

A Survey of Blockchain-based Schemes for Data Sharing and Exchange

Rui Song, Bin Xiao, *Senior Member, IEEE*, Yubo Song, Songtao Guo, *Senior Member, IEEE*,
and Yuanyuan Yang, *Fellow, IEEE*

Abstract—Data immutability, transparency and decentralization of blockchain make it widely used in various fields, such as Internet of things, finance, energy and healthcare. With the advent of the big data era, various companies and organizations urgently need data from other parties for data analysis and mining to provide better services. Therefore, data sharing and data exchange have become an enormous industry. Traditional centralized data platforms face many problems, such as privacy leakage, high transaction costs and lack of interoperability. Introducing blockchain into this field can address these problems, while providing decentralized data storage and exchange, access control, identity authentication and copyright protection. Although many impressive blockchain-based schemes for data sharing or data exchange scenarios have been presented in recent years, there is still a lack of review and summary of work in this area.

In this paper, we conduct a detailed survey of blockchain-based data sharing and data exchange platforms, discussing the latest technical architectures and research results in this field. In particular, we first survey the current blockchain-based data sharing solutions and provide a detailed analysis of system architecture, access control, interoperability, and security. We then review blockchain-based data exchange systems and data marketplaces, discussing trading process, monetization, copyright protection and other related topics.

Index Terms—Blockchain, big data, data sharing, data exchange.

1 INTRODUCTION

IN the era of big data, data is one of the most valuable assets. Commercial organizations use big data to analyze users' preferences and consumption requirements, and thus provide better services. Government departments also need to collect and analyze data extensively to support decision-making and provide better essential services to the public. Researchers and engineers in academia and industry also rely on big data to build more efficient mathematical models [1]. However, it is difficult for any entity to meet these requirements by relying only on its own data. Therefore, data sharing and exchange are gradually becoming a huge industry [2]. Specifically, *data sharing* refers to the process of sharing digital information between different organizations or individuals, while *data exchange* means the act of trading data commodities between data owners and demanders [3]. Data sharing and exchange between commercial companies, government departments and social organizations can empower them to provide better services, improve efficiency and reap benefits. There are currently many data sharing platforms or marketplaces around the world that focus on data sharing and exchange [4], [5].

In most current practices, centralized data sharing platforms or exchange marketplaces act as authoritative third

parties for transactions, should be trusted unconditionally by all participants in the exchange. However, relying on such a centralized platform can pose many problems. First, the success of transactions depends entirely on the honesty and trustworthiness of these platforms. In other words, participants are completely unable to counteract the potential malicious behaviors of platforms. It is even more difficult to terminate the transactions and recover losses promptly. Second, even if the platforms themselves do not behave maliciously, malicious attacks against them may damage participants' interests, or make them fail to provide services properly. Third, as service providers, these platforms may require participants to pay transaction fees, which is totally reasonable. But for a large number of small transactions, the transaction fee may be even higher than the value of the transaction contents, which could limit the use of these platforms to a great extent [6].

Therefore, some studies have proposed schemes that replace these centralized platforms with blockchain-based systems. *Blockchain* was first introduced as a distributed ledger, and has become popular with the subsequent rise in cryptocurrency platforms such as Bitcoin [7] and Ethereum [8]. The inherent properties of blockchain, such as decentralization, immutability and auditability, make it a natural fit for building secure and robust data-related applications. As a result, blockchain is gradually being applied to other areas besides cryptocurrencies, such as the Internet of things [9], digital finance [10], energy [11] and eHealth [12].

The advantages of blockchain are apparent: first, there is no authoritative trusted third party in the blockchain as a decentralized system. The legitimacy and security of data in the blockchain are guaranteed by the consensus algorithm.

-
- R. Song and B. Xiao are with the Department of Computing, The Hong Kong Polytechnic University, Hong Kong SAR, China.
E-mail: csbxiao@comp.polyu.edu.hk
 - Y. Song is with the School of Cyber Science and Engineering, Southeast University, Nanjing, China.
 - S. Guo is with the College of Computer Science, Chongqing University, Chongqing, China.
 - Y. Yang is with the Department of Computer Science, Stony Brook University, Stony Brook, New York, United States.

Second, the meta-information of all transactions is recorded in the blockchain, which means that the blockchain is an auditable public system, and any participant can audit transactions through the blockchain. Third, blockchain is able to prevent tampering with data. Blockchain formed by hash pointer connections can effectively prevent illegal modification of any confirmed block, since any modification could be easily detected and located.

Nevertheless, implementing a blockchain-based system for data sharing and exchange is still challenging, as the following difficulties need to be addressed. First, the widely used blockchain consensus protocols all take a long time to ensure data consistency among nodes, which leads to the general inability of blockchain-based systems to meet the demand for high-frequency transactions. Second, since all consensus nodes in the network theoretically need to keep a copy of the blockchain locally, storing data directly on the blockchain requires exceptionally high storage and network communication costs. In addition, since data on the blockchain is public, protecting the privacy and security of the stored data is also an important issue [13].

Some previous work has discussed the combination of blockchain and big data. Deepa et al. investigate the application of blockchain to the collection, storage, analysis, and privacy protection of big data [14]. Their survey focuses on the application layer, reviewing the use of blockchain in big data applications in different vertical domains. Xie et al. survey the work of blockchain in the cloud storage and exchange markets [15]. They discuss issues such as security, privacy, and exchange management in the cloud. Berdik et al. provide a comprehensive review of blockchain-based information systems, and give some solutions to problems such as interoperability, system efficiency, fault tolerance and data integrity [16]. However, a systematic review of blockchain-based data sharing and exchange schemes is still missing.

In this paper, the application of blockchain in the area of data sharing and exchange is discussed. This paper selects papers that have received more attention and citations in this area, and provides a comprehensive survey and compilation of their contributions. In summary, the main contributions of this paper are as follows.

- We compile papers on blockchain-based data sharing and exchange schemes since 2015, select papers with significantly higher citations, and systematically analyze these studies.
- We analyze data sharing systems using blockchain technology, investigate their architecture and design principles, and discuss data privacy and security issues.
- We survey existing blockchain-based data marketplaces and data exchange platforms, and discuss exchange process, monetization, copyright, and other related topics of high interest.

The structure of this survey is shown in Figure 1 and organized as follows. Section 2 introduces the background knowledge about blockchain. Section 4 summarizes the current schemes for blockchain-based data sharing systems. Section 5 reviews research on building data marketplaces

and exchange platforms using blockchain. Finally, section 7 concludes this article.

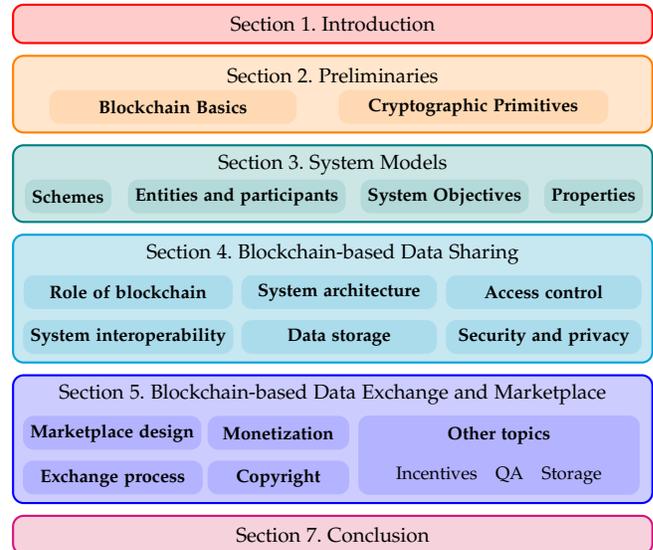


Fig. 1. The structure of this survey.

2 PRELIMINARIES

2.1 Blockchain Basics

2.1.1 Basic concepts and general structure

Blockchain is a technology that originated in the field of cryptocurrencies, which can be traced as far back as the release of Bitcoin in 2009 [7], [17]. Essentially, blockchain is a decentralized ledger that uses peer-to-peer networks for data transfer at the network communication layer [18], [19]. A blockchain system has no centralized server as a *trusted third party* (TTP), and all transactions are verified by specific consensus protocols.

In a blockchain-based cryptocurrency system, transfers between accounts are broadcast across the network in the form of transactions. Precisely, each block consists of a block header and a block body. The block header contains fields such as *block height*, *timestamp*, *hash*, *previous hash*, *Merkle root*, *difficulty*, and *nonce*. On the other hand, the block body consists of a series of transactions organized into a Merkle tree, the root of which is written to the block header [20]. In the blockchain, each block is connected to the previous block by a hash pointer. As new data and transactions are generated, miners in the system pack and generate new blocks, which are then verified by the consensus algorithms and added to the blockchain.

All nodes in a blockchain system keep copies of the blockchain locally. These nodes utilize consensus protocols to synchronize the system state and reach a consensus on newly generated blocks [21]. Tampering with blocks that have been verified and added to the blockchain requires a considerable cost and is very hard to accomplish. This hash pointer structure leads to the fact that trying to tamper with any data in any block requires modifying all subsequent blocks, which is almost computationally impossible [22].

Blockchain is designed to guarantee *data consistency*, *tamper resistance*, *transparency*, *auditability*, and *anonymity*.

To participate in blockchain transactions, users only need to generate a public-private key pair locally. Thus, users are fully capable of generating multiple addresses or even enabling new ones for every transaction, which further enhances the anonymity of blockchain systems.

2.1.2 Types of blockchain

According to the participants involved, blockchains can be classified as public, private, and consortium blockchains [22].

Public blockchains allow any node to join the blockchain's peer-to-peer network without permission from other nodes. All nodes are able to publish transactions, generate and verify blocks freely [8]. Meanwhile, all data in public blockchains can be viewed and verified by all. Most of the cryptocurrencies, such as Bitcoin and Ethereum, use public blockchain as the backbone [7], [8].

Private blockchains, unlike public ones, allow only authorized nodes to join the network. Only authorized entities are allowed to modify and contribute to private blockchains [23]. Although retaining some of the decentralization properties, private blockchains limit nodes to relatively small ranges, which are often managed by the same central authorities. On the other hand, private chains are able to reach consensus at a lower cost and with greater efficiency. Thus they are widely used in business applications [24].

Consortium blockchains allow members of some groups, companies, or collectives to participate in the system [25], [26]. A consortium blockchain is a partially centralized system where only a limited number of selected organizations can modify and view data in the blockchain.

2.1.3 Consensus mechanism

The consensus mechanism is a vital part of a blockchain system, which negotiates and synchronizes the state of nodes by specific algorithms and provides consistency of local copies of the blockchain across nodes. Consensus protocols mainly used by blockchain systems are introduced below.

The *Proof-of-work* (PoW) mechanism is widely used by many cryptocurrencies [7]. In blockchain systems with the PoW mechanism, nodes called miners compete to calculate computational puzzles that can only be computed by brute force. For example, in Bitcoin, miners need to find a nonce that makes the hash value of the block header less than a specific threshold. Anyone who first completes the calculation receives coins as an incentive.

The *Proof-of-stake* (PoS) mechanism has also been used in several systems, such as Peercoin [27]. Ethereum has already migrated from PoW to PoS in 2022. A PoS system needs to select a node to verify and generate the next block in each time slot [21]. The selection is based on the stake held by each node, such as the balance of users in cryptocurrencies.

Implementations of *Byzantine fault tolerance*, such as *Practical Byzantine Fault Tolerance* (PBFT), have also been used in some blockchain systems, such as Hyperledger Fabric [28]. PBFT is able to provide $(n-1)/3$ fault tolerance when there are n nodes in the system, i.e., allowing up to $(n-1)/3$ nodes in the system to be malicious or crashed [29]. PBFT is a replication algorithm that accomplishes consensus by replicating node states in fixed time slots. Since nodes in PBFT systems need to communicate with each other during

consensus, all nodes should know other nodes in the network. Also, the system's performance degrades rapidly as the number of nodes increases [30].

2.1.4 Smart contract

Smart contracts are discussed long before the emergence of blockchain. It is first designed as an automated transaction protocol capable of handling various possible scenarios arising from a transaction based on predefined rules [31], [32]. Blockchain technology has further facilitated the development of smart contracts, as blockchain guarantees mandatory contract execution. Many blockchain systems have already supported smart contracts, such as Ethereum, EOS, and Hyperledger Fabric [8], [28], [33]. Smart contracts abstract the terms agreed upon by the counterparties into executable computer programs. All possible scenarios that may occur in transactions are implemented in the form of control flows in the chain codes. Smart contracts cannot be modified once deployed on the blockchain, and the execution intermediates and results of contracts are also stored in the blockchain [34].

2.2 Cryptographic primitives

We briefly describe the cryptographic primitives that may be covered in the literature surveyed in this paper, to help the reader understand the technical principles of the surveyed schemes.

2.2.1 Identity based encryption

When a sender wants to send an encrypted message to a recipient based on a public-key cryptography scheme, it must first obtain the recipient's public key. It either asks the recipient for its public key certificate or queries the recipient's public key from some public directory. *Identity-based encryption* (IBE) eliminates this reliance on the key exchange. In addition, IBE can also be used to perform searches on encrypted data, as we will see in Section 4.6.

In IBE, any sequence can be used as a public key, such as the recipient's email address. In this case, the sender can send an encrypted message to the receiver as long as it knows the email address. Formally speaking, an identity based encryption scheme is a tuple of four efficient algorithms:

- Setup $(mpk, msk) \stackrel{R}{\leftarrow} S()$: generates a master public key mpk and a master secret key msk ;
- Key generation $sk_{id} \stackrel{R}{\leftarrow} G(msk, id)$: generates a secret key sk_{id} for an identity id ;
- Encryption $c \stackrel{R}{\leftarrow} E(mpk, id, m)$: encrypts the message m using the mpk and identity id ;
- Decryption $m \leftarrow D(sk_{id}, c)$: decrypts the ciphertext c using the secret key sk_{id} .

2.2.2 Attribute based encryption

Similar to IBE, *attribute based encryption* (ABE) also attempts to eliminate the reliance on the key exchange. The difference is that instead of identity, ABE uses the receiver's attributes as encryption and decryption primitives. Strictly speaking, ABE is an encryption scheme defined by predicates. It also contains four efficient algorithms:

- Setup $(mpk, msk) \stackrel{R}{\leftarrow} S()$: generates a master public key mpk and a master secret key msk ;
- Key generation $sk_p \stackrel{R}{\leftarrow} G(msk, p)$: generates a secret key sk_p for a predicate p ;
- Encryption $c \stackrel{R}{\leftarrow} E(mpk, x, m)$: encrypts the message m using the mpk and an attribute x ;
- Decryption $m \leftarrow D(sk_p, c)$: decrypts the ciphertext c using the secret key sk_p . If x satisfies the predicate p , i.e., $p(x) = 1$, the algorithm returns the message m .

3 SYSTEM MODELS

3.1 Data sharing and exchange schemes

Although both involve transferring and sharing data assets among multiple parties, data sharing does not have exactly the same connotation as data exchange. In a typical *data sharing* scheme, one or more individuals or organizations, as data holders, publish data assets to a designated platform [35]. The data requestors, in turn, can request access to some or all of the data assets from the platform after obtaining the appropriate access permissions. It is important to note that in many data sharing schemes, data requestors and data holders may intersect, or even consist of precisely the same set of participants [36]. Essentially, a data sharing scheme aims to efficiently share data assets among multiple parties that are used for each participant's computational or analytical needs to enhance the system's overall interoperability. Therefore, most current data sharing solutions focus on access control, interoperability and data storage of the systems, and make technique selections based on different application scenarios and security requirements [37], [38].

On the other hand, *data exchange* focuses on trading data between two or more counterparties [39]. In the most basic data exchange scenario, the seller of the data holds specific data assets, which is precisely what the buyer needs. The buyer wants to pay a certain amount of currencies or virtual assets in exchange for the raw content or ownership of that data asset. Thus, while access control and data storage still need to be examined, in most data exchange schemes, designers are more concerned with data privacy, copyright, and exchange fairness [40], [41]. In addition, many studies have also evaluated pricing and incentives in the data exchange process related to trading behaviors.

In general, data sharing and exchange schemes share many fundamental building blocks, such as data storage, access control, and privacy preservation. On top of this, data exchange solutions often need to add mechanisms for trading and pricing data assets, as well as the protection of digital copyrights.

3.2 Entities and participants

In most data sharing and data exchange schemes, participants can be abstracted and grouped into the following types of entities:

- *Data providers*: data owners and providers. In data sharing systems, providers provide data with the purpose of sharing data assets with other requesters. While in data exchange systems, the providers' purpose is generally to exchange data for financial benefits.

- *Data requesters*: data consumers. They intend to obtain the data assets they need from data sharing or exchange systems. In data sharing systems, requesters can obtain data for free, as long as they meet specific access control requirements. On the other hand, requesters must pay for data assets in data exchange systems.
- *Blockchain nodes*: generally refer to miners or full nodes of the blockchain. They are responsible for maintaining the regular operation of the blockchain system, including creating and packaging transactions, generating new blocks, network synchronization, and reaching consensus.
- *Platform maintainers*: system maintainers providing data sharing or exchange functions. In different system implementations, the platforms may include different roles, such as *controllers* responsible for access control [42], *re-encryption agents* responsible for data encryption and decryption [43], *queriers* responsible for data indexing and querying, and *data connectors* responsible for data storage and access. For systems using private storage, data connectors store data locally. For systems using cloud or distributed storage, data connectors are responsible for the communication and data transmission between data sharing platforms and the data storage.

3.3 System Objectives

Different implementations and solutions may have different goals and requirements. Examining the form of the data shared or traded, we can classify existing data sharing and exchange schemes into the following types.

- *Raw data packages*: a significant portion of data sharing and exchange systems do not provide data structuring and indexing, but rather provide a simple sharing platform for packets only [44], [45]. In these systems, the data providers upload the raw data they have to the platform, and the requesters request the raw data directly from the platform. Most of this data does not follow a specific schema or have a consistent format.
- *Data queries*: some systems provide a more advanced service by specifying a schema for the data uploaded to the platform and requiring the data provider to provide the data exactly as requested. When data is received, the platform indexes, manages and stores them according to a particular structure. Data requesters can launch queries to the platform according to specific rules, and the platform will return data results that meet the requirements [46].
- *Data interfaces*: instead of single queries, some systems provide data access interfaces. Based on the management and storage of indexed data, these platforms build data access functions into a series of programming interfaces, then provide them as services to data requestors. Data requesters can continuously obtain updated data by interfacing with the API interface of the sharing system [47].

3.4 Properties

From a high-level perspective, most existing data sharing and exchange systems focus on the following core attributes.

- *Privacy*: as data-intensive systems, almost all current data sharing and exchange systems promise to preserve data privacy. However, the specific ways and targets of privacy protection vary for different system implementations and application scenarios. Most systems guarantee that only authenticated participants can access specific data assets by employing data encryption and access control techniques. We will cover this section in more detail in Section 4.6.
- *Integrity*: these systems must provide integrity protection during data storage and transmission. Specifically, they should ensure that data assets cannot be tampered with, or can quickly discover when data tempering happens. Existing schemes generally use digital signatures and cryptographic commitment schemes to protect data integrity during storage and transmission.
- *Availability*: data sharing platforms must provide guarantees of data availability. More specifically, these platforms need to guarantee the data's timeliness, and consistency. In addition, these platforms must ensure that they can recover quickly from system crashes and provide data access functions properly.

On top of this, a significant number of systems promise to provide the following additional attributes.

- *Interoperability*: some systems are able to provide data interoperability. They ensure that data can work together across multiple systems by unifying data formats, communication protocols, and interfaces. On the other hand, some systems attempt to provide a higher level of interoperability by organizing and indexing data to provide useful data to data requesters. We will cover this part in detail in Section 4.4.
- *Response time*: when the volume of data grows with the number of participants, the query and response time for data access has to be taken into account. Some systems reduce query response time by optimizing storage and index structures, while others introduce some cryptographic primitives to improve computational efficiency.
- *Exchange fairness*: in data exchange schemes, exchange fairness should be provided to ensure that neither the data provider (i.e., the seller) nor the data requester (i.e., the buyer) can cheat the other and gain illegal benefits. Precisely, the seller cannot deliver non-conforming data assets to gain benefits, while the buyer must pay for the assets after receiving them.

4 BLOCKCHAIN-BASED DATA SHARING

As big data penetrates various fields, more and more data needs to be appropriately stored, managed, shared and used [48]. On the other hand, the need for privacy protection

essentially limits the sharing and utilization of data, which in turn leads to the emergence of data silos [36]. The introduction of blockchain has effectively alleviated this problem, as blockchain allows multiple participants to guarantee consistency in the content and state of shared data, while not relying on traditional centralized trusted entities [49], [50]. Compared to traditional systems, blockchain-based data sharing systems can provide flexible data interoperability while protecting data privacy, thereby helping individual data holders to make the most of their data assets. In recent years, many researchers have attempted to utilize blockchain for data sharing in healthcare, smart driving, IoT, finance, and other areas [37], [38], [51].

In this section, we will discuss blockchain-based data sharing schemes, including blockchain architecture, sharing scheme design, access control, and data interoperability. In addition, security and data privacy issues of these data sharing schemes will also be considered. Table 1 summarizes influential studies on blockchain-based data sharing schemes since 2015.

4.1 The role of blockchain in data sharing

Essentially, blockchain can be regarded as a stable data intermediary used to replace traditional third-party entities' role in practical applications. In real-world scenarios, data may be stored in databases in different hosts or domains and managed by different protocols. To share data among multiple entities, domains or organizations, intermediaries are inevitably needed for data coordination. Blockchain is ideally suited to be applied in such scenarios [35].

4.1.1 Access control

Access control is a very important issue for data sharing, and blockchain systems can effectively improve this aspect. In traditional systems, access control is usually the responsibility of the data owner. When data queriers request data from the data owner through an intermediary, the data owner uses various identification protocols and permission control mechanisms to authenticate the queriers' identities and privileges, and then decides whether to grant access. Blockchain can provide a more intensive mechanism for data access control, taking over identification and authentication by recording all participants' identities and privileges [65], or supporting fine-grained access control by binding each piece of data with the corresponding authorization information [66].

4.1.2 Data availability

Users often use multiple services provided by different companies or organizations. During this process, a large amount of data is generated, which should be managed by themselves. For example, a patient may have consulted multiple hospitals, so his medical records may be scattered across various databases of different hospitals. To assess his health status, a patient can only get electronic records from hospitals and aggregate them. Blockchain-based systems can effectively avoid this dilemma. Multiple service providers in the same industry could access such a data sharing system, which makes users able to get more comprehensive data through direct access. In addition, such data

TABLE 1
Comparative analysis of blockchain-based data sharing schemes.

Author(s)	Ref.	Year	Application Scenario	Technical Details				Storage		
				Permissioned	Privacy	Access Control	Interoperability	Private	Cloud	Local
Zyskind et al.	[35]	2015	Personal Data	✓	✓	✓			✓	
Peterson et al.	[52]	2016	Healthcare	✓				✓		
Azaria et al.	[53]	2016	Healthcare			✓	✓	✓		✓
Xia et al.	[46], [54]	2017	EMR Sharing	✓		✓	✓		✓	
Shafagh et al.	[44]	2017	IoT	✓	✓	✓			✓	
Dorri et al.	[45]	2017	Smart Vehicle	✓	✓		✓			✓
Gordon et al.	[49]	2018	Healthcare	✓	✓		✓			✓
Zhang et al.	[42]	2018	Healthcare	✓	✓			✓		
Chowdhury et al.	[55]	2018	Personal Data	✓	✓	✓		✓		
Cui et al.	[43]	2018	File Sharing	✓	✓	✓			✓	
Ali et al.	[47]	2018	PingER	✓				✓		
Zhang et al.	[56]	2018	Healthcare			✓	✓			✓
Kang et al.	[57]	2018	Smart Vehicle	✓	✓		✓	✓		
Theodouli et al.	[58]	2018	Healthcare			✓	✓	✓		✓
Sultana et al.	[59]	2020	IoT			✓	✓			✓
Xiao et al.	[60]	2020	Personal Data		✓	✓			✓	
Feng et al.	[61]	2021	5G Drones	✓		✓			✓	
Yu et al.	[62]	2021	IoT	✓		✓	✓			✓
Tan et al.	[63]	2021	COVID-19 EMR	✓	✓	✓				✓
Li et al.	[64]	2022	IoT	✓			✓			✓

can be used as a reference for new service providers. For example, patients can generate reports of their health status from previous records, and provide more comprehensive information to the doctors later. Users can also protect their data privacy and freely choose the way and granularity of data disclosure [67].

4.1.3 Identification and authentication

Blockchain can also help create a unified identification and authentication system. A user may be tagged in different ways in different systems. For example, medical records in different hospitals may use completely different identifiers to refer to the same patient. When multiple healthcare institutes try to exchange cases or medicine data, they have to convert the identifiers to integrate the data for the same patients. Blockchain can effectively solve this problem with a solution inspired by cryptocurrencies. Different data owners can use the public key or public key address as a unique identifier of a user, so that any user can be uniquely identified across multiple systems [68]. Combining with the access control mechanisms, users can access all their relevant data and be confident that no one else can retrieve their private information [69].

4.1.4 Data interoperability

Another advantage of blockchain is its ability to improve data interoperability. Content and service providers are eager to obtain data resources across different domains. In order to achieve data interoperability, it is necessary to first establish data sharing connections between multiple data owners, and then standardize the data formats and transmission protocols. For data interoperability requirements, the traditional approach is for individual data owners to

open data access interfaces and accept data requests. This traditional method has many problems. First, data queriers may need to use offline means to find the required data and facilitate data sharing. In addition, due to the difference in data format and communication standards between different organizations, the owner and the querier need to agree in advance on the format and standard of the data or further process the data afterwards [13]. Blockchain can help circumvent these problems. By connecting entities of the same business to a sharing platform based on blockchain, the expense of establishing data sharing connections can be effectively reduced. By setting up regulations about data formats and communication protocols, blockchain systems can also effectively improve the efficiency and availability of data sharing [70].

4.2 System architecture

To design a blockchain-based data sharing system, one first needs to consider how to choose the proper blockchain architecture. Underlying blockchain could directly determine the interaction mechanism between participants and affect the data flow in the up-layer data sharing systems [13].

4.2.1 Schemes based on private or consortium blockchains

A considerable number of researchers have proposed data sharing systems based on private or consortium blockchains. This system circumvents the inefficiency and high energy consumption of the PoW mechanism, but is only suitable for permissioned blockchains and pre-specified nodes. BBDS and MedShare proposed by Xia et al. tightly integrate the underlying private blockchains with data sharing mechanism in the application layer [46], [54].

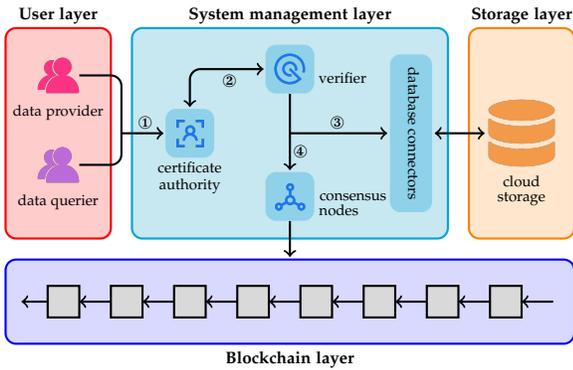


Fig. 2. Typical framework of data sharing systems based on permissioned blockchain.

Shafagh et al. discuss data sharing in IoT scenarios [44]. Their solution uses a permissioned blockchain for data management and access control. To decouple data access from the underlying blockchain, this system introduces a virtual blockchain layer between storage layer and control layer. Zhang et al. proposed a novel multi-blockchain architecture that combines private blockchain and consortium blockchain in their system [56]. The private blockchains in the system are responsible for data storage of each institution, while the consortium blockchain is used to coordinate data sharing between institutions.

Figure 2 summarizes and illustrates a typical framework of permissioned blockchain-based data sharing systems, and abstracts the common structure of the existing work. In a nutshell, in a permissioned blockchain-based data sharing system, the data provider is authenticated through an authority party. The provider’s data is then verified by a verifier and stored in a data storage module via the database connector. Meanwhile, the corresponding metadata will also be stored on the blockchain for query and as a depository evidence. Similarly, when a data requester wants to access a specific chunk of data, it first authenticates its access through an authority party, and then accesses the data through the database connector.

Most of the existing works only give conceptual designs of self-built permissioned blockchains [47], [52]. For self-built permissioned blockchains, a notable problem is that traditional consensus mechanisms such as PoW and PoS cannot be utilized in these systems. This is because these consensus protocols designed for cryptocurrencies often lack efficiency, and require all or most of the nodes in the network to jointly participate in the consensus process [71].

4.2.2 Schemes based on public blockchains or smart contracts

The widespread use of smart contracts has given rise to many data sharing solutions based on them. Among these solutions, the scheme of Azaria et al. is very representative and has inspired several studies in this field [53], [72], [73]. MedRec proposed by Azaria et al. introduces Ethereum smart contracts to record the relationship between patients and medical institutions [53]. These smart contracts associate medical records and cases with access permissions for data queriers. Specifically, two smart contracts are used for

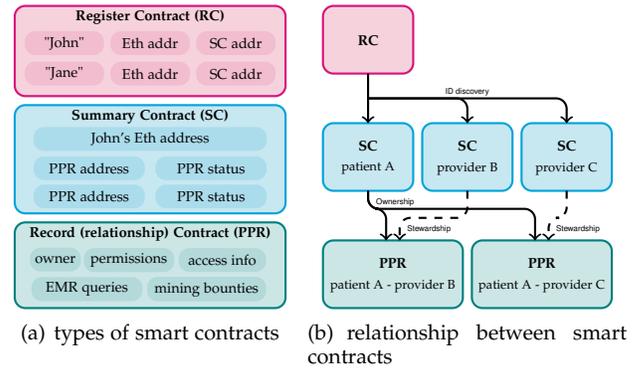


Fig. 3. Smart contracts in MedRec-like data sharing systems [53], [72].

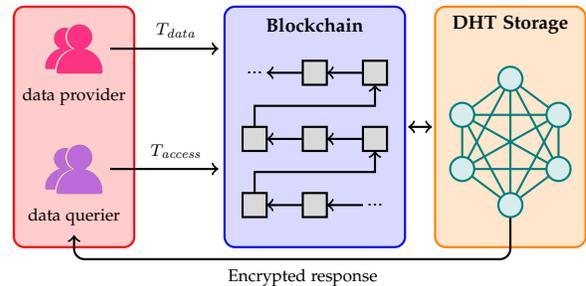


Fig. 4. Transaction-based data access control and permission management [35].

user registration and access control. In addition, another smart contract is designed for operation recording and misbehavior penalization. The structure and relationship of the smart contracts for these schemes are shown in Figure 3. The PrivacyGuard proposed by Xiao et al. introduces *Trusted Execution Environment (TEE)* to execute smart contracts outside the blockchain [60]. PrivacyGuard publishes smart contracts for access control and data management on blockchain, while placing the data storage and computation in trusted environment outside the main blockchain.

The smart contract-based system appoints the transaction and consensus process to the public permissionless blockchains. One only needs to design smart contracts for specific applications and data sharing requirements. However, there are also some problems with smart contract-based systems. the most significant one is the high deployment and invocation costs of smart contracts [74]. To make matters worse, these costs also change rapidly with the dramatically fluctuating prices of cryptocurrencies [75].

4.3 Access control

Traditional techniques for access control use tokens, passwords or keys to authenticate access requests, and grant appropriate read or write permissions to data queriers [76]. Access control schemes such as *Role-based Access Control (RBAC)* and *Attribute-based Access Control (ABAC)* have been discussed by many researchers [77]. These traditional access control approaches lack tracking and auditing of access requests, and cannot control specific entities at a fine-grained level [78].

To address these problems, an early attempt proposed by Zyskind et al. uses blockchain for permission management

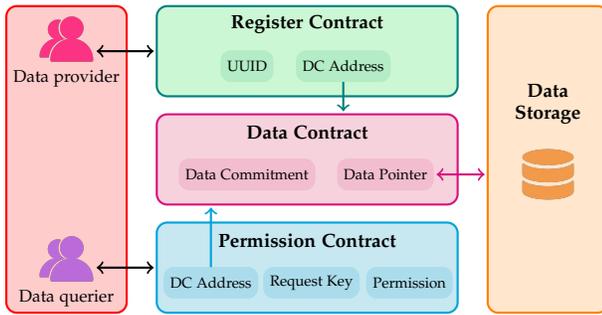


Fig. 5. Smart Contract-based data access control and permission management [53], [58], [59].

and access control in data sharing systems, as shown in Figure 4 [35]. They address the problem that it is difficult to revoke or modify assigned permissions in traditional access control systems, and use blockchain transactions to define access permissions. When a user generates an access request to specific data, the system generates a new access control transaction for the $\langle \text{user}, \text{data} \rangle$ pair and publishes it to the blockchain. Shafagh et al. introduce an access control scheme based on redesigned access control transactions [44]. For all data retrieval requests, the storage nodes first access the blockchain to obtain the access permissions of data queriers, and then fetch appropriate data chunks for queriers. The scheme proposed by Chowdhury et al. combines access control with data encryption, and uses a permission-based blockchain to authorize all access requests [55].

Unlike the above transaction-based access control strategies, Azaria et al. introduce smart contracts for access control [53]. The data pointer and associated access permissions are stored in the contracts, which specify the data access granularity and privileges. Sultana et al. also propose a similar scheme to manage data access permissions through the so-called *Access Control Contract* (ACC) [59]. Figure 5 outlines the structure and execution logic of these smart contract-based access control schemes.

In a nutshell, blockchain-based access control is a good solution for authentication and authorization of access requests in data sharing systems. Compared to traditional RBAC or ABAC paradigms, blockchain-based access control can achieve fine-grained access control requirements more simply [79].

4.4 System interoperability

Interoperability is an essential characteristic for data sharing systems. Interoperability refers to the ability of different organizations, institutions, or data entities to work together [80]. Interoperability consists of two dimensions: syntactic interoperability, which regulates data formats, communication protocols, and software interfaces in data sharing systems; and semantic interoperability, which ensures that data exchange and sharing can provide useful and available data to the participants [81].

There has been much work examining the interoperability of data sharing systems in different industries and different business scenarios. One of the industries that has been focused on is the healthcare industry [82]. Several

countries and organizations are promoting the development of standards related to healthcare data interoperability, such as the *Fast Healthcare Interoperability Resources* (FHIR) for the exchange of *Electronic Health Record* (EHR) [83]. Gordon et al. discuss the contribution of blockchain to interoperability in healthcare data sharing and conceptually discuss the challenges of interoperability in healthcare [13]. Zhang et al. examine the application of blockchain to EHR sharing with FHIR as a unified data format standard [56].

Another area of interest is IoT. BaDS from Zhang et al. discusses data flow and interoperability between IoT data and cloud storage [84]. Similarly, Manzoor et al. discuss the relationship between IoT systems and cloud-based storage in data sharing domain [85]. Unlike the above studies, the system from Sultana et al. focuses on communication and data sharing within IoT devices to achieve interoperability as well as availability of data sharing in IoT networks [59]. the work of Dorri et al. focuses on autonomous driving and proposes blockchain-based data sharing for smart vehicle networking [45]. They specifically analyze the data flow and interoperability between smart vehicles, manufacturers, software providers, cloud storage, and other edge devices in the network. Similarly, Wu et al. also propose a blockchain-based data sharing system for smart cars and connected vehicles. They focus on data sharing and usage between smart cars, edge nodes, and cloud storage nodes [57].

Building blockchain-based data sharing systems in these fields can effectively enhance interoperability in the process of data circulation. A well-designed system can enhance data interoperability without significantly reforming the existing data storage infrastructure. Meanwhile, due to the properties of blockchain, such systems can also provide a good balance between system availability and data security [70], [86].

4.5 Data storage

For the accumulating data volume and growing data sharing needs, data storage architecture and schemes are mandatory considerations when designing blockchain-based data sharing systems. The storage architecture of big data is not only related to the management and usage of data, but also closely related to the availability, interoperability and security of sharing systems [87]. Depending on the characteristics of shared data in different scenarios and the specific data sharing requirements, existing blockchain-based data sharing systems use different storage infrastructures, such as local storage, private storage and cloud storage, as shown in Figure 6.

4.5.1 User local storage

For real-time or sensitive data, users prefer local storage with their own devices and then expose the access interfaces to sharing systems. The smart vehicle data system proposed by Dorri et al. adopts a distributed data storage architecture [45]. Similarly, the data sharing framework proposed by Singh et al. also takes advantage of vehicle local storage [88]. Ali et al. introduce a system for *Ping end-to-end reporting* (PingER) data sharing and storage [47]. It stores the Ping reports in DHT network and only stores metadata such as indexes on the blockchain.

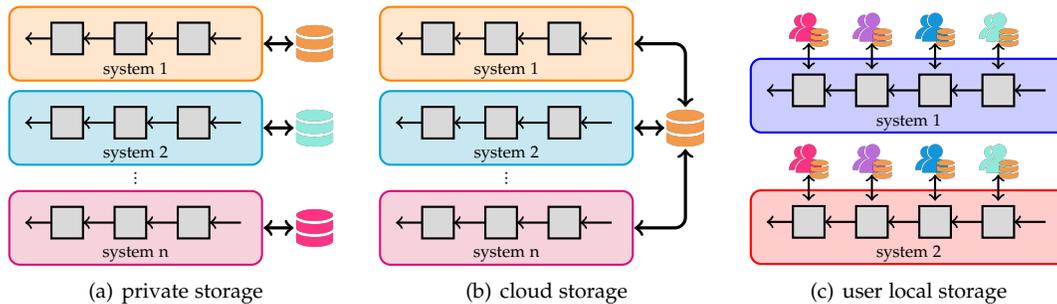


Fig. 6. Data storage schemes commonly used in data sharing systems.

There are not many data sharing systems using local data storage, since appointing data storage to the users themselves can greatly reduce the availability of these systems. On the other hand, additional mechanisms should be introduced to these systems to ensure the integrity and availability of the locally stored data [45], [88].

4.5.2 Private storage

To circumvent the problems from local storage, one can make the data sharing systems take the responsibility of data storage. Specifically, many existing systems rely on centralized databases for private data storage. The medical data sharing system designed by Zhang et al. uses private storage led by hospitals [42]. In this system, each hospital autonomously manages the locally generated medical records and cases. To share these records among multiple hospitals, this system introduces a consortium blockchain for coordination. The MedRec system from Azaria et al. uses a more flexible data storage architecture [53]. Their system contains a database manager that abstracts all access to private databases. The database manager, as a data access interface, can theoretically support various types of databases. Data sharing systems with private storage can achieve a balance between efficiency, security, and availability. In addition, private storage schemes are suitable for existing databases used by data owners [53].

4.5.3 Cloud storage

Data owners prefer to use cloud storage for their accumulating data volumes and large data storage requirements. This is because cloud storage can both reduce the cost of database construction and maintenance, and facilitate data exchange and circulation. The system proposed by Liang et al. collects data from data sources such as wearable devices and medical devices [89]. These data are then synchronized to cloud storage for data sharing with other medical institutions. The BBDS and MedShare systems from Xia et al. both use cloud storage as the underlying storage infrastructure [46]. It can protect data privacy and security by permission group settings and encryption schemes. The IoT data system proposed by Shafagh et al. uses a routing layer to decouple the data storage layer from the blockchain layer [44]. It can also support multiple storage architectures including cloud storage. Similarly, the MedBlock system by Fan et al. can integrate private databases or cloud storage flexibly [90].

4.6 Data security and privacy

Decentralized data storage and blockchain-based data sharing schemes improve the availability and interoperability of shared data. However, inappropriate data storage and management methods can also expose data to the risk of leakage and misuse [91]. Due to the multiple parties involved in the data sharing process, it is challenging to preserve data privacy under complicated application scenarios [73]. In fact, an essential factor that is currently limiting the growth of data sharing in the industry is the need for data security and privacy protection.

4.6.1 Privacy preserving schemes

Several researchers have already examined the privacy-preserving methods in blockchain-based data sharing systems. Currently, the main idea is to use various cryptographic schemes to protect data privacy without compromising data availability.

The simplest and most common way to protect data privacy is to encrypt the original data and have the key managed and distributed by a specific administrator in the system. Combining this with access control and permission management schemes, the administrator can control the granularity and scope of access to specific data by granting and withdrawing keys [43], [54]. However, the use of traditional encryption schemes also poses significant problems that can make it difficult to manage the life cycle of keys. This is because in most cases, keys that have already been distributed are difficult to be revoked, and thus the data have to be re-encrypted to invalidate the invalid keys.

To solve this problem, some researchers have introduced several novel encryption schemes, such as IBE and ABE schemes, as introduced in Section 2.2. BaDS proposed by Zhang et al. uses *Ciphertext-policy Attribute-based Encryption* (CP-ABE) and bilinear pairing to preserve data privacy and provide access control in IoT data sharing scenarios [84]. In this system, all users are assigned keys associated with their attributes. After the data is encrypted, if a user's attributes satisfy a specific predicate, he/she can decrypt the data accordingly. Zheng et al. use the Paillier cryptosystem to protect the privacy of shared data [92]. Their system is able to provide security for data sharing and data exchange systems with cloud storage infrastructures. Also, Zhang et al. use bilinear pairing to perform privacy protection for medical data sharing [42].

In contrast to naive encryption schemes, these pairing-based or attribute-based encryption schemes can circumvent

the system design challenges associated with key distribution and provide more flexible data sharing schemes.

4.6.2 Searchable encryption

While data encryption can certainly protect the data privacy and security, on the other hand, it can also affect the usability and interoperability of the data sharing systems [93]. After processing the original data with cryptography primitives such as encryption or cryptographic commitment, it becomes difficult for data queriers to directly retrieve the desired data by keyword or index querying [94].

One straightforward idea to solve this problem is to introduce encryption algorithms with some homomorphic properties, so that the encrypted data still has some degree of retrievability. For example, Cai et al. propose a searchable encryption scheme for distributed data storage and blockchain-based data sharing [95]. Their scheme combines cryptographic incremental hashing techniques to achieve trusted private keyword retrieval in decentralized storage scenarios. Wang et al. combine ABE with bilinear pairing to implement a searchable encryption scheme based on Ethereum blockchain [66]. Similarly, Zhang et al. and Ma et al. both implement blockchain-based data sharing schemes supporting keyword retrieval using CP-ABE and bilinear pairing [42], [96]. Figure 7 shows a typical sequence diagram of the trusted data sharing scheme based on searchable encryption [97].

5 BLOCKCHAIN-BASED DATA EXCHANGE AND MARKETPLACE

In the era of big data, data is increasingly becoming a valuable asset for businesses, governments and various organizations. Some data holders believe that their data has commercial value and therefore want to disclose the access to it for a fee. Meanwhile, others who demand data are willing to spend money to buy it. This creates a supply and demand relationship for digital commodities.

As a result, there are already many centralized data exchange marketplaces operated by governments or private companies [39]. These marketplaces are similar to real-world trading markets, providing platforms for data exchange and channels for payment [98]. However, centralized data marketplaces may face many copyright and privacy issues. Malicious buyers may resell or directly publish the purchased data on the Internet after acquiring these datasets from the owner [39]. In addition, as centralized platforms or systems, these data marketplaces are also vulnerable to system downtime or malicious attacks [99].

Blockchain-based decentralized data marketplace could solve the above problems. Blockchain is able to record logs of transactions and meta-information of related data. Since blockchain can be tamper-proof in transactions, all exchange participants can audit the trading process and the data flow status at any time afterwards. In addition, the decentralized nature of blockchain is also able to prevent *single points of failure* (SPoF) and malicious attacks against marketplaces.

In this section, we will discuss related work in the area of blockchain-based data exchange and marketplace, including data marketplace design, exchange process, payment and pricing, and dispute resolution. Table 2 summarizes

the schemes and implementations of blockchain-based data exchange platforms and marketplaces since 2017.

5.1 Marketplace design

Similar to blockchain-based data sharing systems, the core of a data exchange marketplace is a blockchain-based distributed network. Data owners can submit their data to the marketplace as commodities for sale, while consumers can go to the marketplace to retrieve or browse data assets. The blockchain-based data exchange marketplace is responsible for managing and coordinating the data delivery and payment process, and provides dispute arbitration and log auditing functions.

The most straightforward idea is to design the marketplace as a simple intermediary for transactions, but not to impose controls on the data assets. Chen et al. propose a solution for a blockchain-based data exchange marketplace [100]. Their system is centered on blockchain and develops exchange specifications, including data formats, transmission protocols and data quality standards through a role called *Congress*. Travizano et al. propose a decentralized data marketplace called *Wibson*, which also consists of buyers, sellers, and a notary responsible for verifying data quality and providing conflict arbitration [106]. Xiong et al.'s system uses Ethereum smart contracts to match desired data of buyers and submitted data from sellers [107]. The focus of all these solutions is to completely eliminate the direct connection between the trading market and the data commodities, ensuring that the marketplace does not intercept any data during the transaction process.

Contrary to this idea, some systems tend to fully control all of the data assets in the marketplaces. When a data owner submits a dataset to the marketplace, he can only specify limited rules such as quotes, data access conditions and used policy. The marketplace will dump the data into a data farm and manage all the meta-information of the data. Buyers can search and purchase data, but can only access it in the specified way provided by the marketplace. Thus, a user cannot freely copy and secondarily distribute the purchased data, which protects the data copyright and privacy. Dai et al. implement a data exchange ecosystem called *SDTE*, which is based on Ethereum and Intel *SGX* [112]. In *SDTE*, the buyer also does not have direct access to the original data, but only to the analysis or processing results of the specific data items. Guan et al. propose a smart contract-based data exchange system that supports trading on both raw data and data statistical results [116]. Further, Hynes et al. propose a decentralized data marketplace called *Sterling* that enables secure trading and utilization of sensitive data using smart contracts [104].

As can be seen, current blockchain-based data exchange systems include both data buyers and sellers as participants. However, some systems may introduce roles such as data agents or proxies, who collect data from data sellers and manage them centrally [100], [101]. These agents can also provide some additional data-based services to buyers, such as statistics, filtering and retrieval of data, etc. Many of the existing systems also include arbitration entities to resolve disputes in case of conflicting exchanges.

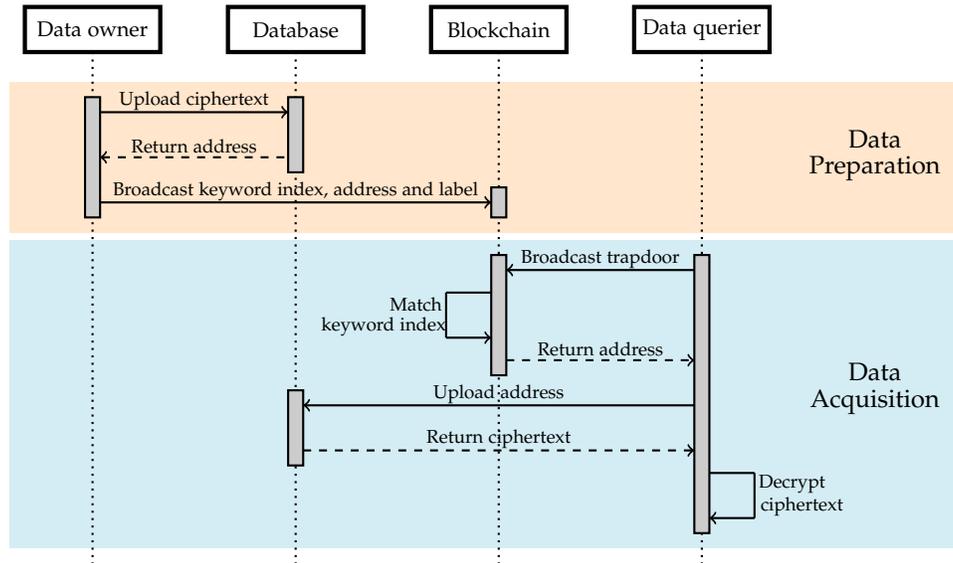


Fig. 7. The sequence diagram of trusted data sharing by Ma et al. [96].

5.2 Data exchange process

Depending on the marketplace architecture, the data exchange process varies from system to system. The differences mainly lie in the way data commodities are published, the transaction flow between buyers and sellers, the responsibility of blockchain in data exchange and the follow-up matters after data exchange, etc.

Chen et al. propose a blockchain and smart contract-based big data marketplace and elaborate the detailed process of data exchange [100]. Their system splits the data exchange process into two parts: dataset distribution and data service provision. First, a data owner can publish bidding documents on the blockchain, and select a data publisher who meets the requirements. The data publisher acquires the datasets from the owner, cleans them, and then provides data services based on these datasets. Smart contracts are introduced to provide data APIs and access constraints for the datasets, and record access logs of the data consumers. Similarly, the system by Nasonov et al. imposes enhanced administrative constraints on the data on exchange [101]. Data owners can submit raw data and quotes to the marketplace, along with buyer access policies. The data marketplace takes over control of all data and stores the data submitted by the owners in its own data farm. When a buyer purchases a data commodity from the marketplace, he can only access the data according to the access policies specified in advance by the data owner. Dai et al.’s SDTE system provides a secure transaction protocol [112]. In their ecosystem, neither the data broker nor the buyer has access to the raw data, but only to the analysis result of the datasets.

Unlike the above systems, Ramachandran’s system provides a more lightweight exchange process, as shown in Figure 8 [41]. In their system, sellers store the product descriptions in a *distributed file storage* (DFS) framework such as IPFS [117] or Storj [118]. Thereafter, sellers publish the stored identifiers to the blockchain for buyers to browse and select. Once a digital good is selected, the buyer and seller

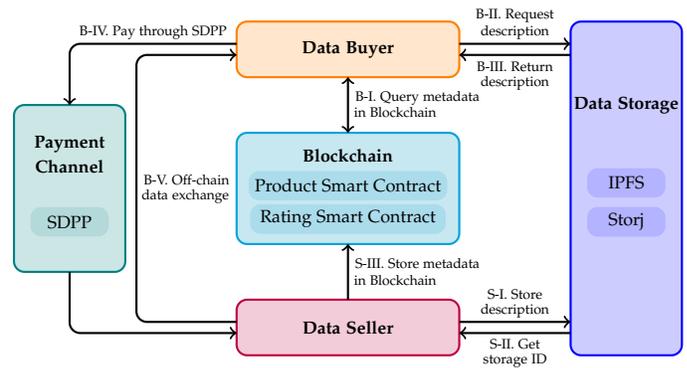


Fig. 8. System flow of a typical decentralized data marketplace [41].

trade the goods off-chain using a *Streaming Data Payment Protocol* (SDPP) [119], and record the transaction on the blockchain. Similarly, Travizano’s Wibson protocol provides a more relaxed data exchange process [106]. In Wibson, buyers and sellers determine the intent to exchange through a series of smart contracts and protocol communications, and use a notary to ensure that the transaction is conducted correctly. Wibson in effect provides a flexible template for data transaction protocols for the participants to follow, and all participants can organize their own transactions.

Looking at the existing data exchange systems and protocols, one can see two different directions of development. A portion of lightweight protocols tend to provide demand matching and payment channels to both data buyers and sellers [41]. These designs tend to use smart contracts to ensure proper execution of the protocols, and intervene in the exchange and handle disputes at the right time [106], [109]. Other relatively heavyweight systems tend to take over all raw data from the sellers and provide buyers with higher quality, more consistent data services through unified data management [104], [112]. Such systems impose stricter constraints and controls on data, and are more suitable

TABLE 2
Comparative analysis of blockchain-based data exchange systems and marketplaces.

Author(s)	Ref.	Year	Application Scenario	Technical Details				Marketplace Design		
				Access Control	Privacy	Fairness	Integrity	Data Quality	Copyright	Monetization
Chen et al.	[100]	2017	Big Data		✓				✓	
Nasonov. et al.	[101]	2018	Big Data	✓	✓		✓			✓
Ozyilmaz et al.	[102]	2018	IoT						✓	✓
Yang et al.	[103]	2018	Smart City Edge Computing		✓			✓		
Hynes et al.	[104]	2018	Healthcare		✓				✓	✓
Ramachandran et al.	[41]	2018	Smart City	✓	✓		✓		✓	
Park et al.	[105]	2018	IoT				✓	✓		
Travizano et al.	[106]	2018	Big Data		✓				✓	
Savelyev et al.	[40]	2018	Digital Assets					✓	✓	
Xiong et al.	[107]	2019	Machine Learning	✓					✓	
Bajoudah et al.	[108]	2019	IoT			✓				
Banerjee et al.	[109]	2019	Big Data		✓	✓				
Chen et al.	[110]	2019	Big Data		✓	✓				
Liu et al.	[111]	2019	IoT						✓	
Dai et al.	[112]	2020	Big Data	✓						✓
Hu et al.	[113]	2021	Big Data						✓	✓
Abdellatif et al.	[114]	2021	Edge Computing Healthcare		✓					✓
Liu et al.	[115]	2022	IoT	✓	✓			✓		

for data commodities involving privacy or copyright issues [100].

5.3 Commodity monetization

Another important topic in data exchange systems is the issue of monetization for commodities. Unlike data sharing systems, in these exchange systems, data sellers sell the data as commodities to buyers for the purpose of receiving direct payment.

5.3.1 Payment methods

Blockchain systems are first used in the cryptocurrency area and are therefore naturally suited as a method for payment and measurement of price. In fact, many blockchain-based data marketplaces and exchange systems directly adopt cryptocurrencies as settlement currencies [108], [110]. On top of this, the introduction of smart contracts enables transaction details and payment processes to be agreed upon beforehand, and is therefore adopted by many exchange systems [105]. However, Almost all cryptocurrencies take tens of seconds or even minutes to package transactions and require an additional period of confirmation. To shorten the transaction time, a protocol called payment channel is proposed and implemented [120]. Once the payment channel is established on the blockchain, it allows both sides of a transaction to quickly complete transactions off-chain multiple times [121].

In the area of data exchanges, the system of Ramachandran et al. introduces a protocol called *Streaming Data Payment Protocol* (SDPP) [41], [119]. SDPP is a payment protocol suitable for real-time data exchange and micro-payments, which uses TCP connections between clients and servers

to transfer data. In addition, SDPP uses a cryptocurrency-based payment channel and a distributed ledger to keep a record of all transactions. Similarly, the IDMoB system by Özyilmaz et al. also uses payment channels to process data exchanges [102].

5.3.2 Commodity pricing strategies

The pricing of digital commodities is also an important issue. Think back to how we price the goods we trade in real life. Usually, sellers set a price for their goods in advance, then put them on the market, and buyers who approve of that price will buy the goods. Consumers could be flexible to choose their purchasing strategies for over or under-pricing to maximize their benefits. For a fully competitive market, the price of a good will eventually reach a reasonable range, and the supply and demand in the market will guide buyers and sellers to adjust their price expectations.

The situation with data exchange is a little different. First, it is difficult for both buyers and sellers to objectively evaluate the value of a specific dataset. Especially in some scenarios, individual data items may seem worthless, but a huge dataset aggregated from small data chunks could contain unlimited business information. Second, unlike traditional physical markets, data marketplaces are not an information-parity trading environment. It is difficult for a buyer to make an objective assessment of the quality and value of a dataset before acquiring it [41].

Therefore, some data exchange schemes also discuss the issue of commodity pricing. the system of Ramachandran et al. proposes a strategy for assessing the quality of data commodities [41]. This strategy is not based on a direct evaluation of specific commodities, but lends an implicit

prediction of the reliability of commodities by evaluation on related sellers. Hynes et al.'s Sterling system draws on techniques related to data valuation to give a scheme for evaluating the value of datasets for machine learning [104]. They use an approximation of the influence functions to circumvent the excessive re-training cost on the blockchain. Liu et al. propose a pricing strategy for blockchain-based data exchange systems [111]. They formulate a two-stage Stackelberg game to provide a pricing strategy for digital goods, and use theories from economics and game theory to analyze the equilibrium of the pricing scheme.

5.4 Digital copyright

As we have mentioned before, digital commodities can be copied and redistributed. In the past, for analog signal media such as video or music tapes, continuous copying could lead to a constant deterioration of data quality. However, similar problems are almost impossible with copies of digital media. Thus, trading data commodities on the Internet directly means losing control over them. And all data exchange platforms or marketplaces are bound to encounter copyright issues [40]. *Digital Rights Management* (DRM) is a widely adopted means of copyright protection, but it is not a perfect solution. Many data and service providers have attacked DRM as significantly degrading the quality of service. This is because DRM increases the complexity of the data commodities, and requires verification of the data's copyright before each user can use it [122].

5.4.1 Copyright management and protection

The use of blockchain for copyright management and traceability is a natural idea since blockchain is naturally traceable. However, the first problem that needs to be solved is how to bind data assets to a commitment or unique ID on the blockchain bidirectionally. Indexing from metadata on the blockchain to a data asset is simple since we can store the URL, cryptographic promise, and signature in a blockchain transaction or smart contract. Reverse binding, however, is a difficult task because it is hard to locate specific metadata on the blockchain by relying only on specific data chunks.

To further clarify this issue, let's take a few examples. Savelyev et al. propose a blockchain-based copyright management scheme [40]. In their design, a cryptographic hash function is used to hash user identity, data meta-information, and timestamp into a unique data identifier. Thus, when an illegal copy of a particular data commodity appears on the network, it can be traced back to the illegal copier by the data identifier. Similarly, a service called *Ascribe.io* generates a unique encrypted ID for all digital goods, which is published to the blockchain to identify the ownership of the data commodities [123]. These cases may seem to solve the copyright problem. However, as long as the attacker modifies the data content slightly, the identifier generated by the system using cryptographic hash functions or other methods could hardly match the data under the chain successfully. Not to mention that the pirate may split, reorganize, transform, and perform a series of operations on the data. After these operations, the data will be wholly unrecognizable and demanding to be located using the identifier [124].

For this reason, some researchers have proposed adding additional data to the original data to provide an index for reverse binding. Using digital watermarking is a natural idea, as done by Meng et al. [125] and Zhao et al. [126]. Mehta et al. use a technique called perceptual hashing to automatically detect image copies and data piracy [127]. Their system claims to be able to detect images after rotation, cropping, and grayscale conversion. In addition, there are copyright protection mechanisms for other mediums of exchange, such as Xiao et al. who study copyright protection for intellectual property [128], Jing et al. for code [129] and Liang et al. for arithmetic circuits based on homomorphic encryption [130]. The logic of all these studies is actually the same at a high level, which is to extract the inherent features of a particular data medium and store these features on the blockchain as metadata and identifiers.

It is easy to see that whether based on watermarking, perceptual hashing, homomorphic encryption or feature analysis, the data mediums targeted by these systems are fixed and one-sided. It is still demanding to find a universal copyright protection scheme targeting any data type and medium.

5.4.2 Control or limitation over raw data

All the measures above focus on tracing and confirming rights after copyright infringement. However, instead of defending copyrights after piracy has occurred, it is better not to directly trade the raw data at the very beginning. Nasonov et al. advocate strong control over all data involved in the marketplace [101]. In fact, their system does not just provide a simple data exchange platform for buyers and sellers, but collects all the data for sale into a centralized data farm and sells the sellers the right to use the data instead of ownership. Similarly, Dai et al.'s SDTE system takes over control of the data and only process or analyze the data according to the buyer's needs [112]. Chen et al. introduce two ideas for protecting copyright [100]. One is the use of APIs to encapsulate the datasets for sale. The second is the use of honey pots to generate a series of artificial tuples to track the flow of data copies for all buyers.

Although there is awareness of the importance of copyright, there is currently no comprehensive discussion or viable solution to this problem in the data exchange area. On the one hand, strong controls over raw data do prevent piracy, but they put more computational pressure on the data marketplaces. The shift in providers of data services from sellers to data marketplaces could place greater requirements on data marketplaces [101]. On the other hand, by adding copyright information to the data traded, one can provide a basis in case of disputes. However, such protection is often used as a remedy for victims after copyrights have already been infringed, but cannot prevent copyright infringement in advance [41].

5.5 Other topics

5.5.1 Incentives and punishments

To maintain a data exchange platform and make participants willing to trade data in the marketplace, reasonable incentives need to be considered. Good incentives not only

help expand marketplaces' user base, but also motivate participants to buy and sell data honestly [102], [131]. Sterling, a trading system for machine learning datasets proposed by Hynes et al. is able to provide fine-grained data payments [104]. Specifically, this system enables the ability to charge for a single prediction and distribute payments among multiple data sellers based on their contribution to the prediction. Data marketplaces may also need to introduce appropriate penalty mechanisms to deter possible malicious behaviors by participants. The buyer may refuse or delay payment after receiving the data commodities, or resell the purchased commodities to a third party for illegal gains. For disagreements caused by payment issues, the platform can use smart contracts to bind buyers' behavior during the transaction and punish dishonest ones [132]. A seller may provide inferior data commodities or datasets which are inconsistent with their descriptions given in advance. In this case, a complaint can be filed by the buyer to the platform to redeem the payment, and additional penalties may be imposed on the dishonest seller [133].

5.5.2 QA and reputation records

It is difficult for consumers to confirm the quality of commodities before receiving them, and this is the same for data exchange. Some sellers may offer data assets that do not meet buyers' requirements, or even provide data containing malicious content to deceive buyers [104].

There are two main ideas to solve this problem. One is to introduce appropriate mechanisms for evaluating the content of data commodities. The evaluation contains several aspects, including the value of the data itself, whether the data meets the buyer's requirement, and whether the data matches the bid price given by the seller [41]. It should be noted that only those marketplaces and data exchange systems that have strong control over the traded data can effectively perform QA towards data commodities. Those platforms that only provide channels for data exchange are unable to manage and review the data, since they do not have direct access to the data itself [101]. Another idea is to track the trading history of all participants and give a rating to the reputation or credit level of all traders [41], [112]. Think back to our experience of buying goods on an e-commerce platform where we have no prior knowledge of the quality and value of the goods, but we can browse other consumers' reviews of the seller and get a side-by-side view of the merchant's creditworthiness. A similar rating system can provide a useful reference for buyers and sellers to help them make decisions in data exchange systems.

5.5.3 Existing commercial platforms

Commercial companies are particularly interested in data analysis in the era of big data, and data exchange can provide a wealth of resources and information for decision-making in various public and private sectors. A number of commercial organizations have already entered this field and have established several data marketplaces or exchange platforms. We briefly introduce a few of them in this subsection, as shown in Table 3.

- **Datum** is a decentralized data storage and exchange system based on Ethereum smart contracts [134]. The

platform leverages BigChainDB and IPFS to implement a scalable, decentralized and fast data storage network [135].

- **GXChain** is a blockchain-based decentralized data exchange platform launched and commercialized in 2017 [136]. It is based on the DPoS consensus mechanism and IPFS distributed storage architecture, and uses WebAssembly to execute smart contracts. GX-Chain also uses technologies such as *secure multi-party computation* (MPC) and *homomorphic encryption* (HE) to ensure privacy during data transmission, storage and exchange.
- **Databroker** is a blockchain-based peer-to-peer data marketplace with a beta release in 2018 and a full release in 2020 [137]. Currently, Databroker has expanded into many areas such as agriculture, human resources, energy, transportation, economy and supply chain. One of its major strengths is its personalized data matching service, which enables users to search for potential data providers from an extensive global network based on users' requirements.
- **Datapace** started as a marketplace for trading IoT sensor data, but has now expanded to other areas [138]. It is based on the Hyperledger Fabric platform, which guarantees the robustness and security of the system through the PBFT consensus protocol, and uses smart contracts to enable a variety of data exchange and processing functionalities.
- **IOTA** is a blockchain-based data sharing and exchange platform for automotive, supply chain, IoT, digital twins and eHealth, etc [5]. IOTA uses a data structure based on *Directed Acyclic Graphs* (DAG) that enables parallelized transactions. Unlike other data exchange platforms or marketplaces, IOTA has built a blockchain system called Tangle to manage and store meta-information about IoT data itself [139].
- **Streamer** is a real-time data streaming exchange platform, also based on the Ethereum platform [4]. Similar to traditional message queues, Streamer transfers data streams from publishers to subscribers, organizing data transfer and transaction services with a subscription/publishing mechanism.

6 CHALLENGES

Blockchain-based data sharing and data exchange systems have extensively promoted the sharing and circulation of data, empowering the data economy by enhancing interoperability between systems. Existing works have discussed the implementation and application of this field in terms of architecture, models, and technologies. However, we noticed that there are still some problems and difficulties to be solved. In this section, we will summarize some of the challenges and open questions that still exist in this field.

6.1 Platform designs

Most of the work is designed for data sharing or exchange systems for specific application scenarios and data models. However, we noticed that in terms of platform design, the following issues are urgently needed in current application

TABLE 3
Comparison between existing commercial blockchain-based data marketplaces.

Platform	Release	Blockchain	Token	Data storage	Remarks
Datum	2017	Ethereum	DAT	BigChainDB/IPFS	incentives mechanism
GXChain	2017	GXChain	GXS & GXC	IPFS/BaaS	WebAssembly execution MPC and HE
Databroker	2018	Ethereum	DTX	local & private storage	data matching
Datapace	2017	Hyperledger Fabric	TAS	local & private storage	Maniflux-based platform for sensor connection
IOTA	2016	Tangle	IOTA	Tangle	DAG for parallelized transactions
Streamer	2019	Ethereum	DATA	local storage & DHT	streaming data exchange publish/subscribe mode

scenarios, and the existing work discussions are still insufficient.

6.1.1 Data schema and interoperability

At its core, the goal of data sharing and exchange is for multiple data owners to share data assets with each other to enable data interoperability and collaboration. When there is a need for data sharing among multiple entities and their business requirements depend on this shared information, interoperability becomes an issue that cannot be ignored. In order to achieve data interoperability, the entities need first to unify the format, schema, or standard of the data. This allows the participants to define the data in a unified, clear and unambiguous form and makes the shared data usable [140]. Data interoperability is critical in all scenarios of data sharing and exchange, as it directly relates to the usability of the data, which is the ultimate goal of data circulation. Currently, there are some organizations in the medical field trying to unify data standards, such as FHIR for exchanging EHRs [46], [54]. However, unified data templates are still missing in other fields that urgently need data circulation, such as IoT and smart cars.

6.1.2 Data indexing and supply-and-demand matching

Whether it is data sharing or data exchange, matching supply and demand is a very important requirement. In order to obtain the data resources they need, data requesters usually need to initiate a search on the platform, and the platform will return possible query results. However, considering the data privacy and security requirements, in most systems, the data is encrypted and stored. This brings additional difficulties to data retrieval. There are two approaches to solve this problem. One is to store metadata in plaintext in the system to build an index for retrieval and query. The other is to introduce searchable encryption technology to search on ciphertext. Using metadata can indeed efficiently build an index, but there is still a risk of data privacy leakage. Using searchable encryption will greatly reduce the performance of the system, as we will introduce in Section 6.3.

6.1.3 Reward and punishment mechanism

Since data exchange systems involve trading behavior, reasonable pricing strategies and reward and punishment mechanisms are necessary. For an actual marketplace, a practical trading rule design and reward and punishment system can make the mechanism work appropriately in the

long run and ensure that each participant tends to remain honest [133]. This section requires an exhaustive game evaluation and simulation to find each participant's optimal incentive and penalty mechanism and build the upper-level trading rules based on this mechanism. We present some work in this area in Sections 5.3 and 5.5, but the work in this area is still inadequate, and there is much room for improvement.

6.2 Copyright and traceability

Unlike physical goods, the nature of digital assets makes them easy to be copied and re-distributed. For existing data sharing or data exchange systems, the system platform will protect data privacy and copyright through encryption, authentication, and access control during the data circulation process. Even so, for all existing mechanisms, when the sharing or exchange is completed, the original data owners will lose control of the data [128]. The data requester can make arbitrary modifications, copies, and re-distributions of the data. This will significantly undermine the long-term operation of data sharing and exchange platforms. In this case, the holder of the data will not have sufficient motivation to share or sell their data on a public platform, because when the requester receives the data, the original owner will effectively lose all ownership of the data, even if it is only intended to share the right to use the data [123].

On the other hand, the unstable nature of data assets also poses difficulties for the requester. Data requesters, especially data buyers in data exchange systems, need to measure whether the data is worth buying. Unlike physical goods, buyers need not only to examine whether the data content meets the requirements, but often also to examine the data provider and trace the provenance and transformation process of the data. All this information together constitutes the criteria for measuring the value of the data [124]. However, since data can be easily copied, modified, and re-distributed, tracing the source of the data becomes almost impossible, which poses significant difficulties for data exchange.

6.3 Data verification and predicate

For data exchange schemes, most of the exchange actions are based on the description given by the data owner or platform. To ensure the exchange fairness and data privacy, the buyer must be able to have some method to verify the validity and correctness of the data asset. In actual

system design, one of the most common ideas is to provide the buyer with a specific verification method or logic in advance. This creates a data privacy risk, as the disclosure of the validation logic is essentially the same as revealing some information about the data to the buyer. Ensuring that the description of a digital asset matches the asset itself is, therefore, a fundamental challenge in this area. Think back to the experience of buying things in our daily lives. The seller usually shows some features of the product or provides some information describing the product. However, for data exchanges, these existing experiences cannot be applied. This is because all of the above descriptions of data assets provided by the seller reveal information about the data, compromising the privacy and fairness of exchanges. As a result, the current data exchange scheme assumes that buyers have access to validation logic, called predicates, for the exchanged data before the transaction. However, there are several problems with this model.

First, predicates are essentially used to prove the legitimacy and validity of the data assets to the buyer. In other words, the purpose of the predicate is to convince the buyer that this data is indeed what the buyer needs and that it is correct. However, it is difficult to satisfy this need by relying on predicates alone, because predicates can only prove that the data is correct, but hardly prove that the data meets the buyer's needs. Second, all current data trading schemes assume that the buyer has the predicate that matches the data asset before the transaction begins, i.e., the predicate is treated as a priori knowledge of both parties to the exchange. However, this is far from reality. In the real world, if the predicate is relied upon to verify the validity of the data, then the predicate itself needs to be generated and distributed by a rigorous protocol. If the predicate is used to guarantee the correctness of the data, then who guarantees the correctness of the predicate? In turn, responding to this question may require reliance on some other trusted entity or environment, which in turn will undermine the value of the blockchain in the scheme.

6.4 System Performance

For data-intensive systems, efficiency is a very important metric. However, existing solutions face many performance bottlenecks. To provide retrieval and indexing of data with guaranteed data privacy, many schemes use cryptographic primitives such as searchable encryption or utilize algorithms such as Attribute-Based Encryption (ABE) or Identity-Based Encryption (IBE) for access control [42], [96]. The problem is that most of these primitives are based on pairing algorithms which have considerable overhead. To query large-scale datasets, existing pairing-based algorithms consume tens of seconds. This bottleneck will greatly slow down the response efficiency of services when facing concurrent queries. In addition, some algorithms try to use same-station encryption or secure multi-party computation for data query and aggregation, and the computational overhead of these cryptographic algorithms is even higher [130]. Therefore, reducing the computational overhead of cryptographic algorithms while ensuring the quality of data services is a fundamental challenge for existing solutions.

In data transaction scenarios, existing schemes often use zero-knowledge-based proofs to provide verification of

encrypted data in order to ensure the fairness of transactions. The problem is that concise zero-knowledge proof mechanisms that can be applied to any generic verification logic do not yet exist. Existing zkSNARKs-based Zero-Knowledge Contingent Payment (ZKCP) schemes require a trusted initialization environment, which is more challenging to implement in systems that rely only on the blockchain [141]. In contrast, zkSTARKs schemes that do not require a trusted environment require far more zero-knowledge proof operations and generate far more proof files than zkSNARKs schemes, thus resulting in a relatively long time for both generating and verifying proofs. On the other hand, the Merkle Proof-based mechanism that does not rely on the zero-knowledge system is also less efficient for transactions with larger files or complex verification logic. The problem is that the proof size in such protocols is proportional to the transaction's data size, regardless of whether malicious behavior occurs. In addition, the protocol interaction requires a fixed number of rounds regardless of whether the seller files a complaint, which is extremely wasteful for the optimistic cases that account for the vast majority of transactions. Thus, for fair data transactions, there is still much room for improvement in algorithm design.

7 CONCLUSION

Blockchain provides an effective solution to the problems in data sharing and data exchange processes. However, there is currently no detailed survey on applications of blockchain in this field. To fill this gap, this paper reviews the design of blockchain-based data sharing and data exchange systems, including topics such as system architecture, data transfer process, access control, interoperability, non-fungible token, and data monetization. We believe that this paper can provide comprehensive knowledge about blockchain-based data systems for commercial companies, government departments, and researchers whose business involves big data interaction and analysis.

For the design of data sharing platforms, it is necessary to combine data indexing, access control and connectivity to provide sufficient system interoperability. As for data exchange systems, apart from the above issues, reasonable monetization rules and incentive systems should also be considered. These mechanisms can ensure the long-term stability of these systems and promote all participants to behave honestly. Designing a practical data exchange system is a challenging task which requires detailed game evaluation and simulation to find a balanced reward and incentive scheme for each participant, and then building the upper-level application.

ACKNOWLEDGMENTS

This work was partially supported by the NSFC/RGC Joint Research Scheme N_PolyU529/22 and the HK RGC GRF PolyU No. 15209822

REFERENCES

- [1] S. Sagioglu and D. Sinanc, "Big data: A review," in *2013 international conference on collaboration technologies and systems (CTS)*. IEEE, 2013, pp. 42–47.

- [2] A. Labrinidis and H. V. Jagadish, "Challenges and opportunities with big data," *Proceedings of the VLDB Endowment*, vol. 5, no. 12, pp. 2032–2033, 2012.
- [3] D. L. Heymann, "Data sharing and outbreaks: best practice exemplified," *The Lancet*, vol. 395, no. 10223, pp. 469–470, 2020.
- [4] Streamer. (2021) Streamer. [Online]. Available: <https://streamr.network>.
- [5] M. Divya and N. B. Biradar, "Iota-next generation block chain," *Int. J. Eng. Comput. Sci.*, vol. 7, no. 04, pp. 23 823–23 826, 2018.
- [6] F. Milanovic, D. Pontille, and A. Cambon, "Biobanking and data sharing: a plurality of exchange regimes," *Genomics, Society and Policy*, vol. 3, no. 1, pp. 1–14, 2007.
- [7] S. Nakamoto. (2009) Bitcoin: A peer-to-peer electronic cash system. [Online]. Available: <https://bitcoin.org/bitcoin.pdf>
- [8] G. Wood, "Ethereum: A secure decentralised generalised transaction ledger," *Ethereum project yellow paper*, vol. 151, pp. 1–32, 2014.
- [9] A. Dorri, S. S. Kanhere, and R. Jurdak, "Towards an optimized blockchain for iot," in *2017 IEEE/ACM Second International Conference on Internet-of-Things Design and Implementation (IoTDI)*. IEEE, 2017, pp. 173–178.
- [10] P. Treleaven, R. G. Brown, and D. Yang, "Blockchain technology in finance," *Computer*, vol. 50, no. 9, pp. 14–17, 2017.
- [11] M. Andoni, V. Robu, D. Flynn, S. Abram, D. Geach, D. Jenkins, P. McCallum, and A. Peacock, "Blockchain technology in the energy sector: A systematic review of challenges and opportunities," *Renewable and Sustainable Energy Reviews*, vol. 100, pp. 143–174, 2019.
- [12] T. Kuo, H. Kim, and L. Ohno, "Blockchain distributed ledger technologies for biomedical and health care applications," *Journal of the American Medical Informatics Association*, vol. 24, no. 6, pp. 1211–1220, 2017.
- [13] W. J. Gordon and C. Catalini, "Blockchain technology for health-care: facilitating the transition to patient-driven interoperability," *Computational and structural biotechnology journal*, vol. 16, pp. 224–230, 2018.
- [14] N. Deepa, Q. Pham, D. C. Nguyen, S. Bhattacharya, B. Prabadevi, T. R. Gadekallu, P. K. R. Maddikunta, F. Fang, and P. N. Pathirana, "A survey on blockchain for big data: Approaches, opportunities, and future directions," *arXiv preprint arXiv:2009.00858*, 2020.
- [15] S. Xie, Z. Zheng, W. Chen, J. Wu, H. Dai, and M. Imran, "Blockchain for cloud exchange: A survey," *Computers & Electrical Engineering*, vol. 81, p. 106526, 2020.
- [16] D. Berdik, S. Otoum, N. Schmidt, D. Porter, and Y. Jararweh, "A survey on blockchain for information systems management and security," *Information Processing & Management*, vol. 58, no. 1, p. 102397, 2021.
- [17] U. Mukhopadhyay, A. Skjellum, O. Hambolu, J. Oakley, L. Yu, and R. Brooks, "A brief survey of cryptocurrency systems," in *2016 14th annual conference on privacy, security and trust (PST)*. IEEE, 2016, pp. 745–752.
- [18] N. Rifi, E. Rachkidi, N. Agoulmine, and N. C. Taher, "Towards using blockchain technology for ehealth data access management," in *2017 fourth international conference on advances in biomedical engineering (ICABME)*. IEEE, 2017, pp. 1–4.
- [19] R. Schollmeier, "A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications," in *Proceedings First International Conference on Peer-to-Peer Computing*. IEEE, 2001, pp. 101–102.
- [20] R. C. Merkle, "A certified digital signature," in *Conference on the Theory and Application of Cryptology*. Springer, 1989, pp. 218–238.
- [21] Y. Xiao, N. Zhang, W. Lou, and Y. T. Hou, "A survey of distributed consensus protocols for blockchain networks," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 2, pp. 1432–1465, 2020.
- [22] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "An overview of blockchain technology: Architecture, consensus, and future trends," in *2017 IEEE international congress on big data (BigData congress)*. IEEE, 2017, pp. 557–564.
- [23] S. Rouhani and R. Deters, "Performance analysis of ethereum transactions in private blockchain," in *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*. IEEE, 2017, pp. 70–74.
- [24] S. Pongnumkul, C. Siripanpornchana, and S. Thajchayapong, "Performance analysis of private blockchain platforms in varying workloads," in *2017 26th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2017, pp. 1–6.
- [25] J. Kang, R. Yu, X. Huang, S. Maharjan, Y. Zhang, and E. Hossain, "Enabling localized peer-to-peer electricity trading among plug-in hybrid electric vehicles using consortium blockchains," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 6, pp. 3154–3164, 2017.
- [26] Z. Li, J. Kang, R. Yu, D. Ye, Q. Deng, and Y. Zhang, "Consortium blockchain for secure energy trading in industrial internet of things," *IEEE transactions on industrial informatics*, vol. 14, no. 8, pp. 3690–3700, 2017.
- [27] S. King and S. Nadal, "Ppcoin: Peer-to-peer crypto-currency with proof-of-stake," *self-published paper*, August, vol. 19, p. 1, 2012.
- [28] E. Androulaki, A. Barger, V. Bortnikov, C. Cachin, K. Christidis, A. De Caro, D. Enyeart, C. Ferris, G. Laventman, and Y. Manevich, "Hyperledger fabric: a distributed operating system for permissioned blockchains," in *Proceedings of the thirteenth EuroSys conference*, 2018, pp. 1–15.
- [29] M. Castro and B. Liskov, "Practical byzantine fault tolerance," in *OSDI*, vol. 99, no. 1999, 1999, pp. 173–186.
- [30] M. Vukolić, "The quest for scalable blockchain fabric: Proof-of-work vs. bft replication," in *International workshop on open problems in network security*. Springer, 2015, pp. 112–125.
- [31] N. Szabo. (1996) Smart contracts: Building blocks for digital markets. [Online]. Available: https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/smart_contracts_2.html
- [32] —. (1997) The idea of smart contracts. [Online]. Available: <https://www.fon.hum.uva.nl/rob/Courses/InformationInSpeech/CDROM/Literature/LOTwinterschool2006/szabo.best.vwh.net/idea.html>
- [33] E. IO. (2017) Eos.io technical white paper v2. [Online]. Available: <https://github.com/EOSIO/Documentation>
- [34] N. Atzei, M. Bartoletti, and T. Cimoli, "A survey of attacks on ethereum smart contracts (sok)," in *International conference on principles of security and trust*. Springer, 2017, pp. 164–186.
- [35] G. Zyskind and O. Nathan, "Decentralizing privacy: Using blockchain to protect personal data," in *2015 IEEE Security and Privacy Workshops*. IEEE, 2015, pp. 180–184.
- [36] C. Xu, K. Wang, P. Li, S. Guo, J. Luo, B. Ye, and M. Guo, "Making big data open in edges: A resource-efficient blockchain-based approach," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 4, pp. 870–882, 2018.
- [37] A. Reyna, C. Martín, J. Chen, E. Soler, and M. Díaz, "On blockchain and its integration with iot. challenges and opportunities," *Future generation computer systems*, vol. 88, pp. 173–190, 2018.
- [38] F. Schär, "Decentralized finance: On blockchain-and smart contract-based financial markets," *FRB of St. Louis Review*, 2021.
- [39] F. Stahl, F. Schomm, and G. Vossen, "The data marketplace survey revisited," ERCIS Working Paper, Tech. Rep., 2014.
- [40] A. Savelyev, "Copyright in the blockchain era: Promises and challenges," *Computer law & security review*, vol. 34, no. 3, pp. 550–561, 2018.
- [41] G. S. Ramachandran, R. Radhakrishnan, and B. Krishnamachari, "Towards a decentralized data marketplace for smart cities," in *2018 IEEE International Smart Cities Conference (ISC2)*. IEEE, 2018, pp. 1–8.
- [42] A. Zhang and X. Lin, "Towards secure and privacy-preserving data sharing in e-health systems via consortium blockchain," *Journal of medical systems*, vol. 42, no. 8, pp. 1–18, 2018.
- [43] S. Cui, M. R. Asghar, and G. Russello, "Towards blockchain-based scalable and trustworthy file sharing," in *2018 27th International Conference on Computer Communication and Networks (ICCCN)*. IEEE, 2018, pp. 1–2.
- [44] H. Shafagh, L. Burkhalter, A. Hithnawi, and S. Duquennoy, "Towards blockchain-based auditable storage and sharing of iot data," in *Proceedings of the 2017 on Cloud Computing Security Workshop*, 2017, pp. 45–50.
- [45] A. Dorri, M. Steger, S. S. Kanhere, and R. Jurdak, "Blockchain: A distributed solution to automotive security and privacy," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 119–125, 2017.
- [46] Q. Xia, E. B. Sifah, K. O. Asamoah, J. Gao, X. Du, and M. Guizani, "Medshare: Trust-less medical data sharing among cloud service providers via blockchain," *IEEE Access*, vol. 5, pp. 14 757–14 767, 2017.
- [47] S. Ali, G. Wang, B. White, and R. L. Cottrell, "A blockchain-based decentralized data storage and access framework for pinger," in *2018 17th IEEE International Conference on Trust, Se-*

- curity and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE). IEEE, 2018, pp. 1303–1308.
- [48] E. Karafiloski and A. Mishev, “Blockchain solutions for big data challenges: A literature review,” in *IEEE EUROCON 2017-17th International Conference on Smart Technologies*. IEEE, 2017, pp. 763–768.
- [49] W. Gordon, A. Wright, and A. Landman, “Blockchain in health care: Decoding the hype,” *NEJM Catalyst*, vol. 3, no. 1, 2017.
- [50] D. J. Skiba, “The potential of blockchain in education and health care,” *Nursing education perspectives*, vol. 38, no. 4, pp. 220–221, 2017.
- [51] U. Javaid, M. N. Aman, and B. Sikdar, “Drivman: Driving trust management and data sharing in vanets with blockchain and smart contracts,” in *2019 IEEE 89th Vehicular Technology Conference (VTC2019-Spring)*. IEEE, 2019, pp. 1–5.
- [52] K. Peterson, R. Deeduvanu, P. Kanjamala, and K. Boles, “A blockchain-based approach to health information exchange networks,” in *Proc. NIST Workshop Blockchain Healthcare*, vol. 1, no. 1, 2016, pp. 1–10.
- [53] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, “Medrec: Using blockchain for medical data access and permission management,” in *2016 2nd International Conference on Open and Big Data (OBD)*. IEEE, 2016, pp. 25–30.
- [54] Q. Xia, E. B. Sifah, A. Smahi, S. Amofa, and X. Zhang, “Bbds: Blockchain-based data sharing for electronic medical records in cloud environments,” *Information*, vol. 8, no. 2, p. 44, 2017.
- [55] M. J. M. Chowdhury, A. Colman, M. A. Kabir, J. Han, and P. Sarda, “Blockchain as a notarization service for data sharing with personal data store,” in *2018 17th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/12th IEEE International Conference on Big Data Science and Engineering (TrustCom/BigDataSE)*. IEEE, 2018, pp. 1330–1335.
- [56] P. Zhang, J. White, D. C. Schmidt, G. Lenz, and S. T. Rosenbloom, “Fhircchain: applying blockchain to securely and scalably share clinical data,” *Computational and structural biotechnology journal*, vol. 16, pp. 267–278, 2018.
- [57] J. Kang, R. Yu, X. Huang, M. Wu, S. Maharjan, S. Xie, and Y. Zhang, “Blockchain for secure and efficient data sharing in vehicular edge computing and networks,” *IEEE Internet of Things Journal*, vol. 6, no. 3, pp. 4660–4670, 2018.
- [58] A. Theodouli, S. Arakliotis, K. Moschou, K. Votis, and D. Tzovarvas, “On the design of a blockchain-based system to facilitate healthcare data sharing,” in *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)*. IEEE, 2018, pp. 1374–1379.
- [59] T. Sultana, A. Almogren, M. Akbar, M. Zuair, I. Ullah, and N. Javaid, “Data sharing system integrating access control mechanism using blockchain-based smart contracts for iot devices,” *Applied Sciences*, vol. 10, no. 2, p. 488, 2020.
- [60] Y. Xiao, N. Zhang, J. Li, W. Lou, and Y. T. Hou, “Privacyguard: Enforcing private data usage control with blockchain and attested off-chain contract execution,” in *European Symposium on Research in Computer Security*. Springer, 2020, pp. 610–629.
- [61] C. Feng, K. Yu, A. K. Bashir, Y. D. Al, Y. Lu, S. Chen, and D. Zhang, “Efficient and secure data sharing for 5g flying drones: a blockchain-enabled approach,” *IEEE Network*, vol. 35, no. 1, pp. 130–137, 2021.
- [62] K. Yu, L. Tan, M. Aloqaily, H. Yang, and Y. Jararweh, “Blockchain-enhanced data sharing with traceable and direct revocation in iiot,” *IEEE transactions on industrial informatics*, 2021.
- [63] L. Tan, K. Yu, N. Shi, C. Yang, W. Wei, and H. Lu, “Towards secure and privacy-preserving data sharing for covid-19 medical records: a blockchain-empowered approach,” *IEEE Transactions on Network Science and Engineering*, vol. 9, no. 1, pp. 271–281, 2021.
- [64] T. Li, H. Wang, D. He, and J. Yu, “Blockchain-based privacy-preserving and rewarding private data sharing for iot,” *IEEE Internet of Things Journal*, 2022.
- [65] D. D. F. Maesa, P. Mori, and L. Ricci, “Blockchain based access control,” in *IFIP international conference on distributed applications and interoperable systems*. Springer, 2017, pp. 206–220.
- [66] S. Wang, Y. Zhang, and Y. Zhang, “A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems,” *IEEE Access*, vol. 6, pp. 38 437–38 450, 2018.
- [67] H. Tian, J. He, and Y. Ding, “Medical data management on blockchain with privacy,” *Journal of medical systems*, vol. 43, no. 2, p. 26, 2019.
- [68] A. Mohsin, A. Zaidan, B. Zaidan, O. S. Albahri, A. S. Albahri, M. Alsalem, and K. Mohammed, “Blockchain authentication of network applications: Taxonomy, classification, capabilities, open challenges, motivations, recommendations and future directions,” *Computer Standards & Interfaces*, vol. 64, pp. 41–60, 2019.
- [69] M. T. Hammi, B. Hammi, P. Bellot, and A. Serhrouchni, “Bubbles of trust: A decentralized blockchain-based authentication system for iot,” *Computers & Security*, vol. 78, pp. 126–142, 2018.
- [70] C. Brodersen, B. Kalis, C. Leong, E. Mitchell, E. Pupo, A. Truscott, and L. Accenture, “Blockchain: Securing a new health interoperability experience,” *Accenture LLP*, pp. 1–11, 2016.
- [71] F. Casino, T. K. Dasaklis, and C. Patsakis, “A systematic literature review of blockchain-based applications: current status, classification and open issues,” *Telematics and informatics*, vol. 36, pp. 55–81, 2019.
- [72] A. Ekblaw, A. Azaria, J. D. Halamka, and A. Lippman, “A case study for blockchain in healthcare: A prototype for electronic health records and medical research data,” in *Proceedings of IEEE open & big data conference*, vol. 13, 2016, p. 13.
- [73] C. Esposito, A. De Santis, G. Tortora, H. Chang, and K. R. Choo, “Blockchain: A panacea for healthcare cloud-based data security and privacy?” *IEEE Cloud Computing*, vol. 5, no. 1, pp. 31–37, 2018.
- [74] A. Khatoun, “A blockchain-based smart contract system for healthcare management,” *Electronics*, vol. 9, no. 1, p. 94, 2020.
- [75] S. Xuan, L. Zheng, I. Chung, W. Wang, D. Man, X. Du, W. Yang, and M. Guizani, “An incentive mechanism for data sharing based on blockchain with smart contracts,” *Computers & Electrical Engineering*, vol. 83, p. 106587, 2020.
- [76] P. Samarati and S. C. de Vimercati, “Access control: Policies, models, and mechanisms,” in *International School on Foundations of Security Analysis and Design*. Springer, 2000, pp. 137–196.
- [77] V. C. Hu, D. Ferraiolo, R. Kuhn, A. R. Friedman, A. J. Lang, M. M. Cogdell, A. Schnitzer, K. Sandlin, R. Miller, and K. Scarfone, “Guide to attribute based access control (abac) definition and considerations (draft),” *NIST special publication*, vol. 800, no. 162, 2013.
- [78] S. Osborn, R. Sandhu, and Q. Munawer, “Configuring role-based access control to enforce mandatory and discretionary access control policies,” *ACM Transactions on Information and System Security (TISSEC)*, vol. 3, no. 2, pp. 85–106, 2000.
- [79] A. Ouaddah, A. Abou Elkalam, and A. A. Ouahman, “Towards a novel privacy-preserving access control model based on blockchain technology in iot,” in *Europe and MENA cooperation advances in information and communication technologies*. Springer, 2017, pp. 523–533.
- [80] A. Tolk and J. A. Muguira, “The levels of conceptual interoperability model,” in *Proceedings of the 2003 fall simulation interoperability workshop*, vol. 7. Citeseer, 2003, pp. 1–11.
- [81] D. Chen and N. Daclin, “Framework for enterprise interoperability,” in *Proc. of IFAC Workshop E12N*. Bordeaux, 2006, pp. 77–88.
- [82] A. Shahnaz, U. Qamar, and A. Khalid, “Using blockchain for electronic health records,” *IEEE Access*, vol. 7, pp. 147 782–147 795, 2019.
- [83] J. C. Mandel, D. A. Kreda, K. D. Mandl, I. S. Kohane, and R. B. Ramoni, “Smart on fhirc: a standards-based, interoperable apps platform for electronic health records,” *Journal of the American Medical Informatics Association*, vol. 23, no. 5, pp. 899–908, 2016.
- [84] Y. Zhang, D. He, and K. R. Choo, “Bads: Blockchain-based architecture for data sharing with abs and cp-abe in iot,” *Wireless Communications and Mobile Computing*, vol. 2018, 2018.
- [85] A. Manzoor, M. Liyanage, A. Braeke, S. S. Kanhere, and M. Ylianttila, “Blockchain based proxy re-encryption scheme for secure iot data sharing,” in *2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*. IEEE, 2019, pp. 99–103.
- [86] G. G. Dagher, J. Mohler, M. Milojkovic, and P. B. Marella, “Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology,” *Sustainable cities and society*, vol. 39, pp. 283–297, 2018.
- [87] R. Li, T. Song, B. Mei, H. Li, X. Cheng, and L. Sun, “Blockchain for large-scale internet of things data storage and protection,” *IEEE Transactions on Services Computing*, vol. 12, no. 5, pp. 762–771, 2018.

- [88] M. Singh and S. Kim, "Blockchain based intelligent vehicle data sharing framework," *arXiv preprint arXiv:1708.09721*, 2017.
- [89] X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, "Integrating blockchain for data sharing and collaboration in mobile healthcare applications," in *2017 IEEE 28th annual international symposium on personal, indoor, and mobile radio communications (PIMRC)*. IEEE, 2017, pp. 1–5.
- [90] K. Fan, S. Wang, Y. Ren, H. Li, and Y. Yang, "Medblock: Efficient and secure medical data sharing via blockchain," *Journal of medical systems*, vol. 42, no. 8, pp. 1–11, 2018.
- [91] L. Zhu, Y. Wu, K. Gai, and K. R. Choo, "Controllable and trustworthy blockchain-based cloud data management," *Future Generation Computer Systems*, vol. 91, pp. 527–535, 2019.
- [92] B. Zheng, L. Zhu, M. Shen, F. Gao, C. Zhang, Y. Li, and J. Yang, "Scalable and privacy-preserving data sharing based on blockchain," *Journal of Computer Science and Technology*, vol. 33, no. 3, pp. 557–567, 2018.
- [93] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano, "Public key encryption with keyword search," in *International conference on the theory and applications of cryptographic techniques*. Springer, 2004, pp. 506–522.
- [94] R. Curtmola, J. Garay, S. Kamara, and R. Ostrovsky, "Searchable symmetric encryption: improved definitions and efficient constructions," *Journal of Computer Security*, vol. 19, no. 5, pp. 895–934, 2011.
- [95] C. Cai, X. Yuan, and C. Wang, "Towards trustworthy and private keyword search in encrypted decentralized storage," in *2017 IEEE International Conference on Communications (ICC)*. IEEE, 2017, pp. 1–7.
- [96] X. Ma, C. Wang, and L. Wang, "The data sharing scheme based on blockchain," in *Proceedings of the 2nd ACM International Symposium on Blockchain and Secure Critical Infrastructure*, 2020, pp. 96–105.
- [97] H. Wu, R. Song, K. Lei, and B. Xiao, "Slicer: Verifiable, secure and fair search over encrypted numerical data using blockchain," in *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2022, pp. 1201–1211.
- [98] K. Mišura and M. Žagar, "Data marketplace for internet of things," in *2016 International Conference on Smart Systems and Technologies (SST)*. IEEE, 2016, pp. 255–260.
- [99] C. Li and G. Miklau, "Pricing aggregate queries in a data marketplace." in *WebDB*, 2012, pp. 19–24.
- [100] J. Chen and Y. Xue, "Bootstrapping a blockchain based ecosystem for big data exchange," in *2017 IEEE international congress on big data (bigdata congress)*. IEEE, 2017, pp. 460–463.
- [101] D. Nasonov, A. A. Visheratin, and A. Boukhanovsky, "Blockchain-based transaction integrity in distributed big data marketplace," in *International Conference on Computational Science*. Springer, 2018, pp. 569–577.
- [102] K. R. Özyilmaz, M. Doğan, and A. Yurdakul, "Idmob: Iot data marketplace on blockchain," in *2018 crypto valley conference on blockchain technology (CVCBT)*. IEEE, 2018, pp. 11–19.
- [103] J. Yang, Z. Lu, and J. Wu, "Smart-toy-edge-computing-oriented data exchange based on blockchain," *Journal of Systems Architecture*, vol. 87, pp. 36–48, 2018.
- [104] N. Hynes, D. Dao, D. Yan, R. Cheng, and D. Song, "A demonstration of sterling: A privacy-preserving data marketplace," *Proceedings of the VLDB Endowment*, vol. 11, no. 12, pp. 2086–2089, 2018.
- [105] J. Park, T. Youn, H. Kim, K. Rhee, and S. Shin, "Smart contract-based review system for an iot data marketplace," *Sensors*, vol. 18, no. 10, p. 3577, 2018.
- [106] M. Travizano, C. Sarraute, G. Ajzenman, and M. Minnoni, "Wibson: A decentralized data marketplace," *arXiv preprint arXiv:1812.09966*, 2018.
- [107] W. Xiong and L. Xiong, "Smart contract based data trading mode using blockchain and machine learning," *IEEE Access*, vol. 7, pp. 102 331–102 344, 2019.
- [108] S. Bajoudah, C. Dong, and P. Missier, "Toward a decentralized, trust-less marketplace for brokered iot data trading using blockchain," in *2019 IEEE international conference on blockchain (Blockchain)*. IEEE, 2019, pp. 339–346.
- [109] P. Banerjee and S. Ruj, "Blockchain enabled data marketplace—design and challenges," *arXiv preprint arXiv:1811.11462*, 2018.
- [110] Y. Chen, J. Guo, C. Li, and W. Ren, "Fade: A blockchain-based fair data exchange scheme for big data sharing," *Future Internet*, vol. 11, no. 11, p. 225, 2019.
- [111] K. Liu, X. Qiu, W. Chen, X. Chen, and Z. Zheng, "Optimal pricing mechanism for data market in blockchain-enhanced internet of things," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 9748–9761, 2019.
- [112] W. Dai, C. Dai, K. R. Choo, C. Cui, D. Zou, and H. Jin, "Sdte: A secure blockchain-based data trading ecosystem," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 725–737, 2019.
- [113] D. Hu, Y. Li, L. Pan, M. Li, and S. Zheng, "A blockchain-based trading system for big data," *Computer Networks*, vol. 191, p. 107994, 2021.
- [114] A. A. Abdellatif, L. Samara, A. Mohamed, A. Erbad, C. F. Chiasserini, M. Guizani, M. D. OâĂZConnor, and J. Laughton, "Medge-chain: Leveraging edge computing and blockchain for efficient medical data exchange," *IEEE Internet of Things Journal*, vol. 8, no. 21, pp. 15 762–15 775, 2021.
- [115] Y. Liu, C. Zhang, Y. Yan, X. Zhou, Z. Tian, and J. Zhang, "A semi-centralized trust management model based on blockchain for data exchange in iot system," *IEEE Transactions on Services Computing*, 2022.
- [116] Z. Guan, X. Shao, and Z. Wan, "Secure fair and efficient data trading without third party using blockchain," in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, 2018, pp. 1395–1401.
- [117] J. Benet, "IPFS-content addressed, versioned, P2P file system," *arXiv preprint arXiv:1407.3561*, 2014.
- [118] S. Wilkinson, T. Boshevski, J. Brandoff, and V. Buterin, "Storj a peer-to-peer cloud storage network," 2014.
- [119] R. Radhakrishnan and B. Krishnamachari, "Streaming data payment protocol (sdpp) for the internet of things," in *2018 IEEE International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*. IEEE, 2018, pp. 1679–1684.
- [120] M. Green and I. Miers, "Bolt: Anonymous payment channels for decentralized currencies," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 473–489.
- [121] S. Dziembowski, L. Eckey, S. Faust, and D. Malinowski, "Perun: Virtual payment hubs over cryptocurrencies," in *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2019, pp. 106–123.
- [122] L. T. Blessing and A. Chakrabarti, *DRM: A design reseach methodology*. Springer, 2009.
- [123] N. P. Sheppard, "Can smart contracts learn from digital rights management?" *IEEE Technology and Society Magazine*, vol. 39, no. 1, pp. 69–75, 2020.
- [124] R. Song, S. Gao, Y. Song, and B. Xiao, "ZKDET: A traceable and privacy-preserving data exchange scheme based on non-fungible token and zero-knowledge," in *2022 IEEE 42nd International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2022, pp. 224–234.
- [125] Z. Meng, T. Morizumi, S. Miyata, and H. Kinoshita, "Design scheme of copyright management system based on digital watermarking and blockchain," in *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*, vol. 2. IEEE, 2018, pp. 359–364.
- [126] S. Zhao and D. O'Mahony, "Bmcprotector: A blockchain and smart contract based application for music copyright protection," in *Proceedings of the 2018 International Conference on Blockchain Technology and Application*, 2018, pp. 1–5.
- [127] R. Mehta, N. Kapoor, S. Sourav, and R. Shorey, "Decentralised image sharing and copyright protection using blockchain and perceptual hashes," in *2019 11th International Conference on Communication Systems & Networks (COMSNETS)*. IEEE, 2019, pp. 1–6.
- [128] L. Xiao, W. Huang, Y. Xie, W. Xiao, and K. Li, "A blockchain-based traceable ip copyright protection algorithm," *IEEE Access*, vol. 8, pp. 49 532–49 542, 2020.
- [129] N. Jing, Q. Liu, and V. Sugumaran, "A blockchain-based code copyright management system," *Information Processing & Management*, vol. 58, no. 3, p. 102518, 2021.
- [130] W. Liang, D. Zhang, X. Lei, M. Tang, K. Li, and A. Zomaya, "Circuit copyright blockchain: Blockchain-based homomorphic encryption for ip circuit protection," *IEEE Transactions on Emerging Topics in Computing*, 2020.

- [131] A. Futoransky, C. Sarraute, A. Waissbein, D. Fernandez, M. Travizano, and M. Minnoni, "Secure exchange of digital goods in a decentralized data marketplace," *arXiv preprint arXiv:1907.12625*, 2019.
- [132] P. Gupta, V. Dedeoglu, S. S. Kanhere, and R. Jurdak, "Towards a blockchain powered iot data marketplace," in *2021 International Conference on COMmunication Systems & NETworkS (COMSNETS)*. IEEE, 2021, pp. 366–368.
- [133] S. Sahoo and R. Halder, "Traceability and ownership claim of data on big data marketplace using blockchain technology," *Journal of Information and Telecommunication*, vol. 5, no. 1, pp. 35–61, 2021.
- [134] R. Haenni, "Datum network: The decentralized data marketplace," *DATUM Whitepaper*, vol. 15, 2017.
- [135] T. McConaghy, R. Marques, A. Müller, D. De Jonghe, T. McConaghy, G. McMullen, R. Henderson, S. Bellemare, and A. Granzotto, "Bigchaindb: a scalable blockchain database," *white paper, BigChainDB*, 2016.
- [136] GXChain, "Gxchain whitepaper 3.0," *GXChain Whitepaper*, vol. 10, 2018.
- [137] Databroker. (2021) Databroker. [Online]. Available: <https://www.databroker.global/>
- [138] D. Draskovic and G. Saleh, "Decentralized data marketplace based on blockchain," *White Paper*, 2017.
- [139] W. F. Silvano and R. Marcelino, "Iota tangle: A cryptocurrency to communicate internet-of-things data," *Future Generation Computer Systems*, vol. 112, pp. 307–319, 2020.
- [140] S. Heiler, "Semantic interoperability," *ACM Computing Surveys (CSUR)*, vol. 27, no. 2, pp. 271–273, 1995.
- [141] Y. Li, C. Ye, Y. Hu, I. Morpheus, Y. Guo, C. Zhang, Y. Zhang, Z. Sun, Y. Lu, and H. Wang, "Zkcplus: Optimized fair-exchange protocol supporting practical and flexible data exchange," in *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, 2021, pp. 3002–3021.



Yubo Song received the B.S., M.S., and Ph.D. degrees in communication engineering from Southeast University, Nanjing, China. He is currently an Associate Professor at Southeast University, China. His research interests include physical layer security, data privacy and blockchain. He has published more than 90 papers in top journals and conferences and obtained more than 40 inventions. He participated in many projects of the National Natural Science Foundation of China.



Songtao Guo received the B.S., M.S., and Ph.D. degrees in computer software and theory from Chongqing University, China, in 1999, 2003, and 2008, respectively. He is currently a Full Professor with Chongqing University, China. He has authored or coauthored more than 110 scientific papers in leading refereed journals and conferences. His research interests include wireless sensor networks, wireless ad hoc networks, and parallel and distributed computing. He was the recipient of many research grants as a Principal Investigator from the National Science Foundation of China and Chongqing and the Postdoctoral Science Foundation of China.



Rui Song received the B.Sc. and M.Sc. degree in computer security from Southeast University, Nanjing, China. He is currently working toward the Ph.D. degree with the Department of Computing, The Hong Kong Polytechnic University under the supervision of Dr. Bin Xiao. His research interest include blockchain, data privacy and cryptography.



Bin Xiao is a professor at the Department of Computing, the Hong Kong Polytechnic University, Hong Kong. Prof. Xiao received B.Sc and M.Sc degrees in Electronics Engineering from Fudan University, China, and a Ph.D. degree in computer science from the University of Texas at Dallas, USA. His research interests include AI and network security, data privacy, and blockchain systems. He published more than 200 technical papers in international top journals and conferences. He is currently an Associate

Editor of *IEEE Transactions on Cloud Computing*, *IEEE Internet of Things Journal*, and *IEEE Transactions on Network Science and Engineering*. He has been the associate editor of the *Elsevier Journal of Parallel and Distributed Computing* from 2016 to 2021. He is the vice chair of the IEEE ComSoc CISTC committee. He has been the track co-chair of IEEE ICDCS2022, the symposium track co-chair of IEEE ICC2020, ICC 2018, and Globecom 2017, and the general chair of IEEE SECON 2018. He is a senior member of IEEE, a member of ACM, and CCF.



Yuanyuan Yang received the B.Eng. and M.S. degrees in computer science and engineering from Tsinghua University, Beijing, China, and the M.S.E. and Ph.D. degrees in computer science from Johns Hopkins University, Baltimore, Maryland. She is currently a SUNY Distinguished Professor in the Department of Electrical & Computer Engineering and Department of Computer Science at Stony Brook University, New York, which she joined in 1999. From 2018-2022, she served as a program director in the National

Science Foundation's Directorate of Computer and Information Science and Engineering. She directed the core computer architecture program and was on the management team of several cross-cutting programs. At Stony Brook, she served as the Associate Dean for Diversity and Academic Affairs of College of Engineering and Applied Sciences from 2016-2018, a Division Director of New York State Center of Excellence in Wireless and Information Technology from 2007-2016 and the Graduate Program Director of ECE Department from 2001-2016. She has authored or coauthored more than 500 papers in major journals and refereed conference proceedings and holds seven US patents in her areas of research which include cloud computing, edge computing, quantum computing and mobile computing. She is currently the Editor-in-Chief of *IEEE Transactions on Cloud Computing*. She was an Associate Editor-in-Chief and Associated Editor for *IEEE Transactions on Computers*, and an Associate Editor for *IEEE Transactions on Parallel and Distributed Systems*. She was also the General Chair, Program Chair, or Vice Chair for several major conferences and a Program Committee Member for numerous conferences. She is a Fellow of National Academy of Inventors (NAI) and a Fellow of IEEE.