Storing Blockchain Data in Public Storage

Khikmatillo Tulkinbekov, and Deok-Hwan Kim*
Department of Electronic Engineering, Inha University
Incheon, South Korea
mr.khikmatillo@gmail.com, deokhwan@inha.ac.kr

2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN) | 978-1-7281-6476-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICUFN49451.2021.9528568

***Abstract*—** The idea of blockchain has emerged in many applications, including IoT data processing, over the last few years. Because of the guaranteed security, the blockchain is believed to create the backbone of future technologies as a P2P network. However, applying the blockchain protocol in IoT data computing faces a severe challenge related to extensive data management. This paper presents a BlockStore, the method for storing blockchain data in single public storage while maintaining the security features. The BlockStore also resolves the access management and recovery issues among the blockchain nodes. The preliminary experiments prove that the expected research brings a considerable advantage to IoT data management using blockchain.[1]

Keywords—Blockchain; Edge Computing; IoT Data; Distributed Computing;

I. INTRODUCTION

Blockchain technology brought another revolution in distributed data computing and transmission over the years. It provides the most reliable and trusted protocol for data communication among a heterogenous unknown number of servers. Because of these features, blockchain technology is being studied in many fields for adaptation in various applications. IoT data computing is also one of those fields which are being merged with blockchain technology. Some research has already been performed for the integration of these two fields. However, there is a common problem for the integration problem related to big data handling. The blockchain requires duplicating the whole data throughout the network. Since the IoT data grows exponentially, the duplication process creates a significant delay in data transmission, and the system fails in both performance and utilization. The possible solution for the problem mentioned above would be saving the IoT data in a single storage and sharing only the generated hash in the blockchain. In this scenario, the network wins in space utilization and keeps high performance. But another problem arises for the mentioned solution related to data security. If the data is not shared in the blockchain network, the secure protocol does not apply for separate storage. Malicious users may try to access the shared storage and alter or delete the data in the database. In this scenario, the blockchain adaptation fails for IoT data. This research proposes a new method named BlockStore for securely saving the blockchain data in the shared storage. The proposed method keeps all features claimed by blockchain for the separate database by storing collective signature and referencing each data by the blockchain transactions. In this way, we claim to win in both performance and utilization while adapting blockchain for IoT data. The preliminary experiments proved that the proposed method guarantees secure shared storage for saving the blockchain data.

* Corresponding author

II. BACKGROUND AND MOTIVATIONS

The study on the integration of blockchain and IoT data computing is getting hotter nowadays. The challenges occurred in the implementation phase, creating tremendous opportunities to bring new ideas into the blockchain world. One of those challenges is related to big data computing. The traditional blockchain (like bitcoin [1]) shares the actual data throughout the network and keeps the consistency on the most secure level, as shown in Fig. 1.

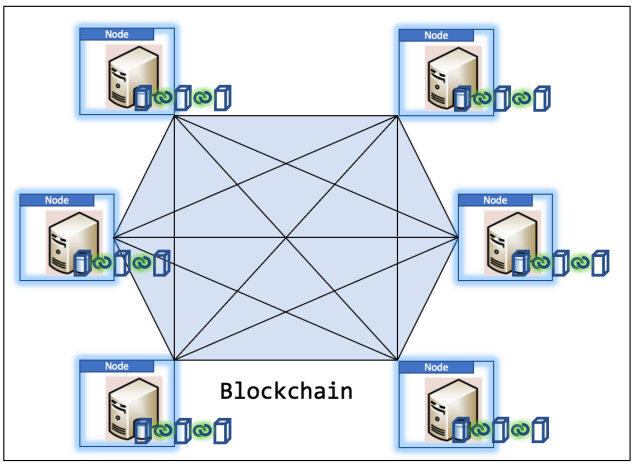

Fig. 1. Traditional blockchain architecture

The blockchain network is constructed by multiple nodes that share the same protocol. Data are stored as a set of records inside the blocks. The blockchain protocol requires creating a new block at a specific interval. For example, bitcoin generates a new block every ten minutes. Recent bitcoin transactions are stored in the temporary block during these ten minutes and inserted into the blockchain as a new block. The blockchain nodes compete to gain the right to create the latest block by solving mathematical computations related to new block generation. The node which finishes the analysis first acquires

the right to add the block to the network. Once the block is broadcasted to the network and approved by other nodes, no changes can be made to the data. Thus, blockchain achieves complete security. In other words, the central rule of the traditional blockchain is, all nodes should have the exact copy of the blockchain and only the longest chain is valid. In one way, this rule gives enormous advantages in terms of security, privacy, accessibility, availability, traceability, and transparency. On the other hand, sharing all the data throughout the networks creates duplicates of the data to all nodes. Bitcoin partially avoids this issue by employing different types of nodes for various purposes. For example, bitcoin employs lightweight nodes serve only for storing wallet information and performing verification and called simplifies payment verification (SPV) nodes. However, the capabilities of these nodes are very limited while they do not have access to the blockchain data. There are several other researches [2], [3], [4], [5] have been performed to improve the blockchain performance. But all share the common limitation with handling big data in the shared network. So, the duplication problem remains as the main vulnerability of the traditional blockchain. This may not be a significant issue for the systems run with small data like money transactions in bitcoin or other cryptocurrencies. But for IoT data creating duplicates and transmission through the network cost a lot in terms of time and space usage. When adopting blockchain technology for IoT data computing systems, these limitations motivated us to develop a new blockchain protocol for lightweight data sharing. The new method named BlockStore introduces secure storage for saving blockchain data.

III. BlockStore Overview

The BlockStore aims to save the blockchain data in shared storage while sharing only the hash among the nodes. The method of separating the data from the network gains advantages as follows:

- A lightweight blockchain network achieves higher scalability.
- Single storage for the whole network reduces the data duplication.

The overall architecture of the BlockStore is shown in Fig. 2 below.

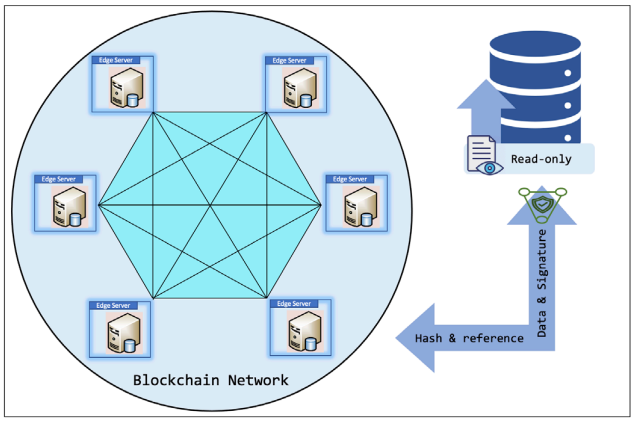

Fig. 2. BlockStore architecture

The right side of the figure shows the shared storage for saving actual data, and the left side shows the blockchain nodes that share the hash references to the data. When the data is first received in the BlockStore, it is directly written in the read-only shared storage, and the hash is generated based on the data content. The hash is written in the new block as a new record together with reference to the actual data in the shared storage. When the blockchain network validates the new block, the block hash will be sent to the shared storage as the collective signature. The data in the shared storage is marked by the collective signature that proves the network validates the data, and no more updates can be performed on the content. The collective signature doubles the data security in the shared storage as the following scenario proves:

"Hash is generated by the data content and saved in the blockchain network. Collective signature is generated based on the set of hashes in the block, so updating the hash invalidates the collective signature. If the malicious user tries to update the data content, the new content generates a new hash, the new hash results in invalidating the collective signature. So, any changes on the data are detected easily using the collective signature."

The BlockStore also gains an advantage in membership management and bootstrapping the nodes. Since only the hash is broadcasted to the network, new nodes only import the hashes from the blockchain network. The actual data in the shared storage is accessed only if the data reads occur. Since the hash size is constant and considerably small compared to other data types, importing the blockchain takes less time for the new nodes.

IV. Preliminary Experiments

We have performed preliminary experiments on BlockStore performance in terms of scalability and storage utilization. For comparison, we used the open-source bitcoin code as the traditional blockchain. Table 1 shows the overall similarities and differences between bitcoin and the BlockStore.

TABLE I. Comparison Table

Criteria	Bitcoin	BlockStore
Blockchain	Traditional	Shared Storage
Digital signature	No	Yes
Duplication	N	P(Hash) * N
Secure	Yes	Yes
Propagation rate	Depends on N	Constant

The most significant difference between bitcoin and the BlockStore is their blockchain architecture. The bitcoin is very first and largest blockchain, so that it can be an example of state-of-art blockchain architecture for comparison. The BlockStore, on the other hand, employs a new method by using shared storage to save the data. To maintain security, the BlockStore also introduces the digital signature to guarantee no changes can be performed on the data after the corresponding

hash is inserted into the blockchain network. Another essential factor that differentiates these blockchain architectures is the data duplication among the nodes. As the bitcoin requires to share the exact copy of the blockchain to all nodes, it results from having duplicates as many as the number of nodes (N). But the BlockStore only shares the hash while keeping the single copy of the actual data. So, the duplication in the BlockStore is reduced to the hash proportion. In terms of security, bitcoin is already proven to be the most secure P2P network nowadays. The BlockStore also maintains the same level of security by employing the digital signature and hash referencing methods. The most significant advantage of the BlockStore over the bitcoin is the constant propagation rate.

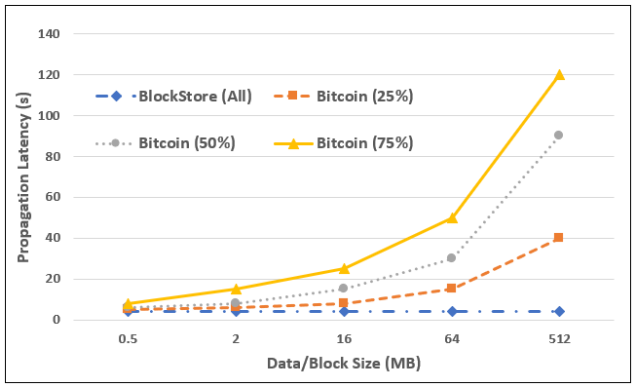

Fig. 3. Propagation latencies for different data/block sizes

The propagation rate refers to the speed of broadcasting data among the networks. This is one of the main criteria for evaluating blockchain scalability. The evaluation results for data propagation of bitcoin and BlockStore are shown in Fig. 2. As mentioned above, the bitcoin stores all the data in the block while the BlockStore only saves the hash and reference combinations. So, the block size and the data size refer to the same data for these two technologies. From the figure, we can observe that the data propagation in bitcoin delays depending on the percentile and the block size. In this experiment, the percentile means the percentage of nodes to broadcast the block. The bitcoin propagation is almost similar to the BlockStore propagation with smaller blocks. As the experiment shows, as the block size increases, the bitcoin propagation latency increases at nearly exponential. And this rate gets bigger for bigger percentiles. In numbers, for 512MB blocks, the bitcoin propagates the block to 75% of the nodes in almost 120 seconds.

When considering the factor of IoT data, the block size is unrealistic to be small. The standard block size is not less than 500MB and is usually bigger. It is possible to create smaller blocks by decreasing the block creation interval, but it increases the propagation period, which is also unacceptable. These experiments prove that the traditional blockchain fails in storing the IoT data and cannot be applied directly. On the other hand, the BlockStore saves only the hash of the data in the blockchain network. The hash size is constant for all data. This means that the block size does not face big changes in size, whatever the data size is. The figure also proves that the BlockStore provides a constant propagation latency for all data sizes and percentiles. These experiments demonstrate that the proposed method is suitable for IoT data management systems when applying blockchain technology.

V. Future Plan

This research focuses on the primary advantages of separating the hash from the actual data in the blockchain. We will extend the proposed method for more factors in both performance and security. Namely, the future work will include the advanced technique for generating the digital signature and enhancing the access control on the shared storage.

VI. Conclusions

This paper proposes a new method for saving the blockchain data in the shared storage and sharing the hash in the network. The proposed method named BlockStore maintains the data security in the shared storage by employing the digital signature and hash-reference indexing methods. The BlockStore achieves a good advantage in storage utilization and scalability by separating hash from the data. The preliminary experiments proved that the BlockStore achieves a constant latency in data propagation while traditional blockchain highly depends on the block size and the propagation percentile.

Acknowledgment

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korean government(MSIT) (No. NRF-2021R1F1A1050750) and in part by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (No.2019-0-00064, Intelligent Mobile Edge Cloud Solution for Connected Car), (No.2019-0-00240, Deep Partition-and-Merge: Merging and Splitting Deep Neural Networks on Smart Embedded Devices for Real-Time Inference).

References

[1] S. Nakomoto, “Bitcoin: E peer-to-peer Electronic Cash System”, Oct. 2008.

[2] K. Lei, M. Du, J. Huang and T. Jin, “Groupchain: Towards a Scalable Public Blockchain in Fog Computing of IoT Services Computing,” in IEEE Transactions on Services Computing, vol. 13, no. 2, pp. 252-262, 1 March-April 2020, doi: 10.1109/TSC.2019.2949801.

[3] M. Zhaofeng, W. Xiaochang, D. K. Jain, H. Khan, G. Hongmin and W. Zhen, “A Blockchain-Based Trusted Data Management Scheme in Edge Computing,” in IEEE Transactions on Industrial Informatics, vol. 16, no. 3, pp. 2013-2021, March 2020, doi: 10.1109/TII.2019.2933482.

[4] H. T. T. Truong, M. Almeida, G. Karame and C. Soriente, “Towards Secure and Decentralized Sharing of IoT Data,” 2019 IEEE International Conference on Blockchain (Blockchain), Atlanta, GA, USA, 2019, pp. 176-183, doi: 10.1109/Blockchain.2019.00031.

[5] M. Hou, T. Kang and L. Guo, “A Blockchain Based Architecture for IoT Data Sharing Systems,” 2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Austin, TX, USA, 2020, pp. 1-6, doi: 10.1109/PerComWorkshops48775.2020.9156107.