A Survey of State-of-the-Art on Blockchains: Theories, Modelings, and Tools

HUAWEI HUANG, WEI KONG, SICONG ZHOU, and ZIBIN ZHENG,
Sun Yat-Sen University, China
SONG GUO, The Hong Kong Polytechnic University, China

To draw a roadmap of current research activities of the blockchain community, we first conduct a brief overview of state-of-the-art blockchain surveys published in the past 5 years. We found that those surveys are basically studying the blockchain-based applications, such as blockchain-assisted Internet of Things (IoT), business applications, security-enabled solutions, and many other applications in diverse fields. However, we think that a comprehensive survey toward the essentials of blockchains by exploiting the state-of-the-art theoretical modelings, analytic models, and useful experiment tools is still missing. To fill this gap, we perform a thorough survey by identifying and classifying the most recent high-quality research outputs that are closely related to the theoretical findings and essential mechanisms of blockchain systems and networks. Several promising open issues are also summarized for future research directions. We hope this survey can serve as a useful guideline for researchers, engineers, and educators about the cutting-edge development of blockchains in the perspectives of theories, modelings, and tools.

CCS Concepts: • **General and reference → Surveys and overviews**; • **Computer systems organization → Dependable and fault-tolerant systems and networks**; • **Security and privacy → Distributed systems security**;

Additional Key Words and Phrases: Blockchain, theoretical modelings, analytic models, experiment tools

ACM Reference format:
Huawei Huang, Wei Kong, Sicong Zhou, Zibin Zheng, and Song Guo. 2021. A Survey of State-of-the-Art on Blockchains: Theories, Modelings, and Tools. *ACM Comput. Surv.* 54, 2, Article 44 (March 2021), 42 pages.
https://doi.org/10.1145/3441692

44

1 INTRODUCTION

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

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).
Authors' addresses: H. Huang, W. Kong, S. Zhou, and Z. Zheng (corresponding author), Sun Yat-Sen University, School of Computer Science and Engineering, 132 East Waihuan Road, Higher Education Mega Center, Guangzhou, Guangdong, 510006, China; emails: {huanghw28, zhzibin}@mail.sysu.edu.cn; S. Guo (corresponding author), The Hong Kong Polytechnic University, Department of Computing, Hong Kong SAR, China; email: song.guo@polyu.edu.hk.

0360-0300/2021/03-ART44 $15.00
https://doi.org/10.1145/3441692

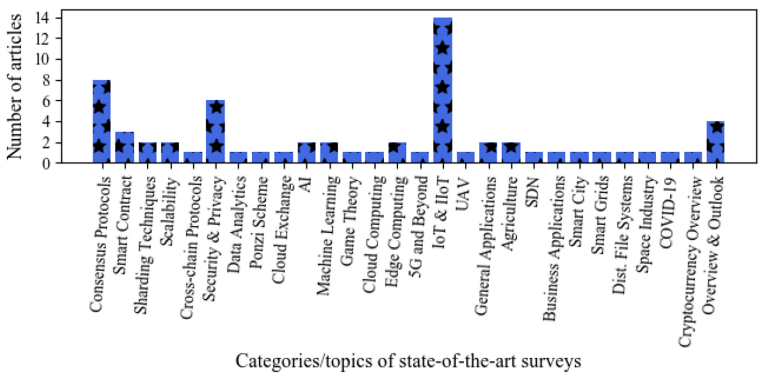

Fig. 1. The categories and number of state-of-the-art blockchain-related surveys published in the past few years.

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 so on. In the following, before discussing the motivation of this survey, we first conduct a brief exposition of the state-of-the-art blockchain survey articles published in the recent few years.

1.1 Taxonomy of State-of-the-Art Blockchain Surveys

To identify the position of our survey, we first collect 67 state-of-the-art blockchain-related survey articles. The numbers of each category of those surveys are shown in Figure 1. We see that the top-three popular topics of blockchain-related survey are IoT and Industrial Internet of Things (IIoT), Consensus Protocols, and Security and privacy. We also classify those existing surveys and their chronological distribution in Figure 1 of the online material, from which we discover that (i) the number of surveys published in each year increases dramatically, and (ii) the diversity of topics also becomes greater following the chronological order. In detail, we summarize the publication years, topics, and other metadata of these surveys in Tables 1 and 2. Basically, those surveys can be classified into the following seven groups. The overall principal of the collection is based on different aspects of blockchain covered in the surveys. In *Group-1*, different abstraction layers of blockchain protocols and intrinsic properties are the main focus. In *Group-2*, the behavior of blockchain's clients are analyzed by means of data mining. In *Group-3*, blockchain as a complicated and rewarding environment is reviewed, with hard choices made by means of AI or game theory. Also, surveys that analyzed the integration of blockchain and decision-making techniques are also classified in this group. In *Group-4*, the integration of blockchain and different communication techniques are reviewed. In *Group-5* and *Group-6*, the applications of blockchain are reviewed. We singled out surveys on IoT applications due to the popularity. *Group-7* are the works of holistic overview of blockchain.

1.1.1 *Blockchain Essentials.*

The first group is related to the essentials of the blockchain. A large number of consensus protocols, algorithms, and mechanisms have been reviewed and summarized in [1–8]. For example, motivated by lack of a comprehensive literature review regarding the consensus protocols for blockchain networks, Wang et al. [3] emphasized on both the system design and the incentive mechanism behind those distributed blockchain consensus protocols such as Byzantine Fault Tolerant (BFT)-based protocols and Nakamoto protocols. From a game-theoretic

Table 1. Taxonomy of Existing Blockchain-Related Surveys (Part 1)

Group	Category	Ref.	Year	Topic
Group-1: **Blockchain Essentials**	Consensus Protocols	Sankar [1]	2017	Consensus protocols on blockchain applications
		Yuan [2]	2018	Blockchain consensus algorithms
		Wang [3]	2018	Consensus mechanisms and mining management in blockchains
		Garay [4]	2018	Consensus taxonomy in blockchain era
		Nguyen [5]	2018	Consensus algorithms used in blockchains
		Wang [6]	2019	Consensus and mining strategy in blockchain networks
		Bano [7]	2019	Consensus in the age of blockchains
		Xiao [8]	2020	Distributed consensus protocols for blockchain networks
	Smart Contract	Atzei [9]	2016	Attacks on Ethereum smart contracts
		Dwived [10]	2019	Blockchain-based smart-contract languages
		Zheng [11]	2020	Challenges, advances, and platforms of smart contracts
	Sharding	Wang [12]	2019	Sharding on blockchains
		Yu [13]	2020	Sharding in blockchains
	Scalability	Pan [14]	2018	Scalability of blockchain technology
		Zhou [15]	2020	Solutions to scalability of blockchain
	Cross-chain	Zamyatin [16]	2019	Cross-ledger communications
		Belchior [17]	2020	Interoperability solutions and problems
	Security and Privacy	Taylor [18]	2019	Blockchain cyber security
		Dasgupta [19]	2019	Security perspective of blockchain
		Ma [20]	2019	Blockchain technology in crowdsourcing services
		Tariq [21]	2019	Security of big data in blockchain-enabled IoT applications
		Feng [22]	2019	Privacy protection in blockchain systems
		Soni [23]	2019	Security, privacy, and potential applications of blockchain
		Xie [24]	2020	Analysis of vulnerabilities, attacks, and other risks of cloud exchange markets
Group-2: **Data Mining**	Data Analytics	Chen [25]	2018	Blockchain data analysis
	Ponzi Scheme	Bartoletti [26]	2020	Dissecting Ponzi schemes on Ethereum
Group-3: **Decision-Making Techniques**	Artificial Intelligence (AI)	Salah [27]	2019	Blockchain for artificial Intelligence
		Zheng [28]	2020	Blockchain and artificial Intelligence
	Machine Learning (ML)	Chen [29]	2018	Privacy and security design when integrating ML and blockchain
		Liu [30]	2020	Blockchain and ML for comm. and networking systems
	Game Theory	Liu [31]	2019	Game theories on blockchain
Group-4: **New Comm. Networking**	Cloud Computing	Park [32]	2017	Blockchain security in cloud computing
	Edge Computing	Xiong [33]	2018	Blockchain meets edge computing
		Yang [34]	2019	Integration of blockchain and edge computing systems
	5G and Beyond	Nguyen [35]	2019	Blockchain for 5G and beyond networks

Table 2. Taxonomy of Existing Blockchain-Related Surveys (Part 2)

Group	Category	Ref.	Year	Topic
Group-5: **IoT & IIoT**	IoT, IIoT	Christidis [36]	2016	Blockchains and smart contracts for IoT
		Ali [37]	2018	Applications of blockchains in IoT
		Fernandez [38]	2018	Usage of blockchain for IoT
		Kouicem [39]	2018	IoT security
		Panarello [40]	2018	Integration of blockchain and IoT
		Dai [41]	2019	Blockchain for IoT
		Wang [42]	2019	Blockchain for IoT
		Nguyen [43]	2019	Integration of blockchain and cloud of things
		Restuccia [44]	2019	Blockchain technology for IoT
		Cao [45]	2019	Challenges in distributed consensus of IoT
		Park [46]	2020	Blockchain technology for green IoT
		Lao [47]	2020	IoT applications in blockchain systems
		Alladi [48]	2019	Blockchain applications in Industry 4.0 and IIoT
		Zhang [49]	2019	5G Beyond for IIoT based on edge intelligence and blockchain
	UAV	Alladi [50]	2020	Blockchain-based UAV applications
Group-6: **Blockchain Applications**	General Applications	Lu [51]	2018	Functions, applications, and open issues of blockchain
		Casino [52]	2019	Current status, classification, and open issues of blockchain apps
	Agriculture	Bermeo [53]	2018	Blockchain technology in agriculture
		Ferrag [54]	2020	Blockchain solutions to Ssecurity and privacy for green agriculture
	SDN	Alharbi [55]	2020	Deployment of blockchains for software defined networks
	Business Apps	Konst. [56]	2018	Blockchain-based business applications
	Smart City	Xie [57]	2019	Blockchain technology applied in smart cities
	Smart Grids	Alladi [58]	2019	Blockchain in use cases of smart grids
		Aderibole [59]	2020	Smart grids based on blockchain technology
	File Systems	Huang [60]	2020	Blockchain-based distributed file systems, IPFS, Filecoin, etc.
	Space Industry	Torky [61]	2020	Blockchain in space industry
	COVID-19	Nguyen [62]	2020	Combat COVID-19 using blockchain and AI-based solutions
Group-7: **General Overview**	Overview & Outlook	Yuan [63]	2016	The state-of-the-art and future trends of blockchain
		Zheng [64]	2017	Architecture, consensus, and future trends of blockchains
		Zheng [65]	2018	Challenges and opportunities of blockchain
		Yuan [66]	2018	Blockchain and cryptocurrencies
		Kolb [67]	2020	Core concepts, challenges, and future directions in blockchains

viewpoint, the authors also studied how such consensus protocols affect the consensus participants in blockchain networks.

During the surveys of smart contracts [9–11], Atzei et al. [9] paid their attention to the security vulnerabilities and programming pitfalls that could be incurred in Ethereum smart contracts. Dwivedi et al. [10] performed a systematic taxonomy on smart-contract languages, while Zheng et al. [11] conducted a survey on the challenges, recent technical advances, and typical platforms of smart contracts.

Sharding techniques are viewed as promising solutions to solving the scalability issue and low-performance problems of blockchains. Several survey articles [12, 13] provide systematic reviews on sharding-based blockchain techniques. For example, Wang et al. [12] focused on the general design flow and critical design challenges of sharding protocols. Next, Yu et al. [13] mainly discussed the intra-consensus security, atomicity of cross-shard transactions, and other advantages of sharding mechanisms.

Regarding scalability, Chen et al. [14] analyzed the scalability technologies in terms of efficiency-improving and function-extension of blockchains, while Zhou et al. [15] compared and classified the existing scalability solutions in the perspective of different layers. Then, Zamyatin et al. [16] conducted a systematic classification of protocols for cross-chain communication. Further, on interoperability, Belchior et al. [17] defined related terms and provided interesting directions. During the investigations [18–23] on security and privacy issues, Taylor et al. [18] reviewed the cyber security space of blockchains including security of blockchain in different directions such as IoT, artificial intelligence (AI) data, and sidechain. Dasgupta et al. [19] discussed general security issues of blockchains from theory to implementation, such as vulnerability, malicious attacks, risks of blockchain applications, and so on. Ma et al. [20] focused on security, privacy, and trust issues in crowdsourcing services. Under the background of big data, Tariq et al. [21] reviewed the security challenges of fog computing-enabled IoT applications, in which blockchain techniques are playing a role of security enabler. In contrast, Refs [22] and [23] emphasized on the privacy issues of blockchain systems and blockchain-based applications.

1.1.2 Data Mining and Analytics. The direction of data analytics for blockchains [24–26] has not yet received too much attention. The existing survey studies are shown as follows. Chen et al. [25] summarized seven typical research issues of data analysis in blockchains, such as entity recognition, privacy identification, network risk parsing, network visualization and portrait, analysis of cryptocurrency market, and the like. Recently, Bartoletti et al. [26] reviewed the Ponzi schemes hiding in Ethereum, aiming to discover the scam behavior and analyze their impact. The authors focused on multiple viewpoints such as the identification methods and the impact of Ponzi schemes to the blockchain ecosystem. Finally, Xie et al. [24] provided an overview on the security and privacy issues, management of transactions, reputation systems of could exchange, where the blockchain technology is used as a key enabler.

1.1.3 Decision-Making Techniques. Blockchains can bring many security advantages for many other fields. On the other hand, blockchain networks also rely on decision-making techniques such as AI [27, 28], machine learning [29, 30], and game theory [31]. This is because the tuning of blockchain network parameters, analysis of user behavior patterns, detection of malicious attacks, identification of market risks, and so on, are playing critical roles for the performance, security, and healthy conditions of blockchain systems and blockchain networks. For example, Salah et al. [27] studied how blockchain technologies benefit key problems of AI. Zheng et al. [28] proposed the concept of *blockchain intelligence* and pointed out the opportunities that both these two terms can benefit each other. Next, Chen et al. [29] discussed the privacy-preserving and secure design of machine learning when blockchain techniques are imported. Liu et al. [30] identified the overview, opportunities, and applications when integrating blockchains and machine learning technologies in the context of communications and networking. Recently, game theoretical solutions [31] have been reviewed when they are applied in blockchain security issues such as malicious attacks and selfish mining, as well as the resource allocation in the management of mining. Both the advantages and disadvantages of game theoretical solutions and models were discussed.

1.1.4 New Communications Networking. First, Park et al. [32] discussed how to take the advantages of blockchains in cloud computing with respect to security solutions. Xiong et al. [33] then investigated how to facilitate blockchain applications in mobile IoT and edge computing environments. Yang et al. [34] identified various perspectives including motivations, frameworks, and functionalities when integrating blockchain with edge computing. Nguyen et al. [35] presented a comprehensive survey when blockchain meets 5G networks and beyond. The authors focused on the opportunities that blockchain can bring for 5G technologies, which include cloud computing, mobile edge computing, SDN/NFV, network slicing, D2D communications, 5G services, and 5G IoT applications.

1.1.5 IoT and IIoT. The blockchain-based applications for IoT [36–47] and IIoT [48, 49] have received the largest amount of attention from both academia and industry. For example, as a pioneer work in this category, Christidis et al. [36] provided a survey about how blockchains and smart contracts promote the IoT applications. Later on, Nguyen et al. [43] presented an investigation of the integration between blockchain technologies and cloud of things with in-depth discussion on backgrounds, motivations, concepts, and architectures. Recently, Park et al. [46] emphasized on the topic of introducing blockchain technologies to the sustainable ecosystem of green IoT. For the IIoT, Zhang et al. [49] discussed the integration of blockchain and edge intelligence to empower a secure IIoT framework in the context of 5G and beyond. In addition, when applying blockchains to the unmanned aerial vehicles (UAV), Alladi et al. [50] reviewed numerous application scenarios covering both commercial and military domains such as network security, surveillance, and the like.

1.1.6 Blockchain Applications. Blockchains have spawned an enormous number of applications in various fields. The research areas covered by the existing surveys on the blockchain-based applications include general applications [51, 52], agriculture [53, 54], Software-defined Networking (SDN) [55], business applications [56], smart city [57], smart grids [58, 59], distributed file systems [60], space industry [61], and COVID-19 [62]. Some of those surveys are reviewed as follows.

Lu et al. [51] performed a literature review on the fundamental features of blockchain-enabled applications. Through the review, the authors expect to outlook the development routine of blockchain technologies. Then, Casino et al. [52] presented a systematic survey of blockchain-enabled applications in the context of multiple sectors and industries. Both the current status and the prospective characteristics of blockchain technologies were identified. In more specific directions, Bermeo et al. [53] proposed a review on the research works focusing on applying blockchain technologies to agriculture. Through an overview on the primary studies published between 2016 and 2018, they found some interesting phenomena such as a large part of relevant papers are solving problems of food supply chain, and Asian community researchers are dominating the blockchain-based agriculture studies. Later on, Ferrag et al. [54] concentrated on the security and privacy issues of green IoT-based agriculture. They also investigated how would blockchain solutions and consensus algorithms be adapted to green IoT-based agriculture. Alharbi [55] then described how blockchain technologies can be integrated into SDN architecture to provide security, confidentiality, and integrity. Konstantinidis et al. [56] discussed the various applications of blockchain technology on the business sectors.

Xie et al. [57] provided a literature review on the smart city services involving blockchain technologies, such as smart citizen, smart healthcare, smart transportation, management of supply chain, and so on. Then, based on the blockchain technology, the two surveys [58, 59] discussed the conceptual model, different use cases, energy trading processes, efficient power generation and distribution strategies, system maintenance and diagnosis for grid facilities, and security and privacy preserving of smart grid domains. Huang et al. [60] reviewed the integration of blockchain-based solutions and the distributed file systems. Taking the Inter-Planetary File System (IPFS) and Swarm as two representative distributed file systems, the authors introduced the principle and structure, as well as the state-of-the-art studies of blockchain-empowered distributed file systems and their utilization scenarios. Next, Torky et al. [61] conducted a systematic discussion on the conceptual exploration to adopt the blockchain technology in space industry. A blockchain-based satellite network, namely SpaceChain, has been initially implemented as a case study of the proposed blockchain-empowered satellite system. As a most timely survey regarding combating the coronavirus (COVID-19), Nguyen et al. [62] presented a comprehensive review on the integrating blockchain and AI technologies while fighting the coronavirus crisis. The roles of blockchain during tackling the pandemic vary in a wide range of applications, such as tracking of population, privacy preserving of citizens, supply chain management, and other tracking services.

1.1.7 General Overview and Outlook. The final group of survey articles [63–67] overviewed the basic concepts of blockchains and cryptocurrencies, the fundamental research challenges, and general issues such as consensus algorithms, solutions to scalability and security, privacy preserving issues, and and the like. Finally, the authors outlooked further potential technical challenges and open issues for shedding light on future studies of blockchain technologies.

Summary of Survey-Article Review: Through the brief review of the state-of-the-art surveys, we have found that the blockchain technologies have been adaptively integrated into a growing range of application sectors. The blockchain theory and technology will bring substantial innovations, incentives, and a great number of application scenarios in diverse fields. Based on the analysis of those survey articles, we believe that there will be more survey articles published in the near future, very likely in the areas of sharding techniques, scalability, interoperability, smart contracts, big data, AI technologies, 5G and Beyond, edge computing, cloud computing, and many other fields.

1.2 Motivation of This Survey

Via the overview, shown in Tables 1 and 2, Figure 1, and Figure 1 of the online supplementary material of the existing blockchain-related surveys, 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. In particular, the following directions need in-depth investigations.

1.2.1 Theories for Improving the Performance of Blockchains. The performance of blockchains includes a number of metrics such as throughput, latency, storage efficiency, reliability, scalability, interoperability, and and so on. Many theories can be devoted to improving the performance metrics of blockchains. For example, the following perspectives are worthy, paying more efforts.

Scalability Solutions. Although blockchain is viewed as a distributed and public database of transactions and has become a platform for decentralized applications, blockchains still face the scalability problem. For example, the system throughput is not scalable with the increasing size of a blockchain network. Thus, the scalability solutions of blockchain still require further studying. The promising solutions to improve the scalability of blockchains include sharding-based and multiple-chain and cross-chain techniques.

New Protocols and Infrastructures. Several classic consensus protocols, such as practical byzantine-fault tolerant (PBFT) protocol and proof-of-work (PoW) protocol, have been widely adopted by popular blockchain systems. However, those classic protocols cannot meet all consensus requirements existing in emerging new blockchains. Thus, it is necessary to review new protocols and infrastructures proposed to serve new scenarios of blockchain-based applications.

1.2.2 Modelings and Techniques for Better Understanding Blockchains. The existing studies on better understanding blockchains that have been reviewed by other surveys mainly focus on the security and privacy issues, and the analysis of the cryptocurrency market, for example, the identification of Ponzi schemes and other scam behaviors. In our opinion, to better understand blockchains, the following wider range of topics should be also emphasized on.

Graph-based Theories. Excepting the classic graph knowledge that have applied to blockchains, such as the Merkel tree and directed acyclic graph (DAG) techniques, the general graph-based analytical techniques are powerful approaches to find insights behind the transactions, smart contracts, and the network structure of blockchains.

Stochastic Modelings, Queueing Theories, and Analytical Models. Several phases of blockchain networks can be described using the stochastic modelings, queueing theories, and analytical models. Based on these theoretical models, researchers can conduct the property analysis of blockchain network, stability analysis, deriving failure probability, modeling of mining procedure, estimating blockchain confirming time, exploring the synchronization process of Bitcoin network and other working principles of blockchains, and understanding how blockchains respond to difference network conditions even malicious attacks.

Data Analytics for Cryptocurrency Blockchains. Security issues of cryptocurrency blockchains and their markets are attracting more and more attention. Although several surveys [24–26] have already studied the Ponzi schemes in Ethereum and other general security and privacy issues of blockchain systems, their surveys mainly emphasized on the identification approaches and the impacts to blockchain systems. In contrast, in our survey, we review the latest studies by exploiting the data analytics techniques to detect the market risks in cryptocurrency ecosystems, where the risks not only include Ponzi schemes, but also take into account the cryptojacking, market manipulation mining, and money-laundering activities. Furthermore, we also review a few studies utilizing data science and stochastic modelings to produce a portrait of cryptoecomonic systems.

1.2.3 Useful Measurements, Datasets, and Experiment Tools for Blockchains. In the aforementioned 66 surveys, we find that there is still no a single article focusing on the performance measurements, datasets, and experiment tools for blockchains. Instead, our survey in this article particularly reviews: (i) performance measurements with respect to throughput, end-to-end confirmation delays of transactions, forking rate, resource utilization, scalability, and the like; and (ii) useful evaluation tools and datasets dedicated to blockchain experiments.

In a summary, by this article, we would like to fill the gap by emphasizing on the cutting-edge theoretical studies, modelings, and useful tools for blockchains. Particularly, we try to include the latest high-quality research outputs that have not been included by other existing survey articles. We believe that this survey can shed new light on the further development of blockchains.

1.3 Contribution of Our Survey

Our survey presented in this article includes the following contributions.

- We conduct a brief classification of existing blockchain surveys to highlight the meaning of our literature review shown in this survey.

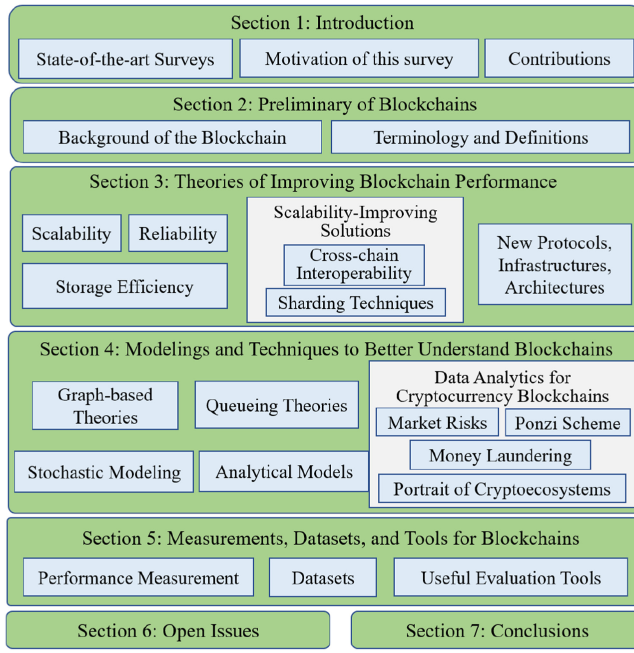

Fig. 2. The structure of this article.

— 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.

The structure of this survey is shown in Figure 2 and organized as follows. Section 2 introduces the preliminaries of blockchains. Section 3 summarizes the state-of-the-art theoretical studies that improve the performance of blockchains. In Section 4, we then review various modelings and analytic models that help understand blockchains. Diverse measurement approaches, datasets, and useful tools for blockchains are overviewed in Section 5. We outlook the open issues in Section 6. Finally, Section 7 concludes this article.

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 article. Due to the frequent use of acronyms in this article, we include an acronym table, i.e., Table 1 in the online supplementary material.

2.1 Prime Blockchain Platforms

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 *Unspent Transaction Output (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 dif-

Table 3. Comparison between Bitcoin and Ethereum

	State Model	Consensus Protocols	Throughput
Bitcoin	UTXO	PoW	3 to 7 TPS[72]
Ethereum1.0	Account/Balance	PoW	7 to 15 TPS[72]
Ethereum2.0	Account/Balance	PoS Sharding	Unknown

ficulty 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, all the other miners can not always 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 [68, 69]. 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 [70] attacks efficiently.

2.1.2 Ethereum. Ethereum [71] 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 and Ethereum is summarized in Table 3.

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 two types of accounts, *external owned accounts* (EOAs) and *contract accounts* (CAs), each controlled by private keys and contract codes, respectively.

2.1.3 Hyperledger Fabric. Hyperledger Fabric [73] 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.

2.1.4 EOSIO. EOSIO [74] is another popular blockchain platform released by a company *block.one* in 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-confirmation latency, low-overhead BFT finality, and so on. These excellent features have 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.

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.

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.

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.

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.

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*.

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.

3 THEORIES TO IMPROVING THE PERFORMANCE OF BLOCKCHAINS

3.1 Latest Theories to Improving Blockchain Performance

Summary of this section is included in Table 4.

3.1.1 Throughput and Latency. Aiming to reduce the confirmation latency of transactions to milliseconds, Hari et al. [75] proposed a high-throughput, low-latency, deterministic confirmation mechanism called ACCEL for accelerating Bitcoin's block confirmation. The key findings of this article includes 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 et al. [76] 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 Transactions Per Second (TPS), the overheads of cross-zone transactions, the confirmation latency of transactions, and so on.

To Bitcoin, low *throughput* and long *transaction confirmation latency* are two critical bottleneck metrics. To overcome these two bottlenecks, Yang et al. [77] 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 Amazon Web Services (AWS). However, the 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 attack 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. [78] proposed a heuristic algorithm named GARET, which is a gas consumption-aware relocation mechanism

Table 4. Latest Theories of Improving the Performance of Blockchains

Emphasis	Ref.	Recognition	Challenge	Methodology
Throughput & Latency	[75]	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.
	[76]	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.
	[77]	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.
	[78]	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.
Storage Efficiency	[79]	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.
	[80]	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.
	[81]	*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	[82]	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.
	[83]	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.

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.

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 a 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. [79] 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 toward running blockchains on IoT devices.

Then, Dai et al. [80] 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% compared with the original Bitcoin.

Under the similar idea, Xu et al. [81] 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 toward data-heavy distributed applications.

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. [82] 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 in 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. [83] 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.

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 section in Table 5.

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. [68] 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 toward 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. [84] 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. [99] 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. [85] proposed a permissioned blockchain system named *SharPer*, which strives 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 author's previous work, they have already proposed a permissioned blockchain, while, in this article, 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, this article 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. [86] 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 compared 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,

Table 5. Latest Scalability Solutions to Improving the Performance of Blockchains

Emphasis	Ref.	Recognition	Methodology
Solutions to Sharding blockchains	[68]	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 toward the open public blockchains.
	[76]	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.
	[84]	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.
	[85]	SharPer	Authors proposed a permissioned blockchain system named *SharPer*, which adopts sharding techniques to improve scalability of cross-shard transactions.
	[86]	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.
	[87]	NRSS	Authors proposed a node-rating based new Sharding scheme, i.e., NRSS, for blockchains, aiming to improve the throughput of committees.
	[88]	OptChain	Authors proposed a new sharding paradigm, called OptChain, mainly used for optimizing the placement of transactions into shards.
	[89]	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.
	[90]	SSChain	Authors proposed a non-reshuffling structure called SSChain, which supports both transaction sharding and state sharding while eliminating huge data-migration across shards.
	[91]	Eumonia	Authors proposed Eumonia, which is a permissionless parallel-chain protocol for realizing a global ordering of blocks.
	[92]	Vulnerability of Sybil attacks	Authors systematically analyzed the vulnerability of Sybil attacks in protocol Elastico.
	[93]	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.
	[94]	CycLedger	Authors proposed a protocol CycLedger to pave a way toward scalability, security, and incentive for sharding blockchains.
Interoperability of multiple-chain systems	[95]	Interoperability architecture	Authors proposed a novel interoperability architecture that supports the cross-chain cooperations among multiple blockchains, and a novel MMR method for the passive cross-chain communications.
	[96]	HyperService	Authors proposed a programming platform that provides interoperability and programmability over multiple heterogeneous blockchains.
	[97]	Protocol *Move*	Authors proposed a programming model for smart-contract developers to create DApps that can interoperate and scale in a multiple-chain environment.
	[98]	Cross-cryptocurrency TX protocol	Authors proposed a decentralized cryptocurrency exchange protocol enabling cross-cryptocurrency transactions based on smart contracts deployed on Ethereum.
	[16]	Cross-chain comm.	Authors conducted a systematic classification of cross-chain communication protocols.

Wang et al. [87] 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. [88] 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 Transaction-to-Shard (T2S)-score and Latency-to-Shard (L2S)-score, 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 has the highest temporal fitness score.

Similar to Ref. [88], Dang et al. [89] 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 toward guarding against the Byzantine-fault malicious coordinators. The experiment results showed that the throughput achieves a few thousand 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. [90] 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 6,500 TPS in a network with 1,800 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. [91] proposed Eumonia, which is a permissionless parallel-chain protocol toward 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 easily attracts 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. [92] systematically formulated a model and performed an analysis w.r.t. the vulnerability of Sybil attacks in the pioneer sharding protocol Elastico [68]. 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.

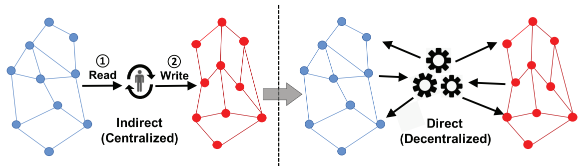

Fig. 3. The illustration of interoperability across blockchains [95]. The left figure demonstrates the indirect way of interoperability that requires a centralized third party. The right figure demonstrates the direct way of interoperability without the presence of any third party.

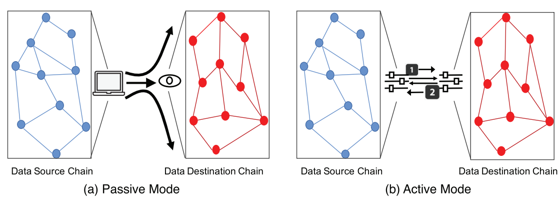

Fig. 4. The interoperability of blockchains [95]. Passive mode is shown in the left figure, in which the source chain is monitored by the destination chain instead of actively sending information to the destination chain as shown in the right figure.

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 unable to be small to ensure the shard functions securely. To improve the sustainability of blockchain sharding, Xu et al. [93] 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. [94] 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.

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 Figure 3. 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. [95] 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 Figure 4. 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, the blockchain interoperability must evolve accordingly. To help realize such new type of interoperability among multiple heterogeneous blockchains, Liu et al. [96] proposed HyperService, which includes two major components, i.e., a programming framework allowing developers to create cross-chain applications; and a universal interoperability protocol toward 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. [97] 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. [98] 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 Ref. [16], in which readers can find a systematic classification of cross-chain communication protocols.

3.3 New Protocols and Infrastructures

This section is summarized in Table 6.

3.3.1 New Protocols for Blockchains. David et al. [100] 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 toward 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. [101] proposed a new protocol named Tendermint, which serves as a new termination mechanism for simplifying BFT consensus protocol.

In Monoxide proposed by Ref. [76], 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 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. [102] 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.

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. [103] 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

Table 6. New Protocols and Infrastructures to Improve the Performance of Blockchains

Emphasis	Ref.	Recognition	Methodology
New Protocols	[100]	Ouroboros Praos	Authors proposed a new secure proof-of-stake protocol named *Ouroboros Praos*, which is proved secure in the semi-synchronous adversarial setting.
	[101]	Tendermint	Authors proposed a new BFT consensus protocol for the wide area network organized by the gossip-based P2P network under adversarial conditions.
	[76]	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.
	[102]	Proof-of-Trust	Authors proposed a novel Proof-of-Trust consensus for the online services of crowdsourcing.
New Infrastructures & Architectures	[103]	StreamChain	Authors proposed to shift the block-based distributed ledgers to a new paradigm of *stream transaction processing* to achieve low end-to-end latencies without affecting throughput much.
	[104]	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.
	[105]	Optimal mining for miners	Authors proposed an edge computing-based blockchain network architecture, aiming to allocate optimal computational resources for miners.
	[106]	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.
	[107]	Non-linear blockchain system	Authors explored three major metrics of blockchains, and devised a non-linear blockchain system.

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 1,500 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. [104] 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. [105] 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. [106] 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. [107] 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 tradeoff among such three properties.

4 VARIOUS MODELINGS AND TECHNIQUES FOR BETTER UNDERSTANDING BLOCKCHAINS

We summarize various analytical models for blockchain networks in Tables 7 and 8.

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 [108], Phantom [109], and Conflux [110], 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. [111] characterized three major activities on Ethereum, i.e., money transfer, the creation of smart contracts, and the invocation of smart contracts. The major contribution of this article 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. Similarly, by processing Bitcoin transaction history, Akcora et al. [112] and Dixon et al. [113] modeled the transfer network into an extreme transaction graph. Through the analysis of chainlet activities [114] 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. [115] 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.

Table 7. Various Modelings, Techniques, and Theories for Better Understanding Blockchains

Category	Emphasis	Ref.	Metrics	Methodology and Implications
Graph-based Theories	Transactions mining	[111]	Cross-graph analysis of Ethereum	Via graph analysis, authors extracted three major activities, i.e., money transfer, smart contracts creation, and smart contracts invocation.
		[115]	Features of transaction graphs	Authors proposed an extendable and computationally efficient method for graph representation learning on Blockchains.
		[116]	Market manipulation patterns	Authors exploited the graph-based data-mining approach to reveal the market manipulation evidence of Bitcoin.
		[119]	Clustering coefficient, assortativity of TX graph	Authors exploited the graph-based analysis to reveal the abnormal transactions of EOSIO.
	Token networks	[117]	Token-transfer distributions	Authors studied the token networks through analyzing smart contracts of Ethereum blockchain based on graph analysis.
		[112, 113]	Extreme chainlet activity	Authors proposed graph-based analysis models for assessing the financial investment risk of Bitcoin.
Stochastic Modelings	Blockchain network analysis	[120]	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	[121]	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	[122–124]	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	[125, 126]	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	[127]	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	[128]	Stationary queue-length distribution	Authors proposed an infinite-server model with random fluid limit for Bitcoin network.
	Mining resources allocation	[129]	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	[130]	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 of the theoretical working principle of blockchain networks.

Table 8. Various Analytics Models for Better Understanding Blockchain Networks

Emphasis	**Ref.**	**Metrics**	**Methodology and Implications**
Applicability of blockchains	[131]	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 permissioned or permissionless, as its technical solution.
	[132]	Scalability, efficiency, and privacy issues in cloud for blockchains	Authors proposed a novel upper bound privacy leakage based approach to identify intermediate datasets 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	[133]	Temporal information and the multiplicity features of Ethereum transactions	Authors proposed an analytical model based on the multiplex network theory for understanding Ethereum transactions.
	[134]	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	[135]	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	[136]	Consistency of a PoW blockchain	Authors derived a neat bound of mining latencies that helps in understanding the consistency of Nakamoto's blockchain consensus in asynchronous networks.
Network connectivity	[137]	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	[138]	Balance among shards, number of TX that would involve multiple shards, the amount of data relocated across shards	Authors studied how sharding impacts Ethereum by firstly modeling Ethereum through graph modeling, and then assessing the three metrics mentioned when partitioning the graph.
Required properties of sharding protocols	[139]	Consistency and scalability	Authors proposed an analytical model to evaluate whether a protocol for sharded distributed ledgers fulfills necessary properties.
Vulnerability by forking attacks	[140]	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	[70]	Robustness parameter, vulnerability probability	Authors studied how to defend and even counterattack the double-spend attacks in PoW blockchains.
Limitations of PBFT-based blockchains	[141]	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	[142]	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.

Furthermore, utilizing the Mt. Gox transaction history, Chen et al. [116] 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. [117] 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 Ref. [111], 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 1,000 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 Ref. [118], where Somin et al. showed that the full transfers seem to obey a power-law distribution. However, the study [117] 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. [119] 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.

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. [120] 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.

Toward the stability and scalability of blockchain systems, Gopalan et al. [121] 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 [84], OmniLedger [69], and Monoxide [76], the failure probability for a committee under Sharding protocol is still unknown. To fill this gap, Hafid et al. [122–124] 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.

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. [125, 126] 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. [127] proposed a theoretical framework integrating the queueing theory and machine learning techniques to have a deep understanding toward 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. [128] 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. [130] 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, the waiting time in memory pool, and so on.

Next, Fang et al. [129] 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 tradeoff between mining energy and queueing delay. Different from the aforementioned work [125–127], the proposed Lyapunov-based algorithm does not need to make any statistical assumptions on the arrivals and services.

4.4 Analytical Models for Blockchain Networks

This section is summarized in Table 8.

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. [131] 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 [132] 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. [133] 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. [134] 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. [135] exploited the Game Theory to find a Nash equilibria while peers are competing mining resources. The proposed approach help researchers better 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 Distributed Denial of Service (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. [143] 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. [136] 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 toward 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. [137] 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. [138] studied how sharding works for Ethereum by modeling Ethereum into a graph. Via partitioning the graph, they evaluated the tradeoff 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. [139] 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. [140] 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 [144] reported that the attacks toward PoW mining-based blockchain systems can be cheap under a specific condition when renting sufficient hashrate capability. Moroz et al. [70] studied how to defend 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. [141] 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.

Table 9. Data Analytics for Better Understanding Cryptocurrency Blockchains

Emphasis	Ref.	Metrics	Methodology and Implications
Cryptojacking detection	[145]	Hardware performance counters	Authors proposed a machine learning-based solution to prevent cryptojacking attacks.
	[146]	Various system resource utilization	Authors proposed an in-browser cryptojacking detection approach (CapJack), based on the latest CapsNet.
Market-manipulation mining	[116]	Various graph characteristics of transaction graph	Authors proposed a mining approach using the exchanges collected from the transaction networks.
Predicting volatility of Bitcoin price	[113]	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	[147]	Various graph characteristics of transaction graph	Authors exploited machine learning models to detect potential money laundering activities from Bitcoin transactions.
Ponzi-scheme detection	[148]	Factors that affect scam persistence	Authors analyzed the demand and supply perspectives of Ponzi schemes on Bitcoin ecosystem.
	[149, 150]	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	[151]	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	[152]	Miner revenue, ASIC value	Authors studied the correlation between the price of mining hardware (ASIC) and the value volatility of underlying cryptocurrency.

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. [142] 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.

4.5 Data Analytics for Cryptocurrency Blockchains

This section is summarized in Table 9.

4.5.1 Market Risks Detection. As aforementioned, Akcora et al. [112] 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 [153] 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. [145] proposed a machine learning-based 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 toward the widespread real-time protection for web users. Similarly, Ning et al. [146] proposed *CapJack*, which is an in-browser cryptojacking detector based on deep capsule network (CapsNet) [154] technology.

As mentioned previously, to detect potential manipulation of Bitcoin market, Chen et al. [116] 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. [113] studied the impact of extreme transaction graph (ETG) activity on the intraday dynamics of the Bitcoin prices. The authors utilized chainlets [114] (sub graphs of transaction graph) for developing their predictive models.

4.5.2 Ponzi Schemes Detection. Ponzi scheme [26], as a classic scam, is taking advantages of mainstream blockchains such as Ethereum. Data mining technologies [155] are widely used for detecting Ponzi schemes. For example, several representative studies are reviewed as follows. Vasek et al. [148] 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 toward a healthier blockchain economic environment, Chen et al. [149, 150] 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.

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. [156] 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 their paper. In contrast, utilizing a transaction dataset collected over three years, Hu et al. [147] 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.

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. [151] presented a practical evidence-based example to show how this manner can be applied to designing cryptoeconomic blockchains.

Yaish et al. [152] discussed the relationship between the cryptocurrency mining and the market price of the special hardware, i.e., the Application Specific Integrated Circuit (ASIC), 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.

5 USEFUL MEASUREMENTS, DATASETS, AND EXPERIMENT TOOLS FOR BLOCKCHAINS

Measurements are summarized in Table 10, and datasets are summarized in Table 11.

Table 10. Various Performance Measurements of Blockchains

Ref.	Target Blockchains	Metrics	Implementation/Experiments/Methodology
[76]	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 1,200 virtual machines, with each owning eight cores and 32 GB memory. In total, 48,000 blockchain nodes were exploited in the testbed.
[77]	General blockchains	Throughput and confirmation latency, scalability under a 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.
[78]	Ethereum	TX throughput, the makespan of transaction latency	The authors 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.
[157]	Bitcoin, Litecoin, Dogecoin, Ethereum	Block interval, block size, and throughput	The authors 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 tradeoffs between the security provisions and performance objectively.
[158]	Hyperledger Fabric	Execution time, latency, throughput, scalability vs. the number of blockchain nodes	The authors presented the performance measurement and analysis toward Hyperledger Fabric version 0.6 and version 1.0.
[159]	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	The authors roposed 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.
[160]	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 tradeoffs while designing the software of blockchains.
[161]	Ethereum	Network size and geographic distribution of Ethereum network nodes	The authors 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.
[162]	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.

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, and so on.

Table 11. Blockchain Dataset Frameworks and Evaluation Tools

Recognition	Target	Ref.	Utilization
XBlock-ETH	Ethereum	[167]	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	[168]	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	[169]	Authors proposed an open-source software platform, named BlockSci, for the analysis of blockchains.
Blockbench	General blockchains	[160]	Authors proposed a benchmarking framework for measuring the data processing capability and performance of different layers of a blockchain system.
NodeFinder	Etheruem nodes	[161]	Authors proposed a measuring tool named NodeFinder to investigate the opaque network characteristics of Ethereum network nodes.
Network simulator for Bitcoin	Bitcoin	[162]	Authors proposed a configurable network simulator for the performance measurements of Bitcoin using lightweight virtualization technologies.

As a pioneer work in this direction, Gervais et al. [157] 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.

Nasir et al. [158] 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 Ref. [76], 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 Ref. [77], 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, and so on.

Later, Zheng et al. [159] proposed a scalable framework for monitoring the real-time performance blockchain systems. This work has evaluated four popular blockchain systems, i.e., Ethereum, Parity [163], Cryptape Inter-enterprise Trust Automation (CITA) [164] and Hyperledger Fabric [165], 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 four popular blockchain systems. Their experimental logs and technique report [166] can be accessed from http://xblock.pro. Recently, Zheng et al. [167] 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 [168] 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.

5.2 Useful Evaluation Tools for Blockchains

Kalodner et al. [169] 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. [160] 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. [161] 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. [162] 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.

6 OPEN ISSUES AND FUTURE DIRECTIONS

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

6.1 Performance-Improving Issues

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.

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.

6.1.3 Cross-Shard Performance. Although a number of committee-based sharding protocols [69, 76, 84, 170] 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.

6.1.4 Cross-Chain Transaction Accelerating Mechanisms. On cross-chain operations, Ref. [95] is essentially a pioneer step toward practical blockchain-based ecosystems. Following this roadmap paved by Ref. [95], 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.

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.

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 Remote Direct Memory Access (RDMA) and high-speed network cards, can be exploited in accelerating the data-access among miners in blockchain networks.

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.

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 of 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.

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.

6.3 Security Issues of Blockchains

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.

6.3.2 Anti-Cryptojacking Mechanisms for Malicious Miners. The Cryptojacking Miners are reportedly existing in web browsers according to Ref. [145]. 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.

6.3.3 Security Issues of Cryptocurrency Blockchains. The security issues of cryptocurrency blockchains, such as double-spend attacks and frauds in smart contracts, have arisen growing attention from both industrial and academic fields. However, little efforts have been committed to the theoretical investigations toward 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.

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.

7 CONCLUSION

Through a brief review of state-of-the-art blockchain surveys at first, 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 survey tried to convey the new protocols, ideas, and solutions that can improve the performance of blockchains, and help people better understand the blockchains in a further level. We believe our survey provides a timely guidance on the theoretical insights of blockchains for researchers, engineers, educators, and generalized readers.

REFERENCES

[1] Lakshmi Siva Sankar, M. Sindhu, and M. Sethumadhavan. 2017. Survey of consensus protocols on blockchain applications. In *Proceedings of the 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE, 1–5.

[2] Y. Yuan, X. Ni, S. Zeng, and F. Wang. 2018. Blockchain consensus algorithms: The state of the art and future trends. *Acta Automatica Sinica* 44, 11 (2018), 2011–2022.

[3] Wenbo Wang, Dinh Thai Hoang, Zehui Xiong, Dusit Niyato, Ping Wang, Peizhao Hu, and Yonggang Wen. 2018. A survey on consensus mechanisms and mining management in blockchain networks. *arXiv preprint arXiv:1805.02707* (2018), 1–33.

[4] Juan A. Garay and Aggelos Kiayias. 2018. SoK: A consensus taxonomy in the blockchain era. *IACR Cryptology Eprint Archive* 2018 (2018), 754.

[5] Giang-Truong Nguyen and Kyungbaek Kim. 2018. A survey about consensus algorithms used in blockchain. *Journal of Information Processing Systems* 14, 1 (2018).

[6] Wenbo Wang, Dinh Thai Hoang, Peizhao Hu, Zehui Xiong, Dusit Niyato, Ping Wang, Yonggang Wen, and Dong In Kim. 2019. A survey on consensus mechanisms and mining strategy management in blockchain networks. *IEEE Access* 7 (2019), 22328–22370.

[7] Shehar Bano, Alberto Sonnino, Mustafa Al-Bassam, Sarah Azouvi, Patrick McCorry, Sarah Meiklejohn, and George Danezis. 2019. SoK: Consensus in the age of blockchains. In *Proceedings of the 1st ACM Conference on Advances in Financial Technologies (AFT'19)*. 183–198.

[8] Yang Xiao, Ning Zhang, Wenjing Lou, and Y. Thomas Hou. 2020. A survey of distributed consensus protocols for blockchain networks. *IEEE Communications Surveys & Tutorials* 22, 2 (2020), 1432–1465.

[9] Nicola Atzei, Massimo Bartoletti, and Tiziana Cimoli. 2016. A survey of attacks on Ethereum smart contracts. *IACR Cryptology Eprint Archive* 2016 (2016), 1007–1030.

[10] Vimal Dwivedi, Vipin Deval, Abhishek Dixit, and Alex Norta. 2019. Blockchain-based smart-contract languages: A systematic literature review. *Preprint arXiv:1710.06372, 2017.*

[11] Zibin Zheng, Shaoan Xie, Hong-Ning Dai, Weili Chen, Xiangping Chen, Jian Weng, and Muhammad Imran. 2020. An overview on smart contracts: Challenges, advances and platforms. *Future Generation Computer Systems* 105 (2020), 475–491.

[12] Gang Wang, Zhijie Jerry Shi, Mark Nixon, and Song Han. 2019. Sok: Sharding on blockchain. In *Proceedings of the 1st ACM Conference on Advances in Financial Technologies (AFT'19)*. 41–61.

[13] G. Yu, X. Wang, K. Yu, W. Ni, J. A. Zhang, and R. P. Liu. 2020. Survey: Sharding in blockchains. *IEEE Access* 8 (2020), 14155–14181. DOI: http://dx.doi.org/10.1109/ACCESS.2020.2965147

[14] Chen Pan, Zhiqiang Liu, Zhen Liu, and Yu Long. 2018. Research on scalability of blockchain technology: Problems and methods. *Journal of Computer Research and Development* 55, 10 (2018), 2099–2110.

[15] Qiheng Zhou, Huawei Huang, Zibin Zheng, and Jing Bian. 2020. Solutions to scalability of blockchain: A survey. *IEEE ACCESS* 8, 1 (December 2020), 16440–16455.

[16] Alexei Zamyatin, Mustafa Al-Bassam, Dionysis Zindros, Eleftherios Kokoris-Kogias, Pedro Moreno-Sanchez, Aggelos Kiayias, and William J. Knottenbelt. 2019. *Sok: Communication Across Distributed Ledgers.* Technical Report. IACR Cryptology ePrint Archive, 2019: 1128.

[17] Rafael Belchior, André Vasconcelos, Sérgio Guerreiro, and Miguel Correia. 2020. A survey on blockchain interoperability: Past, present, and future trends. *arXiv preprint arXiv:2005.14282.*

[18] Paul J. Taylor, Tooska Dargahi, Ali Dehghantanha, Reza M. Parizi, and Kim-Kwang Raymond Choo. 2019. A systematic literature review of blockchain cyber security. *Digital Communications and Networks* (2019).

[19] Dipankar Dasgupta, John M. Shrein, and Kishor Datta Gupta. 2019. A survey of blockchain from security perspective. *Journal of Banking and Financial Technology* 3, 1 (2019), 1–17.

[20] Ying Ma, Yu Sun, Yunjie Lei, Nan Qin, and Junwen Lu. 2019. A survey of blockchain technology on security, privacy, and trust in crowdsourcing services. *World Wide Web* (2019), 1–27.

[21] Noshina Tariq, Muhammad Asim, Feras Al-Obeidat, Muhammad Zubair Farooqi, Thar Baker, Mohammad Hammoudeh, and Ibrahim Ghafir. 2019. The security of big data in fog-enabled IoT applications including blockchain: A survey. *Sensors* 19, 8 (2019), 1788.

[22] Qi Feng, Debiao He, Sherali Zeadally, Muhammad Khurram Khan, and Neeraj Kumar. 2019. A survey on privacy protection in blockchain system. *Journal of Network and Computer Applications* 126 (2019), 45–58.

[23] Sumit Soni and Bharat Bhushan. 2019. A comprehensive survey on blockchain: Working, security analysis, privacy threats and potential applications. In *Proceedings of the 2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT)*, Vol. 1. IEEE, 922–926.

[24] Shaoan Xie, Zibin Zheng, Weili Chen, Jiajing Wu, Hong-Ning Dai, and Muhammad Imran. 2020. Blockchain for cloud exchange: A survey. *Computers & Electrical Engineering* 81 (2020), 106526.

[25] Weili Chen and Zibin Zheng. 2018. Blockchain data analysis: A review of status, trends and challenges. *Journal of Computer Research and Development* 55, 9 (2018), 1853–1870.

[26] Massimo Bartoletti, Salvatore Carta, Tiziana Cimoli, and Roberto Saia. 2020. Dissecting Ponzi schemes on Ethereum: Identification, analysis, and impact. *Future Generation Computer Systems* 102 (2020), 259–277.

[27] Khaled Salah, M. Habib Ur Rehman, Nishara Nizamuddin, and Ala Al-Fuqaha. 2019. Blockchain for AI: Review and open research challenges. *IEEE Access* 7 (2019), 10127–10149.

[28] Zibin Zheng and Hong-Ning Dai. 2020. Blockchain intelligence: When blockchain meets artificial intelligence. *Preprint arXiv:1912.06485, 2019.*

[29] Xuhui Chen, Jinlong Ji, Changqing Luo, Weixian Liao, and Pan Li. 2018. When machine learning meets blockchain: A decentralized, privacy-preserving and secure design. In *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 1178–1187.

[30] Yiming Liu, F Richard Yu, Xi Li, Hong Ji, and Victor CM Leung. 2020. Blockchain and machine learning for communications and networking systems. *IEEE Communications Surveys & Tutorials* (2020).

[31] Ziyao Liu, Nguyen Cong Luong, Wenbo Wang, Dusit Niyato, Ping Wang, Ying-Chang Liang, and Dong In Kim. 2019. A survey on blockchain: A game theoretical perspective. *IEEE Access* 7 (2019), 47615–47643.

[32] Jin Ho Park and Jong Hyuk Park. 2017. Blockchain security in cloud computing: Use cases, challenges, and solutions. *Symmetry* 9, 8 (2017), 164.

[33] Zehui Xiong, Yang Zhang, Dusit Niyato, Ping Wang, and Zhu Han. 2018. When mobile blockchain meets edge computing. *IEEE Communications Magazine* 56, 8 (2018), 33–39.

[34] Ruizhe Yang, F Richard Yu, Pengbo Si, Zhaoxin Yang, and Yanhua Zhang. 2019. Integrated blockchain and edge computing systems: A survey, some research issues, and challenges. *IEEE Communications Surveys & Tutorials* 21, 2 (2019), 1508–1532.

[35] Dinh C. Nguyen, Pubudu N. Pathirana, Ming Ding, and Aruna Seneviratne. 2019. Blockchain for 5G and beyond networks: A state of the art survey. *arXiv preprint arXiv:1912.05062.*

[36] Konstantinos Christidis and Michael Devetsikiotis. 2016. Blockchains and smart contracts for the Internet of Things. *IEEE Access* 4 (2016), 2292–2303.
[37] Muhammad Salek Ali, Massimo Vecchio, Miguel Pincheira, Koustabh Dolui, Fabio Antonelli, and Mubashir Husain Rehmani. 2018. Applications of blockchains in the internet of things: A comprehensive survey. *IEEE Communications Surveys & Tutorials* 21, 2 (2018), 1676–1717.
[38] Tiago M Fernández-Caramés and Paula Fraga-Lamas. 2018. A review on the use of blockchain for the internet of things. *IEEE Access* 6 (2018), 32979–33001.
[39] Djamel Eddine Kouicem, Abdelmadjid Bouabdallah, and Hicham Lakhlef. 2018. Internet of Things security: A top-down survey. *Computer Networks* 141 (2018), 199–221.
[40] Alfonso Panarello, Nachiket Tapas, Giovanni Merlino, Francesco Longo, and Antonio Puliafito. 2018. Blockchain and IoT integration: A systematic survey. *Sensors* 18, 8 (2018), 2575.
[41] Hong-Ning Dai, Zibin Zheng, and Yan Zhang. 2019. Blockchain for internet of things: A survey. *IEEE Internet of Things Journal* 6, 5 (2019), 8076–8094.
[42] Xu Wang, Xuan Zha, Wei Ni, Ren Ping Liu, Y. Jay Guo, Xinxin Niu, and Kangfeng Zheng. 2019. Survey on blockchain for internet of things. *Computer Communications* 136 (2019), 10–29.
[43] Dinh C. Nguyen, Pubudu N. Pathirana, Ming Ding, and Aruna Seneviratne. 2019. Integration of blockchain and cloud of things: Architecture, applications and challenges. *arXiv preprint arXiv:1908.09058*.
[44] Francesco Restuccia, Salvatore D. Kanhere, Tommaso Melodia, and Sajal K. Das. 2019. Blockchain for the internet of things: Present and future. *arXiv preprint arXiv:1903.07448*.
[45] Bin Cao, Yixin Li, Lei Zhang, Long Zhang, Shahid Mumtaz, Zhenyu Zhou, and Mugen Peng. 2019. When Internet of Things meets blockchain: Challenges in distributed consensus. *IEEE Network* 33, 6 (2019), 133–139.
[46] Jong Hyouk Park, Neeraj Kumar, and Pradip Sharma. 2020. Blockchain technology toward green IoT: Opportunities and challenges. *IEEE Network*, 34, 4 (2020), 263–269. DOI: 10.1109/MNET.001.1900526
[47] Laphou Lao, Zecheng Li, Songlin Hou, Bin Xiao, Songtao Guo, and Yuanyuan Yang. 2020. A survey of IoT applications in blockchain systems: Architecture, consensus, and traffic modeling. *ACM Computing Surveys (CSUR)* 53, 1 (2020), 1–32.
[48] Tejasvi Alladi, Vinay Chamola, Reza M. Parizi, and Kim-Kwang Raymond Choo. 2019. Blockchain applications for Industry 4.0 and industrial IoT: A review. *IEEE Access* 7 (2019), 176935–176951.
[49] Ke Zhang, Yongxu Zhu, Sabita Maharjan, and Yan Zhang. 2019. Edge intelligence and blockchain empowered 5G beyond for the industrial internet of things. *IEEE Network* 33, 5 (2019), 12–19.
[50] Tejasvi Alladi, Vinay Chamola, Nishad Sahu, and Mohsen Guizani. 2020. Applications of blockchain in unmanned aerial vehicles: A review. *Vehicular Communications* (February 2020).
[51] Yang Lu. 2018. Blockchain: A survey on functions, applications and open issues. *Journal of Industrial Integration and Management* 3, 04 (2018), 1850015.
[52] Fran Casino, Thomas K. Dasaklis, and Constantinos Patsakis. 2019. A systematic literature review of blockchain-based applications: Current status, classification and open issues. *Telematics and Informatics* 36 (2019), 55–81.
[53] Oscar Bermeo-Almeida, Mario Cardenas-Rodriguez, Teresa Samaniego-Cobo, Enrique Ferruzola-Gómez, Roberto Cabezas-Cabezas, and William Bazán-Vera. 2018. Blockchain in agriculture: A systematic literature review. In *Proceedings of the International Conference on Technologies and Innovation*. Springer, 44–56.
[54] Mohamed Amine Ferrag, Lei Shu, Xing Yang, Abdelouahid Derhab, and Leandros Maglaras. 2020. Security and privacy for green IoT based agriculture review blockchain solutions and challenges. *IEEE Access* 8, 1 (2020), 1–x. DOI: http://dx.doi.org/10.1109/ACCESS.2020.2973178
[55] Talal Alharbi. 2020. Deployment of blockchain technology in software defined networks: A survey. *IEEE Access* 8, 1 (2020), 9146–9156. DOI: http://dx.doi.org/10.1109/ACCESS.2020.2964751
[56] Ioannis Konstantinidis, Georgios Siaminos, Christos Timplalexis, Panagiotis Zervas, Vassilios Peristeras, and Stefan Decker. 2018. Blockchain for business applications: A systematic literature review. In *Proceedings of the International Conference on Business Information Systems*. Springer, 384–399.
[57] Junfeng Xie, Helen Tang, Tao Huang, F. Richard Yu, Renchao Xie, Jiang Liu, and Yunjie Liu. 2019. A survey of blockchain technology applied to smart cities: Research issues and challenges. *IEEE Communications Surveys & Tutorials* 21, 3 (2019), 2794–2830.
[58] Tejasvi Alladi, Vinay Chamola, Joel J.P.C. Rodrigues, and Sergei A. Kozlov. 2019. Blockchain in smart grids: A review on different use cases. *Sensors* 19, 22 (2019), 4862–4886.
[59] Adedayo Aderibole, Aamna Aljarwan, Muhammad Habib Ur Rehman, Hatem H. Zeineldin, Toufic Mezher, Khaled Salah, Ernesto Damiani, and Davor Svetinovic. 2020. Blockchain technology for smart grids: Decentralized NIST conceptual model. *IEEE Access* (2020).
[60] Huawei Huang, Jianru Lin, Baichuan Zheng, Zibin Zheng, and Jing Bian. 2020. When blockchain meets distributed file systems: An overview, challenges, and open issues. *IEEE ACCESS* 8 (March 2020), 50574–50586.

[61] Mohamed Torky, Tarek Gaber, and Aboul Ella Hassanien. 2020. Blockchain in space industry: Challenges and solutions. *arXiv preprint arXiv:2002.12878* (2020).

[62] Dinh Nguyen, Ming Ding, Pubudu N. Pathirana, and Aruna Seneviratne. 2020. Blockchain and AI-based solutions to combat Coronavirus (COVID-19)-like epidemics: A survey. *Techrxiv, No. 12121962* (4 2020). DOI: http://dx.doi.org/10.36227/techrxiv.12121962.v1

[63] Yong Yuan and Fei-Yue Wang. 2016. Blockchain: The state of the art and future trends. *Acta Automatica Sinica* 42, 4 (2016), 481–494.

[64] Zibin Zheng, Shaoan Xie, Hongning Dai, Xiangping Chen, and Huaimin Wang. 2017. An overview of blockchain technology: Architecture, consensus, and future trends. In *Proceedings of the IEEE International Congress on Big Data (BigData Congress)*. 557–564.

[65] Zibin Zheng, Shaoan Xie, Hong-Ning Dai, Xiangping Chen, and Huaimin Wang. 2018. Blockchain challenges and opportunities: A survey. *International Journal of Web and Grid Services* 14, 4 (2018), 352–375.

[66] Yong Yuan and Fei-Yue Wang. 2018. Blockchain and cryptocurrencies: Model, techniques, and applications. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 48, 9 (2018), 1421–1428.

[67] John Kolb, Moustafa AbdelBaky, Randy H. Katz, and David E. Culler. 2020. Core concepts, challenges, and future directions in blockchain: A centralized tutorial. *ACM Computing Surveys (CSUR)* 53, 1 (2020), 1–39.

[68] Loi Luu, Viswesh Narayanan, Chaodong Zheng, Kunal Baweja, Seth Gilbert, and Prateek Saxena. 2016. A secure sharding protocol for open blockchains. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 17–30.

[69] Eleftherios Kokoris-Kogias, Philipp Jovanovic, Linus Gasser, Nicolas Gailly, Ewa Syta, and Bryan Ford. 2018. Omniledger: A secure, scale-out, decentralized ledger via sharding. In *Proceedings of the 2018 IEEE Symposium on Security and Privacy (SP)*. IEEE, 583–598.

[70] Daniel J. Moroz, Daniel J. Aronoff, Neha Narula, and David C. Parkes. 2020. Double-Spend Counterattacks: Threat of Retaliation in Proof-of-Work Systems. arxiv:cs.CR/2002.10736

[71] Gavin Wood et al. 2014. Ethereum: A secure decentralised generalised transaction ledger. *Ethereum Project Yellow Paper* 151 (2014), 1–32.

[72] Ethereum Sharding. Retrieved from https://eth.wiki/sharding/Sharding-FAQs.

[73] Hyperledger Fabric Website. Retrieved from https://hyperledger-fabric.readthedocs.io/en/release-1.4/write_first_app.html.

[74] 2020. EOSIO. *Website*. Retrieved November 2020 from https://eos.io/.

[75] Adiseshu Hari, Murali Kodialam, and TV Lakshman. 2019. ACCEL: Accelerating the Bitcoin blockchain for high-throughput, low-latency applications. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM'19)*. IEEE, 2368–2376.

[76] Jiaping Wang and Hao Wang. 2019. Monoxide: Scale out blockchains with asynchronous consensus zones. In *Proceedings of the 16th USENIX Symposium on Networked Systems Design and Implementation (NSDI)*. 95–112.

[77] Lei Yang, Vivek Bagaria, Gerui Wang, Mohammad Alizadeh, David Tse, Giulia Fanti, and Pramod Viswanath. 2019. Prism: Scaling Bitcoin by 10,000 x. *arXiv preprint arXiv:1909.11261.*

[78] Sangyeon Woo, Jeho Song, Sanghyeok Kim, Youngjae Kim, and Sungyong Park. 2020. GARET: Improving throughput using gas consumption-aware relocation in Ethereum sharding environments. *Cluster Computing* (2020), 1–13.

[79] Doriane Perard, Jérôme Lacan, Yann Bachy, and Jonathan Detchart. 2018. Erasure code-based low storage blockchain node. In *Proceedings of the 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, 1622–1627.

[80] Xiaohai Dai, Jiang Xiao, Wenhui Yang, Chaofan Wang, and Hai Jin. 2019. Jidar: A jigsaw-like data reduction approach without trust assumptions for Bitcoin system. In *Proceedings of the IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 1317–1326.

[81] Yibin Xu and Yangyu Huang. 2020. Segment blockchain: A size reduced storage mechanism for blockchain. *IEEE Access* 8 (2020), 17434–17441.

[82] Ingo Weber, Vincent Gramoli, Alex Ponomarev, Mark Staples, Ralph Holz, An Binh Tran, and Paul Rimba. 2017. On availability for blockchain-based systems. In *Proceedings of the 2017 IEEE 36th Symposium on Reliable Distributed Systems (SRDS)*. IEEE, 64–73.

[83] Peilin Zheng, Zibin Zheng, and Liang Chen. 2019. Selecting reliable blockchain peers via hybrid blockchain reliability prediction. *arXiv preprint arXiv:1910.14614.*

[84] Mahdi Zamani, Mahnush Movahedi, and Mariana Raykova. 2018. Rapidchain: Scaling blockchain via full sharding. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*. 931–948.

[85] Mohammad Javad Amiri, Divyakant Agrawal, and Amr El Abbadi. 2019. SharPer: Sharding permissioned blockchains over network clusters. *arXiv preprint arXiv:1910.00765.*

[86] Sanghyeok Kim, Jeho Song, Sangyeon Woo, Youngjae Kim, and Sungyong Park. 2019. Gas consumption-aware dynamic load balancing in Ethereum sharding environments. In *Proceedings of the IEEE 4th International Workshops on Foundations and Applications of Self* Systems (FAS*W)*. IEEE, 188–193.
[87] Jianrong Wang, Yangyifan Zhou, Xuewei Li, Tianyi Xu, and Tie Qiu. 2019. A node rating based sharding scheme for blockchain. In *Proceedings of the IEEE 25th International Conference on Parallel and Distributed Systems (ICPADS)*. IEEE, 302–309.
[88] Lan N. Nguyen, Truc DT Nguyen, Thang N. Dinh, and My T. Thai. 2019. OptChain: Optimal transactions placement for scalable blockchain sharding. In *Proceedings of IEEE 39th International Conference on Distributed Computing Systems (ICDCS)*. 525–535.
[89] Hung Dang, Tien Tuan Anh Dinh, Dumitrel Loghin, Ee-Chien Chang, Qian Lin, and Beng Chin Ooi. 2019. Towards scaling blockchain systems via sharding. In *Proceedings of the 2019 International Conference on Management of Data*. 123–140.
[90] Huan Chen and Yijie Wang. 2019. Sschain: A full sharding protocol for public blockchain without data migration overhead. *Pervasive and Mobile Computing* 59 (2019), 101055.
[91] Jianyu Niu. 2019. Eunomia: A permissionless parallel chain protocol based on logical clock. *arXiv preprint arXiv:1908.07567*.
[92] Tayebeh Rajab, Mohammad Hossein Manshaei, Mohammad Dakhilalian, Murtuza Jadliwala, and Mohammad Ashiqur Rahman. 2020. On the feasibility of sybil attacks in shard-based permissionless blockchains. *arXiv preprint arXiv:2002.06531*.
[93] Yibin Xu and Yangyu Huang. 2020. An n/2 byzantine node tolerate blockchain sharding approach. *arXiv preprint arXiv:2001.05240*.
[94] Mengqian Zhang, Jichen Li, Zhaohua Chen, Hongyin Chen, and Xiaotie Deng. 2020. Cycledger: A scalable and secure parallel protocol for distributed ledger via sharding. *arXiv preprint arXiv:2001.06778*.
[95] Hai Jin, Xiaohai Dai, and Jiang Xiao. 2018. Towards a novel architecture for enabling interoperability amongst multiple blockchains. In *Proceedings of the 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 1203–1211.
[96] Zhuotao Liu, Yangxi Xiang, Jian Shi, Peng Gao, Haoyu Wang, Xusheng Xiao, Bihan Wen, and Yih-Chun Hu. 2019. Hyperservice: Interoperability and programmability across heterogeneous blockchains. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. 549–566.
[97] Enrique Fynn, Alysson Bessani, and Fernando Pedone. 2020. Smart contracts on the move. *arXiv preprint arXiv:2004.05933*.
[98] Hangyu Tian, Kaiping Xue, Shaohua Li, Jie Xu, Jianqing Liu, and Jun Zhao. 2020. Enabling cross-chain transactions: A decentralized cryptocurrency exchange protocol. *arXiv preprint arXiv:2005.03199*.
[99] Yuefei Gao, Shin Kawai, and Hajime Nobuhara. 2019. Scalable blockchain protocol based on proof of stake and sharding. *Journal of Advanced Computational Intelligence and Intelligent Informatics* 23, 5 (2019), 856–863. DOI: http://dx.doi.org/10.20965/jaciii.2019.p0856
[100] Bernardo David, Peter Gaži, Aggelos Kiayias, and Alexander Russell. 2018. Ouroboros praos: An adaptively-secure, semi-synchronous proof-of-stake blockchain. In *Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer, 66–98.
[101] Ethan Buchman, Jae Kwon, and Zarko Milosevic. 2018. The latest gossip on BFT consensus. *arXiv preprint arXiv:1807.04938*.
[102] Jun Zou, Bin Ye, Lie Qu, Yan Wang, Mehmet A. Orgun, and Lei Li. 2018. A proof-of-trust consensus protocol for enhancing accountability in crowdsourcing services. *IEEE Transactions on Services Computing* 12, 3 (2018), 429–445.
[103] Zsolt István, Alessandro Sorniotti, and Marko Vukolić. 2018. Streamchain: Do blockchains need blocks? In *Proceedings of the 2nd Workshop on Scalable and Resilient Infrastructures for Distributed Ledgers*. 1–6.
[104] Mohammad Javad Amiri, Divyakant Agrawal, and Amr El Abbadi. 2019. CAPER: A cross-application permissioned blockchain. *Proceedings of the VLDB Endowment* 12, 11 (2019), 1385–1398.
[105] Z. Chang, W. Guo, X. Guo, Z. Zhou, and T. Ristaniemi. 2020. Incentive mechanism for edge computing-based blockchain. *IEEE Transactions on Industrial Informatics* 16, 11 (2020), 7105–7114. DOI: http://dx.doi.org/10.1109/TII.2020.2973248
[106] Weilin Zheng, Xu Chen, Zibin Zheng, Xiapu Luo, and Jiahui Cui. 2020. AxeChain: A secure and decentralized blockchain for solving easily-verifiable problems. *arXiv preprint arXiv:2003.13999*.
[107] Lin Chen, Lei Xu, Zhimin Gao, Keshav Kasichainula, and Weidong Shi. 2020. Nonlinear blockchain scalability: A game-theoretic perspective. *arXiv preprint arXiv:2001.08231*.
[108] Yonatan Sompolinsky and Aviv Zohar. 2015. Secure high-rate transaction processing in bitcoin. In *International Conference on Financial Cryptography and Data Security*. Springer, 507–527.

[109] Yonatan Sompolinsky and Aviv Zohar. 2018. PHANTOM: A scalable BlockDAG protocol. *IACR Cryptology Eprint Archive* 2018 (2018), 104.
[110] Chenxing Li, Peilun Li, Dong Zhou, Wei Xu, Fan Long, and Andrew Yao. 2018. Scaling nakamoto consensus to thousands of transactions per second. *arXiv preprint arXiv:1805.03870.*
[111] Ting Chen, Yuxiao Zhu, Zihao Li, Jiachi Chen, Xiaoqi Li, Xiapu Luo, Xiaodong Lin, and Xiaosong Zhange. 2018. Understanding Ethereum via graph analysis. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM).* IEEE, 1484–1492.
[112] Cuneyt Gurcan Akcora, Matthew F. Dixon, Yulia R. Gel, and Murat Kantarcioglu. 2018. Bitcoin risk modeling with blockchain graphs. *Economics Letters* 173 (2018), 138–142.
[113] Matthew F. Dixon, Cuneyt Gurcan Akcora, Yulia R. Gel, and Murat Kantarcioglu. 2019. Blockchain analytics for intraday financial risk modeling. *Digital Finance* 1, 1–4 (2019), 67–89.
[114] Cuneyt G. Akcora, Asim Kumer Dey, Yulia R. Gel, and Murat Kantarcioglu. 2018. Forecasting bitcoin price with graph chainlets. In *Proceedings of the Pacific-Asia Conference on Knowledge Discovery and Data Mining.* Springer, 765–776.
[115] Nazmiye Ceren Abay, Cuneyt Gurcan Akcora, Yulia R. Gel, Murat Kantarcioglu, Umar D. Islambekov, Yahui Tian, and Bhavani Thuraisingham. 2019. Chainnet: Learning on blockchain graphs with topological features. In *Proceedings of the IEEE International Conference on Data Mining (ICDM).* 946–951.
[116] Weili Chen, Jun Wu, Zibin Zheng, Chuan Chen, and Yuren Zhou. 2019. Market manipulation of bitcoin: Evidence from mining the mt. gox transaction network. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM).* 964–972.
[117] Friedhelm Victor and Bianca Katharina Lüders. 2019. Measuring Ethereum-based erc20 token networks. In *Proceedings of the International Conference on Financial Cryptography and Data Security.* Springer, 113–129.
[118] Shahar Somin, Goren Gordon, and Yaniv Altshuler. 2018. Network analysis of erc20 tokens trading on Ethereum blockchain. In *Proceedings of the International Conference on Complex Systems.* Springer, 439–450.
[119] Yijing Zhao, Jieli Liu, Qing Han, Weilin Zheng, and Jiajing Wu. 2020. Exploring EOSIO via graph characterization. *arXiv preprint arXiv:2004.10017.*
[120] Nikolaos Papadis, Sem Borst, Anwar Walid, Mohamed Grissa, and Leandros Tassiulas. 2018. Stochastic models and wide-area network measurements for blockchain design and analysis. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM).* IEEE, 2546–2554.
[121] Aditya Gopalan, Abishek Sankararaman, Anwar Walid, and Sriram Vishwanath. 2020. Stability and scalability of blockchain systems. *arXiv preprint arXiv:2002.02567.*
[122] Abdelatif Hafid, Abdelhakim Senhaji Hafid, and Mustapha Samih. 2019. A probabilistic security analysis of sharding-based blockchain protocols. In *Proceedings of the International Congress on Blockchain and Applications (Blockchain).* 55–60.
[123] Abdelatif Hafid, Abdelhakim Senhaji Hafid, and Mustapha Samih. 2019. A methodology for a probabilistic security analysis of sharding-based blockchain protocols. In *Proceedings of the International Congress on Blockchain and Applications.* Springer, 101–109.
[124] Abdelatif Hafid, Abdelhakim Senhaji Hafid, and Mustapha Samih. 2019. New mathematical model to analyze security of sharding-based blockchain protocols. *IEEE Access* 7 (2019), 185447–185457.
[125] Quan-Lin Li, Jing-Yu Ma, and Yan-Xia Chang. 2018. Blockchain queue theory. In *Proceedings of the International Conference on Computational Social Networks.* Springer, 25–40.
[126] Quan-Lin Li, Jing-Yu Ma, Yan-Xia Chang, Fan-Qi Ma, and Hai-Bo Yu. 2019. Markov processes in blockchain systems. *Computational Social Networks* 6, 1 (2019), 1–28.
[127] Saulo Ricci, Eduardo Ferreira, Daniel Sadoc Menasche, Artur Ziviani, Jose Eduardo Souza, and Alex Borges Vieira. 2019. Learning blockchain delays: A queueing theory approach. *ACM SIGMETRICS Performance Evaluation Review* 46, 3 (2019), 122–125.
[128] Maria Frolkova and Michel Mandjes. 2019. A Bitcoin-inspired infinite-server model with a random fluid limit. *Stochastic Models* 35, 1 (2019), 1–32.
[129] Minghong Fang and Jia Liu. 2020. Toward low-cost and stable blockchain networks. *arXiv preprint arXiv:2002.08027* (2020).
[130] Raheel Ahmed Memon, Jian Ping Li, and Junaid Ahmed. 2019. Simulation model for blockchain systems using queuing theory. *Electronics* 8, 2 (2019), 234.
[131] Karl Wüst and Arthur Gervais. 2018. Do you need a blockchain? In *Proceedings of the 2018 Crypto Valley Conference on Blockchain Technology (CVCBT).* IEEE, 45–54.
[132] Xuyun Zhang, Chang Liu, Surya Nepal, Suraj Pandey, and Jinjun Chen. 2013. A privacy leakage upper bound constraint-based approach for cost-effective privacy preserving of intermediate data sets in cloud. *IEEE Transactions on Parallel and Distributed Systems* 24, 6 (2013), 1192–1202.

[133] Dan Lin, Jiajing Wu, Qi Yuan, and Zibin Zheng. 2020. Modeling and understanding Ethereum transaction records via a complex network approach. *IEEE Transactions on Circuits and Systems II: Express Briefs* 67, 11 (2020), 2737–2741.

[134] José Eduardo de A. Sousa, Vinıcius Oliveira, Júlia Valadares, Alex B. Vieira, Heder S. Bernardino, and Glauber Dias. 2019. An analysis of the fees and pending time correlation in Ethereum. In *Proceedings of the LANOMS*. IFIP, 1–7.

[135] Eitan Altman, Daniel Menasché, Alexandre Reiffers, Mandar Datar, Swapnil Dhamal, Corinne Touati, and Rachid El-Azouzi. 2019. Blockchain competition between miners: A game theoretic perspective. *Frontiers in Blockchain* 2 (2019), 26.

[136] Jun Zhao, Jing Tang, Zengxiang Li, Huaxiong Wang, Kwok-Yan Lam, and kaiping Xue. 2020. An analysis of blockchain consistency in asynchronous networks: Deriving a neat bound. In *Proceedings of the IEEE International Conference on Distributed Computing Systems (ICDCS)*. 1–10.

[137] Yang Xiao, Ning Zhang, Wenjing Lou, and Y. Thomas Hou. 2020. Modeling the impact of network connectivity on consensus security of proof-of-work blockchain. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM'20)*. 1–9.

[138] Enrique Fynn and Fernando Pedone. 2018. Challenges and pitfalls of partitioning blockchains. In *Proceedings of the 2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*. IEEE, 128–133.

[139] Zeta Avarikioti, Eleftherios Kokoris-Kogias, and Roger Wattenhofer. 2019. Divide and scale: Formalization of distributed ledger sharding protocols. *arXiv preprint arXiv:1910.10434*.

[140] Shengling Wang, Chenyu Wang, and Qin Hu. 2019. Corking by forking: Vulnerability analysis of blockchain. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 829–837.

[141] Alysson Bessani, Eduardo Alchieri, João Sousa, André Oliveira, and Fernando Pedone. 2020. From byzantine replication to blockchain: Consensus is only the beginning. *arXiv preprint arXiv:2004.14527*.

[142] Guangsheng Yu, Xuan Zha, Xu Wang, Wei Ni, Kan Yu, J. Andrew Zhang, and Ren Ping Liu. 2020. A unified analytical model for proof-of-x schemes. *Computers & Security* (2020), 101934.

[143] Shuangke Wu, Yanjiao Chen, Minghui Li, Xiangyang Luo, Zhe Liu, and Lan Liu. 2020. Survive and thrive: A stochastic game for DDoS attacks in bitcoin mining pools. *IEEE/ACM Transactions on Networking* 28, 2 (2020), 874–887.

[144] Eric Budish. 2018. *The Economic Limits of Bitcoin and the Blockchain*. Technical Report. National Bureau of Economic Research.

[145] Rashid Tahir, Sultan Durrani, Faizan Ahmed, Hammas Saeed, Fareed Zaffar, and Saqib Ilyas. 2019. The browsers strike back: Countering cryptojacking and parasitic miners on the web. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 703–711.

[146] Rui Ning, Cong Wang, ChunSheng Xin, Jiang Li, Liuwan Zhu, and Hongyi Wu. 2019. CapJack: Capture in-browser crypto-jacking by deep capsule network through behavioral analysis. In *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*. IEEE, 1873–1881.

[147] Yining Hu, Suranga Seneviratne, Kanchana Thilakarathna, Kensuke Fukuda, and Aruna Seneviratne. 2019. Characterizing and detecting money laundering activities on the Bitcoin network. *arXiv preprint arXiv:1912.12060*.

[148] Marie Vasek and Tyler Moore. 2018. Analyzing the Bitcoin Ponzi scheme ecosystem. In *International Conference on Financial Cryptography and Data Security*. Springer, 101–112.

[149] Weili Chen, Zibin Zheng, Jiahui Cui, Edith Ngai, Peilin Zheng, and Yuren Zhou. 2018. Detecting Ponzi schemes on Ethereum: Towards healthier blockchain technology. In *Proceedings of the 2018 World Wide Web Conference (WWW)*. 1409–1418.

[150] Weili Chen, Zibin Zheng, Edith C.-H. Ngai, Peilin Zheng, and Yuren Zhou. 2019. Exploiting blockchain data to detect smart Ponzi schemes on Ethereum. *IEEE Access* 7 (2019), 37575–37586.

[151] Marek Laskowski, Michael Zargham, Hjalmar Turesson, Henry M. Kim, Matt Barlin, Danil Kabanov, and Eden Dhaliwal. 2020. Evidence based decision making in blockchain economic systems: From theory to practice. *arXiv preprint arXiv:2001.03020*.

[152] Aviv Yaish and Aviv Zohar. 2020. Pricing ASICs for cryptocurrency mining. *arXiv preprint arXiv:2002.11064*.

[153] Shayan Eskandari, Andreas Leoutsarakos, Troy Mursch, and Jeremy Clark. 2018. A first look at browser-based cryptojacking. In *Proceedings of the IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. IEEE, 58–66.

[154] Sara Sabour, Nicholas Frosst, and Geoffrey E. Hinton. 2017. Dynamic routing between capsules. In *Advances in Neural Information Processing Systems*. 3856–3866.

[155] Massimo Bartoletti, Barbara Pes, and Sergio Serusi. 2018. Data mining for detecting Bitcoin Ponzi schemes. In *Proceedings of the 2018 Crypto Valley Conference on Blockchain Technology (CVCBT)*. IEEE, 75–84.

[156] Junwoo Seo, Mookyu Park, Haengrok Oh, and Kyungho Lee. 2018. Money laundering in the Bitcoin network: Perspective of mixing services. In *Proceedings of the IEEE International Conference on Information and Communication Technology Convergence (ICTC)*. 1403–1405.

[157] Arthur Gervais, Ghassan O. Karame, Karl Wüst, Vasileios Glykantzis, Hubert Ritzdorf, and Srdjan Capkun. 2016. On the security and performance of proof of work blockchains. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. 3–16.

[158] Qassim Nasir, Ilham A. Qasse, Manar Abu Talib, and Ali Bou Nassif. 2018. Performance analysis of hyperledger fabric platforms. *Security and Communication Networks* 2018 (2018).

[159] Peilin Zheng, Zibin Zheng, Xiapu Luo, Xiangping Chen, and Xuanzhe Liu. 2018. A detailed and real-time performance monitoring framework for blockchain systems. In *Proceedings of the IEEE/ACM 40th International Conference on Software Engineering: Software Engineering in Practice Track (ICSE-SEIP)*. 134–143.

[160] Tien Tuan Anh Dinh, Ji Wang, Gang Chen, Rui Liu, Beng Chin Ooi, and Kian-Lee Tan. 2017. Blockbench: A framework for analyzing private blockchains. In *Proceedings of the 2017 ACM International Conference on Management of Data*. 1085–1100.

[161] Seoung Kyun Kim, Zane Ma, Siddharth Murali, Joshua Mason, Andrew Miller, and Michael Bailey. 2018. Measuring Ethereum network peers. In *Proceedings of the Internet Measurement Conference (IMC'18)*. 91–104.

[162] Lina Alsahan, Noureddine Lasla, and Mohamed M. Abdallah. 2020. Local Bitcoin network simulator for performance evaluation using lightweight virtualization. In *Proceedings of the IEEE International Conference on Informatics, IoT, and Enabling Technologies*. 1–6.

[163] Parity documentation. Retrieved from https://paritytech.github.io/wiki.

[164] CITA Technical Whitepaper. Retrieved from https://github.com/cryptape/cita.

[165] Elli Androulaki, Artem Barger, Vita Bortnikov, Christian Cachin, Konstantinos Christidis, Angelo De Caro, David Enyeart, Christopher Ferris, Gennady Laventman, Yacov Manevich, et al. 2018. Hyperledger fabric: A distributed operating system for permissioned blockchains. In *Proceedings of the 13th EuroSys Conference*. 1–15.

[166] Xblock. 2020. Performance Monitoring. Retrieved February 2020 from http://xblock.pro/performance/.

[167] Peilin Zheng, Zibin Zheng, and Hong-ning Dai. 2019. XBlock-ETH: Extracting and exploring blockchain data from etherem. *arXiv preprint arXiv:1911.00169*.

[168] Weilin Zheng, Zibin Zheng, Hong-Ning Dai, Xu Chen, and Peilin Zheng. 2020. XBlock-EOS: Extracting and exploring blockchain data from EOSIO. *arXiv preprint arXiv:2003.11967*.

[169] Harry Kalodner, Steven Goldfeder, Alishah Chator, Malte Möser, and Arvind Narayanan. 2017. BlockSci: Design and applications of a blockchain analysis platform. *arXiv preprint arXiv:1709.02489* (2017).

[170] Andrew Miller, Yu Xia, Kyle Croman, Elaine Shi, and Dawn Song. 2016. The honey badger of BFT protocols. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security (CCS)*. 31–42.

Received July 2020; revised December 2020; accepted December 2020