single chain, the more nodes that participate in the consensus algorithm, the more time it takes to achieve data consistency for all nodes. The master-slave chain can process data concurrently, so the throughput is better than that of the single chain. From the results, we can see that the throughput of the MSB is generally better than that of a single chain. Experimental results show that the system based on the MSB can effectively provide trusted data to universities.

For the comparison of latency between PBFT and M-PBFT on the master chain, the slave chain uses the same RAFT consensus algorithm. The experimental results are shown in Figure 7.

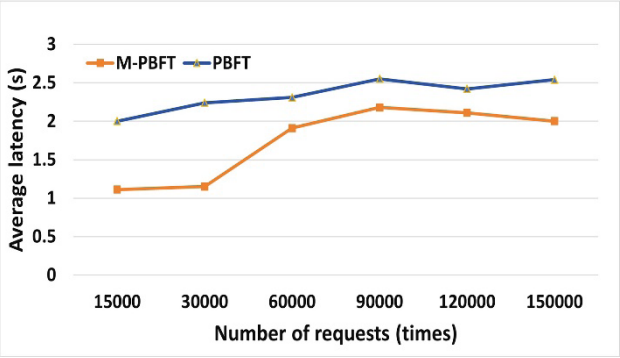



**Figure 7.** Comparison of latency between PBFT and M-PBFT

A different number of requests are sent, and the latency of the PBFT algorithm is always more than 2 seconds. However, the latency of the M-PBFT algorithm is lower than the PBFT algorithm. The reason is the M-PBFT algorithm reduces the communication times of nodes in the commit phase, and the nodes have higher performance. When the number of requests reaches 90000, the latency of the system is relatively stable with both algorithms. The reason is that the two consensus algorithms respectively reach their performance limits.

We compare the data processing capabilities of different numbers of slave chains. In the experiment, 2, 5, 8, and 10 slave chains are respectively used to form the MSB. The system respectively selects 20 nodes from slave chains as the agent nodes. Specific parameters are shown in Table 4.

**Table 4.** Experimental configuration parameters

| Attribute name | Parameter description |
|---|---|
| Consensus algorithm | RAFT/M-PBFT |
| System type | MSB |
| Number of master chain nodes | 20/20/20/20 |
| Number of slave chains | 2/5/8/10 |
| Number of slave chain consensus nodes | 20/8/5/4 |
| Total number of nodes | 60/60/60/60 |

The results of the experiment are shown in Figure 8. It takes 44 seconds for 2 slave chains to handle 15000 requests. Meanwhile, it takes 30 seconds for 10 slave chains to handle 15000 requests. For the same number of requests, the consensus time can be reduced as the number of slave chains increases. The reason is that each slave chain can concurrently complete the local consensus and form blocks.

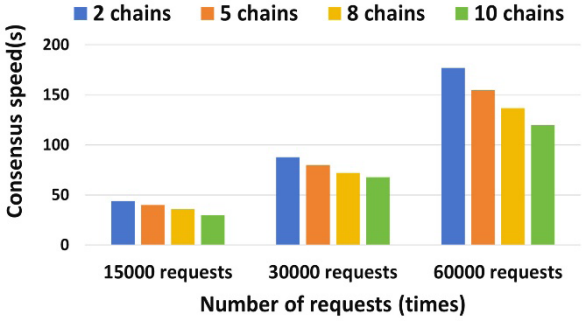



**Figure 8.** Consensus speed with different numbers of slave chains

### 5.2.2 Experiment regarding the Byzantine Node Tolerance

During the operation of the system, we set an agent node as an abnormal node to simulate the failure of the university node. The results are shown in Figure 9.

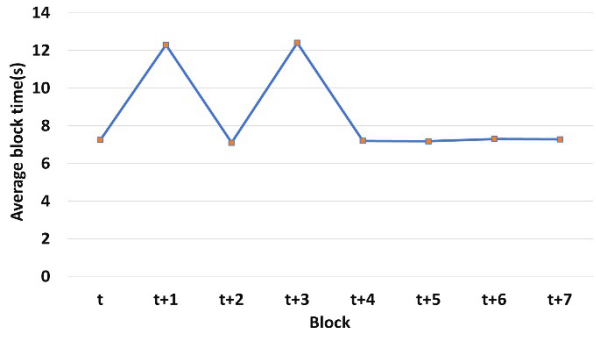



**Figure 9.** The Byzantine node tolerance experiment

From the test experiment, we can see that the system has

a Byzantine node, and the time for generating the next block is increased. Compared with the normal block generation time, the time is increased by about 5 seconds. When a university node is a Byzantine node, the system will reselect another node to participate in the consensus algorithm on the master chain. Although the system increases the time for generating a block when the Byzantine node appears, it can continue to complete the consensus algorithm. Therefore, the M-PBFT algorithm has strong robustness.

In summary, we verify the feasibility and efficiency of the algorithm through various experimental tests. According to the experimental results and analysis, our proposed algorithm can concurrently process large amounts of data and quickly achieve global consistency for EDM.

# 6 Conclusion

In order to achieve efficient data consistency for education data among multiple universities, we propose a double-layer architecture to concurrently handle the multiple domains of education data, in which the hybrid consensus algorithm is used to ensure data consistency. The system realizes the consistency of local data through slave chains using the reputation-based RAFT algorithm. The master chain uses a multi-party optimization PBFT algorithm to exclude Byzantine nodes and achieve global consistency of data. The experiment verifies the feasibility of the method in terms of throughput, latency, robustness, etc. Through the MSB, the system provides credible data for EDM among domains.

# Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant No. 62267002, 62167002), Guangxi Key Research & Development Program (Grant No. Gui Ke AB22080047), Guangxi Key Laboratory of Trusted Software (No. KX202317), and Engineering Research Center for Blockchain Data Management (Ministry of Education).

# References

[1] S. Zhang, J.-H. Lee, Double-Spending with a Sybil Attack in the Bitcoin Decentralized Network, *IEEE Transactions on Industrial Informatics*, Vol. 15, No. 10, pp. 5715-5722, October, 2019.

[2] C. Piao, Y. Hao, J. Yan, X. Jiang, Privacy preserving in blockchain-based government data sharing: A Service-On-Chain (SOC) approach, *Information Processing & Management*, Vol. 58, No. 5, Article No. 102651, September, 2021.

[3] S. Bhagavan, P. Rao, T. Ngo, C3HSB: A Transparent Supply Chain for Multi-cloud and Hybrid Cloud Assets Powered by Blockchain, *2021 IEEE 37th International Conference on Data Engineering Workshops (ICDEW)*, Chania, Crete, Greece, 2021, pp. 100-103.

[4] Y. Yan, C. Wei, X. Guo, X. Lu, X. Zheng, Q. Liu, C. Zhou, X. Song, B. Zhao, H. Zhang, G. Jiang, Confidentiality Support over Financial Grade Consortium Blockchain, *2020 ACM SIGMOD International Conference on Management of Data*, Portland, Oregon, USA, 2020, pp. 2227-2240.

[5] M. Li, S. Shao, Q. Ye, G. Xu, G. Q. Huang, Blockchain-enabled logistics finance execution platform for capital-constrained E-commerce retail, *Robotics and Computer-Integrated Manufacturing*, Vol. 65, Article No. 101962, October, 2020.

[6] Z. Qu, Z. Zhang, M. Zheng, A quantum blockchain-enabled framework for secure private electronic medical records in Internet of Medical Things, *Information Sciences*, Vol. 612, pp. 942-958, October, 2022.

[7] Z. Peng, C. Xu, H. Wang, J. Huang, J. Xu, X. Chu, P2B-Trace: Privacy-Preserving Blockchain-based Contact Tracing to Combat Pandemics, *2021 ACM SIGMOD International Conference on Management of Data*, Xi'an, Shaanxi, China, 2021, pp. 2389-2393.

[8] Y. Ren, D. Huang, W. Wang, X. Yu, BSMD:A Blockchain-Based Secure Storage Mechanism for Big Spatio-Temporal Data, *Future Generation Computer Systems*, Vol. 138, pp. 328-338, January, 2023.

[9] Y. Suo, Y. Wang, S. Luo, Q. Yang, J. Zhao, SV-PBFT: An Efficient and Stable Blockchain PBFT Improved Consensus Algorithm for Vehicle-to-Vehicle Energy Transactions, *Journal of Internet Technology*, Vol. 23, No. 6, pp. 1191-1201, November, 2022.

[10] Y. Ren, Y. Leng, Y. Cheng, J. Wang, Secure Data Storage Based on Blockchain and Coding in Edge Computing, *Mathematical Biosciences and Engineering*, Vol. 16, No. 4, pp. 1874-1892, March, 2019.

[11] J. Zhang, C. Yang, Q. Yang, Y. Lin, Y. Zhang, HGeoHashBase: an optimized storage model of spatial objects for location-based services, *Frontiers of Computer Science*, Vol. 14, No. 1, pp. 208-218, February, 2020.

[12] Y. Ren, Y. Leng, J. Qi, P. K. Sharma, J. Wang, Z. Almakhadmeh, A. Tolba, Multiple cloud storage mechanism based on blockchain in smart homes, *Future Generation Computer Systems*, Vol. 115, pp. 304-313, February, 2021.

[13] R. A. Mishra, A. Kalla, A Braeken, M Liyanage, Privacy Protected Blockchain Based Architecture and Implementation for Sharing of Students' Credentials, *Information Processing & Management*, Vol. 58, No. 3, Article No. 102512, May, 2021.

[14] M. A. Rahman, M. S. Abuludin, L. X. Yuan, M. S. Islam, A. T. Asyhari, EduChain: CIA-Compliant Blockchain for Intelligent Cyber Defense of Microservices in Education Industry 4.0, *IEEE Transactions on Industrial Informatics*, Vol. 18, No. 3, pp. 1930-1938, March, 2022.

[15] X. Li, L. Wei, L. Wang, Y. Ma, C. Zhang, M. Sohail, A Blockchain-based Privacy-preserving Authentication System for Ensuring Multimedia Content Integrity, *International Journal of Intelligent Systems*, Vol. 37, No. 5, pp. 3050-3071, May, 2022.

[16] S. M. Ali, H. Farouk, H. Sharaf, A Blockchain-Based Models for Student Information Systems, *Egyptian*

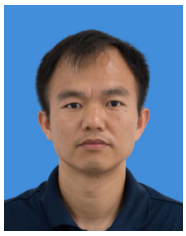
*Informatics Journal*, Vol. 23, No. 2, pp. 187-196, July, 2022.
[17] N. K. Dewangan, P. Chandrakar, S. Kumari, J. Rodrigues, Enhanced privacy-preserving in student certificate management in blockchain and interplanetary file system, *Multimedia Tools and Applications*, Vol. 82, No. 8, pp. 12595-12614, March, 2023.
[18] D. Li, D. Han, Z. Zheng, T.-H. Weng, H. Li, H. Liu, A. Castiglione, K.-C. Li, MOOCsChain: A blockchain-based secure storage and sharing scheme for MOOCs learning, *Computer Standards & Interfaces*, Vol. 81, Article No. 103597, April, 2022.
[19] G. Zhao, H. He, B. Di, J. Chu, StuChain: an efficient blockchain-based student e-portfolio platform integrating hybrid access control approach, *Multimedia Tools and Applications*, Vol. 83, No. 1, pp. 227-251, January, 2024, https://doi.org/10.1007/s11042-023-15560-1
[20] Z. Li, Z. Ma, A blockchain-based credible and secure education experience data management scheme supporting for searchable encryption, *China Communications*, Vol. 18, No. 6, pp. 172-183, June, 2021.
[21] J. Chang, J. Ni, J. Xiao, X. Dai, H. Jin, SynergyChain: A Multichain-Based Data-Sharing Framework With Hierarchical Access Control, *IEEE Internet of Things Journal*, Vol. 9, No. 16, pp. 14767-14778, August, 2022.
[22] S. Guo, F. Wang, N. Zhang, F. Qi, X. Qiu, Master-slave chain based trusted cross-domain authentication mechanism in IoT, *Journal of Network and Computer Applications*, Vol. 172, Article No. 102812, December, 2020.
[23] C. Feng, B. Liu, Z. Guo, K. Yu, Z. Qin, K.-K. R. Choo, Blockchain-Based Cross-Domain Authentication for Intelligent 5G-Enabled Internet of Drones, *IEEE Internet of Things Journal*, Vol. 9, No. 8, pp. 6224-6238, April, 2022.
[24] Y. He, C. Zhang, B. Wu, Y. Yang, K. Xiao, H. Li, Cross-chain Trusted Service Quality Computing Scheme For Multichain-Model-based 5G Network Slicing SLA, *IEEE Internet of Things Journal*, Vol. 10, No. 14, pp. 12126-12139, July, 2023.
[25] A. Xiong, G. Liu, Q. Zhu, A. Jing, S. W. Loke, A notary group-based cross-chain mechanism, *Digital Communications and Networks*, Vol. 8, No. 6, pp. 1059-1067, December, 2022.
[26] X. Huang, Y. Zhang, D. Li, L. Han, A Solution for Bi-layer Energy-Trading Management in Microgrids Using Multiblockchain, *IEEE Internet of Things Journal*, Vol. 9, No. 15, pp. 13886-13900, August, 2022.
[27] F. Bai, T. Shen, Z. Yu, K. Zeng, B. Gong, Trustworthy Blockchain-Empowered Collaborative Edge Computing-as-a-Service Scheduling and Data Sharing in the IIoE, *IEEE Internet of Things Journal*, Vol. 9, No. 16, pp. 14752-14766, August, 2022.
[28] X. Yu, Z. Shu, Q. Li, J. Huang, BC-BLPM: A multi-level security access control model based on blockchain technology, *China Communications*, Vol. 18, No. 2, pp. 110-135, February, 2021.
[29] Y. He, C. Zhang, B. Wu, Y. Yang, K. Xiao, H. Li, A Cross-Chain Trusted Reputation Scheme for a Shared Charging Platform Based on Blockchain, *IEEE Internet of Things Journal*, Vol. 9, No. 11, pp. 7989-8000, June, 2022.
[30] K. Hao, J. Xin, Z. Wang, Z. Yao, G. Wang, Efficient and Secure Data Sharing Scheme on Interoperable Blockchain Database, *IEEE Transactions Big Data*, Vol. 9, No. 4, pp. 1171-1185, August, 2023.
[31] E. Filatovas, M. Marcozzi, L. Mostarda, R. Paulavičius, A MCDM-based framework for blockchain consensus protocol selection, *Expert Systems with Applications*, Vol. 204, Article No. 117609, October, 2022.
[32] X. Tang, X. Lan, L. Li, Y. Zhang, Z. Han, Incentivizing Proof-of-Stake Blockchain for Secured Data Collection in UAV-Assisted IoT: A Multi-Agent Reinforcement Learning Approach, *IEEE Journal on Selected Areas in Communications*, Vol. 40, No. 12, pp. 3470-3484, December, 2022.
[33] D. Marijan, C. Lal, Blockchain verification and validation: Techniques, challenges, and research directions, *Computer Science Review*, Vol. 45, Article No. 100492, August, 2022.
[34] S. Khezr, A. Yassine, R. Benlamri, M. S. Hossain, An Edge Intelligent Blockchain-Based Reputation System for IIoT Data Ecosystem, *IEEE Transactions on Industrial Informatics*, Vol. 18, No. 11, pp. 8346-8355, November, 2022.
[35] J. L. Henning, SPEC CPU2006 benchmark descriptions, *ACM SIGARCH Computer Architecture News*, Vol. 34, No. 4, pp. 1-17, September, 2006.
[36] Xuper, XuperChain, https://xuper.baidu.com/n/ps/opensource
[37] K. Yu, L. Tan, C. Yang, K.-K. R. Choo, A. K. Bashir, J. J. P. C. Rodrigues, T. Sato, A Blockchain-Based Shamir's Threshold Cryptography Scheme for Data Protection in Industrial Internet of Things Settings, *IEEE Internet of Things Journal*, Vol. 9, No. 11, pp. 8154-8167, June, 2022.

## Biographies

**Xianglin Wu** was born in 1984. He is currently pursuing the PhD degree with the School of Computer and Information Security, Guilin University of Electronic Technology, China. His current research interests include blockchain and big data management.

**Tianhao Meng** was born in 1997. He received the B.S. degree from the Guilin University of Electronic Technology, China, in 2021. He is currently pursuing the M.S. degree with the Guilin University of Electronic Technology, China. His research interests include blockchain and big data management.

**Haotian Huang** received the B.S. degree from the Shandong University of Technology, China, in 2020. He obtained the M.S. degree from the Guilin University of Electronic Technology, China, in 2023. His main research interests include database and its query optimization.



**Lianhai Liu** received his PhD degree in Central South University, China in 2019. He is currently an Associate Professor with the School of Computer and Information Security, Guilin University of Electronic Technology, China. His main research interests include protocol design and its optimization, network security applications.

**Jingwei Zhang** received his PhD degree in East China Normal University, China in 2012. He is currently a Professor with the School of Computer and Information Security, Guilin University of Electronic Technology, China. His main research interests include big data management, analysis and smart applications, blockchain platform and its optimization.