Data Privacy Security Mechanism of Industrial Internet of Things Based on Block Chain

Yinggang Xie 1,2, Yuxin Li 1,* and Yunbin Ma 3

Key Laboratory of Information and Communication Systems, Ministry of Information Industry, Beijing Information Science and Technology University, Beijing 100101, China; xieyinggang@bistu.edu.cn
2 Key Laboratory of the Ministry of Education for Optoelectronic Measurement, Technology and Instrument, Beijing Information Science and Technology University, Beijing 100101, China
3 Intelligent Research Center, National Pipeline Network Group Research Institute, Beijing 100101, China; mayb01@pipechina.com.cn
* Correspondence: liyuxin@bistu.edu.cn; Tel.: +86-13691117939

Abstract: In order to solve the problem that data of the industrial Internet of Things (IIoT) is easily tampered with and therefore, the authenticity of data may be questioned, a data-privacy security mechanism of the IIoT based on blockchain is proposed. At the same time, to solve the problem of master node selection and low efficiency in the practical Byzantine fault tolerance (PBFT) algorithm, a reward mechanism based on node behavior is introduced and an improved PBFT algorithm is proposed. The improved PBFT algorithm is more efficient, in line with the application scenarios of the IIoT. Comparative analysis results showed that the proposed blockchain-based data-privacy security mechanism of the IIoT is superior to other models in terms of consensus efficiency, throughput, and block generation speed.

Keywords: IIoT; data privacy; blockchain; PBFT; security mechanism

1. Introduction

In recent years, the Internet of Things (IoT) has been widely used in industry, especially manufacturing, realizing intelligent industrial Internet of Things (IIoT) applications. It is estimated that by 2025, there will be 4.16 billion IoT devices worldwide, which are expected to generate 79.4 ZB of data [1]. How to manage and use IoT data to realize intelligence in an efficient, safe, and economical way has become an important research issue.

As a cutting-edge technology, the emergence of the IIoT enables more industrial entities to achieve interconnection. This has attracted widespread attention from many industries, such as industrial manufacturing, retail, logistics, environmental governance, energy, and medical care. In different industries, the IIoT can facilitate life and improve production efficiency by connecting different types of entities. The IIoT collects real-time data onto connected entities based on different requirements. The data can be used to describe the performance, attributes, and status of the entities to control and monitor different operations on the entities. In the IIoT, uninterrupted manufacturing production is bound to generate a large amount of data, which is collected by the sensors of thousands of entities connected to the IIoT. The obtained data is a precious resource, the security of which needs to be ensured in the process of transmission and sharing. It is therefore increasingly important to ensure data privacy and security in the IIoT [2]. At the same time, trust in the exchange of data that has arisen in accounting for the consumption of water, gas, oil, and metering readings in production systems is an essential part of the relationships in the supply chain.
Blockchain is regarded as a reasonable solution to the above problems. Blockchain can reduce the need for audits and inspections in data transmission in operational technology (OT) environments. Blockchain was originally designed to be applied to the Bitcoin system to provide secure and reliable transactions [3]. However, blockchain still has great challenges in terms of high throughput and scalability, which hinder its development into a general platform suitable for different application scenarios [4]. If the blockchain-based IIoT is used in different scenarios, there are high requirements for TPS (transactions per second). At present, there are different solutions for solving the blockchain scalability problems—off-chain payment networks, BitCoin-NG, and the sharding mechanism, etc. For off-chain payment networks, the typical solution is the Lightning Network, which solves the scalability problem by reducing on-chain redundancy. Bitcoin Cash handles scalability issues by adjusting the block size and block generation speed; BitCoin-NG [5] divides the block into KeyBlock for leader election and MicroBlock for recording transactions, shortening the time of mining and block confirmation. The sharding technology only allows some nodes in the network to process a certain transaction, the network is divided into many shards, and each shard can process different transactions at the same time, which improves the scalability. However, there are mutual constraints between the scalability, centralization, security, and latency of the blockchain. Therefore, when using blockchain as a means of data storage in the IIoT, the scalability cannot be improved blindly, and the balance between these four needs to be considered [6].

2. Background and Related Work

The rapid development of the IIoT has provided many opportunities for industries such as industrial manufacturing, agriculture, environmental monitoring, and safety supervision. [7]. The use of the blockchain as a solution to realize data security in the industrial Internet of Things is that due to the characteristics of distributed connection of the blockchain, a huge P2P transmission network can be formed. By restricting resource utilization and peer-to-peer connections, the IIoT networking architecture can be optimized and decentralized storage can ensure information security and user privacy [8]. Although blockchain has significant benefits, it is difficult for the traditional blockchain to provide better scalability to meet the high transaction throughput requirements of the IIoT. For example, Bitcoin can only confirm 3-4 transactions per second on average, although Ethereum has increased the throughput to an average of 14 transactions per second, but this still cannot meet the transaction processing needs of the IIoT [9].

Lu et al. [10] designed a secure data-sharing architecture with blockchain authorization for distributed multi-parties. The shortcoming of this scheme is that the data demand needs to be broadcast to all blockchain nodes in the whole network, data needs to be collected from various parties, and the process of federated learning needs to be implemented. This consumes a large amount of computing power and it takes a long time to retrieve and train models, affecting the efficiency of data sharing. Chen et al. [11] constructed a data-sharing system composed of a two-layer blockchain. This solution needs to broadcast data demands to the whole network in the blockchain. Excessive broadcast demands may lead to network congestion and demand cannot be responded to in a timely manner. Although the above work is some blockchain-based data-sharing-related research work, it only considers the security of data sharing, not the efficiency of data sharing. This cannot shorten the time of data sharing or increase the value density of shared data.

Based on the above information and with a focus on data security issues in the industrial Internet of Things and the need to ensure data authenticity, this paper proposes a data-privacy security mechanism of the industrial Internet of Things based on blockchain, with the main contributions as follows:
(1) A data-privacy security mechanism for the industrial Internet of Things based on blockchain with a confidential and accurate data-sharing process that takes advantage of the untamability of data stored on the blockchain and the automatic execution of smart contracts is proposed.

(2) An improved algorithm of the PBFT mechanism is proposed, with an added reward mechanism based on node behavior and improved the design of the algorithm process to improve the efficiency of the algorithm.

3. Blockchain-Based IIoT Data Security Sharing Mechanism

In industrial blockchain networks, the existing data-information sharing generally adopts two basic information-sharing architectures: one is the centralized architecture, in which the central organization can realize information interaction as an agent [12], and the other is the peer-to-peer architecture, in which the participants in the network share information [13]. In the distributed network environment, peer-to-peer information-sharing architecture is mostly adopted, but the architecture has the following problems:

(1) The participants in the network have trust problems with each other;
(2) The information shared by participants may be obtained by attackers or other competitors;
(3) There is no fair reward mechanism to ensure that all participants actively share information.

To address the problems existing in the peer-to-peer information-sharing architectures, Vakilinia et al. studied joint approaches and potential privacy challenges for cybersecurity information sharing [14]. The differential privacy technology based on cyber-physical systems (CPS) proposed in [15] solves the privacy and security of information shared by participants. Combining agent-based encryption and attribute-based encryption (BloCyNfo-Share), Badsha et al. proposed a blockchain-based privacy-preserving network security information-sharing scheme. In this scheme, the current organization entrusts an organization to obtain access to the network security information of the entrusted organization, and uses the advantages of blockchain to realize fine-grained access control and apply it in the medical information-sharing scenario [16]. Ref. [17] pointed out that blockchain-enabled information sharing can enhance the collaborative work of different types of supply chains. It enhances collaborative partnerships by ensuring that validated information is available to all members of the chain.

In summary, in the solution of information sharing between participants in the industrial blockchain network, the peer-to-peer information-sharing architecture can be better applied to the scenario of collaborative work of participants in the network after a series of privacy-protection measures. With the emergence of blockchain-based information-sharing schemes, the peer-to-peer information sharing architecture reached a new level [18].

This project combines the untamability of data stored on blockchain and the automatic execution of smart contracts to achieve a confidential and accurate data-sharing process, providing solutions to the problem of data isolation while protecting data security, as shown in Figure 1. By taking advantage of the honesty and reliability of the blockchain, index and verification information of data are recorded through the blockchain network to ensure the confidentiality and non-tampering of stored data.
3.1. Blockchain-Based IIoT Data-Privacy Security Framework

This paper proposes a data-privacy security framework for the industrial Internet of Things based on blockchain, as shown in Figure 2. The Internet of Things is divided into a perception layer, transmission layer, and application layer [19]. In this framework, source data collected by nodes are uploaded to the industrial blockchain network, and technologies such as data blocks, Merkle trees, and timestamps are used to achieve the data security of the industrial control system.

In order to improve the security of the industrial control network, all levels of servers, switches, PLC, and even the underlying equipment in the industrial control network can be connected to the blockchain as participating nodes. Each device node contains the entire database transaction information [20], called a block and a ledger. All device nodes of industrial control system (ICS) form the blockchain and distributed ledger. In order to help the industrial control network and the blockchain reach consensus, the respective blockchains need to establish each device node transaction and industrial control network connection, and they should follow certain rules. These rules are programmed into each blockchain’s client nodes, which are then used to decide whether an incoming transaction is valid and whether it should be relayed into the network.

When blockchain is applied to an ICS network consisting of field-bus connected device nodes, each block contains a Hash sequence and a timestamp combined key [21], forming “blocks” linked to the next block, which also contains ICS transaction data between devices in a network. Multiple “blocks” form a linear sequence. Since “blocks” are identified by hashed sequences of ciphers, and each block also references information from the previous block, it is possible to check whether the sender and receiver’s data has been validated and tampered with through these blocks in the blockchain. When users
trade data in the ICS network, they only need to trade between device nodes because each device node executes a smart contract and records transaction data, and the blockchain replicates itself between the device nodes of the ICS, further demonstrating that any device node in the ICS network can record all transaction data. As a result, device nodes on the ICS network are attached to the blockchain, and only authenticated, mutually agreed transactions are written to the underlying storage of the blockchain, ultimately ensuring the security of the data exchange.

3.2. The Information-Sharing Model Ensures the Consistency of Data among Edge Nodes

The proof-of-work (PoW) algorithm is used by the information-sharing model to ensure the consistency of data among edge nodes. It is also called the mining process. The mining node tries different Nonce values (starting at 0 and incrementing by 1 each time) to find a Hash value that is less than the specified mining difficulty.

As shown in Figure 3, the data-consistency process among edge nodes in the blockchain is as follows:

Step 1: Determine whether new block header assembly is completed in the blockchain. If so, perform Step 2.1; otherwise, perform Step 2.2.

Step 2.1: Double hash the block header SHA256 (SHA256(block header)) to get the hash value $H$, go to Step 3;

Step 2.2: Perform Step 1 after a new block header is generated.

Step 3: Check whether $H$ is less than the network target value. If $H$ is less than the network target value, the block broadcast succeeds (verified by other edge nodes). If $H$ is greater than or equal to the network target value, go to Step 4.

Step 4: change the value of the Nonce and perform Step 2.1.

![Figure 3. Flow chart of data consistency between edge nodes.](image)

4. Improvement of PBFT Algorithm

Since the advent of Bitcoin, blockchain has gradually become a research hotspot in the academic world. As a key technology of blockchain, the consensus algorithm has attracted more and more researchers’ attention. Due to the complex and changeable operating environment of blockchain, it is easy to introduce Byzantine nodes into the system, so the Byzantine fault-tolerant consensus algorithm of blockchain must be overcome.
The common consensus mechanisms in blockchain include proof-of-work (PoW), proof-of-stake (PoS), and delegated proof-of-stake (DPoS). The practical Byzantine fault-tolerant (PBFT), and most of the other consensus mechanisms are derived from these four mechanisms. Table 1 lists the communication overheads, computing overheads, fault tolerances, throughputs, response times, and application platforms of the four algorithms. Compared with PoW and other proof algorithms, the PBFT algorithm has a throughput of thousands of TPS and a response time of seconds, and is considered to be a consensus algorithm suitable for the IoT. However, in a network with N nodes, the PBFT algorithm needs N nodes to broadcast messages twice to the whole network to complete a round of consensus. When the number of nodes in the system increases, the traffic between nodes increases sharply, which brings great pressure on network bandwidth and leads to rapid degradation of system performance. Therefore, the PBFT algorithm is not suitable for a large-scale network environment.

Table 1. Consensus algorithm complexity analysis.

<table>
<thead>
<tr>
<th>Consensus</th>
<th>Communication Overhead</th>
<th>Computing Overhead</th>
<th>Fault Tolerance</th>
<th>Throughput</th>
<th>Response Time</th>
<th>Application Platform</th>
</tr>
</thead>
<tbody>
<tr>
<td>PoW</td>
<td>Low</td>
<td>High</td>
<td>1/2</td>
<td>≥7 TPS</td>
<td>10 min</td>
<td>Bitcoin</td>
</tr>
<tr>
<td>PoS</td>
<td>Low</td>
<td>Medium</td>
<td>1/2</td>
<td>≥25 TPS</td>
<td>1 min</td>
<td>Peercoin</td>
</tr>
<tr>
<td>DPoS</td>
<td>Low</td>
<td>Low</td>
<td>1/2</td>
<td>≥300 TPS</td>
<td>3 s</td>
<td>EOS</td>
</tr>
<tr>
<td>PBFT</td>
<td>High</td>
<td>Low</td>
<td>1/3</td>
<td>≥1000 TPS</td>
<td>Second level</td>
<td>Hyperledger</td>
</tr>
</tbody>
</table>

The consensus mechanism is the core technology of blockchain. It determines whether a new block is verified and who keeps records, affecting the security and reliability of the whole system [22]. The practical Byzantine fault tolerance (PBFT) algorithm is adopted as the consensus algorithm in this paper, but the PBFT algorithm has the following problems:

1. Main node selection problem. Each node of the PBFT algorithm has the same probability of becoming the main node, and they take turns to become the main node and undertake the block task. Selecting the main node in the above way will make some nodes with low performance or poor performance in the consensus process become the main node, which will affect the efficiency of block generation in the system and reduce the system performance.

2. Traffic problems. During the PBFT algorithm’s conformance protocol preparation and submission phases, each slave node broadcasts information to all nodes, resulting in O (n²) traffic in the consensus process, as shown in Figure 4. As a result, the number of node communications increases power with the increase of node number, which leads to the decrease of consensus efficiency and seriously affects system scalability.

![PBFT algorithm flow](image-url)

**Figure 4. PBFT algorithm flow.**

1. Request: C sends a request to primary node 0.
2. Pre-prepare: Master node 0 sorts the request numbers and sends the request numbers to nodes 1, 2, and 3.
3. Prepare: Node 1 and 2 send messages to each other and to master node 0.
4. Commit: Confirm the assignment number of primary node 0 among nodes 0, 1, and 2.
5. Reply: Node 0, 1, 2 Confirm the message and reply to C.
In view of the above problems, ref. [23] improved the PBFT algorithm in terms of consensus strategy, and ensured the coordination and security of the algorithm by dividing nodes and clusters. In terms of node selection, ref. [24] improved the PBFT algorithm by dividing multiple node sets according to node responsibilities, making it suitable for dynamic blockchain with a constantly changing number of nodes. In terms of method innovation, ref. [25] improved the PBFT algorithm by using K-medoids to perform clustering and hierarchical division of nodes, making it suitable for the consensus process involving large-scale consensus nodes. Ref. [26] improved the PBFT algorithm in terms of block structure—by designing multiple main nodes, the problem of high delay when malicious nodes act as main nodes is reduced to a certain extent. Zhong et al. proposed the performance-improved Silkworm [27] algorithm based on proof-of-stake (PoS), which ensured the security and robustness of blockchain. Huang et al. proposed the RBFT [28] consensus algorithm, combined with the improved Raft mechanism and elected leaders to form a committee for PBFT consensus. All of the above improved the PBFT algorithm from different perspectives in order to reduce algorithm complexity and traffic and improve consensus efficiency. However, the node selection algorithm was inserted in the improved algorithm, increasing the algorithm complexity to a certain extent and having the potential risk of reducing consensus efficiency.

4.1. RPBFT Algorithm

Currently, consensus algorithms commonly used in blockchain are proposed to solve the problem of consistency in distributed systems. However, these algorithms take a long time to calculate, consume large amount of resources, and are not suitable for the lightweight Internet of Things. Thus, combined with the application scenario of IoT data-sharing, this paper improves the PBFT algorithm and proposes the RPBFT algorithm, which improves the PBFT algorithm in the following ways:

1. A reward mechanism based on node behavior is added. Rewards or punishments will be reflected by node scores. According to the scores, the RPBFT algorithm selects the more trustworthy nodes as main nodes and consensus nodes to improve the efficiency of the algorithm.

2. In the PBFT algorithm, when the total number of nodes is 3f + 1, no more than f Byzantine nodes can be accommodated [29]. In the RPBFT algorithm, the serial number of the data set of each round of consensus is denoted as S, starting from 0, and a main node is designated for each round of consensus. If consensus cannot be reached, the set S increases until the consensus is reached. The time interval for consensus is t. Once a new block is generated, a new round of consensus is started, and the serial number of the collated set is S = 0. The main process of consensus algorithm is shown in Figure 5. In the figure, n represents the number of nodes participating in the consensus in each round.
4.2. Reward Mechanism Based on Node Behavior

(1) Calculation of node reputation score
In the initial stage, the credit score of each node is set as 100, and the score is based on the behavior change of the node in the consensus process. After completing a round of consensus, the score of each node will be added or subtracted according to its performance in this round of consensus. Bonus points are awarded for nodes that successfully participate in this round of consensus and points are deducted for nodes that do evil or fail in this round of consensus. In terms of the severity of punishment for node behavior, the scheme allows for occasional node failure, that is, when the node fails, the points are reduced less. However, it is absolutely not allowed to do evil to the node. If the node does evil, it will be severely punished, and most of the points of the node will be deducted, so that it is difficult to participate in the consensus process in a short time, putting an end to its continuous evil situation. The correlation coefficients of rewards and punishments in this scheme are set according to the index level of 2, and the score calculation is shown in the following Formula (1).

$$Score_i = Score_{i-1} + 2S_i - 2^3F_i - 2^6E_i, i = 1, 2, 3 ...$$

(1)

where Score represents the current score of the node, S represents whether the node successfully participates in the consensus of this round (if so, S = 1; otherwise, S = 0), F represents whether node failure occurs in this round consensus (if so, F = 1; otherwise, F = 0), and E represents whether the node is guilty in the consensus of this round (if so, E = 1; otherwise, E = 0).
Table 2 below shows the score of the Kth node participating in the consensus of this round. The default score of the previous round is 100.

Table 2. Node scores.

<table>
<thead>
<tr>
<th>Kth</th>
<th>S</th>
<th>F</th>
<th>E</th>
<th>Score</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>102</td>
<td>Node successfully participated</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>92</td>
<td>Node failure</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>36</td>
<td>Node evil</td>
</tr>
</tbody>
</table>

(2) Main node and consensus node selection scheme based on reputation score
The main node is selected. In this paper, the method of selecting the main node is based on the reputation score. The node with the highest score is selected as the main node. When the main node fails or does evil, the node with the second score is selected as the main node according to the view-change protocol, and a new round of consensus is started.

Select a consensus node. The type of blockchain selected in this paper is alliance chain, and the participating nodes have passed the authentication, which strongly ensures the honesty of the nodes. The number of scores is used as the basis of node reliability ranking to further ensure the reliability of nodes. Therefore, in order to improve the efficiency of consensus, some nodes with high scores are selected as consensus nodes in this scheme, and the consensus results of other nodes are synchronized after the conclusion of consensus.

5. Experimental Results and Analysis
This chapter tests the transactions per second (TPS) and delay of the data ledger of the blockchain of the industrial Internet of Things to verify its effectiveness and feasibility. The simulation system uses Java language to simulate the 1 data generation process and 9 consensus execution process in a single machine environment. Operating environment: AMD Ryzen 7 5800H 3.20 GHz CPU, 16 GB memory, CentOS 7 operating system, JDK 1.8.0.

(1) Throughput. TPS in the data blockchain refers to the total number of transactions that the node collects data and uploads. It sends the data summary request to the consensus confirmation and writes the ledger divided by the time. Different block generation times are taken and the test repeated 10 times in each time period; 10 times of data is shown in Figure 6, and the average value of the 10 times is taken as the TPS of the time period. The test results are shown in Figure 7. As can be seen from the figure, the transaction throughput of data blockchain is about 10,000 times/s, which can deal with most scenarios in the real industrial Internet of Things.

Figure 6. Transaction throughput of different blocks.
In order to compare the differences between the RPBFT and PBFT algorithms in TPS, the following experiments were carried out in this paper: with fixed four consensus nodes, 1000, 1500, 2000, 2500, and 3000 transactions were sent, respectively, in the same time interval. The experimental results are shown in Figure 8. The TPS of the RPBFT algorithm was significantly higher than that of the PBFT algorithm. When the transaction volume was 2500, the TPS of the RPBFT algorithm increased by about 210% compared with that of the PBFT algorithm. When the transaction volume is too large and exceeds the system’s processing capacity, threads will be blocked and TPS will decline. However, on the whole, TPS of the RPBFT algorithm is still higher than that of the PBFT algorithm.

(2) Time delay. The delay of blockchain is the time interval between request occurrence and ledger confirmation, which consists of request broadcast transmission time, consensus algorithm execution time, and broadcast confirmation time. According to the generation time of the five blocks mentioned above, the average value of all delays is counted, and the relationship of account delay under different generation times of blocks is shown in Figure 9. In this diagram, the block generation time corresponding to the maximum TPS time delay is the ms level, which can be accepted by most IIoT application scenarios.
In order to compare the difference between the RPBFT and PBFT algorithms in consensus delay, under the condition of four fixed consensus nodes and the same time interval, the system sent comparative experiments of 200, 400, 600, 800, and 1000 transactions, respectively. The experimental results are shown in Figure 10. In the experiment, when the transaction volume increased, the transaction delay of the PBFT algorithm grew rapidly, while the RPBFT algorithm grew slowly. When the transaction volume was 1000, the transaction delay decreased by about 50%. Therefore, when the transaction volume increases, the advantages of the RPBFT algorithm are more obvious and the delay is lower.

6. Conclusions

The decentralized and tamper-proof characteristics of blockchain provide a new solution to the security problems and data authenticity in the data sharing of the industrial Internet of Things. This paper uses reasonable IIoT blockchain network data-privacy security mechanism to guarantee the security of data exchange, improving information security, reducing running cost, and solving the information island. The advantages of the integrated IIoT and blockchain—reducing cost and increasing efficiency—were achieved by information sharing. At the same time, the practical Byzantine fault-tolerant algorithm was improved to improve the efficiency of the algorithm for IIoT data-sharing application scenarios. At present, the development of blockchain only provides some of the application scenarios for the IIoT. In follow-up research, more mature blockchain application scenarios should be developed, with a focus on feasible and practical schemes for application of blockchain in the IIoT.
Author Contributions: Data curation, Y.L.; Formal analysis, Y.M.; Methodology, Y.X. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by Beijing Natural Science Foundation (Grant No. 4192023 and 4202024), Research on the Theory and Technical System of Smart Pipe Network Project (Grant No. JCGL202109), and the Gold-Bridge Funds for Beijing (Grant No. 2021bjjq0017).

Acknowledgments: This work is supported by Beijing Natural Science Foundation (Grant No. 4192023 and 4202024), Research on the Theory and Technical System of Smart Pipe Network Project (Grant No. JCGL202109), and the Gold-Bridge Funds for Beijing (Grant No. 2021bjjq0017).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IoT Internet of Things
IIoT Industrial Internet of Things
TPS Transaction per second
CPS Cyber–physical systems
ICS Industrial control system
PoW Proof-of-work
PBFT Practical Byzantine fault tolerance
OT Operational technology

References

1. IDC. IoT Growth Demands Rethink of Long-Term Storage Strategies, Says IDC. Available online: https://www.idc.com/getdoc.jsp?containerId=prAP46737220 (accessed on 28 July 2020).


