A Review on Scalability of Blockchain

Di Yang
College of Computer
National University of Defense Technology
Changsha, China
469481032@qq.com

Chengnian Long
School of Electronic,
Information and Electrical Engineering
Shanghai Jiaotong University, Shanghai, China
longcn@sjtu.edu.cn

Han Xu
College of Computer
National University of Defense Technology
Changsha, China
xuhan_email@aliyun.com

Shaoliang Peng*
College of Computer Science and
Electronic Engineering
Hunan University, Changsha, China
pengshaoliang@nudt.edu.cn

ABSTRACT
As one of the key technologies of distributed ledgers, blockchain solves the trust problem in open network without relying on any trusted third party. Its decentralized feature has a broad application prospect, but still faces scalability problem. Currently, blockchain scalability bottleneck is mainly in three aspects: performance inefficiency, high confirmation delay, and function extension. For example, Bitcoin can only deal with 7 transactions per second averagely. Obviously, it cannot meet the requirement of current digital payment scenarios, nor can it be carried in other applications such as distributed storage and credit service. What's more, different blockchain systems carry different business and requirements, so scalability is the core issue of the current development of blockchain. This paper introduces the blockchain scalability related technologies from the aspects of improving efficiency and extending functionality of blockchain system, respectively. We summarize four mainstream solutions to improve the performance of blockchain system, including Sharding mechanism, directed acyclic graph based (DAG-based), off-chain payment network and cross-chain technology. In the end, we give some suggestions for further research in blockchain scalability.

CCS Concepts
•General and reference →Surveys and overviews •Networks →Network performance analysis

Keywords
Blockchain; Scalability; Performance; Review

1. INTRODUCTION
Blockchain technology originated in the Bitcoin system [1], and its implementation in an open peer-to-peer(P2P) network really does not depend on the trusted third party payment system. This kind of decentralized characteristics significantly different from existing business payment system, and change the existing system of security trust model. In recent years, blockchain has attracted tremendous attention from practitioners and academics in different disciplines (including law, finance, and computer science) due to its salient features which include distributed structure, immutability, security and privacy [2].

However, the current blockchain system has some scalability bottlenecks:

1) Low transaction efficiency. "Tmall" is the most famous e-commerce shopping platform in China. On November 11, "Tmall" turnover exceeded 38 billion dolloars, and it's order creation peak reaching 544,000 transactions per second(TPS) [3]. As a comparison, current performance of the mainstream blockchain is that bitcoin can only conduct about 7 TPS, and Ethereum is 30 TPS [4]. None of these blockchains can currently support the mainstream transaction usage.

2) High confirmation delay. Because the bitcoin block generation speed is 10 minutes, the confirmation delay is at least 10 minutes, and because of the bifurcation phenomenon, it is customary to consider that the transaction has been confirmed after 60 minutes. Even the better Ethereum also takes 18 seconds to complete the confirmation.

3) Function extension. Different scenarios have different users and needs, so in reality, it is difficult to request a blockchain system to carry all applications in real life. Therefore, we need to implement the interaction between the chains, making blocks not become isolated with each other, so as to realize the real value of interconnection.

This paper introduces and analyzes the blockchain scalability related technologies from the aspects of improving efficiency and extending functionality of blockchain system. Although, there have been some related surveys recently, for example, [5] and [6]. Compared to them, we summarize some projects that have adopted these approaches to improve the performance of blockchain system, and discuss the future research direction.

2. SOLUTIONS
In this section, we will summarize four mainstream solutions to improve the performance of blockchain system, including Sharding mechanism, DAG-based, off-chain payment network and cross-chain technology.

ICBCT'20, March 12–14, 2020, Hilo, HI, USA

ACM ISBN 978-1-4503-7767-6/20/03…$15.00
https://doi.org/10.1145/3390566.3391665

2.1 Sharding Mechanism

Sharding was originally a database partitioning technique that allowed a large database to be split into smaller pieces of data and stored on different servers, allowing it to manage data faster and more efficiently. In 2016, Luu et al. published a paper, which proposed the concept of Sharding in the field of blockchain for the first time [7]. Its general design idea is: Turn each block in the blockchain network into a sub-blockchain, and sub-blockchain can accommodate several (currently 100) collation packaged with transaction data. These collation finally constitute a block on the main chain; Because these collation exists as a whole in a block, and its data must be packaged and generated by a specific miner, essentially the same as the blocks in the existing protocol, so no additional network confirmation is required. That would increase the trading capacity of each block by about 100 times. And this kind of design also has the advantage to expand in the future.

Elastico and Zilliqa are two typical projects using Sharding mechanism. They both adopt PoW to prove as Sharding algorithm, and the scheme during the consensus adopte PBFT algorithm [8]. In order to resist the Sybil attack [9], at the beginning of a consensus, nodes need to conduct simple work prove to get involved in PBFT identity. The criteria for dividing nodes into different sets are based on the result of PoW. By establishing a probability model, it can be obtained that when the sharding size reaches 600, the probability that the attacker can control a sharding (i.e. having more than 1/3 nodes in any shard) is negligible, even if the attacker has 1/3 computing power. The specific process can be abstracted as follows:

a) The node obtains its identity through PoW and divides it into different sets. b) Conduct transaction consensus within each sharding through PBFT algorithm. c) The transaction set after the consensus segmentation and the signature in the consensus process are broadcast to a certain shard. The shard verifies the signature, conducts the consensus segmentation, and packages the consensus into blocks and broadcasts the whole network.

Elastico's solution is based on the UTXO model [10]. By making transactions on the main chain and creating a receipt (with receiptID), the user can store the data in a specified sharding. And the user on the Sharding chain can create a receipt-consuming transaction to spend the receipt given receiptID. Elastico can therefore effectively resist double spend attack during transaction processing, so that when processing a trade, it maps to different Sharding processes via the input to the trade as a baseline.

Zilliqa solution model based on the account model. When the transaction processing, it maps to different shards by the identity of the sender as a baseline. In the consensus process, different sender trading could be mapped to different subdivision (different validation), but the same sender transactions will be handled by the same sharding, so the trusted nodes in the sharding are determined to the state of a particular sender. Therefore, it can resist double spend attack model based on the account.

2.2 DAG-based

Using a DAG as a distributed ledger is not about removing proof-of-work mining, blocks, or transaction fees. It is about leveraging the structural properties of DAGs to potentially solve blockchain's orphan rate problem [11]. The ability of a DAG to withstand this problem and thus improve on scalability is contingent on the additional rules implemented to deal with transaction consistency, and any other design choices made.

The DAG blockchain mainly improves the data layer and consensus layer of bitcoin blockchain, as shown in figure 1.

It follows the P2P network structure of bitcoin blockchain to organize the nodes of the whole network. On the bitcoin blockchain, newly released blocks will be added to the original longest chain, which will be extended indefinitely according to the longest chain considered by all nodes.

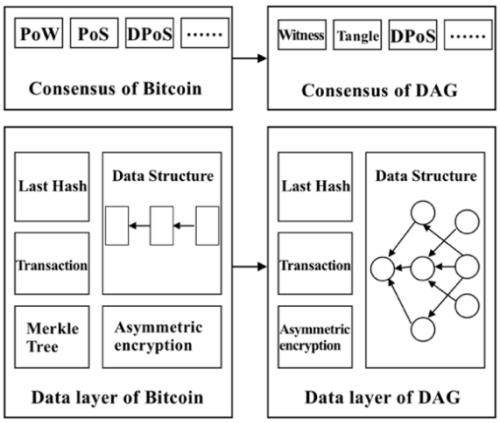

Figure 1. The difference between DAG and Bitcoin

In DAG, when a network node wants to initiate a new transaction, it needs to find two (or more) other transactions in the network to verify and point its newly initiated transaction to these two transactions. The whole network expands gradually. Because of this design, the responsibility of verifying transactions in the whole network is transferred from traditional miners to each node of the network [12]. This design urges every node in the network to actively participate in verifying other transactions.

Table 1. G versus traditional blockchain

Projects	Data structure	Consensus	Open source
Bitcoin	Block	PoW	Yes
Ethereum	Block	Casper	Yes
NXT	Block- DAG	PoS	Yes
Byteball	DAG	Witness	Yes
DagCoin	DAG	Witness	No
Nano	DAG	DPoS	Yes
IOTA	DAG	Tangle	Partial

Table 1 summarizes the data structure, consensus, and open source of several bitcoin blockchains and DAG-based blockchains. In terms of data structure, bitcoin blockchain packages multiple transactions into blocks and forms a stable chain between blocks, while Byteball [13], Dagcoin, Nano and IOTA blockchain, each transaction as well as the parents unit hash value form a basic unit, closely connection between each unit through the hash value, thus forming a directed acyclic graph. In terms of architecture, the listed block chain adopts P2P network structure. In terms of the consensus mechanism, bitcoin adopts the PoW consensus mechanism, while ethereum adopts the mixed Casper consensus mechanism of PoW+PoS. In the DAG diagram block chain, NXT adopts the PoS consensus mechanism, IOTA adopts Tangle consensus mechanism. Dagcoin system is modified based on Byteball's source code, so it also temporarily adopts the same Witness consensus mechanism as Byteball.

2.3 Off-chain Payment Network

Such solutions are generally proposed on the original public chain, such as Bitcoin's Lightning Network [14] and Ethereum's Raiden Network [15], which mainly solve the payment problem.

The two strategies are to keep the underlying blockchain protocol unchanged, put transactions under the chain for execution, and solve scalability problems by changing the way the protocol is used. The part of the off-chain can be implemented with the traditional centralized distributed system, and the performance is scalable. Under this strategy, only coarse-grained ledgers are recorded on the distributed ledger, while truly fine-grained details of bilateral or limited multilateral transactions are not recorded as transactions on the distributed ledger. The disadvantage is that there is a centralized system.

2.3.1 Lightning Network

Lightning network is the earliest scheme to form a payment network through the payment channel under the chain and improve the transaction throughput of the block chain. It consists of a blockchain-based downlink transport network that works at the P2P level, and its availability depends on the creation of a two-way payment channel through which users can conduct seamless cryptocurrency transactions. To create a payment channel, both parties need to set up a multi-signature wallet and store some funds that can only be accessed if both parties provide private keys. After the two sides decide to open a payment channel, they can transfer money back and forth in their wallets. Although the process of establishing a payment channel involves transactions on the chain, all transactions that take place within the channel are on the chain and therefore do not require a global consensus. As a result, these trades can be executed quickly through smart contracts, allowing for higher TPS while paying lower fees.

However, there are some limitations of Lightning networks. If the receiver of both parties is not online, Lightning online payment cannot be made. In order to ensure the security of funds, it is necessary to monitor the payment channel regularly. Lightning networks are not suitable for large payments, because they rely on a large number of multi-signature wallets (basically Shared wallets), so they probably do not have enough balance to act as intermediaries for large payments. Creating and closing payment channels involves on-chain transactions that require manual operation and may incur high transaction costs.

2.3.2 Raiden Network

Model based on Ethernet, Raiden network reference the structure of the Lightning network [15]. Since Raiden Network is a complementary network, it users a channel to deal with a number of deals, some of the encryption algorithm is then used to record and verify the actual transaction data under the chain. Finally, when the Channel is shut down, the transaction data is shut down. Actual cryptocurrency transactions and authentication are carried into the blockchain. In this way, the actual number of transactions on the blockchain will be reduced and transaction costs will be reduced and accelerated.

It works like the bar tab, because you only pay the total amount to the bar at the end of the day, instead of going through the entire payment process every time you buy a drink. Each bar tab is called a channel.

Any particular channel is one-to-one (for example, Alice to Bob) and the channels can be linked together to form a network so that users can pay anyone in the network. The figure 2 shows that even if F has only channel A, anyone in the network can pay F through the channel interaction associated with A.

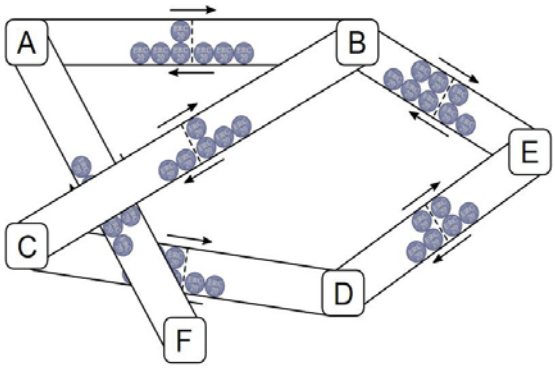

Figure 2. Process of Raiden network [15]

Since only two participants can receive deposits in the payment channel's smart contracts, payment channel transfers are not affected by double spend attacks, making them as secure as transactions on the chain.

2.3.3 Sprites

Miller et al. proposed the scheme of Sprites [16], which is aimed at Lightning network solution efficiency improvements, and the goal is to solve the problem of excessive time cost consumed by users when cross-channel payment is unsuccessful. Obviously, a channel payment protocol can be thought of as a debit card mechanism, that is, a user can provide cross-channel payments to others through the amount of money saved in the channel. A cross-channel fee is an incentive for the user to pay cross channels, similar to the interest on a debit card. However, during the process of cross-channel payment, the funds used for payment are frozen until the cross-channel payment succeeds or fails. In this process, the user cannot use the frozen funds, so the cost paid by the user in the cross-channel payment can be measured by the product of the frozen funds and the frozen time in the payment process.

The payment process of Sprites is similar to that of Raiden networks, and the difference is that cross the channel condition when payment agreement. HTLC is based on the condition of the Hash as published, in the terms of payment condition in the Sprites to Hash the original as the output of the management contract. Once the pay of a node on the path to the Hash as a management contract submitted to Hash the original as, pathway on other nodes can be learned that Hash to the original at the same time as the release of the results, and cross the channel payment completed.

2.3.4 Plasma

Plasma [17] is also known as “blockchains in blockchains”. Anyone can create Plasma on top of the underlying blockchain to support different business needs, such as distributed exchanges, social networks, games, etc.

This is done by using a number of sub-chains. The system connects the sub-chains to the main chain through fraud proofs. It can be used recursively to generate sub-chains and create chain trees. The result is that we can perform many complex transactions on sub-chains, which are faster and charge lower transaction fees because they don't require the entire blockchain to process.

For each sub-chain, it is necessary to create intelligent contracts on the main chain and clarify the rules of the sub-chain, such as block verification of sub-chain, token conversion between parent chain and sub-chain, etc. At the same time, the sub-chain block out the

need to lock in the parent chain funds, for subsequent accountability.

The business data is stored in the sub-chain, whose block hash is synchronized to the parent chain as a credential. Any third party can be responsible for monitoring the operation of the sub-chain. If any fraud is found, the problem block can be submitted to the parent chain for verification to facilitate the accountability and rollback of the problem block.

As long as the user provides the latest transaction results on the sub-chain, the fund can be safely returned to the parent chain. The parent chain does not supervise each block of the sub-chain and cannot confirm the real-time performance of the transaction results provided by the user. The fund refund needs to wait for a challenge period to prevent user fraud. The mechanism is similar to the state channel.

2.3.5 TrueBit

TrueBit is a technique that helps Ethereum perform heavy or complex calculations under the chain [18]. This makes it different from Plasma, which are more useful for increasing the total transaction throughput of the Ethereum blockchain. TrueBit won't allow to complete more transactions, but it will allow Ethereum based applications to handle more complex transactions and still be validated by the main chain.

The basic process of TrueBit is as follows.

a) The user (Task Giver) uploads the code required to be executed (Task) and provides the commission. b) A third party (Solver) discovers the task, thinks the commission is acceptable, performs the task and publishes the result, along with a deposit. c) Another third-party Verifier reexecutes the task and can challenge the Solver if it is found to be fake, again with a security deposit. d) The smart contract on the chain allows Solver and Verifier to play a verification game, and verify the authenticity of the answer through the implementation code provided by the Task Giver on the chain. The one who provides the right answer gets the commission, and the one who falsifies pays the gas required for the whole verification process from the deposit. e) If no one can prove the fraud during the challenge period, Solver will get commission.

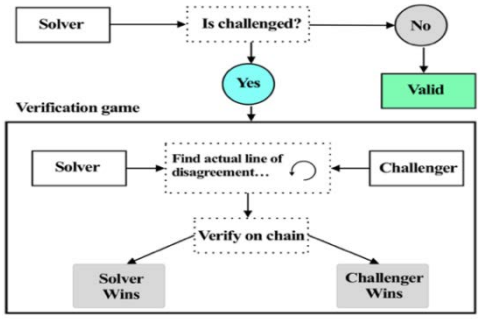

Figure 3. Process of verification game [18]

2.4 Cross-chain Technology

Cross-chain is the technology to solve the problem of the interaction between the chain and chain. The process can be divided into two stages: assets on A chain of locking phase and the corresponding assets on the chain B unlock stage. The main challenges are how to ensure that the asset on chain A is locked, how to determine the asset on chain B is unlocked, and how to guarantee the atomicity of asset locking and unlocking between chain A and B, that is, the corresponding asset between two chains either locks and unlocks successfully or locks and unlocks fail. For the above challenges, different cross-chain technologies were proposed, mainly including 4 categories.

2.4.1 Multi-center Witness

The Witness mechanism uses witness to guarantee the locking and unlocking of assets in different chains. The multi-signature script in the block chain script is mainly used to realize the two-way exchange between the chains. The specific process is as follows: the user transfers the asset of chain A to the multi-signature script address of several witnesses for locking, and the witnesses releases the corresponding asset to the address of the user in chain B after confirmation. This technology is applied in Byteball and DagCoin.

2.4.2 Side Chain/Relay Technology

The side chain is based on anchoring some kind of general certificate on the original chain, just like the dollar anchoring to gold. Side chains are connected to various chains, while other blockchains can exist independently [19].

The representative of side chain technology is BTC Relay [20]. It is considered the first side chain on the blockchain. BTC Relay connected the Ethereum network to the bitcoin network by using Ethereum's smart contracts, allowing users to verify bitcoin transactions on Ethereum. It creates a smaller version of the bitcoin blockchain through Ethereum’s smart contracts, which require access to bitcoins' network data, making decentralization difficult. The specific process is as follows:

a) Alice and Bob use the smart contract to conduct transactions. Alice users BTC coin to exchange Bob's ETH coin, and Bob sends his ETH coin to the smart contract; b) Alice sends BTC coins to Bob's address. c) Alice generates the SPV certificate through the transaction information of bitcoin and inputs the certificate into the contract in the ETH system; d) After the contract is triggered, confirm the SPV certificate, and then release Bob's ETH coin to Alice's address before.

In addition, another typical implementation is Cosmos. It is a heterogeneous network developed by the Tendermint team that supports cross-chain interaction [21]. Cosmos adopts the Tendermint consensus algorithm [22], which is similar to the practical Byzantine fault-tolerant consensus engine [23] with high performance, consistency, and under its strict bifurcated liability system, can prevent malicious participants from making improper operations.

The central and individual Spaces of the Cosmos network can be communicated through the block chain communication (IBC) protocol, which is specific to the block chain network, similar to the UDP or TCP network protocols, as shown in figure 4. Tokens can be transferred safely and quickly from one space to another without the need for exchange liquidity. Instead, all transfers of tokens within the space go through the Cosmos center, which records the total number of tokens held per space. This center separates each space from the other fault Spaces. Because everyone can connect the new space to the Cosmos center, Cosmos is also compatible with future blockchains.

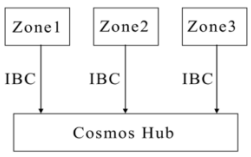

Figure 4. Structure on cosmos relay network

2.4.3 Hash Locking

Hash locking works in the same way as HTLC in Lightning networks. By using Hash preimage as secret and conditional payment, the atomicity of different transactions can be guaranteed without the participation of trusted third parties, so as to realize fair cross-chain exchange.

As shown in figure 5, the process of cross-chain atom exchange is as follows [24]:

a) A generates random number r and calculates its Hash value h, then send h to B.

b) A and B lock up the assets for exchange successively by using HTLC. The locking time of A shall be longer than that of B, that is, T1 < T2. From the perspective of A, in time T1, B can obtain the asset locked by A through public image r, otherwise A can redeem its asset. From the perspective of B, in time T2, A can obtain the asset locked by B through public image r, otherwise B can redeem its asset.

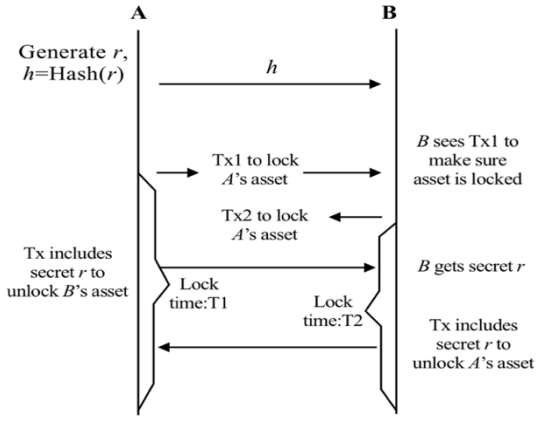

Figure 5. Process of cross-chain atomic swap

c) A obtains the locked assets of B by publishing the preimage r. Meanwhile, B obtains secret r and obtains the locked assets of A on another chain by publishing r.

2.4.4 Distributed Private Key Control Technology

Private assets are mapped to the blockchain of built-in asset templates based on protocols through distributed private key generation and control technology, and new assets are created based on the deployment of new smart contracts which based on cross-chain transaction information. When a registered asset is transferred from the original chain to the cross-chain, the cross-chain node will issue corresponding tokens of equivalent value to the user in the existing contract. In order to ensure that the original chain assets can still trade with each other across the chain, the operation of realizing and unlocking distributed control management is called lock-in and lock-out. Lock-in is the process of implementing distributed control management and asset mapping for all digital assets controlled by keys. The decentralized network needs to be entrusted with the user's private key, and the user holds the private key of the agent asset across the chain. When unlocked, control of the digital asset is returned to the owner.

3. CONCLUSION

Blockchain technology has decentralized, tamper-proof, and programmable features, which makes it have broad application prospects. However, the scalability bottlenecks it faces, including low performance efficiency and difficulty in extending functions, restrict the application of blockchain technology.

On the one hand, blockchain system needs to expand computing and storage capacity through the offline system. On the other hand, the existing system chain needs to dock with the blockchain to solve the problem of information island and tamper-proof. Some suggestions for further research are as follows.

1) Large-scale high-performance P2P network. Around blockchain applications, the original large-scale P2P network is the most important, because blockchain is originally P2P transmission. If there is no breakthrough in network technology, the performance of blockchain system is difficult to improve.

2) Modular security cryptography protocol. Blockchain is distributed with passwords, In other words, blockchain systems embed different security cryptography protocol modules. Therefore, cryptography is also a central issue in blockchain.

3) High-performance programmable computing engine. It is necessary for different smart contracts using different programming languages, and system ecological construction is also very important.

In a word, this paper summarizes the solutions and technologies to solve the scalability problem of blockchain in recent years, discusses the solutions to improve the scalability of blockchain, and discusses the problems and directions that need further research

4. ACKNOWLEDGMENT

This work was supported by National Key R&D Program of China 2017YFB0202602, 2018YFC0910405, 2017YFC1311003, 2016YFC1302500, 2016YFB0200400, 2017YFB0202104; NSFC Grants 61772543, U1435222, 61625202, 61272056; The Funds of Peng Cheng Lab, State Key Laboratory of Chemo/Biosensing and Chemometrics, the Fundamental Research Funds for the Central Universities, and Guangdong Provincial Department of Science and Technology under grant No. 2016B090918122.

5. REFERENCE

[1] Nakamoto, S. 2008. Bitcoin: A peer-to-peer electronic cash system.

[2] Abramaowicz, M. 2016. Cryptocurrency-based law, *Ariz. L. Rev.* vol. 58, p. 359.

[3] Ifeng. 2019. http://finance.ifeng.com/c/7rVjlivAS6v.

[4] Gervais, A., Karame, G. O., Wüst, K. et al. 2016. On the security and performance of proof of work blockchains. *Proc of the 2016 ACM SIGSAC Conf on Computer and Communications Security.* New York:ACM, 2016:3-16

[5] Siris, V., Nikander, P., Voulgaris, S., Fotiou, N., Lagutin, D. Polyzos, G. 2019. Interledger Approaches. *IEEE Access*. PP. 1-1. DOI=10.1109/ACCESS.2019.2926880.

[6] Kim, A., Soohyeong. Kwon, Yongseok. Cho, Sunghyun. 2018. A Survey of Scalability Solutions on Blockchain. 1204-1207. DOI=10.1109/ICTC.2018.8539529.

[7] Luu, L., Narayanan, V., Zheng, C. D., Baweja, K., Gilbert, S., Saxena, P. 2016. A Secure Sharding Protocol For Open Blockchains. DOI=10.1145/2976749.2978389.

[8] Miguel, C., Barbara, L. 1999. Practical byzantine fault tolerance. *In Proceedings of the Third Symposium on Operating Systems Design and Implementation*, vol. 99, New Orleans, USA, 1999, pp. 173–186.

[9] Douceur, J. R. 2002. The sybil atack. *Proc of the Int Workshop on Per-to-Per systems*. Berlin:Springer. 2002: 251-260

[10] Ozyilmaz. Kazim Rifat. Patel. Harsh. Malik. 2018. Ankit Split-Scale: Scaling Bitcoin by Partitioning the UTXO Space. *ICSESS*.

[11] DAGfans. github. https://github.com/DAGfans/TranStudy/

[12] Huang, J., Kong, L., Chen, G., et al. 2019. Towards Secure Industrial IoT: Blockchain System with Credit-Based Consensus Mechanism. *IEEE Transactions on Industrial Informatics*, 2019:1-1.

[13] Byteball. https://www.chainwhy.com/whitepaper/gbytewhitepaper.html

[14] Poon, J., Dryja, T. The bitcoin lightning network:Scalable of f-chain instant payments. https://lightning.network/lightning-network-paper. pdf

[15] Raiden Network. 2019. https://raiden.network/.

[16] Miler, A., Bentov, I., Kumaresan, R. et al. 2018. Sprites:Payment channels that go faster than lightning. https://arxiv.org/pdf/1702.05812.pdf.

[17] Ziegler, Michael, H. 2019. Integration of Fog Computing and Blockchain Technology Using the Plasma Framework, *IEEE ICBC*.

[18] TrueBit. https://learnblockchain.cn/2018/03/24/truebit-whitepaper/

[19] Back, A., Coralo, M., Dashjr, L. et al. Enabling blockchain innovations with pegged sidechains. https://www.blockstream.com/sidechains.pdf

[20] BTC-relay. BTC-relay. http://btcrelay.org

[21] Cosmos Whitepaper. https://cosmos.network/resources/whitepaper

[22] Kwon, J. Tendermint: Consensus without mining, http://tendermint. com/docs/tendermintv.

[23] Miguel, C., Barbara, L. Practical byzantine fault tolerance. 1999. *In Proceedings of the Third Symposium on Operating Systems Design and Implementation*, vol. 99, New Orleans, USA, 1999, pp. 173–186.

[24] Pan, C., Liu, Z. Q., Liu Z. L. 2018. Research on Scalability of Blockchain Technology:Problems and Methods. *Journal of Computer Research and Development*.