Blockchain-based Internet of Things Information Management：Data Sharing and Consensus Algorithm Innovation

Zixi He[1, *], Yefan Wang[2] and Shengrong Li[3]

hzx812404929@163.com[1], yefan_wang_8725@163.com[2], 1810685513@qq.com[3]

Jiangsu University of Science and Technology, Zhenjiang, Jiangsu, China

Abstract. The Internet of Things (IoT) has been rapidly growing, connecting an ever-increasing number of devices and generating massive amounts of data. However, centralized IoT infrastructures face challenges such as single points of failure, privacy concerns, and scalability issues. This paper explores the application of blockchain technology in IoT data sharing and consensus algorithms to address these challenges and explore innovations. Simulation experiments and theoretical analyses demonstrate the effectiveness of the proposed innovations in enhancing blockchain performance, enabling secure data sharing, and incentivizing miner participation, providing new insights for the application of blockchain in IoT information management.

Keywords: Internet of Things, blockchain, data sharing, federated learning, consensus algorithm, coalition game, information management

1 Introduction

The Internet of things (IoT) is a network paradigm that connects objects in the real world to the network. IoT allows devices to collect, process and communicate data without intervention [1]. With the development of IoT, the number of devices connected to the network is growing rapidly. According to the Ericsson's prediction, by 2025, more than 24.9 billion devices will be connected to the network [2]. The exponential explosion of IoT devices has led to an ever-increasing amount of data. On account of this, the magnitude and ability of human society to generate, obtain and process data will usher in a new leap.

Most existing IoT infrastructures are highly dependent on centralized platforms, while centralized IoT networks will face the following challenges: (1) data stored in centralized servers may reveal personal privacy. (2) data stored in a centralized cloud lacks reliability and traceability, is at risk of being deleted or tampered with. (3) with the exponential growth of device number in IoT, centralized servers will lead to large delays when processing a large number of end-to-end communications, which will.

Recently, blockchain technology has been regarded as the main candidate technology for IoT decentralization [3]. In recent years, blockchain technology has received broadly attention jue to its decentralization, immutability and traceability. The data stored on the blockchain needs to be jointly maintained by the whole network, which can effectively transfer value between nodes that lack trust [4]. Moreover, through blockchain technology, applications that used to be able

ICBBEM 2024, March 29-31, Wuhan, People's Republic of China

DOI 10.4108/eai.29-3-2024.2347474

to run only through trusted third-party platforms can now run in a distributed way [5], enabling the market to become a highly decentralized autonomous marketplace.

Numbers of related works form the support and provide a great deal of inspiration for this paper.

(1) Research on the classification and consensus algorithm of blockchain

Blockchain technology can be divided into public chain, consortium blockchain and private chain according to node participation, and can be divided into licensed chain and unlicensed chain according to different permissions [6]. In [7], aiming at the shortcomings of dispersed privilege management, difficult business expansion, centralized reading and writing, and weak transactionality among multiple contracts in complex application scenarios, a multi-blockchain collaboration scheme of consortium chain for complex application scenarios is proposed.

Consensus algorithms are crucial in blockchain technology, by using these algorithms, consortium blockchains are able to reduce communication costs while improving privacy and maintaining data immutability [8]. However, traditional algorithms (i.e. PBFT) have problems such as high complexity and poor scalability. Hence, BFT algorithms have been proposed, such as the high-performance and scalable BFT algorithm proposed in [9] and the component-layer algorithm based on PBFT proposed in [10].

(2) Research on data sharing methods based on blockchain technology

Data sharing in the IoT domain faces unique challenges, among which blockchain-based data sharing methods are strong candidates to address these challenges due to their decentralized, tamper-proof, and traceability features. Smart contracts, especially on the Ethernet platform, offer new possibilities for automated data processing and transactions between IoT devices [11]. They allow IoT big data to circulate without relying on centralized cloud platforms, thus enhancing the potential value of the data. Literature [12] developed a blockchain-based big data sharing framework, designed a low computational complexity collaborative proof-based consensus mechanism, and a filtering and offloading scheme for blockchain transactions to significantly reduce the storage overhead.

(3) Research on distributed secure data sharing based on blockchain and federated learning

As an innovative approach in the field of machine learning, federated learning allows multiple nodes to train models together without sharing the original data, which protects the data privacy of the participating nodes [13]. When federated learning is combined with blockchain technology, it enables distributed and secure sharing of data. In [14], authors design a blockchain-authorized secure data sharing architecture that transforms the data sharing problem into a machine learning problem by merging privacy-preserving federated learning. Further studies verified the feasibility and practical application benefits of implementing the federated learning architecture on blockchain for data sharing in decentralized environments [15].

2 Structure and logic

This paper has made innovative achievements in three aspects: blockchain performance evaluation, distributed data sharing, and the incentive of miner data collection. The simulation results of multi-blockchain performance, the data sharing architecture combining federated learning

and reputation mechanism, and the blockchain consensus mechanism based on coalition formation game provide new ideas for the performance improvement of the blockchain system, data sharing and the innovation of consensus algorithm. The logical architecture of this paper is shown in Figure 1 below:

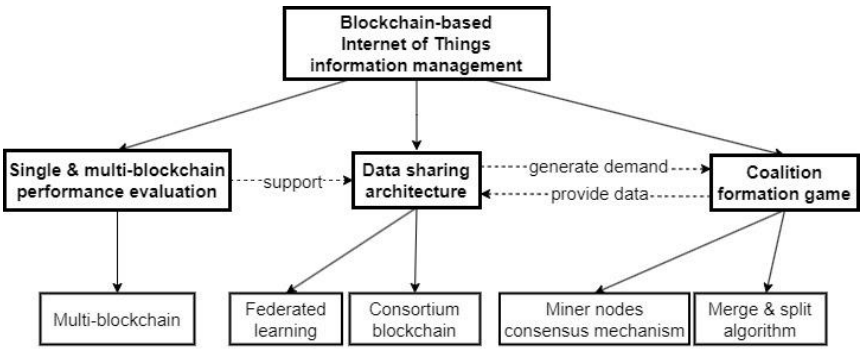

Fig.1. Article logical architecture.

3 Implementation and simulation analysis

3.1 Single & multi-blockchain Performance Simulation and Evaluation

The design of blockchain can adopt single blockchain or multi-blockchain architecture. Single blockchain architecture has only one main chain, and all nodes jointly maintain the same ledger. The multi-blockchain architecture consists of multiple parallel subchains, each of which can have its own independent ledger and consensus mechanism. As the number of nodes and transaction volume increase, the performance of a single blockchain will be limited. In contrast, the multi-blockchain architecture has better scalability. It allows each application to be assigned a separate subchain based on different business requirements. Each subchain can process transactions in parallel and then interact with each other through cross-chain communication protocols.

The simulation leverages the Fisco Bcos blockchain framework, an open-source blockchain platform tailored for enterprise-grade applications. It is designed to provide a consortium blockchain ecosystem with high-performance transaction processing capabilities, which can handle up to thousands of transactions per second, as well as enhanced privacy features through its support of group signature and zero-knowledge proof technologies. The system is configured on a physical computer with specifications including a 2.3GHz Intel i7 CPU, 8GB RAM, and the Ubuntu 18.04 operating system. Within this environment, both single blockchain and multi-blockchain architectures are set up for comparative analysis. The single blockchain is comprised of four nodes, whereas the multi-blockchain architecture includes three parallel blockchains, each with a quartet of nodes, enabling parallel transaction processing. The Fisco Bcos platform utilizes a modified version of the Practical Byzantine Fault Tolerance (PBFT) consensus algorithm, which is well-suited for consortium blockchains where node identities are known and a high transaction throughput is required. Furthermore, the performance metrics, including transaction throughput and average transaction latency of both architectures, are rigorously evaluated using Hyperledger Caliper, a benchmarking tool that provides a standard for assessing blockchain performance across different blockchain systems. In this simulation, the transaction

throughput is defined as the rate at which valid transaction are committed by the blockchain system under test during unit time. The average latency is defined as the average time taken for a transaction's effect to be usable across the network. The average latency contains the time from the point that it is submitted by the transaction node to the point that the result is widely available in the network.

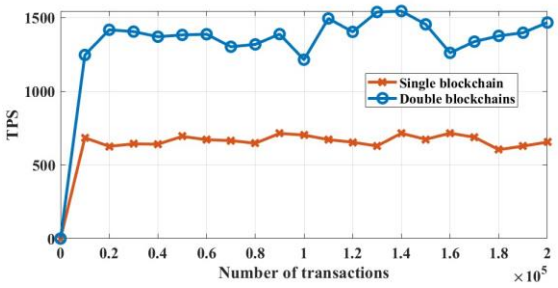

Fig. 2. Throughput as function of number of transactions.

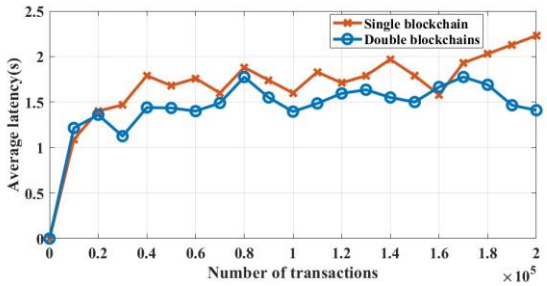

Fig. 3. Average latency as function of number of transactions.

Figure. 2 shows the throughput as function of the number of transactions considering the single blockchain and multi-blockchain. As the number of transactions increases, the throughput of the multi-blockchain system continues increasing, reaching approximately 125 million TPS at 10,000 transactions, meanwhile, single blockchain only reached 70 million TPS, represents a 79% improvement in the transaction processing capacity of the multi-blockchain structure. In the process of increasing the volume of multi-blockchain transactions, despite fluctuations, the TPS highly remained between 125 million and 150 million. In contrast, although the throughput of a single blockchain system also increased initially, it basically stabilized at about 70 million TPS, demonstrating its limitations. Comparing the two architectures, the throughput of the multi-blockchain is basically maintained at 2 to 3 times that of the single blockchain, highlighting its superior scalability when handling larger volumes of transactions.

Figure. 3 depicts the average latency as function of the number of transactions. From Figure. 3, the multi-blockchain consistently maintains lower latency than the blockchain across the entire range of transactions tested. For instance, when transaction volume reaches 200,000, the single blockchain exhibit an average latency of 2.25 seconds, whereas the multi-blockchain system maintains a lower latency of about 1.5 seconds. The increase in latency is modest for the multi-

blockchain system, even as the transaction volume grows, which accentuates its efficiency in managing high transaction loads.

The detailed inspection of two figures above provides empirical evidence of the multi-blockchain architecture's aptitude for sustaining high performance metrics. The multi-blockchain not only surpasses the single blockchain in throughput by a substantial margin but also maintains lower transaction latencies, thereby offering a compelling solution for systems requiring robust scalability and efficiency.

3.2 Data Sharing Architecture

When it comes to data sharing between IoT devices, the data owner may not want to share the raw data with other devices, because malicious devices can spread the data collected by other devices to the network without permission. Federated learning is a burgeoning machine learning scheme, aiming at tackling the problem of data island while preserving data privacy., in which nodes can train models locally based on the original data, and only need to share parameters of the local model without sharing the original data during data sharing. The introduction of federated learning into data sharing in the IoT allows smart devices in the IoT to share data without compromising data privacy. When two data sets have more overlapping user features and less overlapping users, they will be divided horizontally (i.e. the user dimension), and part of data with the same user features but not exactly the same users is extracted for training.

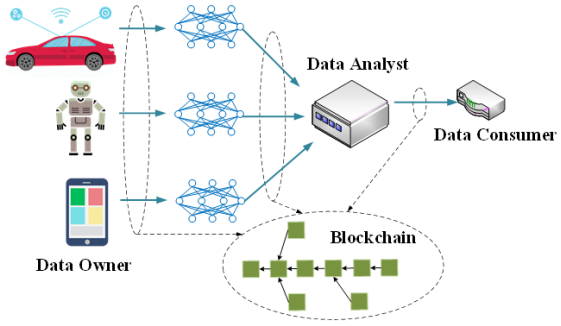

Fig. 4. A data sharing system based on horizontal federated learning and consortium blockchain.

This study constructs a data sharing architecture based on horizontal federated learning and federated chaining. The system model is shown in Figure. 4. The proposed data sharing system framework includes three entities: data owner (DO), data analyst (DA), and data consumer (DC). When DC has a request for data analysis service, DC will make a request for data analysis service to DA. DA will purchase data information from DO based on DC's request for data analysis service. For security reasons, DOs do not want to disclose their raw data. DOs will respond based on DA's purchase information, for example, whether they are willing to participate in this global training or not, and then DA selects some reasonable DOs from the DOs who are willing to share data (the selection can be based on the reputation value). Based on horizontal federated learning, the steps of data sharing between the selected DOs and DA are shown below:

(1) DOs selected by DA will compute the training gradient locally, mask the gradient selection with encryption, differential privacy or secret sharing techniques, and send final result to DA.

(2) DA performs secure aggregation without any knowledge of any DOs' data. The DA then sends the aggregated results back to the DOs.

(3) DOs update their respective models with the decrypted gradients.

Then iterate through the above steps until the loss function converges, completing the whole training process. After DA obtains the final model parameters, it sends the model parameters to DCs, completing a data analysis service transaction. DA pays the corresponding fees based on the contributions of DOs, and at the same time, DA charges DCs for the data analysis service. Transactions between DOs and DAs and between DAs and DCs are recorded on the blockchain.

The consortium blockchain network has received widespread attention for its advantages such as low cost, good scalability and short latency. In this paper, consortium blockchain is used, and all entities involved in the transaction must be licensed. DAs act as consensus nodes to package the transaction and reach a consensus across the network, and the consensus nodes that add blocks at the same time are rewarded.

In addition, previous simulation experiments on multi-blockchain architecture have demonstrated its significant advantages in improving transaction throughput and reducing latency. Notably, multi-blockchain architecture also offers benefits for data privacy and security by allowing the configuration of flexible data access rights and privacy protection schemes for different subchains. This approach not only meets the reasonable data sharing needs within the consortium blockchain but also strictly limits the access scope of sensitive data, achieving a balance between privacy protection and business collaboration. Furthermore, multi-blockchain architecture ensures that the failure of a single subchain will not affect the operation of the entire consortium blockchain, thus enhancing the availability and robustness of the blockchain network. Considering factors such as inter-chain isolation, flexible privacy protection mechanisms, and distributed fault-tolerance capabilities, multi-blockchain architecture collaborative federated learning is likely to be a preferred option for empowering consortium blockchains to achieve further optimization of data security and privacy protection.

3.3 Coalition Formation Game Formulation

In the blockchain network, miner nodes maintain the normal operation of the blockchain by participating in the consensus mechanism, and have the opportunity to obtain a certain consensus reward in return. However, as the scale of the blockchain network continues to expand, the difficulty for miner nodes to obtain consensus rewards is also continuously growing. In order to increase the probability of obtaining rewards, miner nodes can provide more valuable data for the blockchain network by actively participating in the collection of sensory data, so as to enhance their contribution. At the same time, miner nodes can also form a coalition with other miner nodes to share the sensing tasks, so as to obtain economies of scale and further increase the probability of winning the award. Aiming at the above problems, this paper proposes a blockchain consensus mechanism based on coalition formation game. In addition, a merge-and-split algorithm is proposed.

In the considered blockchain network, there exist M miner nodes and the set of miners is denoted as $\mathbb{M} = \{1, 2, ..., m, ..., M\}$. The cooperation of miners choosing reliable workers is modelled as a non-transferable utility (NTU) coalition formation game, denoted as $G = \{M, \Pi, u\}$, where Π is a coalition partition of M and u is a utility function. Each miner selects several workers by their reputations, The number of workers selected by miner m is denoted as $|\mathbb{W}_m|$ the set of

workers selected by miner m is denoted as $\mathbb{W}_m = \{W_{m,1}, W_{m,2}, ..., W_{m,w}, ...W_{m,|\mathbb{W}m|}\}$, where $W_{m,w}$ is the identity number of the w-th worker selected by miner m.

For miner m, the utility function is established as equation (1),

$$um = Pm \times R - Cm - \xi m \times Tm \quad (1)$$

Pm denotes the probability of miner m obtaining the consensus reward, R denotes the total amount of rewards given by the blockchain network, Cm denotes the communication overhead of miner m in forming the coalition, ξm denotes the unit power overhead of miner m in executing the perceptual task, and Tm denotes the time of miner m in performing the perceptual task. The utility of miner m consists of three parts: consensus reward gain, coalition formation overhead and execution of perceptual task overhead. By joining the coalition, miners can share the sensing task, reduce Tm, and increase the award probability Pm, but the formation of the coalition itself will also bring the communication overhead Cm. Therefore, how to balance these three factors, and build the optimal coalition structure, becomes a challenge faced by each miner node.

The rules of merge and split are given as follows:

Merge rules: For any two coalitions, if the new coalition formed by merging them can increase the utility of at least one miner without reducing the utility of other miners, perform merge, following the Pareto order. Specifically, for two coalitions G1 and G2, if the merged coalition G0 satisfies: for at least one miner o' in G0, $uG0(o) > \max\{uG1(o), uG2(o)\}$, and for all other miners o', $uG0(o') \geq \max\{uG1(o'), uG2(o')\}$, then merge is performed.

Split rules: For any coalition, check whether there exists a split such that the new coalition combination Pareto after the split is better than the original coalition. If such a split exists, perform a split. Specifically, for a coalition G0, if there exists a division of G0 {G1, G2, ..., Gk} such that for at least one coalition Gi in the division, the utility uGi(o) of at least one miner o in Gi is higher than uG0(o), and the utility uGi(o') of all other miners o' in Gi is not lower than uG0(o'), then the split is performed.

Algorithm: Coalition formation algorithm for miners in the proposed model:

Input: Player set $\mathbb{M} = \{1, 2, ..., m, ..., M\}$. Workers selected by miners $\mathbb{W}_m = \{W_{m,1}, W_{m,2}, ..., W_{m,|\mathbb{W}m|}\}$, $1 \leq m \leq M$;
Output: The coalition with the highest coalition utility;
1: Initialization: The initial partition of miners Π0, where all the miners are disjoint. Each miner selects several workers;
2: Each coalition computes coalition utility according to utility function (1);
3: Merge mechanism: The Coalition G1 tries to merge with G2 based on the merge rule;
4: Split mechanism: The Coalition G0 tries to split based on the split rule;
5: Until: Merge and split iteration terminates, and the final coalition partition is obtained;
6: Return: The coalition with the highest coalition utility.

The algorithm keeps performing merge and split until no more Pareto improvements can be made, i.e., a Dhp-stable division is reached. The algorithm outputs the coalition division when a Dhp-stable division is reached as a stable solution of the game.

The proposed consensus mechanism, based on coalition formation game theory, significantly enhances the performance and incentive structure of blockchain networks for IoT information

management. By encouraging miners to actively collect and contribute valuable sensory data, the algorithm improves the overall efficiency and reliability of the blockchain system. The mechanism incentivizes miners to provide high-quality data and form coalitions with other miners to share sensing tasks, optimizing resource allocation and increasing their chances of obtaining consensus rewards. The merge-and-split algorithm ensures that miners can efficiently form stable and optimal coalitions, maximizing their utilities by balancing factors such as reward gain, formation overhead, and execution costs. This innovative consensus mechanism promotes active data contribution, enables collaborative mining, and facilitates efficient coalition formation, ultimately enhancing the quality and quantity of data available on the blockchain network. As a result, the performance, reliability, and value of the blockchain system for IoT information management are significantly improved.

4 Conclusions

This paper investigates the application of blockchain technology in IoT data sharing and consensus algorithms from an information management perspective. The following conclusions are drawn: (1) compared with single blockchain, multi-blockchain systems can process transactions more efficiently with lower latency; (2) the data sharing architecture combining horizontal federated learning and coalitional chaining can realize secure and efficient data sharing in the Internet of Things; (3) the consensus mechanism based on coalitional gaming can effectively incentivize miners to participate in the perception task, and improve the quality of data . Future work can further explore the optimal performance of the data sharing architecture, the combination of multi-chain and federated learning, and the fairness and robustness of the federation game model.

References

[1] Pattar S, Buyya R, Venugopal K R, et al. (2018) Searching for the IoT resources: Fundamentals, requirements, comprehensive review, and future directions[J]. IEEE Communications Surveys & Tutorials, 20(3): 2101-2132. https://ieeexplore.ieee.org/abstract/document/8334540.

[2] Ericsson. (2020) Internet of things. https://www.ericsson.com/en/internet-of-things.

[3] Huckle S, Bhattacharya R, White M, et al. (2016) Internet of things, blockchain and shared economy applications[J]. Procedia computer science, 98: 461-466. https://doi.org/10.1016/j.procs.2016.09.074.

[4] Dai H N, Zheng Z, Zhang Y. (2019) Blockchain for Internet of Things: A survey[J]. IEEE internet of things journal, 6(5): 8076-8094. https://ieeexplore.ieee.org/abstract/document/8731639.

[5] Christidis K, Devetsikiotis M. (2016) Blockchains and smart contracts for the internet of things[J]. IEEE access, 4: 2292-2303. https://ieeexplore.ieee.org/abstract/document/7467408.

[6] Zheng Z, Xie S, Dai H N, et al. (2018) Blockchain challenges and opportunities: A survey[J]. International journal of web and grid services, 14(4): 352-375. https://doi.org/10.1504/IJWGS.2018.095647.

[7] Yong Zhang, Zhongyuan Yao, Chao Wang, et al. (2023) Consortium Blockchain Multi-Chain Collaboration Solution for Complex Application Scenarios[J]. Journal of Applied Sciences, 41(04): 601-613. https://kns.cnki.net/kcms/detail/31.1404.n.20230801.1416.006.html

[8] Danzi P, Kalør A E, Stefanović Č, et al. (2019) Delay and communication tradeoffs for blockchain systems with lightweight IoT clients[J]. IEEE Internet of Things Journal, 6(2): 2354-2365. https://ieeexplore.ieee.org/abstract/document/8671694.

[9] Jiang Y, Lian Z. (2019) High Performance and Scalable Byzantine Fault Tolerance[C]//2019 IEEE 3rd Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). IEEE, 2019: 1195-1202. https://ieeexplore.ieee.org/abstract/document/8728972.
[10] Zhang L, Li Q. (2018) Research on Consensus Efficiency Based on Practical Byzantine Fault Tolerance[C]//2018 10th International Conference on Modelling, Identification and Control (ICMIC). IEEE, 2018: 1-6. https://ieeexplore.ieee.org/abstract/document/8529940.
[11] Nguyen D C, Pathirana P N, Ding M, et al. (2019) Blockchain for Secure EHRs Sharing of Mobile Cloud based E-health Systems[J]. IEEE Access. https://ieeexplore.ieee.org/abstract/document/8717579.
[12] Yu Y, Ding Y, Zhao Y, et al. (2018) LRCoin: Leakage-resilient cryptocurrency based on bitcoin for data trading in IoT[J]. IEEE Internet of Things Journal, 6(3): 4702-4710. https://ieeexplore.ieee.org/abstract/document/8513813.
[13] Konečný J, McMahan H B, Yu F X, et al. (2016) Federated learning: Strategies for improving communication efficiency[J]. arXiv preprint arXiv:1610.05492.
[14] Kang J, Xiong Z, Niyato D, et al. (2019) Incentive mechanism for reliable federated learning: A joint optimization approach to combining reputation and contract theory[J]. IEEE Internet of Things Journal, 6(6): 10700-10714. https://ieeexplore.ieee.org/abstract/document/8832210.
[15] Kim H, Park J, Bennis M, et al. (2019) Blockchained on-device federated learning[J]. IEEE Communications Letters, 24(6): 1279-1283. https://ieeexplore.ieee.org/abstract/document/8733825.