2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST) | 978-0-7381-3042-2/20/$31.00 ©2021 IEEE | DOI: 10.1109/ICREST51555.2021.9331254

Increasing Throughput and Reducing Storage Bloating Problem Using IPFS and Dual-Blockchain Method

MD. Soharab Hossain Sohan
Department of Electrical and Electronics Engineering Bangladesh Army International University of Science & Technology (BAIUST)
Cumilla, Bangladesh
soharab@baiust.edu.bd

Minhaz Mahmud
Department of Electrical and Electronics Engineering Bangladesh Army International University of Science & Technology (BAIUST)
Cumilla, Bangladesh
minhaz@baiust.edu.bd

M A Baten Sikder
Department of Electrical and Electronics Engineering Bangladesh Army International University of Science & Technology (BAIUST)
Cumilla Bangladesh
bateneee3@gmail.com

Fakir Sharif Hossain
Department of Electrical and Electronics Engineering Bangladesh Army International University of Science & Technology (BAIUST)
Cumilla Bangladesh
dr.hossain@baiust.edu.bd

Md. Rakibul Hasan *Department of Electrical and Electronics Engineering Khulna University of Engineering & Technology*
Khulna, Bangladesh
rakibul.eeekuet@gmail.com

***Abstract*—Blockchain is a revolutionary technology that has been underlying behind many cryptocurrencies for many years. Due to the scalability issue, this decentralized system lags behind the centralized currency systems, and cannot be adopted by other platforms though it has a lot of unique features. This paper is motivated due to the lack of enough scalability on the existing blockchain technologies. A theoretical method is proposed in this paper to increase throughput and reduce storage dependencies. A distributed storage system IPFS is used to bypass the storing liabilities and to increase throughput. The dual-blockchain method serves the core features of the blockchain by adding the references of the main block into the ledger in place of the original block. The analysis shows that our proposed method can achieve up to 25.8 times greater throughput and almost 1685 times lesser ledger size compared to Bitcoin Core.**

***Index Terms*—Blockchain, Distributed Storage, Throughput, Storage Bloat, Scalability, IPFS, Dual-Blockchain**

I. Introduction

Blockchain is a record-keeping technology that stores transactions and valuable information in the form of blocks and keeps connected within a chain. It has some important features that differentiate it from other centralized systems, such as decentralization, democracy, transparency, privacy, security, and trust-free.

Decentralization of blockchain is served by a peer-to-peer distributed ledger which is verified by all the nodes. Any decision-making situation is handled by the majority number of nodes which is an example of a democratic system. Recorded transactions in the ledger are accessible to all miners that make the blockchain transparent. Although blockchain is transparent, the user's real identity is kept anonymous. Information is stored into blocks and all blocks are connected with each other cryptographically. Manipulation of data makes the chain invalid to the other miners. Hence, the consensus protocol restores the original data in place of manipulated data which makes the system secure. Blockchain is a trust-free system because the necessity of a trusted third party is eliminated by the decentralized infrastructure. Though it has a lot of useful features, scalability is the key barrier that resists the adoption of blockchain technology in other useful applications except cryptocurrency where the main aspects of scalability are throughput and storage. Recently, several researchers have been evaluating many effective innovations in order to make the blockchain more scalable. However, most of the proposed technologies focus on the factors of scalability forsaking other key features such as security, decentralization, etc.

Three important features of blockchain such as decentralization, scalability, and security are collectively known as Trilemma in the blockchain where any blockchain system can have a maximum of two out of those three factors. Unfortunately, there is a lack of such innovations that comprise all the fundamental properties of the blockchain. Therefore, there is a necessity for an appropriate methodology according to the application that meets all the required characteristics.

In this paper, we have proposed a theoretical way of enabling higher Transactions Per Block (TPB) and decreasing the storage bloating problems. In this method, we are using an external distributed file-sharing system. Employing our method blockchain technology can be used in daily life applications as well as in cryptocurrency.

The rest of the paper is organized as follows. The scalability issue is analyzed in Section II from the perspectives of throughput and storage. In Section III related works behind the scalability are discussed. Section IV describes our proposed method. Theoretical analysis and results are observed based on the Bitcoin Core in Section V. Finally, we concluded this paper in Section VI.

II. The Scalability Issue

In this section, we will analyze the scalability issue from the perspectives of throughput and storage.

A. Throughput

Throughput is the rate of valid Transactions Per Second (TPS) that are added to the blockchain. It is a function of the number of TPB and block time. In this manner, TPB are proportional and the block time is inversely proportional to the throughput.

Because of the emergence of blockchain technology, users are handling more and more decentralized transactions over the centralized system. Thus the number of transactions is also increasing day by day. In the year 2020, one block of Bitcoin Core contains 2300 transactions on average [1]. But the blocks are generated on average every 10 minutes, which implies only about 4 TPS. The average throughput of the Bitcoin blockchain is 3.3-7 TPS that is remarkably low. The TPS of Etherium is 20 [2].

Visa International Service Association (VISA), a credit card company, can handle up to 1,667 TPS and PayPal verifies 193 TPS [2]. The TPS of Bitcoin and Etherium are not promising compared to VISA, PayPal, and other centralized financial structures. The current consensus protocol that Bitcoin Core and other decentralized cryptocurrency use can not emulate this transaction rate.

B. Storage

In the traditional blockchain system, recorded data that is added to the block is distributed and stored to all the miners over the world. Whereas more data refers to more blocks and more blocks require more storage in the miner end. Each node has to store the whole blockchain data [3]. And with new miners, they also need to fetch the whole ledger. In terms of Bitcoin, the storage requirement for each miner is over 300 GB for recent years. The rate of increase of this data is exponential regarding time, which leads to an exploding storage bloating problem.

As each miner has to store the whole ledger [3], new miners also need to fetch the whole data to connect with the network. The bigger the ledger will be the more time will be required for the new miner to actively participate in mining. Hence, it is quite impossible to apply blockchain straight away to real business systems where each miner has limited resources. Therefore more studies are required to solve this storage bloating problems with the purpose that blockchain technology can be used in daily life applications as well as cryptocurrency.

III. Related Works for Enhancing Scalability

A. Related Works for Increasing Throughput

Based on the analysis in the previous section, throughput relates to the number of transactions in each block and block interval time. The number of TPB can be increased by the following processes.

1) Increasing the Block Size: It is clear that inserting more and more transactions without changing the block interval in any block increases throughput. Here, the number of transactions defines the size of the block. Bitcoin Cash follows this method and generates a block up to 32 MB size [4]. In this strategy, the throughput is improved significantly, but it also increases the which is not desired.

2) Reducing the Transaction Size: Another way of increasing the transaction number in a single block without changing the block size is by reducing the transaction size. For example, Segregated Witness [5], also known as SegWit reduces 60-70% size of the total transaction by excluding the digital signatures which are required for verification of the transactions from the transaction data and including it at the end of the block. Here, digital signatures are required for transaction verification.

3) Reducing the Number of Transactions Processed by the Nodes: By reducing the number of transactions verified by the nodes, improved throughput can be obtained. This method can be achieved by following off-chain transactions or sharding.

- Off-Chain Transactions: For frequent transactions between the same nodes, off-chain micropayment channels can be created between these nodes. These multi-signature transactions which are made through this channel are stored locally. Only the starting and ending settlement transactions are added to the main blockchain. Lightning Network [6] and Duplex Micropayment Channel [7] are two implementations of this process. Though off-chain transactions can reduce the number of transactions processed by the nodes, it compromises the system security and makes the system complex. Micropayment channels lock a huge amount of digital currency, which can be used on the blockchain only after closing the channels.
- Sharding: In the sharding method, the blockchain is divided into many shards, but these shards remain connected to each other. Each shard consists of multiple nodes of the network and processes a small portion of all the transactions. In this manner, transactions are processed parallelly in different shards. Therefore, the blockchain can be made much more scalable by increasing the number of nodes. Elastico [8] and OmniLedger [9] are implementations of the sharding blockchain system. Despite the fact that sharding increases throughput

by reducing the number of transactions processed by each node, it sacrifices the global consensus. A small number of nodes per shard compromises system security. A large number of nodes per shard affects system performance.

4) Block Interval Can be Reduced by Following Processes: Throughput can also be increased by reducing block interval time keeping the block size constant. If the time required to verify any transaction is reduced, the time required to generate a block can also be reduced which results in a reduced block interval time.

- Fixed Leader: Hyperledger Fabric [10] is a permissioned blockchain. In permissioned blockchain, the participant nodes are selected, whereas in public blockchain like Bitcoin any node can join the network. In hyperledger fabric, only a fixed set of nodes can add and validate the transaction by employing pluggable consensus protocols.
- Single Leader: In Bitcoin-NG [11], the protocol divides the time into epochs. In each epoch, a leader is selected to generate a block through the power of work (PoW) consensus. This selected leader can generate multiple microblocks of transactions until the next leader is selected. Bitcoin-NG also increases throughput, but compromises system security significantly. When a malicious leader is selected, a double-spending attack can easily take place.
- Collective Leader: Another approach to reduce the block interval time is to select multiple leaders or a committee instead of one leader to generate blocks via PoW consensus. To validate these blocks the leaders use the Practical Byzantine Fault Tolerance (PBFT) algorithm. Byzcoin [12] and Solida [13] both employ this mechanism. Although this approach increases throughput, it sacrifices the security of the system. Because the size of the committee is way smaller than the total nodes.

B. Related Works for Decreasing Storage Bloating

Based on the analysis in the previous section, storage is related to the generated data. The basic concept to solve the storage bloating is to pursue off-chain scaling where a large amount of data is stored outside the main blockchain using a distributed storage system. Authors in [14] designed off-chain storage employing the Distributed Hash Table (DHT) where it is not necessary to store the whole data on the chain. Essential data is stored by the DHT while blockchain stores the SHA-256 hashes of the essential data as references. Qiuhong Zheng and others demonstrated a scalable model to alleviate this bloating problem using a distributed storage called InterPlanetary File System (IPFS) [15]. In this model, main transaction data are stored in the IPFS and the block is created using the returned content identifier (CID) as reference. Therefore, the main burden of storing a mass amount of data in the ledger is reduced. IPFS is introduced in another work integrating with Ethereum to design a decentralized service marketplace system called Desema [16]. In this prototype, service metadata and large data are stored in IPFS, and data references are stored in Ethereum. A network coding-based distributed storage (NC-DS) framework was proposed by Mingjun Dai and others in work [17] to save the required storage room. Generated blocks are divided into sub-blocks and then network-coding is applied to encode the partitioned pieces. Then the encoded smaller pieces are distributed to all nodes.

IV. Proposed Method

A. InterPlanetary File System(IPFS)

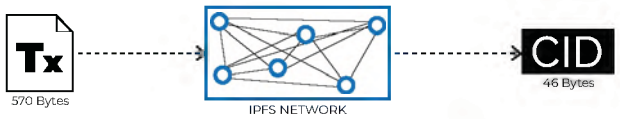

Fig. 1. Block diagram of IPFS.

InterPlanetary File System (IPFS) is a peer-to-peer network for storing and accessing all digital data [18]. It uses content-based addressing rather than the location-based (server) addressing. Content-addressing is achieved by assigning a unique cryptographic hash on the non-similar stored file. That means when a file is uploaded into IPFS, it returns a hash which is also known as the Content-Identifier (CID), shown in Fig. 1. The file inside IPFS partitioned into several pieces of incomplete information and all pieces have corresponding hashes. These pieces are distributed among the peers. CID of the file signifying the root object by connecting all the corresponding hashes and the incomplete pieces of information can be found along the following connected path. When a user asks for any content using a CID, he will discover neighbors containing pieces of that content. These neighbors are called nearby peers. He then accesses the full content by fetching the incomplete information piece by piece. In this paper, IPFS is chosen over other distributed storage for several advantages such as trust-free and open-source. As the pieces are incomplete and distributed to other peers as well, IPFS is said to be incorruptible too. In short, IPFS is a secure storage system that has a high throughput and high storage capacity.

The overall process illustrated in Fig. 2 is described in two consecutive steps. The first one signifies the throughput segment and the second one is the double-blockchain method.

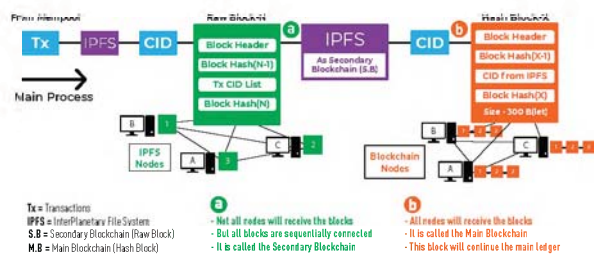

Fig. 2. IPFS and Duel-Blockchain based system.

B. Throughput Segment

At first, the transactions are added to the mempool. All miners pass these transaction data through the IPFS API and

collect the corresponding CID of those data as illustrated in Fig. 1. Assuming a miner, miner-A then generates a raw block with these CIDs by using the Proof of Work (PoW) consensus protocol. He does not distribute this block to the whole network but stores this block locally. Through this process, transactions from the mempool are enlisted in a large way to the raw block, more TPS can be handled within a single block without affecting the block size and block interval time.

C. Dual-Blockchain Segment

As stated earlier, the raw block is not distributed to all the miners as the main blockchain but to the available nodes of the IPFS. Therefore, the immediate raw blocks are cryptographically linked with each other and create a secondary blockchain without distributing it to all the miners. Again, the same miner generates a block with the corresponding CID of the raw block called a hash block and distributes this hash block to the other miners. The rest of the miners of the network verify this block whether this block is valid or not. After verification, this block is added to the main blockchain, and all the miners store this block locally. The size and the number of transactions vary with respect to time, but the size of the CID in the hash block remains constant which is 46 Bytes. Therefore, according to this process, the size of each hash block does not fluctuate rather remains constant. Through this process, the raw block is sent through the IPFS to create a secondary blockchain and then a hash block is created with the CID of the raw block and distributed to the miners to meet the global consensus, illustrated in the latter part of Fig. 2. During this process, the size of each block becomes significantly smaller.

The overall process from mempool to hash block is done at the miner end. Therefore, the complexity is comparatively higher than the traditional Bitcoin mining environment. Replacing an existing complex and vast system with the new framework is quite disastrous.

V. Result and Analysis

A. Result Analysis with respect to Throughput

Assuming the size of the raw block from the throughput segment is 1 MB. The size of each CID of a transaction is 46 Bytes. The size of the header of a block is 80 Bytes [19]. The total number of CID or transactions that can be added to the raw block is $(1048576 - 80) \div 46 = 22793$.

TABLE I
Raw Block Data

Title	Amount & Size
Block Header	80 Bytes [7]
Number of CID	22793 (average)
Size of each CID	46 Bytes

However, the Bitcoin Core has 652313 block heights [1]. The total number of transactions in the Bitcoin Core blockchain is 576822203 [20]. Hence, each block of Bitcoin Core contains an average of 884 transactions.

TABLE II
Bitcoin Core Block Data

Title	Amount & Size
Block Header	80 Bytes [7]
Number of Tx	884 (average)
Average Block Size	**1048576 Bytes or 1 MB**

According to the data from Table I and Table II, the multiplication factor of the transaction per block (TPB) between our proposed method and the Bitcoin Core is $(22793 \div 884) = 25.8$. Fig. 3 describes this comparison visually. In this manner, if we increase the block size, the TPB in our proposed method increases rapidly. By increasing the block size to 8 MB, the total TPB becomes 182359. Assuming the block interval remains 600 seconds, the TPS becomes $(182359 \div 600) = 303$ which beats PayPal's throughput. At this rate, the system throughput can be increased to $[\{(32 \times 1024 \times 1024) - 80\} \div 46] \div 600 = 1215$ by increasing the block size to 32 MB. According to our method, the TPS can compete with VISA. However, in our method, these enhancements in block size do not affect the miner's storage requirement. In another word, these increased sized blocks do not cause storage bloating problems. Because miners have to store just the references for the transaction data.

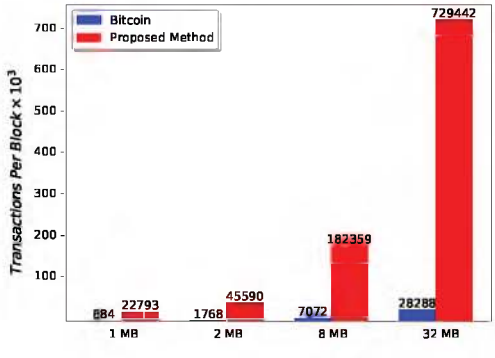

Fig. 3. TPB comparison between Bitcoin Core and proposed method.

B. Result Analysis with respect to Storage

According to the dual-blockchain segment, the hash block contains only one CID of size 46 Bytes. With some additional data such as block header and any authentication references, the total size of the hash block becomes at most 300 Bytes which is shown in Table III. All the miners over the network store this hash block. As per our proposed method, the total size of 652313 blocks [20] which is the total height of the Bitcoin Core is $(652313 \times 300) = 0.1822$ GB.

On the other hand, the total size of 652313 blocks is 303.4 GB [1] in the Bitcoin Core protocol. Each miner has to store this huge file. Now the multiplication factor of storage consumption between the proposed method in this paper and Bitcoin Core is $(0.1822 \div 303.4) = 0.0006$. This ratio will further decrease over time. Fig. 4 shows the size of the blockchain according to our method in MB and Fig. 5 shows the size of the traditional Bitcoin Core blockchain in GB.

TABLE III
HASH BLOCK DATA

Title	Amount & Size
Block Header	80 Bytes [7]
Number of CID	1
Size of each CID	46 Bytes
Reserved Space	174 Bytes
Total Block Size	**300 Bytes**

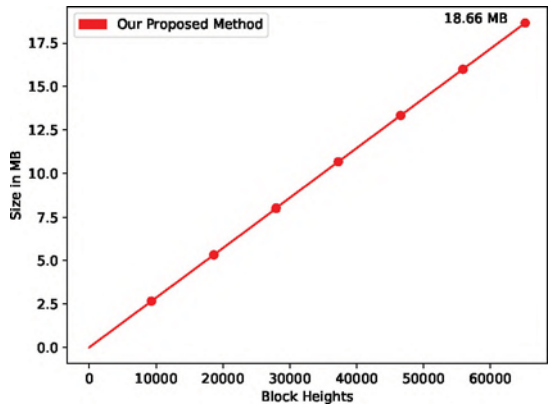

Fig. 4. Proposed blockchain size assumption over the years.

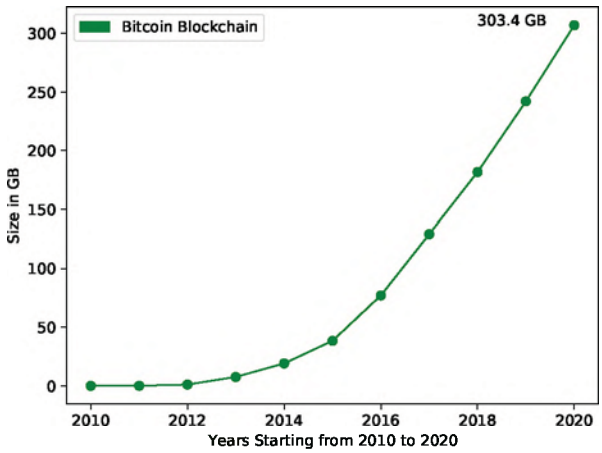

Fig. 5. Bitcoin Core blockchain size over the years.

VI. CONCLUSION

Blockchain shows a lot of potentials, but it also inherits the scalability problem. In this paper, we tried to solve this scalability issue with respect to the throughput and storage requirement. Throughput is a serious matter for every cryptocurrency. In our proposed method, it is possible to stand against PayPal in terms of TPS by increasing the block size to 8 MB. Even our system can compete with VISA's high throughput by increasing the block size to 32 MB. These increments in block size do not cause storage bloating problem. Storage bloating problems, in other words high storage requirements at the miner end, which is one of the main reasons that blockchain can not be adopted to the real business environment. In this paper, the storage requirement for each miner is determined significantly low. Our analysis shows a remarkable opportunity to scale the blockchain technology from the perspective of any application such as cryptocurrency, IoT, healthcare, supply chain management (SCM), and so on. Also, the core purpose of blockchain remains unaffected.

REFERENCES

[1] Bitcoin core charts. Accessed on: Oct. 15, 2020. [Online]. Available: https://charts.bitcoin.com/btc
[2] D. Mechkaroska, V. Dimitrova, and A. Popovska-Mitrovikj, "Analysis of the possibilities for improvement of blockchain technology," in *IEEE 26th Telecommunications Forum (TELFOR), Serbia*, 2018, pp. 1–4.
[3] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," Manubot, Tech. Rep., 2019.
[4] Bitcoin.com. Accessed on: Oct. 15, 2020. [Online]. Available: https://www.bitcoin.com
[5] A. Singh, R. M. Parizi, M. Han, A. Dehghantanha, H. Karimipour, and K.-K. R. Choo, "Public blockchains scalability: An examination of sharding and segregated witness," in *Blockchain Cybersecurity, Trust and Privacy*. Springer, 2020, pp. 203–232.
[6] J. Poon and T. Dryja, "The bitcoin lightning network: Scalable off-chain instant payments," 2016.
[7] C. Decker and R. Wattenhofer, "A fast and scalable payment network with bitcoin duplex micropayment channels," in *Symposium on Self-Stabilizing Systems*. Springer, 2015, pp. 3–18.
[8] L. Luu, V. Narayanan, C. Zheng, K. Baweja, S. Gilbert, and P. Saxena, "A secure sharding protocol for open blockchains," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016, pp. 17–30.
[9] E. Kokoris-Kogias, P. Jovanovic, L. Gasser, N. Gailly, E. Syta, and B. Ford, "Omniledger: A secure, scale-out, decentralized ledger via sharding," in *2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2018, pp. 583–598.
[10] C. Cachin, "Architecture of the hyperledger blockchain fabric," in *Workshop on distributed cryptocurrencies and consensus ledgers*, vol. 310, no. 4, 2016.
[11] I. Eyal, A. E. Gencer, E. G. Sirer, and R. V. Renesse, "Bitcoin-ng: A scalable blockchain protocol," in *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*. Santa Clara, CA: USENIX Association, Mar. 2016, pp. 45–59. [Online]. Available: https://www.usenix.org/conference/nsdi16/technical-sessions/presentation/eyal
[12] E. K. Kogias, P. Jovanovic, N. Gailly, I. Khoffi, L. Gasser, and B. Ford, "Enhancing bitcoin security and performance with strong consistency via collective signing," in *25th USENIX Security Symposium (USENIX Security 16)*. Austin, TX: USENIX Association, Aug. 2016, pp. 279–296. [Online]. Available: https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/kogias
[13] I. Abraham, D. Malkhi, K. Nayak, L. Ren, and A. Spiegelman, "Solida: A blockchain protocol based on reconfigurable byzantine consensus," *arXiv preprint arXiv:1612.02916*, 2016.
[14] G. Zyskind, O. Nathan, and A. . Pentland, "Decentralizing privacy: Using blockchain to protect personal data," in *2015 IEEE Security and Privacy Workshops*, 2015, pp. 180–184.
[15] Q. Zheng, Y. Li, P. Chen, and X. Dong, "An innovative ipfs-based storage model for blockchain," in *2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, 2018, pp. 704–708.
[16] M. Klems, J. Eberhardt, S. Tai, S. Härtlein, S. Buchholz, and A. Tidjani, "Trustless intermediation in blockchain-based decentralized service marketplaces," in *Service-Oriented Computing*, M. Maximilien, A. Vallecillo, J. Wang, and M. Oriol, Eds. Cham: Springer International Publishing, 2017, pp. 731–739.
[17] M. Dai, S. Zhang, H. Wang, and S. Jin, "A low storage room requirement framework for distributed ledger in blockchain," *IEEE Access*, vol. 6, pp. 22 970–22 975, 2018.
[18] J. Benet, "Ipfs-content addressed, versioned, p2p file system," *arXiv preprint arXiv:1407.3561*, 2014.
[19] A. M. Antonopoulos, *Mastering Bitcoin: unlocking digital cryptocurrencies*. O'Reilly Media, Inc., 2014.
[20] Blockchain charts. Accessed on: Oct. 15, 2020. [Online]. Available: https://www.blockchain.com/charts