Mechanisms Design for Blockchain Storage Sustainability

Yunshu Liu, Zhixuan Fang, Man Hon Cheung, Wei Cai, and Jianwei Huang

Abstract—In a blockchain system, ensuring storage sustainability is crucial for the long-term feasibility of the system’s operation. This paper presents a comprehensive review of the state-of-the-art mechanisms designed to address the storage sustainability problem. We first introduce technological mechanisms aimed at reducing miners’ storage costs. Next, we discuss incentive mechanisms that encourage users to pay adequate transaction fees to cover storage costs. Finally, we discuss future challenges and open problems in this field.

I. INTRODUCTION

With the booming of blockchain, both academia and industry have applied the blockchain system to diverse applications [1], such as supply chain, finance, energy, and traffic, as illustrated in Fig. 1. When the fraction of malicious nodes in blockchain system is small (e.g., smaller than 1/3 in Ethereum), the blockchain system can operate and has the following benefits. First, the blockchain system is de-centralized, which does not require a centralized authority. Second, the blockchain maintains a transparent and verifiable ledger for applications, which allows any blockchain user to verify the blockchain data. Moreover, the blockchain system is immutable and nonrepudiable, i.e., no one can tamper with data that is already in the blockchain. Finally, the operation of blockchain does not rely on any third parties.

However, the widespread adoption of blockchain systems has imposed increasingly significant costs on miners, who are the operational nodes of the blockchain. Specifically, these costs include electricity, storage, and communication. We take the two largest cryptocurrencies, Bitcoin and Ethereum, as examples. For the electricity, Bitcoin consumed 0.55% of global electricity in 2020 [2]. For the storage, the current monthly cost of storing the entire Ethereum blockchain is approximately $20 million [3] [4]. For the communication, the synchronization of an Ethereum archival node needs to download 11.8 terabytes of data [5]. Nevertheless, there are some solutions to deal with the electricity and communication costs, e.g., proof-of-stake protocol and compact block design [6], respectively.

The huge storage cost of blockchain has led to the emergence of the storage sustainability problem. On the one hand, the cost of storing the entire blockchain history is a long-term and ever-increasing burden for miners. On the other hand, miners do not receive enough compensation in transaction fees to cover their storage costs [7] [8]. These factors raise concerns about the long-term sustainability of miners operating the blockchain.

The storage sustainability problem undermines the system security in the long run. Such a problem means that miners do not have enough incentive to store the blockchain data and operate the system. Hence, the problem plays an important role in the decline of the number of miners, as observed in Ethereum, where two-thirds of miners dropped out since 2018 [9]. The decline makes the blockchain less decentralized and more susceptible to a 51% attack.

In this article, we discuss the challenges of addressing the storage sustainability problem and review some recently proposed storage sustainable mechanisms from the following two perspectives:

1) The technological approaches to reduce each miner’s storage costs.
2) The economical approaches to encourage users to pay sufficient transaction fees for the storage costs.

To the best of our knowledge, existing surveys have generally neglected the economical approach, which fails to reveal the economic reasons behind the storage sustainability problem.

The rest of the article is organized as follows. First, we present an overview of the blockchain system. Next, we present the recent advances in addressing the problems. Then, we analyze future challenges for designing the storage sustainable mechanism. Finally, we conclude the paper.

II. OVERVIEW OF BLOCKCHAIN ARCHITECTURE

In this section, we first describe the blockchain system in Section II-A. Next, we describe the storage sustainability problem in Section II-B.

Yunshu Liu is with the Chinese University of Hong Kong, Hong Kong and Shenzhen Institute of Artificial Intelligence and Robotics for Society.
Zhixuan Fang is with Tsinghua University and Shanghai Qi Zhi Institute.
Man Hon Cheung is with City University of Hong Kong.
Wei Cai and Jianwei Huang (corresponding author) are with the Chinese University of Hong Kong, Shenzhen and Shenzhen Institute of Artificial Intelligence and Robotics for Society.
A. Blockchain System

In a typical blockchain system like Bitcoin, its operation involves the interaction among the protocol designer, users, and miners as illustrated in Fig. 2. The details are as follows:

First, the protocol designer proposes the blockchain mechanism. Generally, the online community of users and miners collectively serves as the protocol designer. For example, the Bitcoin online community proposed and activated Bitcoin Improvement Proposal 91 to introduce a new transaction format.

Then, users generate transactions (e.g., transfer of cryptocurrency) and set the corresponding transaction fees following the mechanism. Each generated transaction is pending for a miner to include it in a particular block (i.e., the transaction container) on the blockchain, and the transaction fee is a reward for the miner who does so.

Finally, miners select transactions and compete to mine the next block under the proof-of-work (PoW) protocol. We elaborate on this process as follows:

1) Each miner selects a set of pending transactions.
2) Miners compete to solve a cryptographic puzzle. The first miner who solves the puzzle can generate a block (i.e., block \( k + 1 \) in Fig. 2), containing this particular miner’s selected transactions and some auxiliary data. The auxiliary data in Bitcoin includes the hash value of the previous block, timestamp, hash value of the current block, and solution of the puzzle.
3) The miner who generates the block broadcasts the block in the network. Every miner in the network stores the new block in his local copy of blockchain, and transactions in the block are successfully appended to blockchain.
4) The miner who generates the block obtains the block reward from the blockchain and the transaction fees from those users whose transactions are included in the block.

In this article, we focus on the storage sustainable mechanisms under the PoW protocol, since PoW is mature and widely used.

B. Storage Sustainability Problem in Blockchain

The storage sustainability problem refers to the observation that the current transaction fees cannot cover the storage cost of blockchain [7] [8]. Specifically, miners need to store transactions from blockchain users. To compensate miners’ costs, users pay transaction fees to miners. However, it turns out that miners are willing to accept the transaction with a transaction fee insufficient to cover its storage costs in practice. For example, miners like DwarfPool and UUPool accept zero-fee transactions [10]. As a result, users’ current level of transaction fee payment cannot cover miners’ storage costs of transactions. Specifically, in the first six months of 2020, the average monthly transaction fee in Ethereum is $7.32 million [11], which is much smaller than the monthly storage costs of $20 million for all the miners in storing the entire blockchain. Currently, the block reward can help cover the gap between the transaction fee and the storage cost. However, this is only a temporary solution as the block reward generally decreases over time in the blockchain system (e.g., Bitcoin).

In the next section, we summarize the recent mechanisms to address the storage sustainability problem. The first set of works adopts the technological approach to reduce the miner’s storage cost. The second set of works adopts the economic approach and proposes incentive mechanisms to encourage users to pay sufficient transaction fees to cover the storage costs.

III. MECHANISMS FOR SUSTAINABLE BLOCKCHAIN STORAGE

In this section, we present the technological and economic approaches to address the storage sustainability problem in Sections III-A and III-B, respectively.

A. Technological Approach

The technological approach aims to reduce each miner’s storage cost. Specifically, such an approach aims at achieving the following objectives:

- **Efficiency**: The technology approach should be able to cut down each miner’s storage costs in the system.
- **Security**: There is no single point failure in the system, where a single miner leaving the system cannot lead to a system malfunction. Moreover, miners can eventually reach a consensus even in the presence of a fraction of Byzantine or malicious miners (e.g., Ethereum allows a fraction of 1/3).
- **Correctness**: Each miner newly joining the blockchain should be able to get the correct copy of complete blockchain data, even if there are a fraction of Byzantine or malicious miners. This ensures that all miners can reach consensus on the transactions accepted by the blockchain.

Several recent works (e.g., [6], [12], [13]) propose some methods to achieve the above objectives.

*Sharding* is widely studied to reduce the blockchain storage cost, as illustrated in Fig. 3. To achieve the efficiency objective, the sharding system (e.g., [12]) partitions miners into multiple...
groups. Each group corresponds to a shard and each shard only processes and stores a part of transactions and all the shards work in parallel. Given $N$ shards, a miner in a shard generally only bares $1/N$ fraction of storage consumption compared with a miner in the original system (although some miners may join multiple shards and bear higher storage costs). Hence, when there are a sufficient number of shards, each miner’s storage cost is much smaller than in a no-shard system. To meet the security requirement, the sharding method generally adopts a sophisticated algorithm to assign each miner to a random shard and select a committee to vote for the final consensus, so that Byzantine or malicious miners cannot jeopardize the system security in any shard. As long as each shard is secure, the sharding method satisfies the data correctness property.

Although sharding is a promising approach, the cross-shard transaction is generally slow. As illustrated in Fig. 3, the cross-shard transaction involves multiple shards (e.g., a miner in shard 2 transfers some cryptocurrency to the miner in shard 3). The system needs cross-shard communication to process these transactions and requires confirmation from multiple shards to complete the transaction. An important future direction is to design a cross-shard communication protocol that supports fast cross-shard transactions.

Another approach to cut down the storage consumption is state pruning, which is used by some works (e.g., [13]) and the Ethereum protocol. To achieve the efficiency objective, this method reduces the data amount by removing the obsolete data. Take the current Ethereum practice as an example, the pruning method can cut down the storage consumption by one magnitude. On Oct. 2022, the archival node size is 11.8 terabytes while a full node (i.e., after pruned) is 0.95 terabytes [14]. To meet the security requirement, the pruning method only removes the obsolete data and keeps the latest data. Hence the miners can reach consensus the same way as in the existing blockchain protocol. To satisfy the data correctness property, the pruning method also introduces the “transaction replay” to recreate the complete data from the pruned data. However, such a transaction replay process generally takes a long time. Hence designing a quick transaction replay method for the pruned data is an important future direction.

Moreover, some prior studies (e.g., [6]) proposed the compact block and compact transaction design to cut down the size of a transaction or a block. These studies generally designed new data structures to cut down the size of the necessary information in a transaction or a block while satisfying the security and correctness property.

Overall, the technological approach does not reveal the fundamental economic reason behind the storage sustainability problem. Therefore, in the next subsection, we will consider the economical approach.

B. Economical Approach

In our recent works [7] [8], we analyzed the reason for storage sustainability in the existing blockchain protocol and proposed several economic mechanisms that mitigate the storage sustainability problem.

1) Reason of Storage Sustainability: We first use a two-miner example to illustrate how miners include transactions in blockchain, as illustrated in Fig. 4.

- During the process of mining block 1: There are three transactions $tx_1$, $tx_2$, and $tx_3$, pending to be included in blockchain. Suppose that miner 1 chooses to include $tx_2$ and $tx_3$ to his block to be mined, while miner 2 chooses to include $tx_1$. Assume that miner 2 succeeds in generating the block 1 and includes $tx_1$ to the blockchain as part of the newly generated block 1. Then both miners need to store block 1 and the transaction within it, according to the blockchain protocol discussed in Section II-A.

- During the process of mining block 2: As $tx_1$ is already included in the blockchain, there are two pending transactions $tx_2$ and $tx_3$. Suppose that miner 1 chooses to include $tx_2$ and $tx_3$, while miner 2 chooses to include no transaction. Assume that miner 1 succeeds in generating the block 2 and includes both $tx_2$ and $tx_3$ to the blockchain. Both miners store the newly generated block 2 according to the blockchain protocol.

The analysis of miners’ behavior reveals the negative externality of transaction selection. When each miner selects transactions, such as miner 1 selecting $tx_2$ and $tx_3$ when mining block 2 in Fig. 4, it will generate storage costs for...
This article has been accepted for inclusion in a future issue of this magazine.

Stage I: Mechanism Design

Stage II: Transaction Proposal

Stage III: Transaction Selection

Fig. 5: Three-stage Stackelberg game.

all other miners in the blockchain. This is the consequence of the replicated nature of blockchain, where each miner is supposed to store all the data from the chain’s inception to the present. Neglecting such a negative externality is the fundamental economic reason behind the storage sustainability problem. Specifically, it causes each miner to be willing to accept any transaction with a fee higher than his own storage cost, which is only a small fraction of all miners’ storage costs. Hence, we design new incentive mechanisms to address such a problem.

2) Incentive Mechanisms: The incentive mechanism should satisfy the following properties:

- **Storage sustainability**: The mechanism should encourage users to pay sufficient transaction fees for the storage costs in blockchain. This gives miners enough incentive to store the data and stay in the system.
- **Individual rationality**: The mechanism should satisfy users’ individual rationality, meaning that each user still gets a non-negative payoff when he pays sufficient transaction fees for storage costs.
- **Social optimality**: The mechanism should achieve the social optimality, meaning that it would select the outcome that is preferred by the system as a whole. Ideally, the storage sustainability requirement does not cause any social welfare loss.

In our recent works [7] [8], we proposed incentive mechanisms to achieve these three properties. Based on the description in Section II-A, the protocol designer, users, and miners make decisions in a sequential manner. Hence, we model their sequential decision process as a three-stage Stackelberg game, illustrated in Fig. 5.

1) **Stage I (Mechanism Design)**: The protocol designer first designs the incentive mechanism of the blockchain. Specifically, the mechanism design is an optimization problem, where the objective function is to maximize the social welfare (i.e., achieve the social optimality), subject to the constraint of ensuring storage sustainability (i.e., transaction fees are higher than storage costs). The incentive mechanism should satisfy the individual rationality property, where the optimal mechanism ensures that a user’s utility from a transaction is non-negative.

2) **Stage II (Transaction Proposal)**: Following the mechanism announced in Stage I, each user decides on the transaction generation rate and corresponding transaction fees to maximize his payoff, which trades off the transaction fee payment and the waiting time. Specifically, if a user pays a higher transaction fee, he will experience a lower transaction delay, as miners generally first select top-fee-per-byte transactions.

3) **Stage III (Transaction Selection)**: After users’ transaction proposals, each miner selects a set of transactions to be included in each newly generated block. Each miner aims to maximize his payoff, which trades off the transaction fee rewards and the storage costs.

We consider two scenarios when designing the incentive mechanism.

In the heavy-user scenario, each user generates a sequence of transactions. Specifically, each user decides the transaction generation rate and optimizes the fraction of transactions at different fee levels. This setting applies to the scenario where users continuously generate transactions (e.g., smart contracts).

We proposed a Fee choice and Waiting Tax (FWT) mechanism in [7]. Specifically, the FWT mechanism assigns fee choices and the waiting tax to users. The fee choices are no less than the transaction’s storage costs. Hence, the optimal fee choices encourage users to pay sufficient transaction fees at the system equilibrium and address the storage sustainability problem. The optimal waiting tax lets users be more conserva-tive in transaction generation at the equilibrium, which avoids the transaction congestion and achieves the social optimality.

In the light-user scenario, each user proposes one transaction. Specifically, each user decides his transaction proposing probability for his single transaction and the corresponding transaction fee. This often happens in the case where light users propose transactions not very frequently (e.g., the majority of users in Ethereum are not active on a daily basis [15]).

We proposed a Fee lower bound and Transaction Expiration Time (FTET) mechanism in [8] to achieve storage sustainability. Specifically, the FTET assigns the transaction fee lower bound and transaction expiration time to users. The optimal transaction fee lower bound is sufficiently high to encourage users to pay sufficient transaction fees to address the storage sustainability problem. The optimal transaction expiration time sets an upper bound on the transaction waiting time. Hence, when many users experience long transaction waiting times, users prefer the transaction expiration. This helps the system’s Nash equilibrium achieve the social optimality.

IV. FUTURE CHALLENGES IN STORAGE SUSTAINABILITY APPROACH

Despite the recent efforts described above, there still exist some future challenges in both technological and economical...
aspects to address the storage sustainability problem.

1) Multi-shard Storage in Shard Blockchain: Some existing shard blockchain [12] requires some miners in the system to store multiple shards of data to verify cross-shard transactions to ensure the data correctness. This requirement significantly cuts down the benefit of sharding. Moreover, there is generally a lack of proper incentive mechanisms to encourage miners to store multi-shard data. Hence, few miners are willing to store multi-shard blockchain data, which may lead to the failure of verifying cross-shard transactions.

There are two approaches to address this problem. One approach is to design a sharding system that does not require any miners to store the multi-shard data. However, this can be challenging given the potential presence of Byzantine or malicious miners. The other approach is for the protocol designer to design an incentive mechanism to encourage miners to store multi-shard blockchain data. For example, if a miner can present the proof-of-storage of multi-shard data, he can get some bonus from the blockchain system.

2) Incomplete Network Information in Incentive Mechanism: Generally, the interaction of users and miners is realized without complete network information for any of the participants. For example, the protocol designer (i.e., online blockchain community) is unaware of each miner’s actual storage cost of running the blockchain protocol. However, such information is crucial for the protocol designer to determine whether the system meets the storage sustainability requirement. Moreover, the parameters of a social-optimal mechanism are related to the user’s tolerance on waiting time (e.g., [7] [8]), which is often unknown to the protocol designer.

For such a challenge, the protocol designer needs to design proper market mechanisms, such as auctions, to elicit the private information of users and miners. In such a mechanism, users act as bidders proposing transaction fees and miners act as auctioneers selling the blockchain’s storage space for transactions. However, designing such an auction is challenging as the mechanism needs to be incentive compatible to bidders (i.e., users), robust to strategic auctioneers (i.e., miners), and secure under malicious behaviors.

3) Compensation for Future Miners’ Storage Costs: In the blockchain system, miners may come and leave the network at any time. Therefore, one question that arises is how users can pay for the storage costs of a future incoming miner.

One possible solution is to design and implement a storage rent mechanism. For example, Ethereum online community has proposed a similar approach in “EIP-1682”. In the storage rent mechanism, the blockchain system can estimate the storage costs per byte in each time slot based on the number of miners and set the storage renting fee. Then each user can choose either to pay the storage renting fee to keep his data in the blockchain or remove the data. However, the storage rent mechanism still faces some challenges. From the economical perspective, a miner may have the incentive to make clones and claim more storage rent, hence the storage rent mechanism should discourage each miner from making clones. From the technological perspective, every user needs to interact with the mechanism hence an efficient implementation is critically important.

V. CONCLUSION

Blockchain is expected to play a crucial role in many applications, for ensuring the transaction security with decentralization, transparency and verifiability, immutability, and independence of third parties properties. In this article, we reviewed mechanisms helping achieve storage sustainability from both technological and economical perspectives. From the technological perspective, we introduced the sharding, state pruning, and compact design methods. From the economical perspective, we revealed that the storage negative externality is the key reason for storage sustainability. We further introduced a fee choice and waiting tax incentive mechanism for heavy users and a fee lower bound and transaction expiration time incentive mechanism for light users to address the storage sustainability problem. Finally, we discussed some future challenges for achieving storage sustainability and their possible solutions.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China (Project 62271434), the Shenzhen Science and Technology Program (Project JCYJ20210324120010132 and JCYJ20210324124205016), the Guangdong Basic and Applied Basic Research Founda-tion (Project 2021B1515120008), the Shenzhen Key Lab of Crowd Intelligence Empowered Low-Carbon Energy Network (No. ZDSYS20220606100601002), the Shenzhen Institute of Artificial Intelligence and Robotics for Society, and the City University of Hong Kong’s Research Grants (Project 9610493 and 7005685).

REFERENCES

This article has been accepted for inclusion in a future issue of this magazine.

This article has been accepted for inclusion in a future issue of this magazine.


AUTHOR BIOGRAPHIES

YUNSHU LIU [S’20] (ly018@ie.cuhk.edu.hk) is currently pursuing a Ph.D. degree at the Department of Information Engineering, Chinese University of Hong Kong. His research interests include network economics, game theory, and control theory, with the current emphasis on blockchain and its applications.

ZHIXUAN FANG [M’18] (zfang@mail.tsinghua.edu.cn) is a tenure-track assistant professor at the Institute for Interdisciplinary Information Sciences (IIS) at Tsinghua University, Beijing, China. He is also with Shanghai Qi Zhi Institute. He received his Ph.D. degree in computer science from Tsinghua University in 2018. His research interests are in the design and analysis of multi-agent systems and networked systems.

MAN HON CHEUNG (mhcheung@cityu.edu.hk) is an Assistant Professor in the Department of Computer Science at the City University of Hong Kong. He received his Ph.D. degree in Electrical and Computer Engineering from the University of British Columbia in 2012. His research interests are in the topic of network economics and optimization, including wireless networking, crowdfunding, and blockchain systems.

WEI CAI [SM’22] (caiewei@cuhk.edu.cn) is currently an Assistant Professor of Computer Engineering, School of Science and Engineering at The Chinese University of Hong Kong, Shenzhen, China. He received his Ph.D. degree in Electrical and Computer Engineering from the University of British Columbia in 2016. His research interests are in the topic of human-centered computing for metaverse, including blockchain, digital game, Web 3.0, and computational art.

JIANWEI HUANG [F’16] (jianweihuang@cuhk.edu.cn) is a Presidential Chair Professor and Associate Vice President of the Chinese University of Hong Kong, Shenzhen. He is also the Vice President of Shenzhen Institute of Artificial Intelligence and Robotics for Society. He is an IEEE Fellow, an IEEE ComSoc Distinguished Lecturer, and a Clarivate Analytics Highly Cited Researcher. More information at https://jianwei.cuhk.edu.cn/.