SCIENCE CHINA
Information Sciences

• POSITION PAPER •

May 2022, Vol. 65 153101:1–153101:11
https://doi.org/10.1007/s11432-020-3183-0

Towards trustworthy blockchain systems in the era of "Internet of value": development, challenges, and future trends

Hai JIN & Jiang XIAO*

National Engineering Research Center for Big Data Technology and System, Services Computing Technology and System Lab, Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, *China*

Received 5 November 2020/Revised 20 December 2020/Accepted 3 February 2021/Published online 18 October 2021

Abstract Since the advent of cryptocurrencies such as Bitcoin, blockchain, as their underlying technologies, has drawn a massive amount of attention from both academia and the industry. This ever-evolving technology inherits the "genes" of distributed systems, offering significant advantages of immutability, transparency, auditability, and tamper-resistance. These benefits help blockchain re-establish public confidence, and hold the significant promise of reliable information sharing and value transfer. Therefore, blockchain has become the foundation of crucial strategic deployments in countries across the world, and the fundamental basis for building the next generation Web 3.0 — "Internet of value". In this article, we will start with unraveling the essential ingredients of blockchain technology, and showing the characteristics of each of these ingredients in the context of distributed systems. We will then present the core technical challenges that need to be addressed prior to unleashing its full potential, including its performance, scalability, and cross-chain interoperability. Finally, we will introduce the recent developments of blockchain systems, and discuss the future trends of the blockchain ecosystem.

Keywords blockchain, distributed systems, internet of value, challenges, development

Citation Jin H, Xiao J. Towards trustworthy blockchain systems in the era of "Internet of value": development, challenges, and future trends. Sci China Inf Sci, 2022, 65(5): 153101, https://doi.org/10.1007/s11432-020-3183-0

1 Introduction

Today blockchain technology is widely acknowledged as the fundamental basis of building the "Internet of value" in the new era of Web 3.0. In a world where data increases exponentially over easily accessible mediums, to make information reliable and facilitate trustworthy value transferring services is of paramount importance. It leads to a rapid shift from exchanging information in the previous Web 2.0 era — the "Internet of information". The core value propositions associated with blockchain lie in its ability to support building tamper-resistant information chains, facilitating data sharing, and empowering value exchange. This technologically fosters a trustworthy environment where you cannot do evil, achieves a high degree of credibility among trustless participating peers, and tremendously reduces the cost of trust.

In the road map of the evolving development of distributed systems, Blockchain, an innovative paradigm with trusted ingredients, has drawn a massive amount of attention from both industry and the research community. Blockchain made its first appearance as the underlying technology of Bitcoin as a form of cryptocurrency[1)] during the 2008 financial crisis. For Bitcoin, it consists of a large number of computers worldwide to validate and record all financial transactions in a consensually shared ledger. Thus, it can allow transactions to have a public "witness" since transactions are added to an immutable distributed ledger with authenticity (i.e., immune to counterfeiting). Unlike past distributed systems where the "trust" established between transacting parties depends on a large number of intermediaries that authenticate the information, which is extremely vulnerable to a single point of failure, e.g., cyber-attacks

* Corresponding author (email: jiangxiao@hust.edu.cn)

1) Bitcoin whitepaper. https://bitcoin.org/bitcoin.pdf/.

or third-party interventions. Since its inception, blockchain technology is expected to play a pivotal role in coping with the erosion of public trust, allowing people to trust each other and transact in a peer-to-peer (P2P) fashion, making the need for third parties such as central banks obsolete. The cover story of an October 2015 issue of the Economist magazine defined the blockchain as "the trust machine" [1]. As such, it transforms the conventional way of establishing trust with a centralized infrastructure through consensus and complex computer code.

By itself, blockchain is a new generation of distributed systems by a conjunction of essential ingredients, including consensus protocols, distributed storage, P2P networks, cryptography, and smart contracts. Through recording transactions among parties that are mutually distrusted, it encompasses the beneficial features of decentralization, immutability, tamper-resistance, traceability, and collaborative maintenance. The distributed nodes in a blockchain system form a P2P network to complete transactions with each other. Each node in the blockchain is authenticated through cryptographic algorithms and owns an anonymous digital identity, which simultaneously enables both user privacy and system security. A consensus protocol determines the rules by which the blockchain data is generated and updated. The openness of blockchain enables its participants to freely access and interact with each other with little friction, ensuring immediate, across-the-board transparency. It provides an innovative technical means to allow new distributed models of value transfer and new forms of P2P human collaborations.

Being the foundation of crucial strategic deployments in countries across the world, blockchain technology has brought immense impacts on society and the global economy. The past decade has shown its great promises for delivering automaticity and efficiency for the digital currency — "Blockchain 1.0". Evolving far beyond the underpinnings of Bitcoin as a cryptocurrency, Ethereum opens up a new era of "Blockchain 2.0" in the form of irreversible real-life agreements enabled by programmable smart contracts. The application of blockchain technology is undergoing a transformation from a single field to a diversified area: from the earliest finance related to the digital currency, expanding to a diverse set of industry sectors, including faster and leaner global trade finance, cross-border commitment, transportation and logistics, superior transparency and traceability in the manufacturing supply chain, increased automation in public healthcare, and e-government services.

Despite a growing number of blockchain systems and various industry-level or enterprise-level blockchain alliances launched to disrupt all the relevant sectors, the current blockchain technology is still in a very early stage, where challenges to overcome are similar to the opportunities offered. A list of challenging issues such as performance and scalability may jeopardize the fast deployment of blockchain systems. Designing secure consensus mechanisms, efficient network protocols, and scalable storage models is expected to need significant future work. Nowadays, there is an increasing number of blockchains launched, in the context of a variety of multi-chain application scenarios, including the transfer of cross-chain assets, interactive authentication, distributed collaboration, cross-chain data sharing, cross-chain real-time clearing, multi-role user identity authentication, and so forth. However, the heterogeneity of multiple blockchains leads to a "lack of interoperability", e.g., unable to communicate with or verify information on the other chains. This "blockchain data-island" situation will significantly worsen, further eroding user experience and hindering the industry growth. It is an emerging issue on which the blockchain community is working actively [2–4].

It is foreseeable that the evolving urgent requirements can lead the future trends in blockchain technology. For example, it is desirable to design an interactive and scalable blockchain data model for diversified storage and query demands, while effectively ensuring the privacy of data owners and data control rights. Meanwhile, it is also important to design an efficient blockchain intercommunication protocol and hierarchical topology model to enhance communication efficiency, as well as to support high-performance real-time processing functions. This will also require standardization and political will to create a regulatory framework conducive to the evolving blockchain network.

The remainder of this paper is structured as follows. First, the evolving development of distributed systems and the essential ingredients of blockchain technology are outlined in Section 2. Then the incorporated challenges are explained with in-depth analysis in Section 3. We then aim to provide an impartial discussion about the current status and increasing needs of the blockchain technology in Section 4. Finally, Section 5 presents an outlook on the future trends of the blockchain ecosystem, and Section 6 concludes the paper.

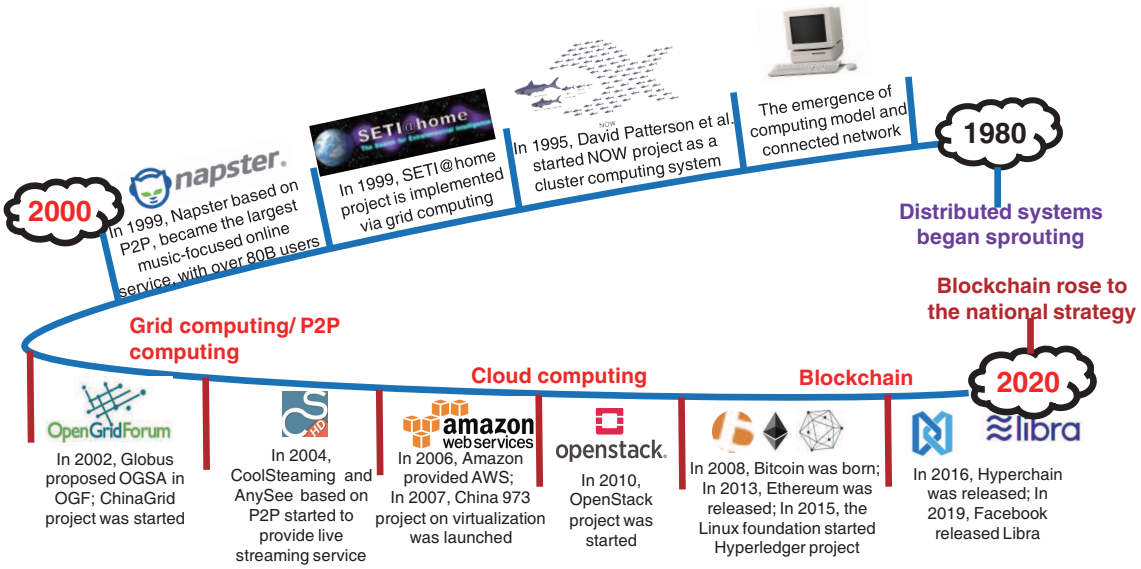

Figure 1 (Color online) A historical perspective of distributed systems.

2 Historical evolution of distributed systems

The blooming development of the Internet for remote access and the proliferation of commodity servers have become the dominant forces to advocate the evolution of distributed systems infrastructure, as shown in Figure 1. The first real and widespread distributed systems began to appear in the 1980s, consisting of multiple machines located physically apart and connected over a network. These nodes work together to fulfill a task, or provide the same service or application to clients, forming a distributed systems cluster. Such a cluster computing infrastructure had roots in the Berkeley network of workstations (NOW) project developed in 1995 by Patterson et al. [5]. The NOW project's initial focus was on improving virtual memory and file system performance, on achieving cheap, highly available, and scalable file storage, and on providing multiple CPUs for parallel computing. It harnesses greater processing power and more advanced analytics to satisfy the needs of both desktop computing and applications that require a hundredfold more computing sources than any single machine within that building can provide.

In 1999, the SETI@home (search for extraterrestrial intelligence) project for analyzing radio telescope signals from space was implemented via grid computing [6]. It was the first large-scale public involvement for scientific research, involving millions of computers provided by the general public. The concept of open grid services architecture (OGSA) was conceived in the 2002 Globus Alliance paper during the Global Grid Forum (OGF) for supporting the interoperable grid requirements [7]. The ChinaGrid project aims to provide a ubiquitous, high-performance, and highly reliable grid service for research and education purposes [8]. This project supports a diverse set of grid applications along with three categories, namely the computation grid (e-science), the information service grid (e-info), and the instrument sharing grid (e-instrument).

In 2006, cloud computing launched with the pioneering Amazon web services (AWS) project of Amazon[2)]. The initial aim of AWS is to offer IT services and resources to the market in the form of web services. The cloud infrastructure enables a pay-as-you-go pricing model that customers need not plan for servers and other IT infrastructure, and therefore providing and releasing resources on-demand, and charged based on the type of resources and peruse. In this way, as a cloud service providers AWS can instantly spin up hundreds or thousands of servers in minutes and deliver results faster. Since then, the control paradigm of computing has shifted from distributed to centralized, which differs cloud computing from grid computing. In 2010, OpenStack launched an open-source cloud operating system to establish a cloud computing management platform with ease of implementation, large-scale deployment, and prosperous and unified standards [9]. OpenStack provided infrastructure-as-a-service (IaaS) solutions through a variety of industry services. Each service provides users with an integrated API, ensuring that users can choose pluggable service components depending on their needs.

2) Amazon web services. https://aws.amazon.com/.

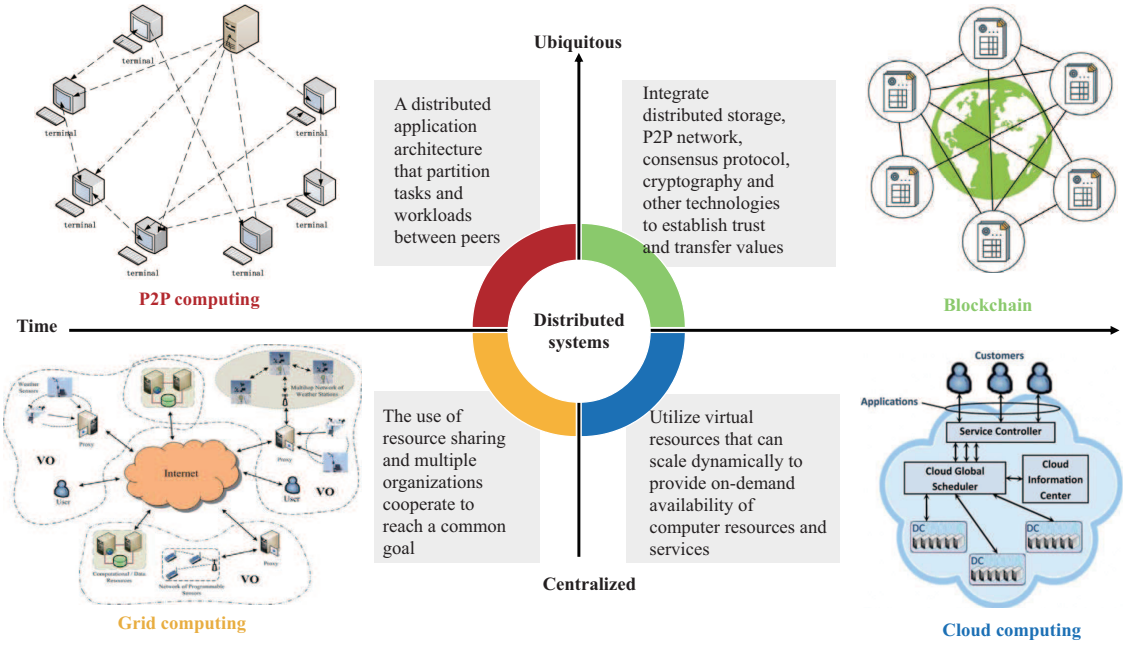

Figure 2 (Color online) The dominant computing paradigms of distributed systems.

The advent of a new wave of computing driven by the blockchain is reshaping the landscape of distributed systems design, operations, and management. The concept of blockchain was first coined as the underlying technology of Bitcoin in 2008. Bitcoin establishes itself as a relatively stable and the most successful cryptocurrency to date. In late 2013, Vitalik Buterin started Ethereum — the second-largest public blockchain platform to date through immutable, programmatic smart contracts, i.e., computer programs that self-execute when certain conditions are met[3)]. The decentralized nature and cryptographic security make the Ethereum network well protected against possible hacking attacks and fraudulent activities. These program-based contractual obligations can enforce stakeholder-agreed rules and process automatically. Promising reducing transaction costs associated with contracting and establishing trust between two parties, Ethereum has gained increasing interest from both academia and industry. Concentrating on advancing the scalability and performance for business sectors, the Linux Foundation originated the enterprise-grade permissioned Hyperledger project [10] in 2015. In 2009, Facebook announced the initial version release of a new blockchain-based digital currency called Libra[4)].

Broadly speaking, the evolution of distributed systems has gone through three different stages. The first is the traditional cluster model that depends on centralized control. A reconstruction of homogeneous nodes formed the cluster. It was followed by the emergence of grid computing that used distributed control to achieve collaborative work in a distributed environment. Thus, a critical issue in grid computing was collaborative problem-solving. Alternatively, P2P computing was fully distributed and autonomous. Both cloud computing and blockchain evolved from the foundation of grid computing and P2P computing in the past decade. Blockchain participants are located physically apart from each other and connected on a network. Each participant operating a full node maintains a complete copy of a ledger that is updated with new transactions as they occur.

The development roadmap of distributed systems tends to follow two paths, as depicted in Figure 2. One path emphasizes centralized control and resource management. For example, the coordinated and aggregate use of diverse resources in grid computing led by virtual organizations. In contrast, the control of the back-end infrastructure in cloud computing is limited to the cloud vendor. The other path is motivated by the ubiquitous dispersion of widespread resources. Both are necessary, and successful advances in distributed computing synthesize elements from both camps.

Blockchain is an innovative distributed computing paradigm with extended features in distributed systems, consisting of a fusion of consensus protocol, distributed storage, and peer-to-peer networks. Table 1 presents a comparison between blockchain and traditional distributed systems. The most essential

3) Ethereum whitepaper. https://ethereum.github.io/yellowpaper/paper.pdf.
4) Libra whitepaper. https://libra.org/en-US/white-paper/?noredirect=1.

Table 1 Blockchain vs. traditional distributed systems

		Blockchain system	Traditional distributed system
Consensus protocol	Architecture	Decentralization	Centralization
	Fault tolerance	Byzantine fault tolerance (BFT)	Crash fault tolerance (CFT)
	Performance	Low throughput (<100 TPS)	High throughput (K∼10K TPS)
Data storage	Architecture	Decentralization	Centralization
	Storge cost/reliability	Full-copy, high reliability	Multi-copy, low cost
	Data model	Chain-based, DAG-based	Diverse data structure (key value, graph, document, etc.)
	Functionality	Simple	Rich
Network communication	Architecture	P2P	C/S, P2P
	Node identity	Pseudonymous	Revealed
	Broadcast	Gossip, Kamedia	Tree-based
	Topology	Unstructured, structured	Centralized, unstructured, structured, semi-structured

ingredient is the distributed consensus protocol, which aims to update and synchronize the blockchain networks' data. It is used to determine the ordering of transactions in an adversarial environment (i.e., assuming not every participant is honest). Nevertheless, it is non-trivial to ensure data consistency, especially reconciling the assumptions of no honest, faulty, or malicious participating peers. In general, existing blockchain consensus protocols can be divided into two categories: lottery-based and voting-based [11,12]. The former category has good scalability but poor performance, with proof-of-work (PoW) as its representative. While this lottery-based consensus encounters performance bottlenecks, the classical BFT-based (Byzantine fault tolerance) consensus is introduced to blockchain. The BFT makes voting decisions among peers to achieve high performance, thus also termed as voting-based consensus.

Inherited from cloud computing, the data is structured in an authenticated block and consecutively chained together using cryptography, verified by Merkle tree, and distributed stored in multiple nodes. On a basic level, a blockchain [4] is a kind of cryptographical sequence of blocks, where each block consists of multiple transactions with a limited storage capacity (e.g., 2 MB in Bitcoin). The first block is called the genesis block, and the following blocks are linked chronologically using cryptographic hashes. The basic data structure of a block involves the data (i.e., multiple transactions), a chaining-hash derived from the immediately preceding one, and a block-hash for identification. Note that block-hash is a unique indicator for every block which can ensure the tamper-resistance of block data by malicious users. Specifically, any manipulation on the block data will inherently change the block-hash. In the case of Bitcoin, the PoW consensus protocol is integrated to enhance the safety of the system, making it highly resistant to cyberattacks.

In general, the blockchain data is replicated in each node to make the storage fully distributed. With local storage, each node can verify the legitimacy of new transactions and blocks independently. The blockchain storage can be separated into two parts: on-chain ledger and off-chain state. The former is recorded on the blockchain directly, including the bodies of transactions and blocks. The latter is constructed and updated based on the on-chain ledger, such as the UTXO set in Bitcoin, which aims to accelerate blockchain's data operations.

The geographically dispersed nodes need to communicate with each other to synchronize the ledger data and transaction proposals. Since there is no centralized server in the network, the nodes communicate with each other in a P2P manner. The P2P protocols adopted by the blockchain network include two kinds: unstructured P2P and structured P2P. Different nodes communicate with each other at random in the former, which provides low efficiency at low costs, such as Gossip adopted in Bitcoin. On the contrary, different nodes in the latter communicate with each other subject to a constrained topology structure. Although it improves the communication efficiency, it also brings high overhead to maintain the topology structure, with Kamedia-like P2P protocol in Ethereum as an example.

The fundamental problems of distributed systems demand blockchain to exploit the abovementioned major ingredients of innovative characteristics. The first one refers to a distributed consensus protocol, i.e., how to ensure data consistency, guarantee throughput, and empower high-frequency concurrent processing. It follows by the performance of the blockchain network with a large scale of nodes, i.e., how to achieve the upgrade of the topology model, the optimization of the transmission protocol, and the efficiency improvement of data synchronization. Storage scalability is another vital issue for blockchain.

For example, how to sustain incremental data storage, how to achieve trusted data verification and transparent data query. Indeed, it is extremely difficult to meet these system-level goals simultaneously. In Section 3, the challenges faced by blockchain technology are presented in detail.

3 Key challenges facing blockchain technology

As an emerging computing paradigm of distributed systems, blockchain has to face the inherent classical distributed problems with divergent new features. The 2019 Gartner hype cycle of blockchain technology[5)] reports that although distributed ledger and cryptocurrency mining have reached maturity, the other core techniques like blockchain managed services, distributed storage in blockchain, and blockchain interoperability remain at the initial stage, or the rising stage. The gap between the proof-of-concept (POC) and wide adoption is still relatively large. In this section, we will present the key challenging issues that have not yet been effectively resolved.

Let us first take a close look at the consensus protocol of blockchain. It stands for the root cause of the lack of scalability for mainstream blockchain systems. Based on the PoW consensus, the generation time of each block in Bitcoin is around 10 min, and the throughput is far lower than the VISA. On behalf of the lottery-based consensus, such energy-intensive proof mechanism makes it difficult to efficiently obtain consistency. In this light, the proof-of-stake (PoS) is an energy-efficient substitute that committing stake (e.g., tokens) for validating block while its security leaves as an unresolved issue. BFT-based consensus provides a high performance among a small number of nodes, e.g., practical Byzantine fault tolerance (PBFT). However, the growth of network size and strong dynamics make the BFT consensus unapplicable for large-scale blockchain deployment.

The blockchain P2P network is a key factor restricting the performance of the blockchain, which will significantly affect the broadcast speed of the block. For example, in the Ethereum network, it takes 10 s to synchronize a block to the entire system. Such a high network delay dramatically affects the efficiency of the Ethereum consensus. On the other hand, the block broadcast delay occupies too much block interval, which will lead to the poor performance of the blockchain system, and vulnerability to double-spending attacks. Current advancements in the blockchain network can be divided into two categories. One type is a fully distributed unstructured P2P network protocol that optimizes the Bitcoin network by shortening the network diameter. However, this solution will bring huge computational overhead, because each node needs to repeatedly calculate the network distance between it and all other nodes. The second category is that Ethereum adopts a fully distributed structured P2P network protocol. Although it increases the connectivity of the entire network, the growth of Ethereum nodes will bring high maintenance costs.

The continuous growth of data storage overhead impedes the current blockchain development. At present, mainstream blockchain systems inevitably face the storage problem of increasing data volume. Take Bitcoin, the public chain with the largest number of users in the world, as an example. To ensure immutability and system security, any full node needs to save the historical on-chain data for verification. As of May 2019, the total amount of data on Bitcoin is about 200 GB. Similarly, the storage capacity of an Ethereum full node is about 600 GB. As time goes by, the data storage overhead of the blockchain will continue to increase, resulting in excessive load and reduced efficiency of network nodes, further restricting the development of blockchain technology. At the same time, the cost of data storage is prohibitive. For example, on Ethereum, it costs about 3.76 ETH to store 1 MB of data, over USD 1385, at the current market price. Obviously, such a high price is unacceptable in real-world commercial use.

Moreover, with the improvement of blockchain performance, the amount of data processing per unit time has increased sharply, which brings more severe challenges in storage overhead to the underlying data layer. For example, when the throughput of Bitcoin is 7 TPS (transaction per second), the data stored by the full node each year is 50 GB, and the TPS can be increased to 1000. At that time, the storage overhead will be as high as 7500 GB. The centralization issue is another challenging issue when blockchain requires enormous computing power as the network size increases. Those nodes with sufficient computing and storage capacities will dominate the system and are likely to become the major parties. This defeats the original merits of a blockchain. Therefore, it is necessary to establish a flexible and scalable blockchain storage mechanism while improving the overall performance.

5) Hype cycle for blockchain technologies. https://www.gartner.com/en/newsroom/press-releases/2019-10-08-gartner-2019-hype-cycle-shows-most-blockchain-technologies-are-still-five-to-10-years-away-from-transformational-impact.

Recently, interoperability has emerged as a considerable obstacle that impedes blockchain deployments [3]. Different blockchain platforms create data islands between different blockchains. For example, different blockchains use different data structures, different consensus protocols, and different security assurance mechanisms. To function effectively, blockchains must make trade-offs between their performance, scalability, and security. The nature of these trade-offs necessitates purpose-built blockchains with specific strengths, limitations, and use-cases in mind. For instance, a blockchain may sacrifice some degree of decentralization or security to achieve a higher TPS; another may sacrifice the performance for a higher level of decentralization and security, vice versa. Therefore, it is urgent to overcome the blockchain interoperability bottleneck. Multiple blockchains demand to cooperate, establish mutual trust, and to realize information sharing and value release.

The current stage of blockchain has shifted from a single chain advancement — concentrates on designing the consensus mechanisms, network protocol, and distributed storage model, to multi-chain collaboration and intercommunication. The main tasks of this new stage involve achieving multi-party collaborative processing, through interdependence to reach efficient and secure consensus, and providing a flexible and extensible underlying storage. It strengthens the synergy of each structure in the ecosystem, thereby improving the capacity, scalability, and safety. This undergoes an entirely different route from the original independent development of the single chain.

4 Recent development of significance

With significant benefits of blockchain for societal impact insight, the global blockchain market size is expected to grow from USD 3.0 billion in 2020 to USD 39.7 billion by 2025, at an impressive compound annual growth rate (CAGR) of 67.3% during 2020-2025[6)]. Currently, there exist 71735 repositories of blockchain on the GitHub development platform, which continues to grow at a rapid pace[7)]. However, the challenges mentioned above potentially raise technical complexities and operational overhead for enterprises to configure suitable blockchain infrastructures and build upper-layer applications. It causes the hype around blockchain to slow down. According to the statistics released by Gartner, a real blockchain-led transformation of the economy and government is still five to ten years away. The tech giants and start-ups actively engage in embracing the actual game-changing opportunities of blockchain by figuring out their strategies.

It is noteworthy that blockchain has now become a platform technology divided into two types (as shown in Figure 3): permissionless blockchains where anyone can participate (i.e., without the need to trust the centralized authorities), and permissioned blockchains where participants need prior approval and authentication. Compared to permissioned blockchains, permissionless blockchains are open and fully transparent with complete freedom. Each user can read and modify the ledger data freely. However, today, growing interest is gaining in permissioned blockchains as it supports an authenticated ecosystem of enterprise-level participants via access control mechanisms and cryptographic privacy protections.

Bitcoin represents the prominent application of Blockchain 1.0 — "stateless: transaction-optimized" digital cryptocurrency era. As an ancestor, Bitcoin starts the boom of the permissionless blockchain family. Many alternative digital currencies (a.k.a. altcoin) have emerged with increasing usage rates due to their low transaction speed and energy-intensive computation. For instance, Litecoin [13] is a commerce substitution of Bitcoin that provides a shorter block generation interval and faster transaction confirmation time. To further expand the application fields, Ethereum opens up a new era of Blockchain 2.0 — "stateful: logic-optimized" smart contracts-enabled value transferring with programmable transactions. Integrated with the Internet of things (IoT) devices, IOTA [14] is the first released project that introduces a DAG (directed acyclic graph) data structure into the underlying storage of blockchain. IOTA can process a couple of transactions concurrently and thus promote overall system performance. Conflux [15] is a DAG-based public blockchain platform that positions itself to provide a scalable and secure protocol for future decentralized applications (DApps). As the first project taking the sharding technique, Zilliqa bears out the possibility to incorporate sharding into the blockchain [16]. However, there still lacks practically applicable permissionless blockchain systems, which should provide high performance and guarantee security at the same time.

6) https://www.marketsandmarkets.com/Market-Reports/blockchain-technology-market-90100890.html.
7) https://github.com/search?q=blockchain.

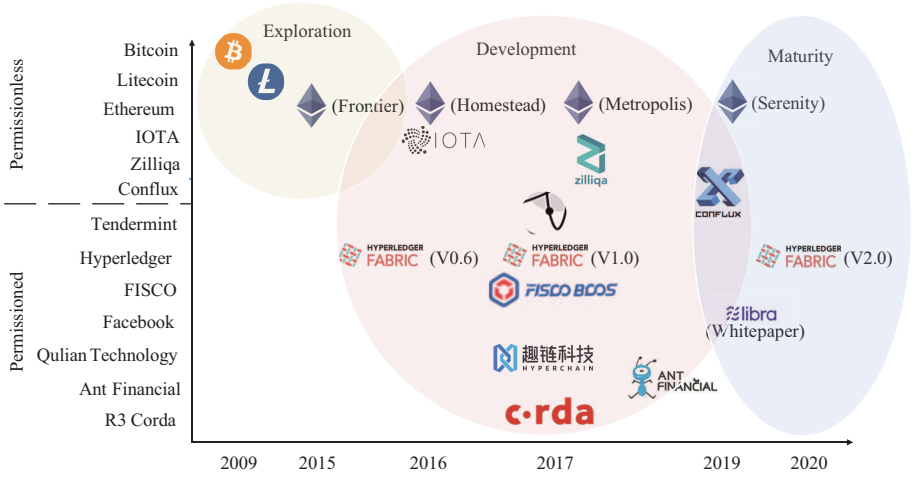

Figure 3 (Color online) Mainstream permissionless and permissioned blockchain platforms.

As a complement to the permissionless blockchain, permissioned blockchain projects promise highly efficient consensus and security among a trusted consortium with enrollment certificates. Almost all the permissioned blockchains strive to provide a more friendly user interface and cost-effective deployment. Specifically, the read and write permissions of pre-authorized entities are controlled by certain restrictions. As a pioneering spotlight project, Hyperledger Fabric can deliver high throughput of 20000 TPS, making it prominent for enterprise use cases. By incorporating the advantageous features of both Ethereum and Hyperledger Fabric, FISCO-BCOS[8)] is launched to develop secure and efficient enterprise-level blockchain solutions. It comprises over 80 financial institutions and FinTech companies, and serves as a pragmatic research platform [17]. R3's Corda[9)] is a competitive consortium platform involving more than 300 worldwide firms to revolutionize multiple industries, such as to build full trade lifecycle solutions for digital asset marketplaces. Hyperchain[10)] makes full use of cloud platforms, which provides a more usable and flexible blockchain service. Ant financial unveiled AntChain [18] to help strengthen transparency and build trust in more than 50 use cases, ranging from cross-border remittance, charitable donations, to agricultural product traceability. It is noteworthy that AntChain allows an average upload rate of over 100 million digital assets per day. The Cosmos[11)] blockchain network is launched with the ultimate goal of creating an Internet of blockchains via IBC (inter-blockchain communication) protocol and the Tendermint BFT protocol. As the star project in 2019, Libra is proposed by Facebook to reshape the existing payment methods.

The transformative potentials of blockchain will also give rise to new platform-level players to gain traction in the new ecosystems. This, in return, will require deep collaboration between technology innovators, government, and regulators.

5 Future directions

Recent blockchain developments have shown that the technology has significant potential of forming a trustworthy cornerstone for the Internet of Value. Blockchain technology is now a vital enabler of the development of digital economies, and the trustworthy social system structure, transitioning from concept to reality. Nevertheless, as new application areas are emerging, more stringent requirements come forth to address certain practical aspects of this technology. Further improvement in performance, scalability, and perfect functions have become the new development goals of blockchain systems. When multiple heterogeneous blockchain systems exist in parallel, they face the core requirements of cross-chain collaboration and data sharing. They urgently need to upgrade the design principles, e.g., storage structure, communication protocols, and consensus algorithms. This section anticipates the potential research trends for the foreseeable future.

8) FISCO BCOS whitepaper. https://github.com/FISCO-BCOS/whitepaper.
9) Corda whitepaper. https://www.corda.net/content/corda-platform-whitepaper.pdf.
10) Hyperchain. https://www.hyperchain.cn.
11) Cosmos network. https://cosmos.network.

Innovation on the data model: from chain-based to chainless. At present, the underlying data storage model of mainstream blockchain systems generally adopts a chain-based structure. That is, several transactions are packaged and stored according to a timestamp sequence in a block as the primary storage unit. Only one block of data can be processed at a time, making it challenging to support complex state operations and dynamic data storage. Meanwhile, existing chain-based blockchain systems adopt consensus protocols to ensure data consistency, such as proof of work and proof of stake, leading to more computing overhead, serious resource consumption, and slow processing speed. High concurrent user access further restricts the performance and scalability of these systems. To solve the dilemma that the processing speed of traditional chain-based blockchain cannot meet the large-scale commercial use, researchers entail a topologically ordered DAG structure that uses fine-grained transactions instead of blocks [19]. Unlike the serialized verification processing mechanism with the chain-based structure, the schematic DAG-based blockchain can process multiple transactions in parallel, and has significant advantages of high concurrent processing efficiency and low verification computation overhead.

Underlying hybrid storage. Notably, the current blockchain employing a full-copy design differs it from the conventional distributed storage that is always assumed with relatively few copies of data, making it unable to accommodate the users with diverse storage capacities. According to the relevance of multi-source blockchain data, it is also desirable to improve query functionalities and support high-performance analytics [20]. A hybrid blockchain storage model can achieve fast and efficient retrieval and dynamic real-time update, forming a novel multi-chain data management framework. New data integration methods and advanced data analytics can ensure the credibility of cross-chain interactions, offer numerous possibilities to rethink, and improve the functional scale of blockchain technology.

Accelerating the blockchain network. Today's blockchain network is based on the P2P protocol. It demands the blockchain network to deal with an explosive amount of transaction requests to meet the performance needs of a wide range of applications. More architectural solutions are on the horizon to improve the capabilities of the blockchain network. Therefore, the blockchain network will be more extensive and tolerate the future Internet. More DApps will be built on the blockchain network by building trust, reducing risk, decreasing costs, and speeding up transactions. Organizations will build applications on the blockchain technology, moving beyond the proof-of-concept stage into widespread adoption as they start to realize the cost savings and enhanced security they can achieve.

Robust governance models for the blockchain network. Robust and efficient regulation is crucial to the future success of the blockchain ecosystem. This can naturally benefit a broad range of industrial and social sectors. Typical scenarios involve the rapidly evolving new financial technology (FinTech) area, where the systemic financial risks can be automatically compliant with blockchain-based regulated sandboxes, pre-emptively solving problems before they arise [17]. In the global outbreak of COVID-19, blockchain plays an essential data surveillance role to facilitate the circulation of traceable anti-epidemic materials circulation and transparent monetary donations, and to build up a reliable infectious disease reporting system to avoid the spread of rumors. By applying artificial intelligence (AI) techniques, we can gain penetrating insights into the blockchain network to enable intelligent authentication and authorization. Essentially, AI-based blockchain network governance models can be used for continuously monitoring, evaluation, and diagnosis. Meanwhile, it is essential to apply cybersecurity technologies for maintaining the necessary adherence to regulatory standards, governance strategies, and process compliance.

Directly interoperable multi-chain. Nowadays, the blockchain ecosystem comprises many blockchains, where each one is developed to cater to the emerging use cases with distinct performance and functionality requirements. Riding such a recent trend in a world of heterogeneous multi-chain that possesses different consensus protocols and different data structures, developing interoperable blockchain mechanisms can reduce the impact of fragmentation, and thus receive much attention from academia and industry. The focus of interoperability is on the development of a cross-chain architecture and interoperability as a service solution, making them more useful, user-friendly, efficient, and scalable [21]. At this stage, the multi-chain interoperability remains open research issues. To be specific, the state-of-the-art highly relies on a designated notary or sidechains to indirectly connect two blockchains. However, these indirect proposals may raise security concerns and degrade the overall efficiency. Therefore, the effective "direct" interconnection between multiple blockchains is strongly favored beyond the existing niche-focused solutions, which allows the data and value to flow more freely throughout the blockchain ecosystem.

Hybrid blockchains. Current multi-chain systems can support the interoperability of homogeneous blockchains, whole architecture and functions are similar to each other. A step toward mainstream adoption of blockchain technology should be a hybrid of heterogeneous multi-chain, performing their own duties along with different responsibilities. Such a hybrid ecosystem lies between public and consortium blockchains and offers the benefits of both. It can thus support the connection of multiple heterogeneous blockchain systems, with more operation types, functional interactions, and diverse demands. The ultimate state should be that retains the beneficial features of each individual blockchain, and enables more types of interoperability, deriving diversified application scenarios and promoting wider cross-chain collaboration.

Building a symbiotic relationship with other advanced IT technologies. At present, blockchain gradually merges with multiple underlying architectures, such as IoT and cloud computing, which serve as the foundation for many innovative applications. These developments fuel a technological push toward innovation and a driving force to move quickly. It stimulates the growth of interests from government authorities, entrepreneurs, and developers to reap the benefits of the convergence of these key technologies involved in digital transformation, including AI, blockchain, edge computing, and IoT. For example, the BaaS (blockchain-as-a-service) platform formed by the integration of blockchain and cloud computing infrastructure provides blockchain services for financial institutions in the economic sector.

6 Conclusion

Blockchain is a new paradigm of distributed computing, which can effectively achieve consistent data storage, tamper-proof traceability, and gradual trust transmission in a multi-entity 'weak trust' environment. It not only inherits the technical essence of traditional distributed systems, but also poses some new challenges that are leaving the significant potential for further innovation. The major bottlenecks include huge storage costs, unsatisfactory performance, and lack of interoperability. The core technology of the blockchain is still on the eve of breakthroughs, that is, five to ten years away from the transformative impact. The future blockchain ecosystem is envisioned to be fueled by chainless data model, underlying hybrid storage, network acceleration, penetrating regulation, and directly interoperable multi-chain.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant No. 62072197).

References

1 Berkley J. The promise of the blockchain: the trust machine. The Economist, 2015
2 Herlihy M. Atomic cross-chain swaps. In: Proceedings of ACM Symposium on Principles of Distributed Computing (PODC), Egham, 2018. 245–254
3 Liu Z T, Xiang Y X, Shi J, et al. Hyperservice: interoperability and programmability across heterogeneous blockchains. In: Proceedings of ACM SIGSAC Conference on Computer and Communications Security (CCS), London, 2019. 549–266
4 Andreas M. Mastering Bitcoin: Unlocking Digital Cryptocurrencies. Sebastopol: O'Reilly Media, 2014
5 Culler D E, Anderson T E, Patterson D A. System Support for Distributed Supercomputing on A Network of Workstations (now). AFRL-IF-RS-TR-1999-226 Final Technical Report, 1999
6 Anderson D P, Cobb J, Korpela E, et al. SETI@home: an experiment in public-resource computing. Commun ACM, 2002, 45: 56–61
7 Foster I, Kesselman C, Nick J M, et al. The physiology of the grid: an open grid services architecture for distributed systems integration. 2002. https://users.cs.northwestern.edu/~srg/Papers/04-25-02/grid.pdf
8 Jin H. ChinaGrid: making grid computing a reality. In: Proceedings of International Conference on Asian Digital Libraries (ICADL), Shanghai, 2004. 13–24
9 Sefraoui O, Aissaoui M, Eleuldj M. Openstack: toward an open-source solution for cloud computing. Int J Comput Appl, 2012, 55: 38–42
10 Androulaki E, Barger A, Bortnikov V, et al. Hyperledger fabric: a distributed operating system for permissioned blockchains. In: Proceedings of the 30th EuroSys Conference, New York, 2018. 1–15
11 Wang W B, Hoang D T, Hu P Z, et al. A survey on consensus mechanisms and mining strategy management in blockchain networks. IEEE Access, 2019, 7: 22328–22370
12 Xiao Y, Zhang N, Lou W J, et al. A survey of distributed consensus protocols for blockchain networks. IEEE Commun Surv Tut, 2020, 22: 1432–1465
13 Andrews A. Litecoin: The Complete Guide to Understanding Litecoin Cryptocurrency and Litecoin Mining. New York: Platinum Press LLC, 2019

14 Alexander R. IOTA-introduction to the tangle technology: everything you need to know about the revolutionary blockchain alternative. 2018

15 Li C X, Li P L, Zhou D, et al. A decentralized blockchain with high throughput and fast confirmation. In: Proceedings of USENIX Annual Technical Conference (ATC), Boston, 2020. 515–528

16 Luu L, Narayanan V, Zheng C D, et al. A secure sharding protocol for open blockchains. In: Proceedings of ACM SIGSAC Conference on Computer and Communications Security (CCS), Vienna, 2016. 17–30

17 Ji Y M, Gu W H, Chen F, et al. SEBF: a single-chain based extension model of blockchain for fintech. In: Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI), Yokohama, 2020. 4497–4504

18 Qi Y, Xiao J. Fintech: AI powers financial services to improve people's lives. Commun ACM, 2018, 61: 65–69

19 Benčić F M, Žarko I P. Distributed ledger technology: blockchain compared to directed acyclic graph. In: Proceedings of the 38th International Conference on Distributed Computing Systems (ICDCS), Vienna, 2018. 1569–1570

20 Dinh T T A, Liu R, Zhang M H, et al. Untangling blockchain: a data processing view of blockchain systems. IEEE Trans Knowl Data Eng, 2018, 30: 1366–1385

21 Jin H, Dai X H, Xiao J. Towards a novel architecture for enabling interoperability amongst multiple blockchains. In: Proceedings of the 38th International Conference on Distributed Computing Systems (ICDCS), Vienna, 2018. 1203–1211