Wuhui Chen
Zibin Zheng
Huawei Huang *Editors*

Blockchain Scalability

Springer

Wuhui Chen • Zibin Zheng • Huawei Huang
Editors

Blockchain Scalability

Springer

Editors
Wuhui Chen
GuangDong Engineering Technology
Research Center of Blockchain
Sun Yat-sen University
Guangdong, China

Zibin Zheng
GuangDong Engineering Technology
Research Center of Blockchain
Sun Yat-sen University
Guangdong, China

Huawei Huang
GuangDong Engineering Technology
Research Center of Blockchain
Sun Yat-sen University
Guangdong, China

ISBN 978-981-99-1058-8 ISBN 978-981-99-1059-5 (eBook)
https://doi.org/10.1007/978-981-99-1059-5

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

Blockchain technology has been the subject of extensive attention from government, enterprise, and academia due to its potential to create a trusted, decentralized environment for a variety of applications. However, the current blockchain faces a significant scalability bottleneck that limits its ability to meet the demands of large-scale practical applications. The bottleneck is mainly characterized by low performance efficiency and difficulty in functional extension, which pose significant challenges for realizing the full potential of blockchain technology.

Over the past few years, substantial progress has been made in blockchain scalability technologies. Various methods have been developed to improve the performance of blockchain or enable cross-chain technology for interoperability. However, the research in this field is still in its early stages of development.

This book aims to provide a comprehensive and state-of-the-art resource for researchers, engineers, policymakers, and students interested in understanding and addressing the scalability bottleneck problem in blockchain technology. The book adopts an approach that is based on the existing large-scale application scenarios, which provides readers with a comprehensive analysis of blockchain scalability issues, key technologies, and future directions. The book covers various areas related to blockchain scalability, including the root of blockchain scalability problems, mainstream blockchain performance, the classification of existing scalability problem solutions, exciting sharding-based approaches, open issues, and future directions to scale blockchain.

The book's comprehensive coverage of blockchain scalability issues and solutions makes it a valuable resource for anyone interested in understanding and addressing the scalability bottleneck problem in blockchain technology. We hope that this book will contribute to the realization of the full potential of blockchain technology by providing a holistic view of the challenges and opportunities in this field.

Guangdong, China

Wuhui Chen
Zibin Zheng
Huawei Huang

Acknowledgments

We would like to express our sincere appreciation to all those who have contributed to the completion of this book on blockchain scalability.

First, we extend our gratitude to the contributors who have shared their valuable expertise and insights on this complex and rapidly evolving topic. Their contributions have been instrumental in creating a comprehensive and up-to-date resource on the subject.

We are also grateful to the editorial team at Springer for their guidance and support throughout the publishing process. Their professionalism, expertise, and commitment to excellence have been crucial in making this book a reality.

Finally, we express our appreciation to the readers of this book. We hope that the book will be a valuable resource for researchers, engineers, policymakers, and others working in the area of blockchain scalability. We also hope that the book will inspire further research and innovation in this exciting and important field.

We acknowledge that the subject of blockchain and its applications is complex and rapidly evolving, and that there are many potential conflicts of interest that may arise in the course of researching, writing, and publishing a book on this topic. We have made every effort to disclose any conflicts of interest that we are aware of, and we welcome feedback and input from readers and other stakeholders on any potential biases or conflicts that may be present in the work. This book is partially supported by the National Key R&D Program of China (No. 2020YFB1006001), the National Natural Science Foundation of China under project (62032025, 62272496, 62172453), Fundamental Research Funds for the Central Universities, Sun Yat-sen University (Grant No. 23lgbj019), the National Natural Science Foundation of Guangdong province (2022A1515010154), the Major Key Project of PCL (PCL2021A06), the Program for Guangdong Introducing Innovative and Entrepreneurial Teams (2017ZT07X355), and the Pearl River Talent Recruitment Program (No. 2019QN01X130).

Contents

Chapter 1
Blockchain Scalability Fundamentals

Huawei Huang, Wei Kong, Sicong Zhou, Zibin Zheng, and Song Guo

1.1 Overview

Centralized security mechanisms are prone to Single Point of Failure, meaning that once a centralized component is compromised, the whole system would cease to function. The decentralization of blockchain can eliminate such concern without the need of a trusted third party. With the benefit of decentralized characteristics, blockchains have been deeply diving into multiple applications that are closely related to every aspect of our daily life, such as cryptocurrencies, business applications, smart city, Internet-of-Things (IoT) applications, and etc. The blockchain

H. Huang · Z. Zheng (✉)
GuangDong Engineering Technology Research Center of Blockchain, Sun Yat-sen University, Guangdong, China
e-mail: huanghw28@mail.sysu.edu.cn; zhzibin@mail.sysu.edu.cn

W. Kong · S. Zhou
School of Computer Science and Engineering, Sun Yat-Sen University, Guangzhou, China

S. Guo (✉)
Department of Computing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong, China
e-mail: song.guo@polyu.edu.hk

W. Chen et al. (eds.), *Blockchain Scalability*,
https://doi.org/10.1007/978-981-99-1059-5_1

theory and technology will bring substantial innovations, incentives, and a great number of application scenarios in diverse fields.

We have found that a survey of the state-of-the-art theories, modelings and useful tools that can (i) improve the performance of blockchains, and (ii) help better understand blockchains, is still missing. As a result, we made in-depth investigations on these directions and presented in this chapter includes the following contributions.

- We introduces the preliminaries of blockchains.
- We then present a comprehensive investigation on the state-of-the-art theoretical modelings, analytics models, performance measurements, and useful experiment tools for blockchains, blockchain networks, and blockchain systems.
- Several promising directions and open issues for future studies are also envisioned finally.

1.2 Preliminaries of Blockchains

Blockchain is a promising paradigm for content distribution and distributed consensus over P2P networks. In this section, we present the basic concepts, definitions and terminologies of blockchains appeared in this chapter. Due to the frequent use of acronyms in this book, we include an acronym table, i.e., Table 1 in the online supplementary material.

1.2.1 Prime Blockchain Platforms

1.2.1.1 Bitcoin

Bitcoin is viewed as the blockchain system that executes the first cryptocurrency. It builds upon two major techniques, i.e., *Nakamoto Consensus* and *UTXO Model*, which are introduced as follows.

Nakamoto Consensus To achieve an agreement of blocks, Bitcoin adopts the Nakamoto Consensus, in which miners generate new blocks by solving a puzzle. In such a puzzle-solving process, also referred to as mining, miners need to calculate a nonce value that fits the required difficulty level. Through changing the difficulty, Bitcoin system can maintain a stable rate of block-generation, which is about one block per 10 minutes. When a miner generates a new block, it broadcasts this message to all the other miners in the network. If others receive this new block, they add this block to their local chain. If all of the other miners receive this new block timely, the length of the main chain increases by one. However, because of the network delays, not always all the other miners can receive a new block in time. When a miner generates a block before it receives the previous one, a fork yields. Bitcoin addresses this issue by following the rule of longest chain.

UTXO Model The Unspent Transaction Output (UTXO) model is adopted by cryptocurrencies like Bitcoin, and other popular blockchain systems [1, 2]. A UTXO is a set of digital money, each represents a chain of ownership between the owners and the receivers based on the cryptography technologies. In a blockchain, the overall UTXOs form a set, in which each element denotes the unspent output of a transaction, and can be used as an input for a future transaction. A client may own multiple UTXOs, and the total coin of this client is calculated by summing up all associated UTXOs. Using this model, blockchains can prevent the double-spend [3] attacks efficiently.

1.2.1.2 Ethereum

Ethereum [4] is an open-source blockchain platform enabling the function of smart contract. As the token in Ethereum, *Ether* is rewarded to the miners who conducted computation to secure the consensus of the blockchain. Ethereum executes on decentralized Ethereum Virtual Machines (EVMs), in which scripts are running on a network consisting of public Ethereum nodes. Comparing with Bitcoin, the EVM's instruction set is believed Turing-complete. Ethereum also introduces an internal pricing mechanism, called *gas*. A unit of gas measures the amount of computational effort needed to execute operations in a transaction. Thus, gas mechanism is useful to restrain the spam in smart contracts. Ethereum 2.0 is an upgraded version based on the original Ethereum. The upgrades include a transition from PoW to Proof-of-Stake (PoS), and a throughput-improving based on sharding technologies. The comparison between Bitcoin & Ethereum is summarized in Table 1.1.

Account/Balance Model Unlike Bitcoin where states are composed by UTXOs, Ethereum adopts a more common and straightforward model that is used by banks, the Account/Balance Model. In every account, an incrementing counter of transaction execution, nonce, is implemented to prevent double spending attacks, which serves as a complement for the model's simple structure. There are basically 2 types of accounts, *external owned accounts* (EOAs) and *contract accounts* (CAs), each controlled by private keys and contract codes, respectively.

Table 1.1 Comparison between Bitcoin and Ethereum

	State model	Consensus protocols	Throughput
Bitcoin	UTXO	PoW	3 to 7 TPS[5]
Ethereum1.0	Account/balance	PoW	7 to 15 TPS[5]
Ethereum2.0	Account/balance	PoS sharding	Unknown

1.2.1.3 Hyperledger Fabric

Hyperledger Fabric [6] is a popular permissioned blockchain platform for industrial use. In industry, goals are quite different from cryptocurrency systems. Greater significance is attached to lower maintenance cost, higher throughput performance and permission control. For a node in a permissioned setting, other nodes, though untrusted, the identities are known. With different levels of trust among users, different consensus protocols can be customized for fault tolerant.

1.2.1.4 EOSIO

EOSIO [7] is another popular blockchain platform released by a company *block.one* on 2018. Different from Bitcoin and Ethereum, the smart contracts of EOSIO don't need to pay transaction fees. Its throughput is claimed to reach millions of transactions per second. Furthermore, EOSIO also enables low block-confirmatoin latency, low-overhead BFT finality, and etc. These excellent features has attracted a large-number of users and developers to quickly and easily deploy decentralized applications in a governed blockchain. For example, in total 89,800,000 EOSIO blocks have been generated in less than one and a half years since its first launching.

1.2.2 Consensus Mechanism

The consensus mechanism in blockchains is for fault-tolerant to achieve an agreement on the same state of the blockchain network, such as a single state of all transactions in a cryptocurrency blockchain. Popular proof-based consensus protocols include PoW and PoS. In PoW, miners compete with each other to solve a puzzle that is difficult to produce a result but easy to verify the result by others. Once a miner yields a required nonce value through a huge number of attempts, it gets paid a certain cryptocurrencies for creating a new block. In contrast, PoS doesn't have miners. Instead, the new block is forged by *validators* selected randomly within a committee. The probability to be chosen as a validator is linearly related to the size of its stake. PoW and PoS are both adopted as consensus protocols for the security of cryptocurrencies. The former is based on the CPU power, and the latter on the coin age. Therefore, PoS is with lower energy-cost and less likely to be attacked by the 51% attack.

1.2.3 Scalability of Blockchains

Blockchain as a distributed and public database of transactions has become a platform for decentralized applications. Despite its increasing popularity, blockchain

technology faces the scalability problem: throughput does not scale with the increasing network size. Thus, scalable blockchain protocols that can solve the scalability issues are still in an urgent need. Many different directions, such as *Off-chain*, *DAG*, and *Sharding* techniques, have been exploited to address the scalability of blockchains. Here, we present several representative terms related to scalability.

1.2.3.1 Off-Chain Techniques

Contrary to the on-chain transactions that are dealt with on the blockchain and visible to all nodes of the blockchain network, the off-chain transactions are processed outside the blockchain through a third-party guarantor who endorses the correctness of the transaction. The on-chain transactions incur longer latencies since the confirmation of an on-chain transaction has to take different steps. In contrast, the off-chain techniques can instantly execute the off-chain transactions because those transactions don't need to wait on the queue as on an on-chain network.

1.2.3.2 DAG

Mathematically, a DAG is a finite directed graph where no directed cycles exist. In the context of blockchain, DAG is viewed as a revolutionized technology that can upgrade blockchain to a new generation. This is because DAG is blockless, and all transactions link to multiple other transactions following a topological order on a DAG network. Thus, data can move directly between network participants. This results in a faster, cheaper and more scalable solution for blockchains. In fact, the bottleneck of blockchains mainly relies on the structure of blocks. Thus, probably the blockless DAG could be a promising solution to improve the scalability of blockchains substantially.

1.2.3.3 Sharding Technique

The consensus protocol of Bitcoin, i.e., Nakamoto Consensus, has significant drawbacks on the performance of transaction throughput and network scalability. To address these issues, *sharding* technique is one of the outstanding approaches, which improves the throughput and scalability by partitioning the blockchain network into several small shards such that each can process a bunch of unconfirmed transactions in parallel to generate medium blocks. Such medium blocks are then merged together in a final block. Basically, sharding technique includes *Network Sharding*, *Transaction Sharding* and *State Sharding*.

1.2.3.4 Cross-Shard Transactions

One shortcoming of sharding technique is that the malicious network nodes residing in the same shard may collude with each other, resulting in security issues. Therefore, the sharding-based protocols exploits *reshuffling* strategy to address such security threats. However, reshuffling brings the *cross-shard* data migration. Thus, how to efficiently handle the cross-shard transactions becomes an emerging topic in the context of sharding blockchain.

1.3 Theories to Improving the Performance of Blockchains

1.3.1 Latest Theories to Improving Blockchain Performance

Summary of this subsection is included in Table 1.2.

1.3.1.1 Throughput and Latency

Aiming to reduce the confirmation latency of transactions to milliseconds, Hari et al. [8] proposed a high-throughput, low-latency, deterministic confirmation mechanism called ACCEL for accelerating Bitcoin's block confirmation. The key findings include how to identify the singular blocks, and how to use singular blocks to reduce the confirmation delay. Once the confirmation delay is reduced, the throughput increases accordingly.

Two obstacles have hindered the scalability of the cryptocurrency systems. The first one is the low throughput, and the other one is the requirement for every node to duplicate the communication, storage, and state representation of the entire blockchain network. Wang and Wang [9] studied how to solve the above obstacles. Without weakening decentralization and security, the proposed Monoxide technique offers a linear scale-out ability by partitioning the workload. And they preserved the simplicity of the blockchain system and amplified its capacity. The authors also proposed a novel *Chu-ko-nu* mining mechanism, which ensures the cross-zone atomicity, efficiency and security of the blockchain system with thousands of independent zones. Then, the authors have conducted experiments to evaluate the scalability performance of the proposed Monoxide with respect to TPS, the overheads of cross-zone transactions, the confirmation latency of transactions, etc.

To bitcoin, low *throughput* and long *transaction confirmation latency* are two critical bottleneck metrics. To overcome these two bottlenecks, Yang et al. [10] designed a new blockchain protocol called Prism, which achieves a scalable throughput as high as 70,000 transactions per second, while ensuring a full security of bitcoin. The project of Prism is open-sourced in Github. The instances of Prism can be flexibly deployed on commercial cloud platform such as AWS. However, the

Table 1.2 Latest theories of improving the performance of blockchains

Emphasis	Ref.	Recognition	Challenge	Methodology
Throughput and latency	[8]	ACCEL: reduce the confirmation delay of blocks	Most of the blockchain applications desire fast confirmation of their transactions	Authors proposed a high-throughput, low-latency, deterministic confirmation mechanism, aiming to accelerate Bitcoin's block confirmation.
	[9]	Monoxide	Scalability issues, and efficient processing of cross-shard transactions	The proposed Monoxide offers a linear scale-out by partitioning workloads. Particularly, *Chu-ko-nu* mining mechanism enables the cross-zone atomicity, efficiency and security of the system.
	[10]	Prism	Low transaction throughput and large transaction confirmation of bitcoin	Authors proposed a new blockchain protocol, i.e., Prism, aiming to achieve a scalable throughput with a full security of bitcoin.
	[11]	GARET	How to place transactions to shards considering the complexity of transactions or the workload generated by transactions	Authors proposed a gas consumption-aware relocation mechanism for improving throughput in sharding-based Ethereum.

(continued)

Table 1.2 (continued)

Emphasis	Ref.	Recognition	Challenge	Methodology
Storage efficiency	[12]	Erasure code-based	How to reduce the storage consumption of blockchains	Authors proposed a new type of low-storage blockchain nodes using erasure code theory to reduce the storage space of blockchains.
	[13]	Jidar: data-reduction strategy	How to reduce the data consumption of bitcoin's blocks	Authors proposed a data reduction strategy for Bitcoin namely Jidar, in which each node only has to store the transactions of interest and the related Merkle branches from the complete blocks.
	[14]	*Segment blockchain*	To reduce the storage of blockchain systems while maintaining the decentralization without sacrificing security	Authors proposed a data-reduced storage mechanism named *segment blockchain* such that each node only has to store a segment of the blockchain.
Reliability analysis	[15]	Availability of blockchains	The availability of read and write on blockchains is uneven	Authors studied the availability for blockchain-based systems, where the read and write availability is conflict to each other.
	[16]	Reliability prediction	The reliability of blockchain peers is unknown	Authors proposed H-BRP to predict the reliability of blockchain peers by extracting their reliability parameters.

authors also admitted that although the proposed Prism has a high throughput, its confirming latency still maintains as large as 10 seconds since there is only a single *voter chain* in Prism. A promising solution is to introduce a large number of such voter chains, each of which is not necessarily secure. Even though every voter chain is under attacking with a probability as high as 30%, the successful rate of attacking a half number of all voter chains is still theoretically very low. Thus, the authors believed that using multiple voter chains would be a good solution to reducing the confirmation latency while not sacrificing system security.

Considering that Ethereum simply allocates transactions to shards according to their account addresses rather than relying on the workload or the complexity of transactions, the resource consumption of transactions in each shard is unbalanced. In consequence, the network transaction throughput is affected and becomes low. To solve this problem, Woo et al. [11] proposed a heuristic algorithm named GARET, which is a gas consumption-aware relocation mechanism for improving throughput in sharding-based Ethereum environments. In particular, the proposed GARET can relocate transaction workloads of each shard according to the gas consumption. The experiment results show that GARET achieves a higher transactions throughput and a lower transaction latency compared with existing techniques.

1.3.1.2 Storage Efficiency

The transactions generated at real-time make the size of blockchains keep growing. For example, the storage efficiency of original-version Bitcoin has received much criticism since it requires to store the full transaction history in each Bitcoin peer. Although some revised protocols advocate that only the full-size nodes store the entire copy of whole ledger, the transactions still consume a large storage space in those full-size nodes. To alleviate this problem, several pioneer studies proposed storage-efficient solutions for blockchain networks. For example, By exploiting the erasure code-based approach, Perard et al. [12] proposed a low-storage blockchain mechanism, aiming to achieve a low requirement of storage for blockchains. The new low-storage nodes only have to store the linearly encoded fragments of each block. The original blockchain data can be easily recovered by retrieving fragments from other nodes under the erasure-code framework. Thus, this type of blockchain nodes allows blockchain clients to reduce the storage capacity. The authors also tested their system on the low-configuration Raspberry Pi to show the effectiveness, which demonstrates the possibility towards running blockchains on IoT devices.

Then, Dai et al. [13] proposed Jidar, which is a data reduction strategy for Bitcoin. In Jidar, each node only has to store the transactions of interest and the related Merkle branches from the complete blocks. All nodes verify transactions collaboratively by a query mechanism. This approach seems very promising to the storage efficiency of Bitcoin. Their experiments show that the proposed Jidar can reduce the storage overhead of each peer to about 1% comparing with the original Bitcoin.

Under the similar idea, Xu et al. [14] reduced the storage of blockchains using a *segment blockchain* mechanism, in which each node only needs to store a piece of blockchain segment. The authors also proved that the proposed mechanism endures a failure probability $(\phi/n)^m$ if an adversary party commits a collusion with less than a number ϕ of nodes and each segment is stored by a number m of nodes. This theoretical result is useful for the storage design of blockchains when developing a particular segment mechanism towards data-heavy distributed applications.

1.3.1.3 Reliability of Blockchains

As a decentralized mechanism for data protection, the reliability of blockchains plays an important role in data falsification. The following works studied the fundamental supporting mechanisms to achieve data falsification prevention. The availability of blockchains is a key factor for blockchain-based distributed applications (DApps). However, such availability guarantees of blockchain systems are unknown. To this end, Weber et al. [15] studied the availability limitations of two popular blockchains, i.e., Bitcoin and Ethereum. The authors found that the availability of reading and writing operations are conflict to each other. Through measuring and analyzing the transactions of Ethereum, they observed that the DApps could be stuck in an uncertain state while transactions are pending in a blockchain system. This observation suggests that maybe blockchains should support some built-in transaction-abort options for DApps. The authors finally presented techniques that can alleviate the availability limitations of Ethereum and Bitcoin blockchains.

In public blockchains, the system clients join the blockchain network basically through a third-party peer. Thus, the reliability of the selected blockchain peer is critical to the security of clients in terms of both resource-efficiency and monetary issues. To enable clients evaluate and choose the reliable blockchain peers, Zheng et al. [16] proposed a hybrid reliability prediction model for blockchains named H-BRP, which is able to predict the reliability of blockchain peers by extracting their reliability parameters.

1.3.2 Scalability-Improving Solutions

One of the critical bottlenecks of today's blockchain systems is the scalability. For example, the throughput of a blockchain is not scalable when the network size grows. To address this dilemma, a number of scalability approaches have been proposed. In this part, we conduct an overview of the most recent solutions with respect to Sharding techniques, interoperability among multiple blockchains, and other solutions. We summarize this subsection in Table 1.3.

Table 1.3 Latest scalability solutions to improving the performance of blockchains

Emphasis	Ref.	Recognition	Methodology
Solutions to sharding blockchains	[1]	Elastico	Authors proposed a new distributed agreement protocol for the permission-less blockchains, called Elastico, which is viewed as the first secure candidate for a sharding protocol towards the open public blockchains.
	[9]	Monoxide	The proposed Monoxide enables the system to handle transactions through a number of independent zones. This scheme is essentially following the principle of sharding mechanism.
	[17]	Rapidchain	Authors proposed a new sharding-based protocol for public blockchains that achieves non-linearly increase of intra-committee communications with the number of committee members.
	[18]	SharPer	Authors proposed a permissioned blockchain system named *SharPer*, which adopts sharding techniques to improve scalability of cross-shard transactions.
	[19]	D-GAS	Authors proposed a dynamic load balancing mechanism for Ethereum shards, i.e., D-GAS. It reallocates Tx accounts by their gas consumption on each shard.
	[20]	NRSS	Authors proposed a node-rating based new Sharding scheme, i.e., NRSS, for blockchains, aiming to improve the throughput of committees.
	[21]	OptChain	Authors proposed a new sharding paradigm, called OptChain, mainly used for optimizing the placement of transactions into shards.

(continued)

Table 1.3 (continued)

Emphasis	Ref.	Recognition	Methodology
	[22]	Sharding-based scaling system	Authors proposed an efficient shard-formation protocol that assigns nodes into shards securely, and a distributed transaction protocol that can guard against malicious Byzantine fault coordinators.
	[23]	SSChain	Authors proposed a non-reshuffling structure called SSChain, which supports both transaction sharding and state sharding while eliminating huge data-migration across shards.
	[24]	Eumonia	Authors proposed Eumonia, which is a permissionless parallel-chain protocol for realizing a global ordering of blocks.
	[25]	Vulnerability of Sybil attacks	Authors systematically analyzed the vulnerability of Sybil attacks in protocol Elastico.
	[26]	n/2 BFT Sharding approach	Authors proposed a new blockchain sharding approach that can tolerate up to 1/2 of the Byzantine nodes within a shard.
	[27]	CycLedger	Authors proposed a protocol CycLedger to pave a way towards scalability, security and incentive for sharding blockchains.
Interoperability of multiple-chain systems	[28]	Interoperability architecture	Authors proposed a novel interoperability architecture that supports the cross-chain cooperations among multiple blockchains, and a novel Monitor Multiplexing Reading (MMR) method for the passive cross-chain communications.
	[29]	HyperService	Authors proposed a programming platform that provides interoperability and programmability over multiple heterogeneous blockchains.

	[30]	Protocol *Move*	Authors proposed a programming model for smart-contract developers to create DApps that can interoperate and scale in a multiple-chain environment.
	[31]	Cross-cryptocurrency TX protocol	Authors proposed a decentralized cryptocurrency exchange protocol enabling cross-cryptocurrency transactions based on smart contracts deployed on Ethereum.
	[32]	Cross-chain comm.	Authors conducted a systematic classification of cross-chain communication protocols.

1.3.2.1 Solutions to Sharding Blockchains

Bitcoin's transaction throughput does not scale well. The solutions that use classical Byzantine consensus protocols do not work in an open environment like cryptocurrencies. To solve the above problems, Luu et al. [1] proposed a new distributed agreement protocol for the permission-less blockchains, called *Elastico*, which is viewed as the first secure candidate for a sharding protocol towards the open public blockchains that tolerate a constant fraction of byzantine-fault network nodes. The key idea in Elastico is to partition the network into smaller committees, each of which processes a disjoint set of transactions or a *shard*. The number of committees grows linearly in the total computational power of the network. Using Elastico, the blockchain's transaction throughput increases almost linearly with the computational power of the network.

Some early-stage sharding blockchain protocols (e.g., Elastico) improve the scalability by enforcing multiple groups of committees work in parallel. However, this manner still requires a large amount of communication for verifying every transaction linearly increasing with the number of nodes within a committee. Thus, the benefit of sharding policy was not fully employed. As an improved solution, Zamani et al. [17] proposed a Byzantine-resilient sharding-based protocol, namely Rapidchain, for permissionless blockchains. Taking the advantage of block pipelining, RapidChain improves the throughput by using a sound intra-committee consensus. The authors also developed an efficient cross-shard verification method to avoid the broadcast messages flooding in the holistic network.

To enforce the throughput scaling with the network size, Gao et al. [33] proposed a scalable blockchain protocol, which leverages both sharding and Proof-of-Stake consensus techniques. Their experiments were performed in an Amazon EC2-based simulation network. Although the results showed that the throughput of the proposed protocol increases following the network size, the performance was still not so high, for example, the maximum throughput was 36 transactions per second and the transaction latency was around 27 seconds.

Aiming to improve the efficiency of cross-shard transactions, Amiri et al. [18] proposed a permissioned blockchain system named *SharPer*, which is strive for the scalability of blockchains by dividing and reallocating different data shards to various network clusters. The major contributions of the proposed SharPer include the related algorithm and protocol associated to such SharPer model. In the Amiri previous work, they have already proposed a permissioned blockchain, upon which the authors extended it by introducing a consensus protocol in the processing of both intra-shard and cross-shard transactions. Finally, SharPer was devised by adopting sharding techniques. One of the important contributions is that SharPer can be used in the networks where there are a high percentage of non-faulty nodes. Furthermore, SharPer also contributes a flattened consensus protocol w.r.t the order of cross-shard transactions among all involved clusters.

Considering that the Ethereum places each group of transactions on a shard by their account addresses, the workloads and complexity of transactions in shards are apparently unbalanced. This manner further damages the network throughput.

To address this uneven problem, Kim et al. [19] proposed D-GAS, which is a dynamic load balancing mechanism for Ethereum shards. Using such D-GAS, the transaction workloads of accounts on each shard can be reallocated according to their gas consumption. The target is to maximize the throughput of those transactions. The evaluation results showed that the proposed D-GAS achieved at most a 12% superiority of transaction throughput and a 74% lower transaction latency comparing with other existing techniques.

The random sharding strategy causes imbalanced performance gaps among different committees in a blockchain network. Those gaps yield a bottleneck of transaction throughput. Thus, Wang et al. [20] proposed a new sharding policy for blockchains named NRSS, which exploits node rating to assess network nodes according to their performance of transaction verifications. After such evaluation, all network nodes will be reallocated to different committees aiming at filling the previous imbalanced performance gaps. Through the experiments conducted on a local blockchain system, the results showed that NRSS improves throughput by around 32% under sharding techniques.

Sharding has been proposed to mainly improve the scalability and the throughput performance of blockchains. A good sharding policy should minimize the cross-shard communications as much as possible. A classic design of sharding is the *Transactions Sharding*. However, such Transactions Sharding exploits the *random sharding* policy, which leads to a dilemma that most transactions are cross-shard. To this end, Nguyen et al. [21] proposed a new sharding paradigm differing from the random sharding, called OptChain, which can minimize the number of cross-shard transactions. The authors achieved their goal through the following two aspects. First they designed two metrics, named T2S-score (Transaction-to-Shard) and L2S-score (Latency-to-Shard), respectively. T2S-score aims to measure how likely a transaction should be placed into a shard, while L2S-score is used to measure the confirmation latency when placing a transaction into a shard. Next, they utilized a well-known PageRank analysis to calculate T2S-score and proposed a mathematical model to estimate L2S-score. Finally, how does the proposed OptChain place transactions into shards based on the combination of T2S and L2S scores? In brief, they introduced another metric composed of both T2S and L2S, called *temporal fitness* score. For a given transaction u and a shard S_i, OptChain figures the temporal fitness score for the pair $\langle u, S_i \rangle$. Then, OptChain just puts transaction u into the shard that is with the highest temporal fitness score.

Similar to [21], Dang et al. [22] proposed a new shard-formation protocol, in which the nodes of different shards are re-assigned into different committees to reach a certain safety degree. In addition, they also proposed a coordination protocol to handle the cross-shard transactions towards guarding against the Byzantine-fault malicious coordinators. The experiment results showed that the throughput achieves a few thousands of TPS in both a local cluster with 100 nodes and a large-scale Google cloud platform testbed.

Considering that the reshuffling operations lead to huge data migration in the sharding-based protocols, Chen et al. [23] devised a non-reshuffling structure called SSChain. Such new sharding-based protocol can avoid the overhead of

data migration while enabling both transaction sharding and state sharding. Their evaluation results showed that SSChain achieves at least 6500 TPS in a network with 1800 nodes and no periodical data-migration needed.

Multiple chains can help increase the throughput of the blockchain. However, one issue under multiple-chain system must be solved. That is, the logical ordering of blocks generated should be guaranteed, because the correct logical order is critical to the confirmation of transactions. To this end, Niu et al. [24] proposed Eumonia, which is a permissionless parallel-chain protocol towards a global ordering of blocks. The authors implemented Eunomia by exploiting a fine-grained UTXO sharding model, in which the conflicted transactions can be well handled, and such protocol is proved as Simple Payment Verification (SPV) friendly.

Although the sharding techniques have received much interests recently, it should be noticed that the committee organization is easily to attract Sybil attacks, in which a malicious node can compromise the consensus by creating multiple dummy committee members in the vote phase of the consensus protocol. To address such Sybil attacks, Rajab et al. [25] systematically formulated a model and performed an analysis w.r.t the vulnerability of Sybil attacks in the pioneer sharding protocol Elastico [1]. The authors found that the blockchain nodes that have high hash-computing power are capable to manipulate Elastico protocol using a large number of Sybil IDs. The other two conditions of Sybil attacks were derived and evaluated by numerical simulations.

The traditional Sharding blockchain protocols can only endure up to 1/3 Byzantine-fault nodes within a shard. This weak BFT feature makes the number of nodes inside a shard cannot be small to ensure the shard functions securely. To improve the sustainability of blockchain sharding, Xu et al. [26] proposed a new BFT sharding approach that can tolerate at most 1/2 Byzantine-fault nodes existing inside a shard. This approach benefits the throughput of decentralized databases.

Although the existing sharding-based protocols, e.g., Elastico, OminiLedger and RapaidChain, have gained a lot of attention, they still have some drawbacks. For example, the mutual connections among all honest nodes require a big amount of communication resources. Furthermore, there is no an incentive mechanism driven nodes to participate in sharding protocol actively. To solve those problems, Zhang et al. [27] proposed *CycLedger*, which is a protocol designed for the sharding-based distributed ledger towards scalability, reliable security, and incentives. Such the proposed CycLedger is able to select a leader and a subset of nodes for each committee that handle the intra-shard consensus and the synchronization with other committees. A semi-commitment strategy and a recovery processing scheme were also proposed to deal with system crashing. In addition, the authors also proposed a reputation-based incentive policy to encourage nodes behaving honestly.

1.3.2.2 Multiple-Chain and Cross-Chain: Interoperability Amongst Multiple Blockchains

The interoperability of blockchains plays a significant role for the cross-chain transactions. Such interoperability mainly includes the effective communications and data exchange amongst multiple blockchains, as shown in Fig. 1.1. A lot of theoretical and practical issues of this direction need urgent solutions. Some representative studies are reviewed as follows.

To enable rich functionalities and capabilities for the future blockchain ecosystems, Jin et al. [28] proposed a novel interoperability architecture that supports the cross-chain cooperation among multiple blockchains, such as bitcoin and Ethereum. The authors classified the interoperability of multiple-chain ecosystems into passive and active modes, which are shown in Fig. 1.2. Then, the authors introduced a particular method, called Monitor Multiplexing Reading (MMR), dedicated to the passive cross-chain communications.

Following the widespread adoption of smart contracts, the roles of blockchains have been upgraded from token exchanges into programmable state machines. Thus,

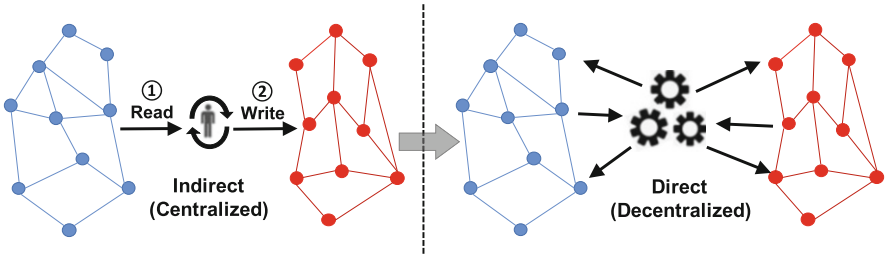

Fig. 1.1 The illustration of interoperability across blockchains [28]. The left figure demonstrates the indirect way of interoperability that requires a centralized third party. The right figure demonstrates the direct way of interoperability without the presence of any third party

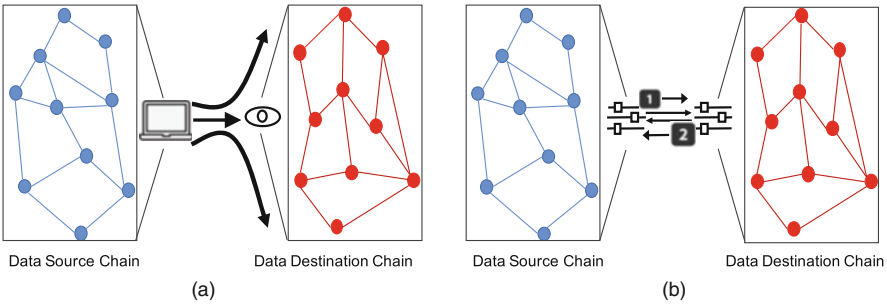

Fig. 1.2 The interoperability of blockchains [28]. Passive mode is shown in the left figure, in which the source chain is monitored by the destination chain instead of actively sending information to the destination chain as shown in the right figure. (**a**) Passive mode. (**b**) Active mode

the blockchain interoperability must evolve accordingly. To help realize such new type of interoperability among multiple heterogeneous blockchains, Liu et al. [29] proposed HyperService, which includes two major components, i.e., a programming framework allowing developers to create cross-chain applications; and a universal interoperability protocol towards secure implementation of DApps on blockchains. The authors implemented a 35,000-line prototype to prove the practicality of HyperService. Using the prototype, the end-to-end delays of cross-chain DApps, and the aggregated platform throughput can be measured conveniently.

In an ecosystem that consists of multiple blockchains, interoperability among those difference blockchains is an essential issue. To help the smart-contract developers build DApps, Fynn et al. [30] proposed a practical *Move* protocol that works for multiple blockchains. The basic idea of such protocol is to support a move operation enabling to move objects and smart contracts from one blockchain to another. Recently, to enable cross-cryptocurrency transactions, Tian et al. [31] proposed a decentralized cryptocurrency exchange strategy implemented on Ethereum through smart contracts. Additionally, a great number of studies of cross-chain communications are included in [32], in which readers can find a systematic classification of cross-chain communication protocols.

1.3.3 New Protocols and Infrastructures

This subsection is summarized in Table 1.4.

1.3.3.1 New Protocols for Blockchains

David et al. [34] proposed a provably secure PoS protocol named *Ouroboros Praos*, which particularly exploits forward secure digital signatures and a verifiable random function such that the proposed Ouroboros Praos can endure any corruption towards any participants from an adversary in a given message delivery delay.

In blockchain systems, a node only connects to a small number of neighbor nodes. Mutual communications are achieved by gossip-like P2P messages. Based on such P2P gossip communications, Buchman et al. [35] proposed a new protocol named Tendermint, which serves as a new termination mechanism for simplifying BFT consensus protocol.

In Monoxide proposed by Wang and Wang [9], the authors have devised a novel proof-of-work scheme, named *Chu-ko-nu mining*. This new proof protocol encourages a miner to create multiple blocks in different zones simultaneously with a single PoW solving effort. This mechanism makes the effective mining power in each zone is almost equal to the level of the total physical mining power in the entire network. Thus, Chu-ko-nu mining increases the attack threshold for each zone to 50%. Furthermore, Chu-ko-nu mining can improve the energy consumption spent

Table 1.4 New protocols and infrastructures to improving the performance of blockchains

Emphasis	Ref.	Recognition	Methodology
New protocols	[34]	Ouroboros praos	Authors proposed a new secure proof-of-stake protocol named *Ouroboros Praos*, which is proved secure in the semi-synchronous adversarial setting.
	[35]	Tendermint	Authors proposed a new BFT consensus protocol for the wide area network organized by the gossip-based P2P network under adversarial conditions.
	[9]	Chu-ko-nu mining	Authors proposed a novel proof-of-work scheme, named *Chu-ko-nu mining*, which incentivizes miners to create multiple blocks in different zones with only a single PoW mining.
	[36]	Proof-of-trust (PoT)	Authors proposed a novel proof-of-trust consensus for the online services of crowdsourcing.
New infrastructures and architectures	[37]	StreamChain	Authors proposed to shift the block-based distributed ledgers to a new paradigm of *stream transaction processing* to achieve a low end-to-end latencies without much affecting throughput.
	[38]	CAPER: cross-app trans. handling	Authors proposed a permissioned blockchain named CAPER that can well manage both the internal and the cross-application transactions for distributed applications.
	[39]	Optimal mining for miners	Authors proposed an edge computing-based blockchain network architecture, aiming to allocate optimal computational resources for miners.
	[40]	AxeChain: useful mining	Authors proposed a new framework for practical PoW blockchains called AxeChain, which can spend computing power of blockchains to solve arbitrary practical problems submitted by system clients.
	[41]	Non-linear blockchain system	Authors explored three major metrics of blockchains, and devised a non-linear blockchain system.

on mining new blocks because a lot of more blocks can be produced in each round of normal PoW mining.

The online services of crowdsourcing face a challenge to find a suitable consensus protocol. By leveraging the advantages of the blockchain such as the traceability of service contracts, Zou et al. [36] proposed a new consensus protocol, named *Proof-of-Trust* (PoT) consensus, for crowdsourcing and the general online service industries. Basically, such PoT consensus protocol leverages a trust management of all service participants, and it works as a hybrid blockchain architecture in which a consortium blockchain integrates with a public service network.

1.3.3.2 New Infrastructures and Architectures for Blockchains

Conventionally, block-based data structure is adopted by permissionless blockchain systems as blocks can efficiently amortize the cost of cryptography. However, the benefits of blocks are saturated in today's permissioned blockchains since the block-processing introduces large batching latencies. To the distributed ledgers that are neither geo-distributed nor Pow-required, István et al. [37] proposed to shift the traditional block-based data structure into the paradigm of *stream-like transaction processing*. The premier advantage of such paradigm shift is to largely shrink the end-to-end latencies for permissioned blockchains. The authors developed a prototype of their concept based on Hyperledger Fabric. The results showed that the end-to-end latencies achieved sub-10 ms and the throughput was close to 1500 TPS.

Permissioned blockchains have a number of limitations, such as poor performance, privacy leaking, and inefficient cross-application transaction handling mechanism. To address those issues, Amiri et al. [38] proposed CAPER, which a permissioned blockchain that can well deal with the cross-application transactions for distributed applications. In particular, CAPER constructs its blockchain ledger using DAG and handles the cross-application transactions by adopting three specific consensus protocols, i.e., a global consensus using a separate set of orders, a hierarchical consensus protocol, and a *one-level* consensus protocol. Then, Chang et al. [39] proposed an edge computing-based blockchain architecture, in which edge-computing providers supply computational resources for blockchain miners. The authors then formulated a two-phase stackelberg game for the proposed architecture, aiming to find the Stackelberg equilibrium of the theoretical optimal mining scheme. Next, Zheng et al. [40] proposed a new infrastructure for practical PoW blockchains called AxeChain, which aims to exploit the precious computing power of miners to solve arbitrary practical problems submitted by system users. The authors also analyzed the trade-off between energy consumption and security guarantees of such AxeChain. This study opens up a new direction for pursing high energy efficiency of meaningful PoW protocols.

With the non-linear (e.g., graphical) structure adopted by blockchain networks, researchers are becoming interested in the performance improvement brought by new data structures. To find insights under such non-linear blockchain systems,

Chen et al. [41] performed a systematic analysis by taking three critical metrics into account, i.e., *full verification*, *scalability*, and *finality-duration*. The authors revealed that it is impossible to achieve a blockchain that enables those three metrics at the same time. Any blockchain designers must consider the trade-off among such three properties.

1.4 Various Modelings and Techniques for Better Understanding Blockchains

We summarize various analytical models for blockchain networks in Tables 1.5 and 1.6.

1.4.1 Graph-Based Theories

The graphs are widely used in blockchain networks. For example, Merkel Tree has been adopted by Bitcoin, and several blockchain protocols, such as Ghost [42], Phantom [43], and Conflux [44], constructed their blocks using the directed acyclic graph (DAG) technique. Different from those generalized graph structures, we review the most recent studies that exploit the graph theories for better understanding blockchains in this part.

Since the transactions in blockchains are easily structured into graphs, the graph theories and graph-based data mining techniques are viewed as good tools to discover the interesting findings beyond the graphs of blockchain networks. Some representative recent studies are reviewed as follows.

Leveraging the techniques of graph analysis, Chen et al. [45] characterized three major activities on Ethereum, i.e., money transfer, the creation of smart contracts, and the invocation of smart contracts. The major contribution is that it performed the first systematic investigation and proposed new approaches based on cross-graph analysis, which can address two security issues existing in Ethereum: attack forensics and anomaly detection. Particularly, w.r.t the graph theory, the authors mainly concentrated on the following two aspects:

1. *Graph Construction*: They identified four types of transactions that are not related to money transfer, smart contract creation, or smart contract invocation.
2. *Graph Analysis*: Then, they divided the remaining transactions into three groups according to the activities they triggered, i.e., money flow graph (MFG), smart contract creation graph (CCG) and contract invocation graph (CIG).

Via this manner, the authors delivered many useful insights of transactions that are helpful to address the security issues of Ethereum.

Table 1.5 Various modelings, techniques and theories for better understanding blockchains

Category	Emphasis	Ref.	Metrics	Methodology and implications
Graph-based theoriess	Transactions mining	[45]	Cross-graph analysis of Ethereum	Via graph analysis, authors extracted three major activities, i.e., money transfer, smart contracts creation, and smart contracts invocation.
		[49]	Features of transaction graphs	Proposed an extendable and computationally efficient method for graph representation learning on Blockchains.
		[50]	Market manipulation patterns	Authors exploited the graph-based data-mining approach to reveal the market manipulation evidence of Bitcoin.
		[53]	Clustering coefficient, assortativity of TX graph	Authors exploited the graph-based analysis to reveal the abnormal transactions of EOSIO.
	Token networks	[51]	Token-transfer distributions	Authors studied the token networks through analyzing smart contracts of Ethereum blockchain based on graph analysis.
		[46, 47]	Extreme chainlet activity	Authors proposed graph-based analysis models for assessing the financial investment risk of Bitcoin.

Stochastic modelings	Blockchain network analysis	[54]	Block completion rates, and the probability of a successful adversarial attack	Authors derived stochastic models to capture critical blockchain properties, and to evaluate the impact of blockchain propagation latency on key performance metrics. This study provides us useful insights of design issues of blockchain networks.
	Stability analysis	[55]	Time to consistency, cycle length, consistency fraction, age of information	Authors proposed a network model which can identify the stochastic stability of blockchain systems.
	Failure probability analysis	[56–58]	Failure probability of a committee, sums of upper-bounded hypergeometric and binomial distributions for each epoch	Authors proposed a probabilistic model to derive the security analysis under Sharding blockchain protocols. This study can tell how to keep the failure probability smaller than a defined threshold for a specific sharding protocol.
Queueing theories	Mining procedure and block-generation	[59, 60]	The average number of TX in the arrival queue and in a block, and average confirmation time of TX	Authors developed a Markovian batch-service queueing system to express the mining process and the generation of new blocks in miners pool.
	Block-confirmation time	[61]	The residual lifetime of a block till the next block is confirmed	Authors proposed a theoretical framework to deeply understand the transaction confirmation time, by integrating the queueing theory and machine learning techniques.
	Synchronization process of Bitcoin network	[62]	Stationary queue-length distribution	Authors proposed an infinite-server model with random fluid limit for Bitcoin network.

(continued)

Table 1.5 (continued)

Category	Emphasis	Ref.	Metrics	Methodology and implications
	Mining resources allocation	[63]	Mining resource for miners, queueing stability	Authors proposed a Lyapunov optimization-based queueing analytical model to study the allocation of mining resources for the PoW-based blockchain networks.
	Blockchain's theoretical working principles	[64]	number of TX per block, mining interval of each block, memory pool size, waiting time, number of unconfirmed TX	Authors proposed a queueing theory-based model to have a better understanding the theoretical working principle of blockchain networks.

Table 1.6 Various analytics models for better understanding blockchain networks

Emphasis	Ref.	Metrics	Methodology and implications
Applicability of blockchains	[65]	Public verifiability, transparency, privacy, integrity, redundancy, and trust anchor	Authors proposed the first structured analytical methodology that can help decide whether a particular application system indeed needs a blockchain, either a permissioned or permissionless, as its technical solution.
	[66]	Scalability, efficiency and privacy issues in cloud for blockchains	Authors proposes a novel upper bound privacy leakage based approach to identify intermediate data sets partitioned and distributed in cloud for encryption. This approach can significantly improve the scalability and efficiency of data processing for privacy preserving in cloud.
Exploration of ethereum transactions	[67]	Temporal information and the multiplicity features of Ethereum transactions	Authors proposed an analytical model based on the multiplex network theory for understanding Ethereum transactions.
	[68]	Pending time of Ethereum transactions	Authors conducted a characterization study of the Ethereum by focusing on the pending time, and attempted to find the correlation between pending time and fee-related parameters of Ethereum.
Modeling the competition over multiple miners	[69]	Competing mining resources of miners of a cryptocurrency blockchain	Authors exploited the Game Theory to find a Nash equilibria while peers are competing mining resources.
A neat bound of consistency latency	[70]	Consistency of a PoW blockchain	Authors derived a neat bound of mining latencies that helps understand the consistency of Nakamoto's blockchain consensus in asynchronous networks.

(continued)

Table 1.6 (continued)

Emphasis	Ref.	Metrics	Methodology and implications
Network connectivity	[71]	Consensus security	Authors proposed an analytical model to evaluate the impact of network connectivity on the consensus security of PoW blockchain under different adversary models.
How ethereum responds to sharding	[72]	Balance among shards, number of TX that would involve multiple shards, the amount of data relocated across shards	Authors studied how sharding impact Ethereum by firstly modeling Ethereum through graph modeling, and then assessing the three metrics mentioned when partitioning the graph.
Required properties of sharding protocols	[73]	Consistency and scalability	Authors proposed an analytical model to evaluate whether a protocol for sharded distributed ledgers fulfills necessary properties.
Vulnerability by forking attacks	[74]	Hashrate power, net cost of an attack	Authors proposed fine-grained vulnerability analytical model of blockchain networks incurred by intentional forking attacks taking the advantages of large deviation theory.
Counterattack to double-spend attacks	[3]	Robustness parameter, vulnerability probability	Authors studied how to defense and even counterattack the double-spend attacks in PoW blockchains.
Limitations of PBFT-based blockchains	[75]	Performance of blockchain applications, Persistence, Possibility of forks	Authors studied and identified several misalignments between the requirements of permissioned blockchains and the classic BFT protocols.
Unified analysis of different PoX consensus schemes	[76]	Resource sensitivity, system convergence, and resource Fairness	Authors proposed a new Markov model to unify the analysis of the steady-state for weighted resource distribution of different PoX-based Blockchains.

Similarly, by processing Bitcoin transaction history, Akcora et al. [46] and Dixon et al. [47] modeled the transfer network into an extreme transaction graph. Through the analysis of chainlet activities [48] in the constructed graph, they proposed to use GARCH-based forecasting models to identify the financial risk of Bitcoin market for cryptocurrency users.

An emerging research direction associated with blockchain-based cryptocurrencies is to understand the network dynamics behind graphs of those blockchains, such as the transaction graph. This is because people are wondering what the connection between the price of a cryptocurrency and the dynamics of the overlying transaction graph is. To answer such a question, Abay et al. [49] proposed Chainnet, which is a computationally lightweight method to learning the graph features of blockchains. The authors also disclosed several insightful findings. For example, it is the topological feature of transaction graph that impacts the prediction of Bitcoin price dynamics, rather than the degree distribution of the transaction graph.

Furthermore, utilizing the Mt. Gox transaction history, Chen et al. [50] also exploited the graph-based data-mining approach to dig the market manipulation of Bitcoin. The authors constructed three graphs, i.e., extreme high graph (EHG), extreme low graph (ELG), and normal graph (NMG), based on the initial processing of transaction dataset. Then, they discovered many correlations between market manipulation patterns and the price of Bitcoin.

On the other direction, based on *address graphs*, Victor et al. [51] studied the ERC20 token networks through analyzing smart contracts of Ethereum blockchain. Different from other graph-based approaches, the authors focused on their attention on the address graphs, i.e., token networks. With all network addresses, each token network is viewed as an overlay graph of the entire Ethereum network addresses. Similar to [45], the authors presented the relationship between transactions by exploiting graph-based analysis, in which the arrows can denote the invoking functions between transactions and smart contracts, and the token transfers between transactions as well. The findings presented by this study help us have a well understanding of token networks in terms of time-varying characteristics, such as the usage patterns of the blockchain system. An interesting finding is that around 90% of all transfers stem from the top 1000 token contracts. That is to say, only less than 10% of token recipients have transferred their tokens. This finding is contrary to the viewpoint proposed by Somin et al. [52], where Somin et al. showed that the full transfers seem to obey a power-law distribution. However, the study [51] indicated that those transfers in token networks likely do not follow a power law. The authors attributed such the observations to the following three possible reasons: (1) most of the token users don't have incentives to transfer their tokens. Instead, they just simply hold tokens; (2) the majority of inactive tokens are treated as something like unwanted spam; (3) a small portion, i.e., approximately 8%, of users intended to sell their tokens to a market exchange.

Recently, Zhao et al. [53] explored the account creation, account vote, money transfer and contract authorization activities of early-stage EOSIO transactions through graph-based metric analysis. Their study revealed abnormal transactions like voting gangs and frauds.

1.4.2 Stochastic Modelings

The latencies of block transfer and processing are generally existing in blockchain networks since the large number of miner nodes are geographically distributed. Such delays increase the probability of forking and the vulnerability to malicious attacks. Thus, it is critical to know how would the network dynamics caused by the block propagation latencies and the fluctuation of hashing power of miners impact the blockchain performance such as block generation rate. To find the connection between those factors, Papadis et al. [54] developed stochastic models to derive the blockchain evolution in a wide-area network. Their results showed us practical insights for the design issues of blockchains, for example, how to change the difficulty of mining in the PoW consensus while guaranteeing an expected block generation rate or an immunity level of adversarial attacks. The authors then performed analytical studies and simulations to evaluate the accuracy of their models. This stochastic analysis opens up a door for us to have a deeper understanding of dynamics in a blockchain network.

Towards the stability and scalability of blockchain systems, Gopalan et al. [55] also proposed a stochastic model for a blockchain system. During their modeling, a structural asymptotic property called *one-endedness* was identified. The authors also proved that a blockchain system is one-ended if it is stochastically stable. The upper and lower bounds of the stability region were also studied. The authors found that the stability bounds are closely related to the conductance of the P2P blockchain network. Those findings are very insightful such that researchers can assess the scalability of blockchain systems deployed on large-scale P2P networks.

Although Sharding protocol is viewed as a very promising solution to solving the scalability of blockchains and adopted by multiple well-known blockchains such as RapidChain [17], OmniLedger [2], and Monoxide [9], the failure probability for a committee under Sharding protocol is still unknown. To fill this gap, Hafid et al. [56–58] proposed a stochastic model to capture the security analysis under Sharding-based blockchains using a probabilistic approach. With the proposed mathematical model, the upper bound of the failure probability was derived for a committee. In particular, three probability inequalities were used in their model, i.e., Chebyshev, Hoeffding, and Chvátal. The authors claim that the proposed stochastic model can be used to analyze the security of any Sharding-based protocol.

1.4.3 Queueing Theories for Blockchain Systems

In blockchain networks, several stages of mining processing and the generation of new blocks can be formulated as queueing systems, such as the transaction-arrival queue, the transaction-confirmation queue, and the block-verification queue. Thus, a growing number of studies are exploiting the queueing theory to disclose the mining

and consensus mechanisms of blockchains. Some recent representative works are reviewed as follows.

To develop a queueing theory of blockchain systems, Li et al. [59, 60] devised a batch-service queueing system to describe the mining and the creating of new blocks in miners' pool. For the blockchain queueing system, the authors exploited the type GI/M/1 continuous-time Markov process. Then, they derived the stable condition and the stationary probability matrix of the queueing system utilizing the matrix-geometric techniques.

Then, viewing that the confirmation delay of Bitcoin transactions are larger than conventional credit card systems, Ricci et al. [61] proposed a theoretical framework integrating the queueing theory and machine learning techniques to have a deep understanding towards the transaction confirmation time. The reason the authors chose the queueing theory for their study is that a queueing model is suitable to see insights into how the different blockchain parameters affect the transaction latencies. Their measurement results showed that the Bitcoin users experience a delay that is slightly larger than the residual time of a block confirmation.

Frolkova et al. [62] formulated the synchronization process of Bitcoin network as an infinite-server model. The authors derived a closed-form for the model that can be used to capture the queue stationary distribution. Furthermore, they also proposed a random-style fluid limit under service latencies.

On the other hand, to evaluate and optimize the performance of blockchain-based systems, Memon et al. [64] proposed a simulation model by exploiting queueing theory. In the proposed model, the authors constructed an M/M/1 queue for the memory pool, and an M/M/c queue for the mining pool, respectively. This model can capture multiple critical statistics metrics of blockchain networks, such as the number of transactions every new block, the mining interval of a block, transactions throughput, and the waiting time in memory pool, etc.

Next, Fang et al. [63] proposed a queueing analytical model to allocate mining resources for the general PoW-based blockchain networks. The authors formulated the queueing model using Lyapunov optimization techniques. Based on such stochastic theory, a dynamic allocation algorithm was designed to find a trade-off between mining energy and queueing delay. Different from the aforementioned work [59–61], the proposed Lyapunov-based algorithm does not need to make any statistical assumptions on the arrivals and services.

1.4.4 Analytical Models for Blockchain Networks

This subsection is summarized in Table 1.6.

For the people considering whether a blockchain system is needed for his/her business, a notable fact is that blockchain is not always applicable to all real-life use cases. To help analyze whether blockchain is appropriate to a specific application scenario, Wust et al. [65] provided the first structured analytical methodology and applied it to analyzing three representative scenarios, i.e., supply chain management,

interbank payments, and decentralized autonomous organizations. The other article [66] proposes a novel upper bound privacy leakage based approach to identify intermediate data sets partitioned and distributed in cloud for encryption. This approach can significantly improve the scalability and efficiency of data processing for privacy preserving in cloud. This study provides insights of scalability, efficiency and privacy issues in cloud for blockchain.

Although Ethereum has gained much popularity since its debut in 2014, the systematically analysis of Ethereum transactions still suffers from insufficient explorations. Therefore, Lin et al. [67] proposed to model the transactions using the techniques of multiplex network. The authors then devised several random-walk strategies for graph representation of the transactions network. This study could help us better understand the temporal data and the multiplicity features of Ethereum transactions.

To better understand the network features of an Ethereum transaction, Sousa et al. [68] focused on the pending time, which is defined as the latency counting from the time a transaction is observed to the time this transaction is packed into the blockchain. The authors tried to find the correlations between such pending time with the fee-related parameters such as gas and gas price. Surprisingly, their data-driven empirical analysis results showed that the correlation between those two factors has no clear clue. This finding is counterintuitive.

To achieve a consensus about the state of blockchains, miners have to compete with each other by invoking a certain proof mechanism, say PoW. Such competition among miners is the key module to public blockchains such as Bitcoin. To model the competition over multiple miners of a cryptocurrency blockchain, Altman et al. [69] exploited the Game Theory to find a Nash equilibria while peers are competing mining resources. The proposed approach help researchers well understand such competition. However, the authors also mentioned that they didn't study the punishment and cooperation between miners over the repeated games. Those open topics will be very interesting for future studies.

Besides competitions among individual miners, there are also competitions among mining pools. Malicious pools can pull off DDoS attacks to overload the victim pools' manager with invalid share submissions. The delay in verifying extra share submissions potentially impairs the hash power of the victim pool and thus undermines the potential reward for pool miners. Knowing that the chance of getting a reward is smaller, miners in the victim pools would migrate to another mining pools, which would further weaken the victim pools. To better understand this kind of competition, Wu et al. [77] proposed a stochastic game-theoretic model in a two-mining-pool case. The authors used Q-learning algorithm to find the Nash equilibrium and maximize the long-term payoffs. The experiment showed that the smaller mining pool is more likely to attack the larger one. Also, mining pools tend to adopt lower attack level when the DDoS attack cost increases.

To ensure the consistency of PoW blockchain in an asynchronous network, Zhao et al. [70] performed an analysis and derived a neat bound around $\frac{2\mu}{\ln(\mu/\nu)}$, where $\mu + \nu = 1$, with μ and ν denoting the fraction of computation power dominated by the honest and adversarial miners, respectively. Such a neat bound of

mining latencies is helpful to us to well understand the consistency of Nakamoto's blockchain consensus in asynchronous networks.

Bitcoin's consensus security is built upon the assumption of honest-majority. Under this assumption, the blockchain system is thought secure only if the majority of miners are honest while voting towards a global consensus. Recent researches believe that network connectivity, the forks of a blockchain, and the strategy of mining are major factors that impact the security of consensus in Bitcoin blockchain. To provide pioneering concrete modelings and analysis, Xiao et al. [71] proposed an analytical model to evaluate the network connectivity on the consensus security of PoW blockchains. To validate the effectiveness of the proposed analytical model, the authors applied it to two adversary scenarios, i.e., *honest-but-potentially-colluding*, and *selfish mining* models.

Although Sharding is viewed as a prevalent technique for improving the scalability to blockchain systems, several essential questions are: what we can expect from and what price is required to pay for introducing Sharding technique to Ethereum? To answer those questions, Fynn et al. [72] studied how sharding works for Ethereum by modeling Ethereum into a graph. Via partitioning the graph, they evaluated the trade-off between the edge-cut and balance. Several practical insights have been disclosed. For example, three major components, e..g, computation, storage and bandwidth, are playing a critical role when partitioning Ethereum; A good design of incentives is also necessary for adopting sharding mechanism.

As mentioned multiple times, sharding technique is viewed as a promising solution to improving the scalability of blockchains. However, the properties of a sharded blockchain under a fully adaptive adversary are still unknown. To this end, Avarikioti et al. [73] defined the *consistency* and *scalability* for sharded blockchain protocol. The limitations of security and efficiency of sharding protocols were also derived. Then, they analyzed these two properties on the context of multiple popular sharding-based protocols such as *OmniLedger*, *RapidChain*, *Elastico*, and *Monoxide*. Several interesting conclusions have been drawn. For example, the authors thought that Elastico and Momoxide failed to guarantee the balance between consistency and scalability properties, while OmniLedger and RapidChain fulfill all requirements of a robust sharded blockchain protocol.

Forking attacks has become the normal threats faced by the blockchain market. The related existing studies mainly focus on the detection of such attacks through transactions. However, this manner cannot prevent the forking attacks from happening. To resist the forking attacks, Wang et al. [74] studied the fine-grained vulnerability of blockchain networks caused by intentional forks using the large deviation theory. This study can help set the robustness parameters for a blockchain network since the vulnerability analysis provides the correlation between robust level and the vulnerability probability. In detail, the authors found that it is much more cost-efficient to set the robust level parameters than to spend the computational capability used to lower the attack probability.

The existing economic analysis [78] reported that the attacks towards PoW mining-based blockchain systems can be cheap under a specific condition when renting sufficient hashrate capability. Moroz et al. [3] studied how to defense the

double-spend attacks in an interesting reverse direction. The authors found that the counterattack of victims can lead to a classic game-theoretic *War of Attrition* model. This study showed us the double-spend attacks on some PoW-based blockchains are actually cheap. However, the defense or even counterattack to such double-spend attacks is possible when victims are owning the same capacity as the attacker.

Although BFT protocols have attracted a lot of attention, there are still a number of fundamental limitations unaddressed while running blockchain applications based on the classical BFT protocols. Those limitations include one related to low performance issues, and two correlated to the gaps between the state machine replication and blockchain models (i.e., the lack of strong persistence guarantees and the occurrence of forks). To identify those limitations, Bessani et al. [75] first studied them using a digital coin blockchain App called SmartCoin, and a popular BFT replication library called BFT-SMART, then they discussed how to tackle these limitations in a protocol-agnostic manner. The authors also implemented an experimental platform of permissioned blockchain, namely SmartChain. Their evaluation results showed that SmartChain can address the limitations aforementioned and significantly improve the performance of a blockchain application.

The Nakamoto protocol is designed to solve the Byzantine Generals Problem for permissionless Blockchains. However, a general analytical model is still missing for capturing the steady-state profit of each miner against the competitors. To this end, Yu et al. [76] studied the weighted resource distribution of proof-based consensus engines, referred to as Proof-of-X (PoX), in large-scale networks. The proposed Markov model attempts to unify the analysis of different PoX mechanisms considering three new unified metrics, i.e., resource sensitivity, system convergence, and resource fairness.

1.4.5 Data Analytics for Cryptocurrency Blockchains

This subsection is summarized in Table 1.7.

1.4.5.1 Market Risks Detection

As aforementioned, Akcora et al. [46] proposed a graph-based predictive model to forecast the investment risk of Bitcoin market. On the other hand, with the tremendously increasing price of cryptocurrencies such as Bitcoin, hackers are imminently utilizing any available computational resources to participate in mining. Thus, any web users face severe risks from the cryptocurrency-hungry hackers. For example, the *cryptojacking* attacks [87] have raised growing attention. In such type of attacks, a mining script is embedded secretly by a hacker without notice from the user. When the script is loaded, the mining will begin in the background of the system and a large portion of hardware resources are requisitioned for mining. To tackle the cryptojacking attacks, Tahir et al. [79] proposed a machine learning-based

Table 1.7 Data analytics for better understanding cryptocurrency blockchains

Emphasis	Ref.	Metrics	Methodology and implications
Cryptojacking detection	[79]	Hardware performance counters	Authors proposed a machine learning-based solution to prevent cryptojacking attacks.
	[80]	Various system resource utilization	Authors proposed an in-browser cryptojacking detection approach (CapJack), based on the latest CapsNet.
Market-manipulation mining	[50]	Various graph characteristics of transaction graph	Authors proposed a mining approach using the exchanges collected from the transaction networks.
Predicting volatility of Bitcoin price	[47]	Various graph characteristics of extreme chainlets	Authors proposed a graph-based analytic model to predict the intraday financial risk of Bitcoin market.
Money-laundering detection	[81]	Various graph characteristics of transaction graph	Authors exploited machine learning models to detect potential money laundering activities from Bitcoin transactions.
Ponzi-scheme detection	[82]	Factors that affect scam persistence	Authors analyzed the demand and supply perspectives of Ponzi schemes on Bitcoin ecosystem.
	[83, 84]	Account and code features of smart contracts	Authors detected Ponzi schemes for Ethereum based on data mining and machine learning approaches.
Design problem of cryptoeconomic systems	[85]	Price of XNS token, Subsidy of App developers	Authors presented a practical evidence-based example to show how data science and stochastic modeling can be applied to designing cryptoeconomic blockchains.
Pricing mining hardware	[86]	Miner revenue, ASIC value	Authors studied the correlation between the price of mining hardware (ASIC) and the value volatility of underlying cryptocurrency.

solution, which leverages the hardware performance counters as the critical features and can achieve a high accuracy while classifying the parasitic miners. The authors also built their approach into a browser extension towards the widespread real-time protection for web users. Similarly, Ning et al. [80] proposed *CapJack*, which is an in-browser cryptojacking detector based on deep capsule network (CapsNet) [88] technology.

As mentioned previously, to detect potential manipulation of Bitcoin market, Chen et al. [50] proposed a graph-based mining to study the evidence from the transaction network built based on Mt. Gox transaction history. The findings of this study suggests that the cryptocurrency market requires regulation.

To predict drastic price fluctuation of Bitcoin, Dixon et al. [47] studied the impact of extreme transaction graph (ETG) activity on the intraday dynamics of the Bitcoin prices. The authors utilized chainlets [48] (sub graphs of transaction graph) for developing their predictive models.

1.4.5.2 Ponzi Schemes Detection

Ponzi scheme [89], as a classic scam, is taking advantages of mainstream blockchains such as Ethereum. Data mining technologies [90] are widely used for detecting Ponzi schemes. For example, several representative studies are reviewed as follows. Vasek et al. [82] analyzed the demand and supply Ponzi schemes on Bitcoin ecosystem. The authors were interested at the reasons that make those Ponzi frauds succeeded in attracting victims, and the lifetime of those scams. To detect such Ponzi schemes towards a healthier blockchain economic environment, Chen et al. [83, 84] proposed a machine learning-based classification model by exploiting data mining on smart contracts of Ethereum. The experimental results showed that the proposed detection model can even identify Ponzi schemes at the very beginning when those schemes are created.

1.4.5.3 Money-Laundering Detection

Although Bitcoin has received enormous attention, it is also criticized for being carried out criminal financial activities such as ponzi schemes and money laundering. For example, Seo et al. [91] mentioned that money laundering conducted in the underground market can be detected using the Bitcoin mixing services. However, they didn't present an essential anti-money laundering strategy. In contrast, utilizing a transaction dataset collected over three years, Hu et al. [81] performed in-depth detection for discovering money laundering activities on Bitcoin network. To identify the money laundering transactions from the regular ones, the authors proposed four types of classifiers based on the graph features appeared on the transaction graph, i.e., immediate neighbors, deepwalk embeddings, node2vec embeddings and decision tree-based.

1.4.5.4 Portrait of Cryptoeconomic Systems

It is not common to introduce data science and stochastic simulation modelings into the design problem of cryptoeconomic engineering. Laskowski et al. [85] presented a practical evidence-based example to show how this manner can be applied to designing cryptoeconomic blockchains.

Yaish et al. [86] discussed the relationship between the cryptocurrency mining and the market price of the special hardware (ASICs) that supports PoW consensus. The authors showed that the decreasing volatility of Bitcoin's price has a counterintuitive negative impact to the value of mining hardware. This is because miners are not financially incentivized to participate in mining, when Bitcoin becomes widely adopted thus making its volatility decrease. This study also revealed that a mining hardware ASIC could be imitated by bonds and underlying cryptocurrencies such as bitcoins.

1.5 Useful Measurements, Datasets and Experiment Tools for Blockchains

Measurements are summarized in Table 1.8, and datasets are summarized in Table 1.9.

1.5.1 Performance Measurements and Datasets for Blockchains

Although diverse blockchains have been proposed in recent years, very few efforts have been devoted to measuring the performance of different blockchain systems. Thus, this part reviews the representative studies of performance measurements for blockchains. The measurement metrics include throughput, security, scalability, etc.

As a pioneer work in this direction, Gervais et al. [92] proposed a quantitative framework, using which they studied the security and performance of several PoW blockchains, such as Bitcoin, Litecoin, Dogecoin and Ethereum. The authors focused on multiple metrics of security model, e.g., stale block rate, mining power, mining costs, the number of block confirmations, propagation ability, and the impact of eclipse attacks. They also conducted extensive simulations for the four blockchains aforementioned with respect to the impact of block interval, the impact of block size, and throughput. Via the evaluation of network parameters about the security of PoW blockchains, researchers can compare the security performance objectively, and thus help them appropriately make optimal adversarial strategies and the security provisions of PoW blockchains.

Table 1.8 Various performance measurements of blockchains

Ref.	Target blockchains	Metrics	Implementation/experiments/methodology
[9]	General mining-based blockchains, e.g., bitcoin and ethereum	TPS, the overheads of cross-zone transactions, the confirmation latency of transactions, etc.	Monoxide was implemented utilizing C++. RocksDB was used to store blocks and TX. The real-world testing system was deployed on a distributed configuration consisting of 1200 virtual machines, with each owning 8 cores and 32 GB memory. In total 48,000 blockchain nodes were exploited in the testbed.
[10]	General blockchains	Throughput and confirmation latency, scalability under different number of clients, forking rate, and resource utilization (CPU, network bandwidth)	Prism testbed is deployed on Amazon EC2 instances each with 16 CPU cores, 16 GB RAM, 400 GB NVMe SSD, and a 10 Gbps network interface. In total 100 Prism client instances are connected into a topology in random 4-regular graph.
[11]	Ethereum	TX throughput, the makespan of transaction latency	The proposed GARET algorithm was measured to outperform existing techniques by up to 12% in TX throughput, and decrease the makespan of TX latency by about 74% under various conditions in sharding ethereum.
[92]	Bitcoin, litecoin, dogecoin, ethereum	Block interval, block size, and throughput	Proposed a quantitative framework, using which they studied the security and performance of several PoW blockchains. Via the evaluation of network parameters about the security of PoW blockchains, researchers can make trade-offs between the security provisions and performance objectively.
[93]	Hyperledger fabric	Execution time, latency, throughput, scalability vs the number of blockchain nodes	Presented the performance measurement and analysis towards Hyperledger Fabric version 0.6 and version 1.0.

Ref.	Target blockchains	Metrics	Implementation/experiments/methodology
[94]	Ethereum, parity, CITA, hyperledger fabric	TPS, average response delay, transactions per CPU, TX per memory second, TX per disk I/O and TX per network data	Proposed a scalable framework for monitoring the real-time performance blockchain systems. The authors evaluated four popular blockchain systems, i.e., ethereum, parity, CITA and hyperledger fabric.
[95]	Private blockchains	Throughput and latency, scalability, fault tolerance and security, and other micro measurements, e.g., CPU utilization, network utilization, etc.	The authors proposed blockbench for measuring and analyzing the multiple performance of private blockchain systems. Through this blockbench, the authors revealed several insightful bottlenecks and trade-offs while designing the software of blockchains.
[96]	Ethereum	Network size and geographic distribution of ethereum network nodes	Proposed a network monitoring tool named NodeFinder, which is designed to find the unusual network properties of ethereum network nodes in the underlying P2P network perspective.
[97]	Bitcoin network	TPS, network latency, number of forks, and mining rewards	The authors proposed a local Bitcoin network simulator to study the performance of bitcoin under different network conditions including various topologies, network latencies, packet loss rates, and mining difficulties.

Table 1.9 Blockchain dataset frameworks and evaluation tools

Recognition	Target	Ref.	Utilization
XBlock-ETH	Ethereum	[102]	Authors released a new open-source dataset framework for analysis of ethereum, i.e., XBlock-ETH, which includes multiple types of ethereum datasets such as transactions, smart contracts and tokens.
XBlock-EOS	EOS	[103]	Authors proposed a new dataset framework dedicated to EOSIO, named XBlock-EOS, to show how to perform comprehensive statistics and exploration of EOSIO datasets.
BlockSci	General blockchains	[104]	Authors proposed an open-source software platform, named BlockSci, for the analysis of blockchains.
Blockbench	General blockchains	[95]	Authors proposed a benchmarking framework for measuring the data processing capability and performance of different layers of a blockchain system.
NodeFinder	Etheruem nodes	[96]	Authors proposed a measuring tool named NodeFinder, to investigate the opaque network characteristics of ethereum network nodes.
Network simulator for bitcoin	Bitcoin	[97]	Authors proposed a configurable network simulator for the performance measurements of bitcoin using lightweight virtualization technologies.

Nasir et al. [93] conducted performance measurements and discussion of two versions of Hyperledger Fabric. The authors focused on the metrics including execution time, transaction latency, throughput and the scalability versus the number of nodes in blockchain platforms. Several useful insights have been revealed for the two versions of Hyperledger Fabric. As already mentioned previously in [9], the authors evaluated their proposed Monoxide w.r.t the metrics including the scalability of TPS as the number of network zones increase, the overhead of both cross-zone transactions and storage size, the confirmation latency of transactions, and the orphan rate of blocks. In [10], the authors performed rich measurements for their proposed new blockchain protocol Prism under limited network bandwidth and CPU resources. The performance evaluated includes the distribution of block propagation delays, the relationship between block size and mining rate, block size versus assembly time, the expected time to reach consensus on block hash, the expected time to reach consensus on blocks, etc.

Later, Zheng et al. [94] proposed a scalable framework for monitoring the real-time performance blockchain systems. This work has evaluated four popular blockchain systems, i.e., Ethereum, Parity [98], Cryptape Inter-enterprise Trust Automation (CITA) [99] and Hyperledger Fabric [100], in terms of several metrics including *transactions per second*, *average response delay*, *transactions per CPU*, *transactions per memory second*, *transactions per disk I/O* and *transactions per network data*. Such comprehensive performance evaluation results offered us rich viewpoints on the 4 popular blockchain systems. Their experimental logs and technique report [101] can be accessed from http://xblock.pro. Recently, Zheng et al. [102] extended their work and released a new open-source dataset framework, called XBlock-ETH, for the data-driven analysis of Ethereum. XBlock-ETH contains multiple types of Ethereum data such as transactions, smart contracts and tokens. Thus, researchers can extract and explore the data of Ethereum using XBlock-ETH. The authors first collected and cleaned the most recent on-chain dataset from Ethereum. Then, they presented how to perform basic exploration of these datasets to make them best. Like their previous work, those datasets and processing codes can be found from the webpage *xblock.pro* aforementioned. In the other similar work [103] of the same team, authors proposed another new dataset framework dedicated to EOSIO, named XBlock-EOS, which also includes multiple types of rich on-chain/off-chain datasets such as transactions, blocks, smart contracts, internal/external EOS transfer events, tokens, accounts and resource management. To show how to utilize the proposed framework, the authors presented comprehensive statistics and explorations using those datasets, for example, blockchain analysis, smart contract analysis, and cryptocurrency analysis. Finally, this study also discussed future directions of XBlock-EOS in the topics including: (i) data analysis based on off-chain data to provide off-chain user behavior for blockchain developers, (ii) exploring new features of EOSIO data that are different from those of Ethereum, and (iii) conducting a joint analysis of EOSIO with other blockchains.

1.5.2 Useful Evaluation Tools for Blockchains

Kalodner et al. [104] proposed BlockSci, which is designed as an open-source software platform for blockchain analysis. Under the architecture of BlockSci, the raw blockchain data is parsed to produce the core blockchain data including transaction graph, indexes and scripts, which are then provided to the analysis library. Together with the auxiliary data including P2P data, price data and user tags, a client can either directly query or read through a Jupyter notebook interface.

To evaluate the performance of private blockchains, Dinh et al. [95] proposed a benchmarking framework, named Blockbench, which can measure the data processing capability and the performance of various layers of a blockchain system. Using such Blockbench, the authors then performed detailed measurements and analysis of three blockchains, i.e., Ethereum, Parity and Hyperledger. The results disclosed some useful experiences of those three blockchain systems. For example, today's blockchains are not scalable w.r.t data processing workloads, and several bottlenecks should be considered while designing different layers of blockchain in the software engineering perspective.

Ethereum has received enormous attention on the mining challenges, the analytics of smart contracts, and the management of block mining. However, not so many efforts have been spent on the information dissemination in the perspective of P2P networks. To fill this gap, Kim et al. [96] proposed a measuring tool named NodeFinder, which aims to discover the opaque network properties of Ethereum network nodes. Through a three-month long data collection on the P2P network, the authors analyzed and found several unprecedented differences of Ethereum network comparing with other popular P2P networks like BitTorrent, Bitcoin and Gnutella in terms of network size and geographic distribution.

Recently, by exploiting lightweight virtualization technologies, Alsahan et al. [97] developed a configurable network simulator for the performance measurements of Bitcoin. The proposed simulator allows users to configure diverse network conditions, such as blockchain network topology, link delays, and mining difficulties, to emulate the real-world operation environment. Using this simulator, experiments can be performed to measure Bitcoin network under various network conditions. It also supports conducting the tests of security attacks and point of failure simulations. The authors also made this simulator open-source on Github.

1.6 Open Issues and Future Directions

In this section, we envision the open issues and promising directions for future studies.

1.6.1 Performance-Improving Issues

1.6.1.1 Scalability Issues

Scalability is still a severe challenge for most of the blockchain systems. For example, the PBFT consensus protocols issue a $O(n^2)$ number of messages, where n is the number of participants. The large number of messages makes the scalability unrealistic. Therefore, new distributed practical byzantine protocols and theoretical modelings of scalability solutions, such as sidechain, subchain, off-chain, sharding technique, DAG, and even chain-less proposals, are in an urgent need for scalable blockchains.

1.6.1.2 Resilient Mechanisms for Sharding Technique

The sharding technique includes three typical categories, i.e., transaction sharding, network sharding, and state sharding. Via the extensive review on the existing studies of sharding techniques, we found that the resilient mechanisms for sharding blockchains are still missing. Particularly to the state sharding, once the failures occurred on blockchain nodes, how to ensure the correct recovery of the real-time running states in the failed blockchain node(s) is critical to the resilience and robustness of the blockchain.

1.6.1.3 Cross-Shard Performance

Although a number of committee-based sharding protocols [2, 9, 17, 105] have been proposed, those protocols can only endure at most 1/3 adversaries. Thus, more robust byzantine agreement protocols need to be devised. Furthermore, all the sharding-based protocols incur additional cross-shard traffics and latencies because of the cross-shard transactions. Therefore, the cross-shard performance in terms of throughput, latency and other metrics, has to be well guaranteed in future studies. On the other hand, the cross-shard transactions are inherent for the cross-shard protocols. Thus, the pros and cons of such the correlation between different shards are worthy investigating using certain modelings and theories such as graph-based analysis.

1.6.1.4 Cross-Chain Transaction Accelerating Mechanisms

On cross-chain operations, [28] is essentially a pioneer step towards practical blockchain-based ecosystems. Following this roadmap paved by Jin et al. [28], we are exciting to anticipate the subsequent related investigations will appear soon in the near future. For example, although the inter-chain transaction experiments achieve an initial success, we believe that the secure cross-chain transaction

accelerating mechanisms are still on the way. In addition, further improvements are still required for the interoperability among multiple blockchains, such as decentralized load balancing smart contracts for sharded blockchains.

1.6.1.5 Ordering Blocks for Multiple-Chain Protocols

Although multiple-chain techniques can improve the throughput by exploiting the parallel mining of multiple chain instances, how to construct and manage the blocks in all chains in a globally consistent order is still a challenge to the multiple-chain based scalability protocols and solutions.

1.6.1.6 Hardware-Assisted Accelerating Solutions for Blockchain Networks

To improve the performance of blockchains, for example, to reduce the latency of transaction confirmation, some advanced network technologies, such as RDMA (Remote Direct Memory Access) and high-speed network cards, can be exploited in accelerating the data-access among miners in blockchain networks.

1.6.1.7 Performance Optimization in Different Blockchain Network Layers

The blockchain network is built over the P2P networks, which include several typical layers, such as mac layer, routing layer, network layer, and application layer. The BFT-based protocols are essentially working for the network layer. In fact, performance improvements can be achieved by proposing various protocols, algorithms, and theoretical models for other layers of the blockchain network.

1.6.1.8 Blockchain-Assisted BigData Networks

Although big data and blockchain have several performance metrics that are contrary to each other. For example, big data is a centralized management technology with an emphasize on the privacy-preserving oriented to diverse computing environments. The data processed by big data technology should ensure nonredundancy and unstructured architecture in a large-scale computing network. In contrast, blockchain technology builds on a decentralized, transparent and immutable architecture, in which data type is simple, data is structured and highly redundant. Furthermore, the performance of blockchains require scalability and the off-chain computing paradigm. Thus, how to integrate those two technologies together and pursue the mutual benefit for each other is an open issue that is worthy in-depth studies. For example, the potential research topics include how to design a suitable

new blockchain architecture for big data technologies, and how to break the isolated data islands using blockchains while guaranteeing the privacy issues of big data.

1.6.2 Issues for Better Understanding Blockchains Further

Although the state-of-the-art studies have reviewed a lot of modelings and theories for better understanding blockchains, more sophisticated approaches and insightful mechanisms are still needed to help researchers gain a new level of perception over the high-performance blockchain systems. Some interesting directions are summarized here for inspiring more subsequent investigations.

- Exploiting more general queueing theories to capture the real-world arrival process of transactions, mining new blocks, and other queueing-related blockchain phases.
- Performing priority-based service policies while dealing with transactions and new blocks, to meet a predefined security or regulation level.
- Developing more general probabilistic models to characterize the correlations among the multiple performance parameters of blockchain systems.

1.6.3 Security Issues of Blockchains

1.6.3.1 Privacy-Preserving for Blockchains

From the previous overview, we observe that most of the existing works under this category are discussing the blockchain-based security and privacy-preserving applications. The fact is that the security and privacy are also the critical issues of the blockchain itself. For example, the privacy of transactions could be hacked by attackers. However, dedicated studies focusing on those issues are still insufficient.

1.6.3.2 Anti-cryptojacking Mechanisms for Malicious Miners

The Cryptojacking Miners are reportedly existing in web browsers according to [79]. This type of malicious codes is commandeering the hardware resources such as computational capability and memory of web users. Thus, the anti-cryptojacking mechanisms and strategies are necessary to develop for protecting normal browser users.

1.6.3.3 Security Issues of Cryptocurrency Blockchains

The security issues of cryptocurrency blockchains, such as double-spend attacks, frauds in smart contracts, have arisen growing attention from both industrial and academic fields. However, little efforts have been committed to the theoretical investigations towards the security issues of cryptocurrency blockchains. For example, the exploration of punishment and cooperation between miners over multiple chains is an interesting topic for cryptocurrency blockchains. Thus, we expect to see broader perspectives of modeling the behaviors of both attackers and counterattackers in the context of monetary blockchain attacks.

1.6.4 Powerful Experimental Platforms for Blockchains

To most of the beginners in the field of the blockchain, they have a problem about lack of powerful simulation/emulation tools for verifying their new ideas or protocols. Therefore, the powerful simulation/emulation platforms that are easy to deploy scalable testbeds for the experiments would be very helpful to the research community.

Tailor-made experiment platforms based on existing blockchain systems are also needed. Building a blockchain system from scratch, or learning from the implementation of existing blockchain systems by reading codes, these are some time-consuming yet not rewarding tasks for researchers. A platform that enable us to tweak a variety of aspects of interest in existing blockchain systems can potentially be very helpful to the research community as well.

1.7 Conclusion

Through investigations, we found that a dedicated survey focusing on the theoretical modelings, analytical models and useful experiment tools for blockchains is still missing. To fill this gap, we then conducted a comprehensive survey of the state-of-the-art on blockchains, particularly in the perspectives of theories, modelings, and measurement/evaluation tools. The taxonomy of each topic presented in this chapter tried to convey the new protocols, ideas, and solutions that can improve the scalability of blockchains, and help people better understand the blockchains in a further level. We believe our work provides a timely guidance on the theoretical insights of blockchains for researchers, engineers, educators, and generalized readers.

Acknowledgments This work was supported in part by the Key-Area Research and Development Program of Guangdong Province (No. 2019B020214006), the National Natural Science Foundation of China (No. 62032025, No. 61902445, No. 61872310), the Fundamental Research Funds for the Central Universities of China (No.19lgpy222), the Guangdong Basic and Applied Basic Research Foundation (No. 2019A1515011798), the Hong Kong RGC Research Impact Fund (RIF) (No. R5060-19, No. R5034-18), General Research Fund (GRF) (No. 152221/19E), and the Collaborative Research Fund (CRF) (No. C5026-18G).

References

1. Luu L, Narayanan V, Zheng C, Baweja K, Gilbert S, Saxena P (2016) A secure sharding protocol for open blockchains. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, pp 17–30
2. Kokoris-Kogias E, Jovanovic P, Gasser L, Gailly N, Syta E, Ford B (2018) Omniledger: a secure, scale-out, decentralized ledger via sharding. In: 2018 IEEE symposium on security and privacy (SP). IEEE, Piscataway, pp 583–598
3. Moroz DJ, Aronoff DJ, Narula N, Parkes DC (2020) Double-spend counterattacks: threat of retaliation in proof-of-work systems
4. Wood G, et al (2014) Ethereum: a secure decentralised generalised transaction ledger. Ethereum project yellow paper 151:1–32
5. Ethereum sharding. https://eth.wiki/sharding/Sharding-FAQs
6. Hyperledger fabric website. https://hyperledger-fabric.readthedocs.io/en/release-1.4/write_first_app.html
7. EOSIO (2020). https://eos.io/
8. Hari A, Kodialam M, Lakshman T (2019) Accel: accelerating the bitcoin blockchain for high-throughput, low-latency applications. In: IEEE conference on computer communications (INFOCOM'19). IEEE, Piscataway, pp 2368–2376
9. Wang J, Wang H (2019) Monoxide: scale out blockchains with asynchronous consensus zones. In: Proceedings of 16th USENIX symposium on networked systems design and implementation (NSDI), pp 95–112
10. Yang L, Bagaria V, Wang G, Alizadeh M, Tse D, Fanti G, Viswanath P (2019) Prism: scaling bitcoin by 10,000 x. arXiv:190911261
11. Woo S, Song J, Kim S, Kim Y, Park S (2020) GARET: improving throughput using gas consumption-aware relocation in ethereum sharding environments. Cluster Comput. 23:2235–2247
12. Perard D, Lacan J, Bachy Y, Detchart J (2018) Erasure code-based low storage blockchain node. In: 2018 IEEE international conference on internet of things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData). IEEE, Piscataway, pp 1622–1627
13. Dai X, Xiao J, Yang W, Wang C, Jin H (2019) Jidar: a jigsaw-like data reduction approach without trust assumptions for bitcoin system. In: IEEE 39th international conference on distributed computing systems (ICDCS). IEEE, Piscataway, pp 1317–1326
14. Xu Y, Huang Y (2020) Segment blockchain: a size reduced storage mechanism for blockchain. IEEE Access 8:17434–17441
15. Weber I, Gramoli V, Ponomarev A, Staples M, Holz R, Tran AB, Rimba P (2017) On availability for blockchain-based systems. In: 2017 IEEE 36th symposium on reliable distributed systems (SRDS). IEEE, Piscataway, pp 64–73
16. Zheng P, Zheng Z, Chen L (2019) Selecting reliable blockchain peers via hybrid blockchain reliability prediction. arXiv:191014614
17. Zamani M, Movahedi M, Raykova M (2018) Rapidchain: scaling blockchain via full sharding. In: Proceedings of the 2018 ACM SIGSAC conference on computer and communications security, pp 931–948

18. Amiri MJ, Agrawal D, Abbadi AE (2019) Sharper: sharding permissioned blockchains over network clusters. arXiv:191000765
19. Kim S, Song J, Woo S, Kim Y, Park S (2019) Gas consumption-aware dynamic load balancing in ethereum sharding environments. In: IEEE 4th International Workshops on foundations and applications of self* systems (FAS*W). IEEE, Piscataway, pp 188–193
20. Wang J, Zhou Y, Li X, Xu T, Qiu T (2019) A node rating based sharding scheme for blockchain. In: Proceedings of IEEE 25th international conference on parallel and distributed systems (ICPADS). IEEE, Piscataway, pp 302–309
21. Nguyen LN, Nguyen TD, Dinh TN, Thai MT (2019) Optchain: optimal transactions placement for scalable blockchain sharding. In: Proceedings of IEEE 39th international conference on distributed computing systems (ICDCS), pp 525–535
22. Dang H, Dinh TTA, Loghin D, Chang EC, Lin Q, Ooi BC (2019) Towards scaling blockchain systems via sharding. In: Proceedings of the 2019 international conference on management of data, pp 123–140
23. Chen H, Wang Y (2019) SSChain: a full sharding protocol for public blockchain without data migration overhead. Pervasive Mobile Comput 59:101055
24. Niu J (2019) Eunomia: a permissionless parallel chain protocol based on logical clock. arXiv:190807567
25. Rajab T, Manshaei MH, Dakhilalian M, Jadliwala M, Rahman MA (2020) On the feasibility of sybil attacks in shard-based permissionless blockchains. arXiv:200206531
26. Xu Y, Huang Y (2020) An n/2 byzantine node tolerate blockchain sharding approach. arXiv:200105240
27. Zhang M, Li J, Chen Z, Chen H, Deng X (2020) Cycledger: a scalable and secure parallel protocol for distributed ledger via sharding. arXiv:200106778
28. Jin H, Dai X, Xiao J (2018) Towards a novel architecture for enabling interoperability amongst multiple blockchains. In: 2018 IEEE 38th international conference on distributed computing systems (ICDCS). IEEE, Piscataway, pp 1203–1211
29. Liu Z, Xiang Y, Shi J, Gao P, Wang H, Xiao X, Wen B, Hu YC (2019) Hyperservice: interoperability and programmability across heterogeneous blockchains. In: Proceedings of the 2019 ACM SIGSAC conference on computer and communications security, pp 549–566
30. Fynn E, Bessani A, Pedone F (2020) Smart contracts on the move. arXiv:200405933
31. Tian H, Xue K, Li S, Xu J, Liu J, Zhao J (2020) Enabling cross-chain transactions: a decentralized cryptocurrency exchange protocol. arXiv:200503199
32. Zamyatin A, Al-Bassam M, Zindros D, Kokoris-Kogias E, Moreno-Sanchez P, Kiayias A, Knottenbelt WJ (2019) SoK: communication across distributed ledgers. Tech. Rep., IACR Cryptology ePrint Archive, 2019:1128
33. Gao Y, Kawai S, Nobuhara H (2019) Scalable blockchain protocol based on proof of stake and sharding. J Adv Comput Intell Intell Inf 23(5):856–863. https://doi.org/10.20965/jaciii.2019.p0856
34. David B, Gaži P, Kiayias A, Russell A (2018) Ouroboros praos: an adaptively-secure, semi-synchronous proof-of-stake blockchain. In: Annual international conference on the theory and applications of cryptographic techniques. Springer, Berlin, pp 66–98
35. Buchman E, Kwon J, Milosevic Z (2018) The latest gossip on bft consensus. arXiv:180704938
36. Zou J, Ye B, Qu L, Wang Y, Orgun MA, Li L (2018) A proof-of-trust consensus protocol for enhancing accountability in crowdsourcing services. IEEE Trans Serv Comput 12:429–445
37. István Z, Sorniotti A, Vukolić M (2018) Streamchain: do blockchains need blocks? In: Proceedings of the 2nd workshop on scalable and resilient infrastructures for distributed ledgers, pp 1–6
38. Amiri MJ, Agrawal D, Abbadi AE (2019) Caper: a cross-application permissioned blockchain. Proc VLDB Endowment 12(11):1385–1398
39. Chang Z, Guo W, Guo X, Zhou Z, Ristaniemi T (2020) Incentive mechanism for edge computing-based blockchain. IEEE Trans Ind Inf 16(11):7105–7114. https://doi.org/10.1109/TII.2020.2973248

40. Zheng W, Chen X, Zheng Z, Luo X, Cui J (2020) AxeChain: a secure and decentralized blockchain for solving easily-verifiable problems. arXiv:200313999
41. Chen L, Xu L, Gao Z, Kasichainula K, Shi W (2020) Nonlinear blockchain scalability: a game-theoretic perspective. arXiv:200108231
42. Sompolinsky Y, Zohar A (2015) Secure high-rate transaction processing in bitcoin. In: International Conference on Financial Cryptography and Data Security. Springer, Berlin, pp 507–527
43. Sompolinsky Y, Zohar A (2018) Phantom: a scalable blockdag protocol. IACR Cryptology ePrint Archive 2018:104
44. Li C, Li P, Zhou D, Xu W, Long F, Yao A (2018) Scaling nakamoto consensus to thousands of transactions per second. arXiv:180503870
45. Chen T, Zhu Y, Li Z, Chen J, Li X, Luo X, Lin X, Zhange X (2018) Understanding ethereum via graph analysis. In: Proceedings of IEEE conference on computer communications (INFOCOM). IEEE, Piscataway, pp 1484–1492
46. Akcora CG, Dixon MF, Gel YR, Kantarcioglu M (2018) Bitcoin risk modeling with blockchain graphs. Econ Lett 173:138–142
47. Dixon MF, Akcora CG, Gel YR, Kantarcioglu M (2019) Blockchain analytics for intraday financial risk modeling. Digit Financ 1(1–4):67–89
48. Akcora CG, Dey AK, Gel YR, Kantarcioglu M (2018) Forecasting bitcoin price with graph chainlets. In: Pacific-Asia conference on knowledge discovery and data mining. Springer, Berlin, pp 765–776
49. Abay NC, Akcora CG, Gel YR, Kantarcioglu M, Islambekov UD, Tian Y, Thuraisingham B (2019) Chainnet: learning on blockchain graphs with topological features. In: IEEE international conference on data mining (ICDM), pp 946–951
50. Chen W, Wu J, Zheng Z, Chen C, Zhou Y (2019) Market manipulation of bitcoin: evidence from mining the Mt. Gox transaction network. In: Proceedings of IEEE conference on computer communications (INFOCOM), pp 964–972
51. Victor F, Lüders BK (2019) Measuring ethereum-based ERC20 token networks. In: International conference on financial cryptography and data security. Springer, Berlin, pp 113–129
52. Somin S, Gordon G, Altshuler Y (2018) Network analysis of ERC20 tokens trading on ethereum blockchain. In: International conference on complex systems. Springer, Berlin, pp 439–450
53. Zhao Y, Liu J, Han Q, Zheng W, Wu J (2020) Exploring eosio via graph characterization. arXiv:200410017
54. Papadis N, Borst S, Walid A, Grissa M, Tassiulas L (2018) Stochastic models and wide-area network measurements for blockchain design and analysis. In: Proceedings of IEEE conference on computer communications (INFOCOM). IEEE, Piscataway, pp 2546–2554
55. Gopalan A, Sankararaman A, Walid A, Vishwanath S (2020) Stability and scalability of blockchain systems. arXiv:200202567
56. Hafid A, Hafid AS, Samih M (2019) A probabilistic security analysis of sharding-based blockchain protocols. In: Proceedings of international congress on blockchain and applications (Blockchain), pp 55–60
57. Hafid A, Hafid AS, Samih M (2019) A methodology for a probabilistic security analysis of sharding-based blockchain protocols. In: Proceedings of international congress on blockchain and applications. Springer, Berlin, pp 101–109
58. Hafid A, Hafid AS, Samih M (2019) New mathematical model to analyze security of sharding-based blockchain protocols. IEEE Access 7:185447–185457
59. Li QL, Ma JY, Chang YX (2018) Blockchain queue theory. In: International conference on computational social networks. Springer, Berlin, pp 25–40
60. Li QL, Ma JY, Chang YX, Ma FQ, Yu HB (2019) Markov processes in blockchain systems. Comput Soc Netw 6(1):1–28
61. Ricci S, Ferreira E, Menasche DS, Ziviani A, Souza JE, Vieira AB (2019) Learning blockchain delays: a queueing theory approach. ACM SIGMETRICS Perform Eval Rev 46(3):122–125

62. Frolkova M, Mandjes M (2019) A bitcoin-inspired infinite-server model with a random fluid limit. Stoch Models 35(1):1–32
63. Fang M, Liu J (2020) Toward low-cost and stable blockchain networks. arXiv:200208027
64. Memon RA, Li JP, Ahmed J (2019) Simulation model for blockchain systems using queuing theory. Electronics 8(2):234
65. Wüst K, Gervais A (2018) Do you need a blockchain? In: 2018 crypto valley conference on blockchain technology (CVCBT). IEEE, Piscataway, pp 45–54
66. Zhang X, Liu C, Nepal S, Pandey S, Chen J (2013) A privacy leakage upper bound constraint-based approach for cost-effective privacy preserving of intermediate data sets in cloud. IEEE Trans Parallel Distrib Syst 24(6):1192–1202
67. Lin D, Wu J, Yuan Q, Zheng Z (2020) Modeling and understanding ethereum transaction records via a complex network approach. IEEE Trans Circuits Syst II: Exp Briefs 67(11):2737–2741
68. Sousa JEA, Oliveira V, Valadares J, Vieira AB, Bernardino HS, Dias G (2019) An analysis of the fees and pending time correlation in ethereum. In: Proceedings of LANOMS, IFIP, pp 1–7
69. Altman E, Menasché D, Reiffers A, Datar M, Dhamal S, Touati C, El-Azouzi R (2019) Blockchain competition between miners: a game theoretic perspective. Front Blockchain 2:26
70. Zhao J, Tang J, Li Z, Wang H, Lam KY, Xue k (2020) An analysis of blockchain consistency in asynchronous networks: deriving a neat bound. In: Proceedings of IEEE international conference on distributed computing systems (ICDCS), pp 1–10
71. Xiao Y, Zhang N, Lou W, Hou YT (2020) Modeling the impact of network connectivity on consensus security of proof-of-work blockchain. In: IEEE conference on computer communications (INFOCOM'20), pp 1–9
72. Fynn E, Pedone F (2018) Challenges and pitfalls of partitioning blockchains. In: 2018 48th annual IEEE/IFIP international conference on dependable systems and networks workshops (DSN-W). IEEE, Piscataway, pp 128–133
73. Avarikioti Z, Kokoris-Kogias E, Wattenhofer R (2019) Divide and scale: formalization of distributed ledger sharding protocols. arXiv:191010434
74. Wang S, Wang C, Hu Q (2019) Corking by forking: vulnerability analysis of blockchain. In: Proceedings of IEEE conference on computer communications (INFOCOM). IEEE, Piscataway, pp 829–837
75. Bessani A, Alchieri E, Sousa J, Oliveira A, Pedone F (2020) From byzantine replication to blockchain: consensus is only the beginning. arXiv:200414527
76. Yu G, Zha X, Wang X, Ni W, Yu K, Zhang JA, Liu RP (2020) A unified analytical model for proof-of-x schemes. Comput Secur 96:101934
77. Wu S, Chen Y, Li M, Luo X, Liu Z, Liu L (2020) Survive and thrive: a stochastic game for DDoS attacks in bitcoin mining pools. IEEE/ACM Trans Netw 28(2):874–887
78. Budish E (2018) The economic limits of bitcoin and the blockchain. Tech. Rep., National Bureau of Economic Research
79. Tahir R, Durrani S, Ahmed F, Saeed H, Zaffar F, Ilyas S (2019) The browsers strike back: countering cryptojacking and parasitic miners on the web. In: Proceedings of IEEE conference on computer communications (INFOCOM). IEEE, Piscataway, pp 703–711
80. Ning R, Wang C, Xin C, Li J, Zhu L, Wu H (2019) CapJack: capture in-browser crypto-jacking by deep capsule network through behavioral analysis. In: Proceedings of IEEE conference on computer communications (INFOCOM). IEEE, Piscataway, pp 1873–1881
81. Hu Y, Seneviratne S, Thilakarathna K, Fukuda K, Seneviratne A (2019) Characterizing and detecting money laundering activities on the bitcoin network. arXiv:191212060
82. Vasek M, Moore T (2018) Analyzing the bitcoin ponzi scheme ecosystem. In: International conference on financial cryptography and data security. Springer, Berlin, pp 101–112
83. Chen W, Zheng Z, Cui J, Ngai E, Zheng P, Zhou Y (2018) Detecting ponzi schemes on ethereum: towards healthier blockchain technology. In: Proceedings of the 2018 world wide web conference (WWW), pp 1409–1418

84. Chen W, Zheng Z, Ngai ECH, Zheng P, Zhou Y (2019) Exploiting blockchain data to detect smart ponzi schemes on ethereum. IEEE Access 7:37575–37586
85. Laskowski M, Zargham M, Turesson H, Kim HM, Barlin M, Kabanov D, Dhaliwal E (2020) Evidence based decision making in blockchain economic systems: from theory to practice. arXiv:200103020
86. Yaish A, Zohar A (2020) Pricing ASICs for cryptocurrency mining. arXiv:200211064
87. Eskandari S, Leoutsarakos A, Mursch T, Clark J (2018) A first look at browser-based cryptojacking. In: IEEE European symposium on security and privacy workshops (EuroS&PW). IEEE, Piscataway, pp 58–66
88. Sabour S, Frosst N, Hinton GE (2017) Dynamic routing between capsules. In: Advances in neural information processing systems, pp 3856–3866
89. Bartoletti M, Carta S, Cimoli T, Saia R (2020) Dissecting ponzi schemes on ethereum: identification, analysis, and impact. Future Gener Comput Syst 102:259–277
90. Bartoletti M, Pes B, Serusi S (2018) Data mining for detecting bitcoin ponzi schemes. In: 2018 crypto valley conference on blockchain technology (CVCBT). IEEE, Piscataway, pp 75–84
91. Seo J, Park M, Oh H, Lee K (2018) Money laundering in the bitcoin network: perspective of mixing services. In: Proceedings of IEEE international conference on information and communication technology convergence (ICTC), pp 1403–1405
92. Gervais A, Karame GO, Wüst K, Glykantzis V, Ritzdorf H, Capkun S (2016) On the security and performance of proof of work blockchains. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security, pp 3–16
93. Nasir Q, Qasse IA, Abu Talib M, Nassif AB (2018) Performance analysis of hyperledger fabric platforms. Secur Commun Netw
94. Zheng P, Zheng Z, Luo X, Chen X, Liu X (2018) A detailed and real-time performance monitoring framework for blockchain systems. In: Proceedings of IEEE/ACM 40th international conference on software engineering: software engineering in practice track (ICSE-SEIP), pp 134–143
95. Dinh TTA, Wang J, Chen G, Liu R, Ooi BC, Tan KL (2017) Blockbench: a framework for analyzing private blockchains. In: Proceedings of the 2017 ACM international conference on management of data, pp 1085–1100
96. Kim SK, Ma Z, Murali S, Mason J, Miller A, Bailey M (2018) Measuring ethereum network peers. In: Proceedings of the internet measurement conference (IMC'18), pp 91–104
97. Alsahan L, Lasla N, Abdallah MM (2020) Local bitcoin network simulator for performance evaluation using lightweight virtualization. In: Proceedings of IEEE international conference on informatics, IoT, and enabling technologies, pp 1–6
98. Parity documentation. https://paritytech.github.io/wiki
99. Cita technical whitepaper. https://github.com/cryptape/cita
100. Androulaki E, Barger A, Bortnikov V, Cachin C, Christidis K, De Caro A, Enyeart D, Ferris C, Laventman G, Manevich Y, et al (2018) Hyperledger fabric: a distributed operating system for permissioned blockchains. In: Proceedings of the thirteenth Eurosys conference, pp 1–15
101. Xblock (2020) Performance monitoring. http://xblock.pro/performance/
102. Zheng P, Zheng Z, Dai Hn (2019) XBlock-ETH: extracting and exploring blockchain data from etherem. arXiv:191100169
103. Zheng W, Zheng Z, Dai HN, Chen X, Zheng P (2020) XBlock-EOS: extracting and exploring blockchain data from EOSIO. arXiv:200311967
104. Kalodner H, Goldfeder S, Chator A, Möser M, Narayanan A (2017) BlockSci: design and applications of a blockchain analysis platform. arXiv:170902489
105. Miller A, Xia Y, Croman K, Shi E, Song D (2016) The honey badger of BFT protocols. In: Proceedings of the 2016 ACM SIGSAC conference on computer and communications security (CCS), pp 31–42