Survey | Received 16 February 2023; Accepted 3 July 2023; Published 12 July 2023
https://doi.org/10.55092/blockchain20230007

Interaction mechanism between blockchain and IPFS

Feng Yang[1,*], Zongya Ding[1], Yankuan Yu[1] and Yi Sun[1,2]

[1] Shandong Key Laboratory of Blockchain Finance, School of Computer Science and Technology, Shandong University of Finance and Economics, Jinan 250014, China
[2] Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190, China

* Correspondence author; E-mail: creyf@126.com.

Abstract: Blockchains can provide integrity and authenticity, but their limited storage capacity can be a challenge when it comes to storing large amounts of data. To address this issue, off-chain storage solutions such as the InterPlanetary File System (IPFS) can be utilized. This has led to the emergence of various applications that utilize both blockchain and IPFS. After reviewing a large body of literature utilizing blockchain and IPFS, we found that the coordinated interaction between blockchain and IPFS can help solve many problems and provide many research opportunities. Therefore, this survey paper aims to introduce the interaction mechanism between blockchain and IPFS. We first provide a general overview and comparison of different P2P data networks to help understand why IPFS is suitable as the storage layer for blockchains. Subsequently, we use select applications that leverage blockchain and IPFS to show how the mechanism works and explore new developments in this area. Specifically, we identify research areas and provide a qualitative comparison of these different applications. From the comparison, we derive research goals related to the interaction mechanism between blockchain and IPFS.

Keywords: Blockchain; IPFS; coordinated interaction; data networks; peer-to-peer networks; data availability

1. Introduction

Since Nakamoto [1] introduced the concept of the blockchain in 2008, blockchains have been widely used in various scenarios, including medical treatment, logistics, copyright protection, and finance, due to their security, high transparency, and reliability [2]. As a distributed network, a blockchain utilizes a consensus mechanism to record and synchronize data, ensuring that data is immutable [3]. The inclusion of new data within the blockchain is accomplished through the formation of blocks, which, upon receipt of new data, generate a hash value derived from the relevant data, the index of the block in the chain, the timestamp of data reception, and the hash of the preceding block within the chain. After a node successfully mined the block, copies of the block are disseminated to other nodes for linkage with their respective local chains. Each block is divisible into two sections, the block header and the block body. While the former links to the previous block to form a chain structure, the latter records data information within the network. For more information on blockchains, see [4]. However, blockchains face challenges,

such as limited on-chain storage capacity [5,6], which can limit the amount of data that can be stored per block. This can result in high transaction fees when attempting to store large amounts of data on a blockchain, such as in Bitcoin or Ethereum [7]. Each transaction must be verified and processed by network participants, which can be costly in terms of computing power and energy consumption.

One way to improve blockchains' storage issues is off-chaining transaction data [8]. By moving transaction data off the blockchain and onto an off-chain storage platform, the amount of data stored on the blockchain can be reduced, thereby improving scalability. In that case, an on-chain and off-chain interaction mechanism is required. The on-chain and off-chain interaction mainly include transferring transaction data originally stored in the blockchain to an off-chain storage platform, storing a digest corresponding to the original off-chain data on the blockchain, and utilizing the on-chain digest to validate and retrieve off-chain original data. To guarantee the data consistency, data availability, data authenticity, efficiency and security of data retrieval, it is essential to establish an on-chain and off-chain interaction mechanism.

Many different peer-to-peer (P2P) data networks can be utilized as off-chain storage platforms. P2P data networks offer decentralized data storage, distribution, replication, and exchange, providing an alternative to centralized cloud storage and breaking free from data silos [9]. Popular P2P technologies include Napster [10], Freenet [11], Chord [12], CAN [13], Pastry [14], and BitTorrent [15]. The InterPlanetary File System (IPFS) [16] is the beginning of a new generation of P2P data networks [9] and is often used as a storage layer for blockchains. It is a P2P distributed hypermedia distribution protocol. Computing devices connected by it all have the same file management mode, integrate the distributed system, and have excellent security and high transmission speed [17]. IPFS uses content-based addressing to uniquely identify and retrieve files, assigning each file a unique content-addressed identifier (CID) derived from the file's content. IPFS has a good application prospect in distributed networks. However, the authenticity of data stored on IPFS cannot be guaranteed, as files are uploaded by each node and stored on nodes that can be targeted by hackers, making data tampering a risk [3]. Note, that we introduce IPFS more detailly in a separate section.

Given all the above, it is time to formally introduce the interaction mechanism between blockchain and IPFS, which we will refer to as the interaction mechanism in the following for simplicity. The interaction mechanism provides a secure and efficient way to store and access data. Data is uploaded onto IPFS, and the CID is stored on the blockchain, enabling easy tracing and authentication [3]. The interaction mechanism can fulfill various purposes beyond simply improving storage issues of blockchains and ensuring the authenticity of IPFS data. For example, the utilization of blockchain technology can serve two distinct purposes in relation to IPFS data: facilitating access control [18–20] and generating an audit trail [21]. Moreover, cryptocurrencies can function as an incentive system for P2P data networks, increasing their availability and robustness [9]. Consequently, the interaction mechanism is widely utilized in various areas, e.g., medical data storage [22,23], agricultural products tracking [24], and IoT data privacy [25].

In this paper, we explore the interaction between blockchain and IPFS and how this interaction addresses specific challenges in different research areas with select applications. We conducted a comprehensive review of literature that analyzes or utilizes the interaction mechanism. By comparing different applications and studying relevant literature, we extract five significant and representative areas related to the interaction mechanism that merit further research, including performance, access control,

security, data availability, and data look-up. We provide an introduction to existing research in each of these areas and a comparative overview of some applications that leverage blockchain and IPFS within the last few years with respect to these five areas. While numerous new applications based on blockchain and IPFS have been proposed, we focus on applications that demonstrate distinct characteristics across various scenarios. Our overview emphasizes scientific and technical aspects combined with specific implementation details to extract meaningful insights. Drawing on our findings, we extract new research opportunities concerning the interaction mechanism between blockchain and IPFS. To conduct this survey, we consulted a diverse range of sources, including journal articles, white papers, books, and conference proceedings.

The remainder is structured as follows: First, we present an overview and a comparison of various P2P data networks (Section 2). Subsequently, we delve into technical analysis of the interaction mechanism between blockchain and IPFS (Section 3). Finally, Section 4 concludes this survey.

2. Peer-to-peer data networks

Aside from IPFS, many P2P data networks are in development. Daniel and Tschorsch [9] provided a general overview and a qualitative comparison of IPFS [16] with other unique P2P data networks such as BitTorrent [15], Swarm [26], the Hypercore Protocol [27], SAFE [28], Storj [29], and Arweave [30]. They found that most of the literature on P2P data networks focuses on IPFS, with analyses and utilization of other networks being scarce. The possible reasons for this could include inadequate real-world application, limited user adoption, insufficient implementation, or inadequate clear and organized documentation. Most systems, unlike IPFS, are challenging to understand, access, and prove that they work.

For a better understanding of why IPFS is suitable as the storage layer for blockchains, we briefly introduce IPFS and these other P2P data networks and mainly discuss what advantages IPFS has over traditional file systems and other P2P data networks. For more details on P2P data networks, we refer to [9]. Table 1 summarizes and compares IPFS with the following data networks.

2.1. InterPlanetary File System (IPFS)

The InterPlanetary File System (IPFS) [16] is a cutting-edge P2P file system that utilizes advanced technologies to create a secure and efficient means of storing and sharing files. IPFS addresses many of the limitations of traditional HTTP-based file transfer protocols, such as centralized servers, high bandwidth costs, and a lack of content integrity. By utilizing distributed hash tables (DHTs) and content-addressed storage, IPFS offers a more efficient, secure, and resilient way to store, access, and share files over the Internet.

IPFS employs libp2p [31], a modular P2P networking stack, to facilitate communication between nodes. Additionally, IPFS uses Kademlia [32], which is likely the most extensively utilized DHT, to enable content routing and peer discovery. When a node in IPFS discovers a new node through Kademlia, it attempts to establish a connection and assigns it into a bucket [33]. These connections are managed by a connection management system that trims idle connections once the HighWater threshold is achieved (default 900), until the LowWater threshold is reached (default 600).

Table 1. Summary and Comparison of the Different Data Networks.

System	Main goal	File handling	Security	User base	Token	Mutability	Network
IPFS [16,34]	Decentralized web enabling fast distribution through content addressing	Opportunistic file look-up, a DHT is for the backup look-up; files are split into blocks and their locations are random; passive file replication	Content-addressing, replication, and incentives	Large	Filecoin [35]	IPNS	Hybrid
BitTorrent [15]	Efficient file distribution over the Internet in a decentralized manner	A central component or a DHT is for the backup look-up; file-based storage and the location is random; passive file replication	Meta-data file, replication, and incentives	Large	BTT [36]	-	Unstructured
Swarm [26,37]	Decentralized storage and communication structure supported by complex Ethereum-based incentives	A DHT is for the backup look-up; files are split into blocks and content addressed	Manifests, content-addressing, replication, erasure codes, and incentives	Small	Ethereum [7]	ENS, Feeds	Kademlia [32]
Hypercore [27,38]	Simple sharing of large mutable data objects between selected peers	A DHT is for the backup look-up; file-based storage and the location is random; passive file replication	Public key, meta-data file, and replication	Small	-	Yes	Unstructured
SAFE [28,39]	Autonomous data and communications network using self-encryption and self-authentication to improve privacy and decentralization	A DHT is for the backup look-up; files are split into blocks and content addressed; active file replication	Self-authentication, content-addressing, self-encryption, replication, and incentives	Small	Safecoin	Specific	Kademlia
Storj [29,40]	Decentralized cloud storage that increases file availability using erasure codes	Central file look-up; files are split into segments and their locations are random	Satellite nodes, erasure codes, and incentives	Small	Centralized payments	Yes	Unstructured
Arweave [30,41]	Permanent storage in a blockchain-like structure	Opportunistic file look-up; file-based storage; passive file replication	Blockweave, replication, and incentives	Small	Arweave token	-	Unstructured

IPFS uses Merkle directed acyclic graphs (DAGs), as shown in Figure 1, to create a content-addressed storage system. Merkle DAGs are data structures similar to Merkle trees, but with a few key differences. While Merkle trees are binary trees where every non-leaf node is the hash of its two children, Merkle DAGs allow for multiple parents for a given node. This means that a node can have multiple child nodes, and those child nodes can each be linked to by other nodes in the DAG. In addition to this, Merkle DAGs are not required to be balanced, which means that nodes can have different depths in the DAG, and there may be multiple paths from the root to a leaf node. By using the root of the Merkle DAG to represent content, IPFS provides an efficient and secure way to store and access data.

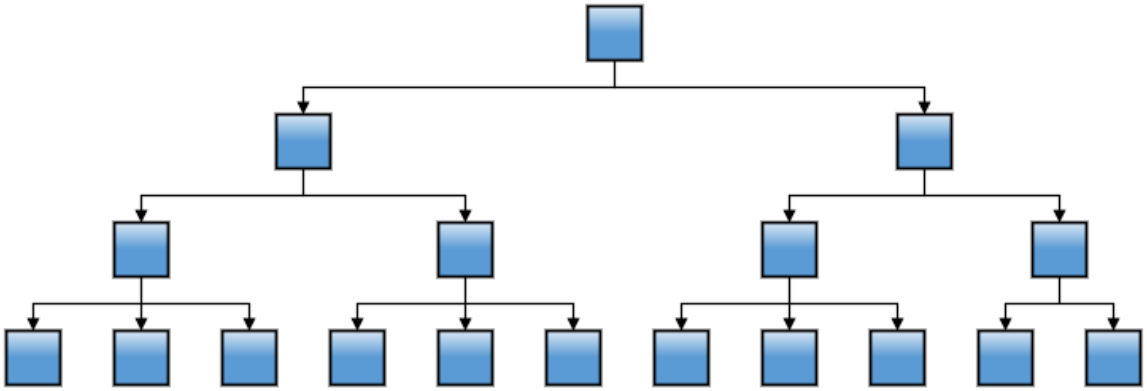

Figure 1. Structure of a Merkle DAG.

To efficiently retrieve data, IPFS uses the Bitswap protocol, which sends a combination of WANT-HAVE and WANT-BLOCK messages to multiple peers in a session. The WANT-HAVE message asks if the peer has the requested block, while the WANT-BLOCK message requests the block directly. Since version 0.5, Bitswap also sends an optimistic WANT-BLOCK message to one peer. If a block is received from a peer, other pending requests for the same block can be canceled with a CANCEL message [42]. This helps to ensure that data is retrieved as quickly and efficiently as possible.

Discussion: IPFS offers numerous advantages over traditional file systems, making it a highly efficient, decentralized, and scalable solution. By leveraging a network of peers rather than relying on a single server, IPFS reduces the load on any single node and makes the system more resilient to failures. Content-addressing and deduplication technologies are integrated in IPFS, which not only makes it easier to replicate and distribute data across the network, but also reduces storage requirements and improves efficiency [9].

IPFS uses a DHT to store information about where files can be found on the network. This enables nodes to quickly locate the nearest copy of a file, making it faster to retrieve files and reducing the time and bandwidth needed to retrieve files from other nodes.

Decentralization is a key feature of IPFS, aligning with the decentralization goals of blockchain technology. As a result, IPFS is a natural fit for storing off-chain data, and its scalability makes it an ideal storage layer for blockchains that may need to handle large amounts of data. Storing CIDs on the blockchain is equivalent to storing the original file on IPFS, but without the need for additional block capacity. Furthermore, IPFS is widely used with over 4000 contributors [43] worldwide and has been deployed in over 2700 Autonomous Systems, covering 152 countries [44]. IPFS infrastructure is highly reliable and handles millions of content retrievals from clients every day.

Overall, IPFS is a robust and efficient file system that offers numerous advantages over traditional file systems, making it an excellent choice for storing off-chain data and serving as the storage layer for blockchains.

2.2. BitTorrent

The BitTorrent protocol [15] is a popular and well-researched [45–47] peer-to-peer file sharing protocol that has been in use for over two decades. Its unique approach allows users to share large files over the internet by breaking them down into smaller pieces, which can be downloaded and uploaded simultaneously by multiple peers. This significantly increases download speeds and reduces the strain on any single user's internet connection. BitTorrent has stood the test of time and is still actively used by millions of people [48] around the world for sharing files. Its success has also served as a model for newer P2P file distribution systems.

Moreover, the BitTorrent Foundation, in partnership with the Tron Foundation, developed BitTorrent Token (BTT) [36], a blockchain-based incentive layer that rewards users for sharing and seeding files. By providing incentives, BTT aims to increase the availability and persistence of files on the network, further enhancing the effectiveness and scalability of the BitTorrent protocol. The use of blockchain technology also adds an additional layer of security and transparency to the protocol. Overall, BitTorrent and BTT remain important tools for efficient and secure file sharing on the internet.

Discussion: Daniel and Tschorsch [9] consider BitTorrent a first-generation P2P data network which can be used to store data, but it primarily concentrates on data sharing other than data storage. While BitTorrent offers many advantages, it also has some significant drawbacks. For example, downloading can be unstable, limiting its widespread use to specific occasions. Additionally, BitTorrent lacks the ability to verify file publishers, making it difficult to ensure the credibility of downloaded content [49]. Furthermore, BitTorrent does not support mutability [9] and uses location-based addressing, which requires a tracker to keep track of where files are located. The dependency on a tracker poses significant challenges when it comes to accessing files in the event of tracker unavailability. The ability for users to modify and delete files introduces potential inconsistencies within the network. Additionally, BitTorrent's reliance on a tracker to coordinate file distribution can lead to network slowdowns and make it susceptible to overloading or crashes. This reliance on a tracker makes it more challenging to access files if the tracker goes down. It can also slow down the network and make it vulnerable to overloading or crashes. The ability for users to modify and delete files can create inconsistencies in the network. In contrast, IPFS offers content addressing, DHT, and data deduplication, making it a faster and more efficient storage and retrieval solution than BitTorrent. Its decentralized network also makes it more resistant to attacks, providing a more reliable storage solution for blockchain applications.

2.3. Swarm

Swarm [26] is a revolutionary decentralized P2P network protocol that aims to provide a decentralized and censorship-resistant platform for storing and distributing data and files within the web3 stack [50]. As an important part of the Ethereum ecosystem, Swarm enables users to securely store and process data on a decentralized network, utilizing innovative features such as incentivization models, versioning, and mutability support.

One of the most notable features of Swarm is its use of the Ethereum Name Service (ENS) [51], which allows for human-readable names to be resolved into Swarm addresses, making it easier for users to locate and access content on the network. This creates a more user-friendly experience for accessing data on the network and improves the overall accessibility of decentralized storage.

Discussion: IPFS outperforms Swarm in several aspects. In terms of flexibility, IPFS is more versatile as it can store any type of data, from structured and unstructured data to large and small files, and even entire web applications. In contrast, Swarm is primarily designed for storing small pieces of data like metadata. IPFS has a larger user base and is more widely adopted than Swarm, while Swarm is a newer technology that adapts the secure and stable network of Ethereum.

Furthermore, Swarm tends to use slightly more CPU, memory, and network resources than IPFS. However, in terms of connectivity, IPFS has a higher average number of connected peers [52]. Another potential issue with Swarm is that it may face storage problems due to the determined storage locations [9].

Overall, while both IPFS and Swarm have their strengths and weaknesses, IPFS appears to be a more flexible, widely used, and established technology.

2.4. Hypercore Protocol

The Hypercore Protocol [27,53] is a cutting-edge decentralized data synchronization protocol that facilitates real-time data sharing and collaboration across multiple devices without the need for centralized servers or cloud services. This protocol is designed with security, efficiency, and decentralization.

Hypercore utilizes advanced cryptographic techniques to ensure that data is kept private and tamper-proof, making it an ideal technology for creating distributed systems and applications. Its ability to handle large-scale data synchronization in a decentralized manner makes it a promising technology for the future of data sharing and collaboration.

Numerous applications have already been built using Hypercore, including collaborative text editors, social networks, and file-sharing tools. As a result, it has proven to be a versatile and adaptable technology that can be used in a wide variety of settings. Given its strengths and versatility, the Hypercore Protocol is likely to continue playing a key role in the evolution of decentralized data sharing and collaboration.

Discussion: Like BitTorrent, Hypercore focuses more on data sharing than data storage. Although it is able to provide fast and secure off-chain storage solutions for blockchains, IPFS outperforms Hypercore in several key areas. For example, IPFS excels in distributing files quickly [54] and efficiently. This is thanks to its content addressing mechanism, which enables it to cache content more efficiently and distribute it more effectively across a network. As a result, IPFS provides faster access and improved performance. In contrast, Hypercore uses key-based addressing, where files are identified by a unique key specific to the node that stores it. This can result in redundancy and slower access times compared to IPFS.

2.5. Secure access for everyone (SAFE)

Secure Access For Everyone (SAFE) [28,55] is a comprehensive network security solution that enables remote workers, contractors, and clients to connect securely to company networks from anywhere. To ensure secure and reliable remote access, SAFE incorporates various security mechanisms such as encryption and authentication. It is noteworthy that SAFE does not rely on any centralized component for authentication, instead using a self-authentication [56] mechanism for added security. In addition, the system employs a self-encryption [57] algorithm that encrypts files using the file itself as the key, which adds another layer of security.

The security of SAFE has been analyzed by Paul *et al*. [58], who focusd on confidentiality, integrity, and availability. The analysis has identified the system's strengths and weaknesses, helping to inform ongoing efforts to improve its security.

Discussion: One potential concern regarding the use of the SAFE network is the possibility of its self-authentication feature being exploited, which could lead to the exposure of personal user data [59]. In comparison, IPFS offers several advantages, including a larger and more active community of developers and contributors. This is evident from the substantial support IPFS has garnered, with 78 contributors and over 22100 stars [60] on GitHub, whereas SAFE currently has only 5 contributors and no stars [61]. The broader adoption and support of IPFS makes it more convenient for developers to use it effectively as the storage layer for blockchains. While SAFE possesses promising features, such as a strong emphasis on privacy and security, it is still a relatively new project that has not yet gained the same level of adoption and community support as IPFS. As the field of decentralized storage continues to evolve, it will be interesting to the development and competition among these and other projects. However, based on current data and metrics, IPFS appears to be better suited for use as the storage layer for blockchains.

2.6. Storj

Storj [29] is a decentralized cloud storage platform that uses a P2P network of storage nodes to provide secure and private storage for its users. One of its unique features is the use of Reed-Solomon erasure codes [62] to ensure data protection and redundancy, even in the event of node failures or attacks.

Discussion: Storj has faced some security concerns. For instance, a study by De Figueiredo *et al*. [63] found that satellite nodes in the network could potentially be exploited as vectors for Denial-of-Service attacks. In addition, Zhang *et al*. [64] uncovered a vulnerability in Storj v2.0 that enabled unencrypted data to be uploaded to storage nodes, potentially framing the storage node owners for illegal content.

While both IPFS and Storj are decentralized storage platforms, IPFS is better suited as the storage layer for blockchains due to its content-based addressing and deduplication system, which is more efficient and secure than Storj's addressing system. IPFS seamlessly integrates with blockchains thanks to its content-addressable storage system and decentralized naming systems like IPNS. On the other hand, Storj lacks these features, making it less suitable for seamless integration with blockchains. Additionally, Storj uses erasure codes, which could add overhead to storing files [9].

Moreover, IPFS benefits from a more mature and active developer community compared to Storj. IPFS is continuously improving with new features and capabilities being added regularly. In contrast, Storj has a smaller user base and fewer developers, resulting in slower progress and fewer updates.

2.7. Arweave

Arweave [30] is a blockchain-based storage network that provides permanent and decentralized data storage at a low cost. The network uses a unique consensus mechanism called "Proof of Access" and a native cryptocurrency, AR, to incentivize miners to maintain the network's security and stability.

To address scalability challenges, Arweave has implemented innovative technical solutions such as blockshadows. Similar to compact blocks [65], blockshadows allow nodes to verify the authenticity of a block without having to download the entire block. This approach makes the verification process faster and more efficient.

Additionally, Arweave has implemented Wildfire, a solution that improves the scalability and efficiency of the network. In the wildfire, nodes within the Arweave network rank their peers according to their generosity and responsiveness. Subsequently, the node prioritizes communication with higher-ranked peers using a gossip-based algorithm. By deploying a network of nodes running wildfire agents, the throughput of the network can be maximized. The structure of the network can be represented as a directed graph with weighted arcs, where the weights are measured in units such as 'bytes per second from last N responses'. Refer to Figure 2 for a visualization of this network.

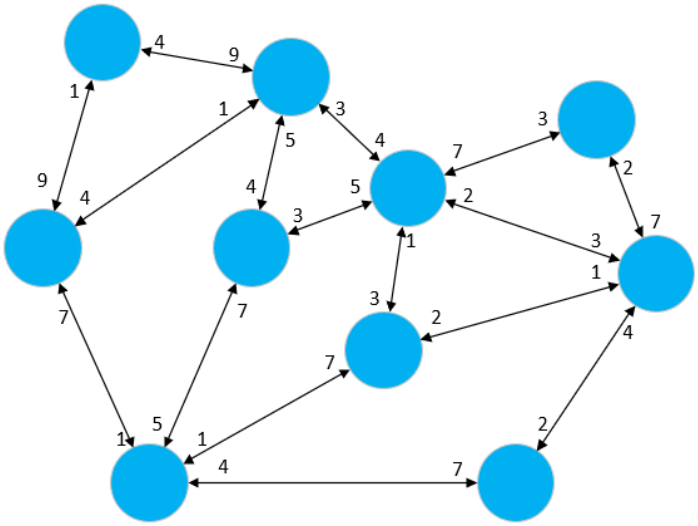

Figure 2. Wildfire as a weighted graph.

Discussion: IPFS has several advantages over Arweave:

Firstly, IPFS uses a content-addressable system that provides faster retrieval times by looking up content by its CID in the distributed network. On the other hand, Arweave's blockchain-based system requires more computational resources to retrieve content, resulting in slower retrieval times. Therefore, IPFS performs better than Arweave in terms of retrieval times.

Secondly, IPFS has been around for longer and has a larger user base and community support than Arweave, providing a more comprehensive set of tools and resources for developers. This advantage makes IPFS more attractive to developers looking to build decentralized applications.

Thirdly, Arweave does not support mutability [9], which means that once content is stored on the network, it cannot be changed or updated. In contrast, IPFS supports mutability, allowing content to be modified or updated.

Fourthly, the Arweave protocol provides on-chain storage on a blockchain-like structure, therefore it possesses similar disadvantages of a blockchain [9].

Lastly, IPFS is more efficient in terms of storage usage than Arweave. IPFS uses content-addressable storage, which stores identical pieces of content only once on the network, leading to more efficient storage usage. In contrast, Arweave's blockchain-based system requires redundant storage, leading to inefficient use of storage resources.

In summary, IPFS outperforms Arweave in several key areas, including faster retrieval times, a larger user base and community support, support for mutability, scalability, and more efficient storage usage. These advantages make IPFS a more attractive option for developers looking to build decentralized applications.

3. Research areas

Blockchains can address the storage issues by leveraging IPFS as an off-chain storage layer. However, many challenges remain, such as difficulty in look-up and single functionality. In addition, IPFS itself has faced several challenges, which still apply to the interaction mechanism. Daniel and Tschorsch [9] identified certain challenges of the new generation of decentralized data storage systems represented by IPFS. This includes performance, confidentiality and access control, security, anonymity, and naming. Although blockchains may provide potential solutions to some of these challenges, but more consideration is needed.

We observe five representative areas of the interaction mechanism, which could provide new research opportunities: performance, access control, security, data availability, and data look-up. Table 2 shows both a summary of existing research and a comparison of twenty applications that leverage blockchain and IPFS.

3.1. Performance

When evaluating a new mechanism, its performance is often the first factor to consider. Simulations or tests can be conducted to investigate its performance, data storage efficiency, feasibility, and how it compares to other methods. Such evaluations can identify new use cases and strengthen claims that a mechanism may replace existing schemes. A data storage model using IPFS and blockchain was introduced by Hao *et al*. [24], demonstrating superior performance compared to current methods. Sun *et al*. [66] developed a secure storage and sharing scheme for electronic medical records by utilizing IPFS and blockchain in conjunction with ciphertext-policy attribute-based encryption (CP-ABE). Their approach was shown to be efficient and feasible through performance analysis and simulation experiments with real data sets. The interaction mechanism has also performed well in other use cases, e.g., distributed e-commerce [3], storage and access control of insurance data [67], a data auction mechanism [68], and a data trading mode [69].

Table 2. Comparison of Applications based on Blockchain and IPFS.

Application	Performance	Access control	Security	Data availability	Data look-up
Storage scheme for agricultural products tracking [24]	Can outperform the existing methods	-	Double chain verification mechanism based on blockchain	-	Data query algorithm
Distributed e-commerce system based on IPFS [3]	Has good storage efficiency and can verify the authenticity of the data	-	IPFS data storage scheme for data authenticity	IPFS backup scheme for data availability	Official search center provided by OpenBazaar and secondary database
Storage and access scheme for electronic medical records [66]	Efficient and feasible	CP-ABE [70]	CP-ABE	-	Verifiable keyword search
Non-repudiation storage and access control scheme of insurance data [67]	Efficient and feasible	CP-ABE	CP-ABE	-	-
Anti-collusion data auction mechanism [68]	Effective	Smart contracts [71]	SmartCheck [72] and Ethereum	-	-
Smart contracts based data trading mode [69]	Successfully achieved the goal	Smart contracts	Off-chain download mechanism	-	-
Logistics information platform [54]	Has greatly improved the block storage capacity reduction	Sub-node query scheme	Encryption and consortium blockchain	-	Sub-node query scheme and unencrypted copies
Method of university education resource sharing [73]	Ensures safe and dependable data storage, reduces storage expenses, and greatly enhances file accessibility speed	-	Consortium blockchain and encryption	-	Smart contracts

Table 2. *Cont.*

Application	Performance	Access control	Security	Data availability	Data look-up
Knowledge sharing mechanism based on blockchain [74]	Can meet the needs	Consortium blockchain	Smart contracts	-	Smart contracts
Electronic medical record sharing framework [75]	Highly achievable with secure storage and efficient sharing, while balancing efficiency and cost	Membership service providers and chaincode mechanism of Hyperledger Fabric [76]	Hyperledger Fabric and symmetric encryption	-	-
Electronic certificate storage system [77]	Solves the storage problem while ensuring the safety, credibility, and traceability	Access permission control scheme, Hyperledger Fabric	Hyperledger Fabric, advanced encryption standard [78], and elliptic curve cryptography	-	Smart contracts
Fine-grained access control scheme for VANET data [79]	Low-performance overhead	HECP-ABE	HECP-ABE and advanced encryption standard	Replication proof, erasure coding, and incentives	-
High-quality educational resource platform based on blockchain [80]	Achieves distributed storage and sharing	Smart contracts	Encryption and smart contracts	-	-
Scheme for secure sharing of business collaboration data [81]	Feasible and can meet the requirements	Authorization token	Paillier [82]	-	-
IDDS [83]	Improves work efficiency and ensures the stability and sharing security of data storage	-	Disease prevention and control algorithm and symmetric encryption	-	-
Blockchain-based traceability system for P2P distribution [84]	Works well and supports normal P2P distribution	-	-	-	-

3.2. Access control

Access control for IPFS using blockchains is an active research direction that has been explored by several researchers [18–20,66,67,74,77,79,85,86]. There are existing proposals for access control using smart contracts [74,75], which are computer programs that run on a shared and replicated ledger [71]. Smart contracts are capable of processing information and managing the transfer of value. Smart contracts built on blockchain technology consist of processing and storage mechanisms for transactions, along with comprehensive state machines that can receive and handle diverse smart contracts. Additionally, some blockchains have distinct technical features that can be leveraged to achieve access control. For example, Hyperledger Fabric [78] can implement fine-grained access control with its membership service providers and chaincode mechanism.

Attribute-based encryption (ABE) [66,67,79] is a widely researched approach to address the challenge of fine-grained access control in data sharing. Sahai and Waters [87] proposed ABE, also called Fuzzy identity-based encryption, in 2005. It is considered the most promising encryption primitive supporting fine-grained access control. With ABE, there's no need for the data owner to know the recipient's identity. Instead, the identity is considered as a series of attributes. The user can decrypt the ciphertext only if their attributes comply with the data owner's access policy. There are mainly two types of ABE: ciphertext-policy attribute-based encryption (CP-ABE) [70] and key-policy attribute-based encryption (KP-ABE) [88]. KP-ABE is mainly utilized in biometric identification systems, while CP-ABE is better suited for encrypted storage systems.

In addition to ABE, Proxy re-encryption is another encryption scheme that enables access control. It is a special type of asymmetric encryption, first proposed by Blaze *et al* [89]. The purpose of proxy re-encryption is mainly to transform the ciphertext that requires a particular private key for decryption to a new ciphertext that can be decrypted by a specified identity private key without exposing any private keys. This approach prevents the proxy from accessing the plaintext, thereby maintaining the confidentiality and security of the data. Figure 3 illustrates the encryption process, which mainly involves three roles: data owner (Alice), Proxy encryptor (Proxy), and data requester (Bob).

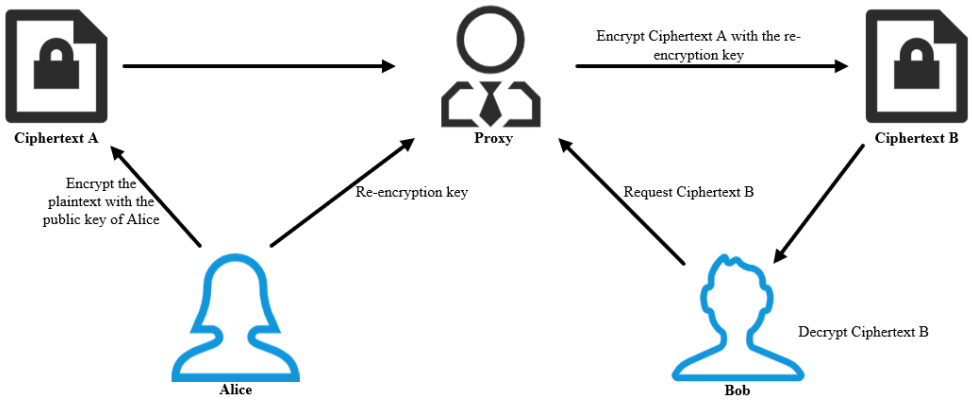

Figure 3. Process of Proxy re-encryption.

Several research papers have introduced novel and complex concepts related to this subject. For example, Miao [54] designed a sub-node query scheme that utilizes different data query methods based on node role permissions. This approach effectively guarantees the confidentiality of data.

3.3. Security

Security research often involves a back-and-forth process of discovering and fixing new vulnerabilities [9]. In the context of blockchain and IPFS, the interaction mechanism ensures the authenticity of the data. However, it's important to note that different types of blockchains offer varying degrees of security. The three primary types of blockchains are public blockchains, private blockchains, and consortium blockchains. Public blockchains are completely decentralized, transparent, and open to all nodes in the network for reading, writing, verifying, and consensus building. Participants can earn economic incentives through probabilistic consensus mechanisms. Private blockchains are characterized by high efficiency, good privacy protection, and low transaction costs, but they have less decentralization and stricter control over node authority compared to public and consortium blockchains. Consortium blockchains offer limited decentralization and allow for flexible access to nodes through authorization. Each node usually has an entity organization, and interest alliances can be formed to maintain the consortium blockchain system. Consortium blockchains perform well in terms of transaction efficiency and privacy security while allowing more flexible access to nodes [83], making them suitable for some systems that require secure data storage using blockchain and IPFS [52,73].

One of the challenges faced by blockchains is the issue of forking. Ideally, nodes involved in the data writing process would broadcast notifications to other nodes upon successfully mining a block. The recipient nodes would then verify the legality of the mined block and subsequently add it to the end of the blockchain while halting the mining process of that particular block. In this case, data consistency and legality across all nodes is guaranteed by the blockchain mechanism. However, delays in data communication caused by factors such as network issues mean that two nodes can generate legitimate block simultaneously and broadcast it to other nodes at the same time. Each node receives the block in a different order, which leads to the phenomenon of inconsistent blockchain data. The blockchain forking problem is to study how to maintain data consistency in a practical environment. Xie [90] proposed that initiating a network-wide vote by a randomly generated arbitration node when a fork occurs, which decides the legitimate branch. Liu [83] proposed a new improvement scheme based on delegated proof of stake (DPoS) in three aspects: reputation points, voting mechanism, rewards and thresholds, which improves the security of consensus and the motivation of voting nodes.

In addition, well-designed smart contracts can be utilized to realize the security of data [74,80]. However, to achieve this security, it’s crucial to ensure that the smart contract code is error-free and not vulnerable to any security threats [91]. One tool that can aid in this effort is SmartCheck [72], an open-source tool for analyzing smart contract code against known errors and vulnerabilities. By utilizing SmartCheck, developers can enhance the security of their smart contracts and ensure that they are protected against potential threats.

Encryption plays a vital role in safeguarding data in distributed systems [9]. Apart from schemes supporting access control we mentioned above, AES algorithms [77,79], ECC schemes [77] and Paillier [81] are some of the popular encryption schemes used. Many other schemes [54,66,67,73,75,80,83] are also utilized to secure the data before storing it in the interaction mechanism. Security research also needs to look at attack vectors, not just those that are known, but also potential new ones [9]. As a result, security research extends beyond blockchain and IPFS to explore new ways of protecting against attacks.

3.4. Data availability

The IPFS network lacks implicit mechanisms to ensure availability, making it vulnerable to node maintenance or failure which can result in reduced availability of stored blocks. Typically, only the data owner and the data requesters possess all the blocks of a file. Active replication is not employed as part of the protocol. If the local data of both the data owner and the data requesters is lost, the file becomes inaccessible. To address this issue, it may be necessary to create multiple copies of blocks to enhance availability. In general, P2P systems face a significant challenge with long-term availability [9].

Incentives can enhance replication mechanisms and guarantee redundancy through financial methods [9]. For storage assurance, Filecoin [65] exists. Filecoin is a public cryptocurrency and digital payment system that operates on a blockchain, enabling network participants to record their commitments. Filecoin leverages a storage and retrieval market to store and retrieve files, and manages deal execution through distributed ledgers that leverage Proof-of-Replication [92] and Proof-of-Space-Time. The storage market is responsible for storing data, while the retrieval market is responsible for retrieving data. Retrieval miners provide data in exchange for Filecoin. To promote collaboration and compensation among clients and retrieval miners, the system employs payment channels to ensure secure data retrieval and micro-payment compensation. These payment channels facilitate the exchange of data in small increments, with compensation provided before the transfer of each new piece of data.

IPFS can work independently from Filecoin [9]. As a result, it is worthwhile to explore alternative approaches for ensuring data availability, beyond relying solely on incentive mechanisms. IPFS utilizes cache-based replication, which occurs naturally in response to requests and relies on volunteers. This replication method is particularly beneficial for popular content. However, active replication can further enhance data availability. To accomplish this, a certain level of coordination and communication becomes necessary [9]. Thus, it is crucial to develop a coordinated and suitable scheme for data replication or backup.

In today's rapidly evolving information technology landscape, data backup has emerged as a critical aspect of ensuring data availability. As the diversity and volume of data continue to grow, traditional backup solutions that rely on storing data on multiple servers can become costly to maintain and manage. Furthermore, if multiple servers are placed in the same physical location, it’s possible to create a single point of failure, potentially resulting in permanent data loss during catastrophic disasters. Therefore, it is imperative to establish a robust backup solution that not only facilitates quick data recovery in the event of a disaster but also ensures seamless fulfillment of data requests from users, even in the event of server failure or data loss [93].

To address these challenges, P2P backup solutions can be explored. In a P2P network, data is distributed across multiple peers as there is no central server to store it. To prevent data loss and ensure availability, backup technologies in P2P networks use redundant storage across multiple peers, improving data sharing speed [94,95]. This approach leverages idle computing and storage resources within the network, eliminating bottlenecks associated with using a single resource and improving data transfer efficiency. There are three backup primary modes in P2P networks [3]: centralized backup, cluster backup, and replication contract backup.

In the centralized backup mode, a unified backup center is established within the P2P network, as depicted in Figure 4. The primary responsibility of the backup center is to backup data for the entire P2P

network. This approach offers the advantage of centralized storage and management of backup data, eliminating the uncertainty of data availability caused by peers in the P2P network being online or offline at any time. However, this approach also has its disadvantages. the backup center is under significant load, and any crash or failure could lead to a substantial amount of data becoming unavailable [3].

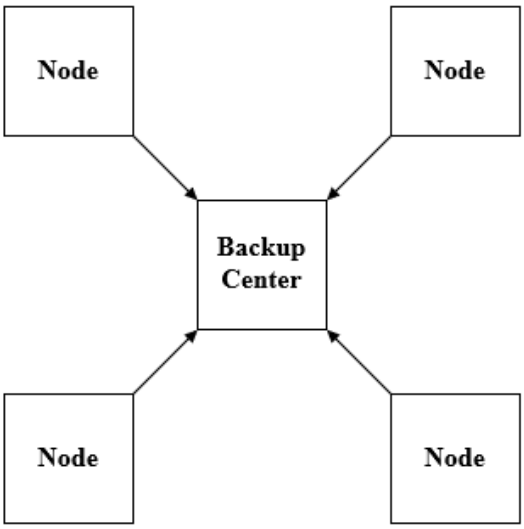

Figure 4. Centralized backup mode.

In the cluster backup mode, the network is divided into small clusters [96], as shown in Figure 5. Each cluster regularly elects a master node to manage and monitor data backup within the cluster. The master node serves as the center of the cluster, and data is backed up on each peer. It has the advantage of effective utilization of the storage space within a cluster. However, here are some drawbacks to this mode. Data needs to be frequently synchronized within a cluster, and peers in a cluster may become isolated from the outside network. If the storage capability of peers in a cluster are insufficient, due to an uneven distribution of online time or limited storage space, the backup results may be inadequate [3]. Therefore, it is important to carefully consider the storage capabilities and distribution of peers before implementing cluster backup mode.

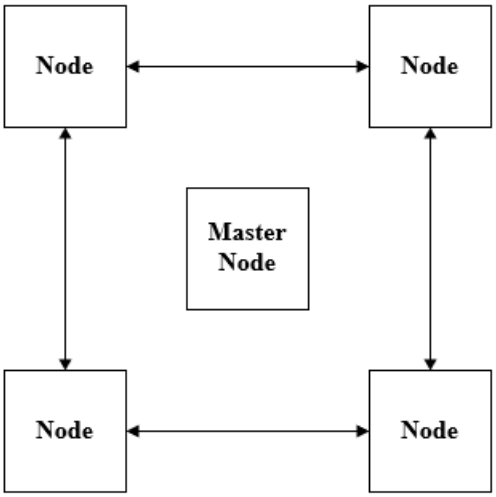

Figure 5. Cluster backup mode.

In the replication contract backup mode, as depicted in Figure 6, a replication contract [97–101] is established between two peers. Both parties store each other's backup data in a peer-to-peer transaction without requiring an incentive mechanism. One of the key advantages of the replication contract backup mode is that the protocol is straightforward, eliminating the need for incentives or frequent communication. Each peer can focus solely on its contract partner, simplifying the process. Additionally, unlike the cluster backup mode, there is no isolation of peers in this mode. However, there is one potential disadvantage to consider. When selecting a backup partner, it is necessary to send a request to the entire network until a suitable backup partner is selected. Nevertheless, if a suitable backup partner is chosen, there is no need for frequent selection of another backup partner [3]. Overall, the replication contract backup mode offers a simple and efficient backup solution, with the potential to overcome some of the challenges associated with traditional backup mechanisms.

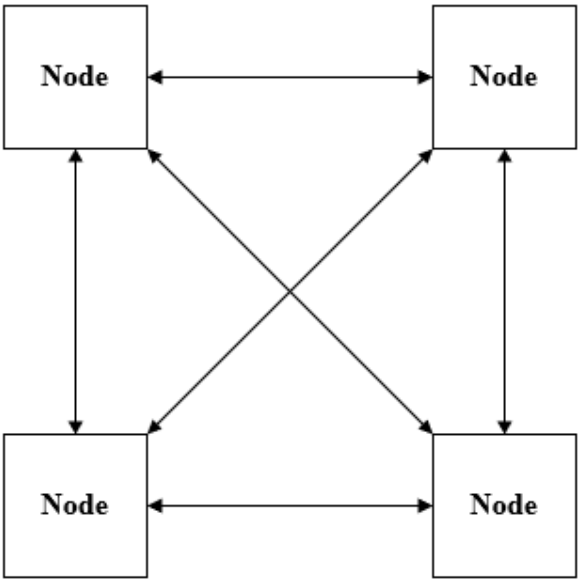

Figure 6. Replication contract backup mode.

Hao [3] proposed an IPFS backup scheme for data availability, which is a replication contract backup scheme. The scheme includes a backup node selection algorithm based on node online time complementation and a replication contract scheme based on peer assistance monitoring. The backup node selection algorithm uses historical online time data to predict the online time of nodes in the future. Based on the online time of all nodes that have backed up the data, the available time of the backup data is calculated. The degree of complementarity between the online time of nodes and the available time of backup data determines the improvement of backup data availability. Candidate backup nodes are ranked based on this degree of complementarity, and the node with the highest score is selected as the backup node in priority. Once the backup node is identified, a replication contract is established between both parties, and a replication contract scheme based on peer-assisted monitoring is used to store each other's data and update the backup data at any time. During the replication contract fulfillment process, a monitoring mechanism based on peer-assisted monitoring is utilized to ensure the fulfillment of the agreement. If the monitoring mechanism detects that the backup node does not fulfill the agreement, the replication contract is canceled, and new backup nodes will be reselected based on the backup node selection algorithm.

3.5. Data look-up

In IPFS, files are split into blocks and organized into a Merkle DAG, where the root node contains enough information to retrieve the entire file. The process of file look-up is opportunistic, meaning that peers are queried without prior knowledge of their possession of the blocks or file. If the opportunistic request fails, a backup lookup is performed using the DHT. The root node is stored on the blockchain, and data look-up on the blockchain is typically done by calling smart contracts [73,74,77]. Some researchers have proposed other intriguing methods, such as the query verification algorithm of Hao *et al*. [24] for a blockchain-based double-chain storage structure, which aims to verify the authenticity of data during retrieval. Additionally, the sub-node query scheme of Miao [54], which we mentioned above, uses different data query methods based on node role permissions, which can improve efficiency when querying different nodes.

4. Conclusion

Blockchains are decentralized digital ledgers that offer data immutability without relying on third-party intermediaries. Yet, blockchains suffer from problems such as limited on-chain storage capacity. To overcome this limitation, IPFS has emerged as the mainstream platform for off-chain storage. Moreover, the integration of blockchain with IPFS can offer a range of benefits beyond just storage. Blockchains can ensure the authenticity and integrity of data stored on IPFS, providing a mechanism for access control, and enabling secure sharing of information. This interaction mechanism between blockchain and IPFS combines the strengths of both technologies, mitigating each other's weaknesses and offering new opportunities for decentralized applications.

In this survey paper, we have examined the interaction mechanism between blockchain and IPFS, focusing on identifying open research areas. From our qualitative comparison of twenty applications that leverage the interaction mechanism, we can conclude that these applications utilize varying solutions to address access control, security, and data look-up challenges. Most notably, encryption schemes supporting fine-grained access control, e.g., CP-ABE, seem to be ubiquitous to effectively regulate data access while maintaining retrieval speed. We also see different measurements to ensure feasible access control and data security, *i.e*., well-designed smart contracts.

Improving the data availability of IPFS beyond incentive mechanisms is an important open task that lacks sufficient research. Especially, suitable data backup schemes are difficult to design or choose in certain scenarios. We can consider taking inspiration from existing data backup approaches or availability mechanisms used by different data networks. In general, the interaction mechanism has become part of the research agenda, either as the foundation for other applications or as a focal point for investigation. Nonetheless, numerous challenges and unresolved queries persist. Thus, we are optimistic that the interaction mechanism between blockchain and IPFS will present a plethora of thrilling research prospects.

Acknowledgments

The work described in this paper was supported by the National Natural Science Foundation of China (U22B2032).

Conflicts of Interests

The authors declare no conflict of interest.

Authors' Contribution

Conceptualisation, F.Y.; Investigation, Z.D.; Writing – original draft, Z.D.; Writing – review & editing, Z.D.; Project administration, F.Y., Z.D.; Supervision, F.Y., Y.S. All authors have read and agreed to the published version of the manuscript.

References

[1] Nakamoto S. Bitcoin: A peer-to-peer electronic cash system. 2009. https://bitcoin.org/bitcoin.pdf (accessed on 10 January 2023).
[2] Pilkington M. Blockchain technology: principles and applications In *Research Handbook on Digital Transformations.* Cheltenham: Edward Elgar Publishing, 2016: 15-18.
[3] Hao JT. Research and implementation of distributed e-commerce system based on IPFS. Master Degree, Beijing University of Posts and Telecommunications, China, 2019.
[4] Tschorsch F, Scheuermann B. Bitcoin and beyond: A technical survey on decentralized digital currencies. *IEEE Commun. Surv.*, 2016, 18(3): 2084-2123.
[5] Gervais A, Karame G O, Wüst K, Glykantzis V, Ritzdorf H, Capkun S. On the security and performance of proof of work blockchains. In *Proceedings of the 23rd ACM SIGSAC Conference on Computer and Communications Security*, Vienna, Austria, October 24, 2016, pp. 3-16.
[6] Joseph B, Miller A, Clark J, Narayanan A, Kroll JA, Felten EW. Sok: Research perspectives and challenges for bitcoin and cryptocurrencies. In *Proceedings of the 36th IEEE Symposium on Security and Privacy*, San Jose, CA, USA, May 17-21, 2015, pp. 104-121.
[7] Wood G. Ethereum. A secure decentralised generalized transaction ledger. Available: http://gavwood.com/Paper.pdf (accessed on 10 January 2023).
[8] Norvill R, Pontiveros BBF, State R, Cullen A. IPFS for reduction of chain size in ethereum. In *Proceedings of the 2018 IEEE International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data*, Halifax, NS, Canada, July 30-August 3, 2018, pp. 1121-1128.
[9] Daniel E, Tschorsch F. Ipfs and friends: a qualitative comparison of next generation peer-to-peer data networks. *IEEE Commun. Surv. Tutor.* 2022, 24(1): 31-52.
[10] Saroiu S, Gummadi P K, Gribble S D. Measurement study of peer-to-peer file sharing systems. In *proceedings of Multimedia Computing and Networking 2002*, San Jose, CA, USA, December 10, 2001, pp. 156-170.
[11] Clarke I, Sandberg O, Wiley B, Hong T W. Freenet: A distributed anonymous information storage and retrieval system. In *Proceedings of the International Workshop on Designing Privacy Enhancing Technologies: Design Issues in Anonymity and Unobservability*, Berkeley, CA, USA, July 25-26, 2000, pp. 46-66.
[12] Stoica I, Morris R, Karger D, Kaashoek M F, Balakrishnan H. Chord: A scalable peer-to-peer lookup service for internet applications. In *Proceedings of the 2001 ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, San Diego, CA, USA, August 27, 2001, pp. 149-160.

[13] Ratnasamy S, Francis P, Handley M, Karp R, Shenker S. A scalable content-addressable network. In *Proceedings of the 2001 ACM Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, San Diego, CA, USA, August, 2001, pp. 161–172.
[14] Rowstron A, Druschel P. Pastry: Scalable, decentralized object location, and routing for large-scale peer-to-peer systems. In *Proceedings of the 2001 IFIP/ACM International Conference on Distributed Systems Platforms*, Heidelberg, Germany, November 12-16, 2001, pp. 329-350.
[15] Cohen B. Incentives build robustness in bittorrent. 2003. Available: http://www.scs.stanford.edu/10au-cs144/sched/readings/bittorrentecon.pdf (accessed on 10 January 2023).
[16] Benet J. IPFS - content addressed, versioned, P2P file system (draft 3). *arXiv* 2014, arXiv: 1407.3561.
[17] Yin L, Wang W. Research on Distributed Data Sharing System Based on IPFS (in Chinese). *IOTT* 2016, 6(006): 60-62.
[18] Wang S, Zhang Y, Zhang Y. A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems. *IEEE Access* 2018, 6: 38437-38450.
[19] Steichen M, Fiz B, Norvill R, Shbair W, State R. Blockchain-based, decentralized access control for IPFS. In *Proceedings of the 2018 IEEE International Conference on Internet of Things and IEEE Green Computing and Communications and IEEE Cyber, Physical and Social Computing and IEEE Smart Data*, Halifax, NS, Canada, July 30-August 3, 2018, pp. 1499-1506.
[20] Battah A A, Madine M M, Alzaabi H, Yaqoob I, Salah K, Jayaraman R. Blockchain-based multi-party authorization for accessing ipfs encrypted data. *IEEE Access* 2020, 8: 196813-196825.
[21] Nyaletey E, Parizi R M, Zhang Q, Choo KKR. Blockipfs - blockchain-enabled interplanetary file system for forensic and trusted data traceability. In *Proceedings of the 2019 International Conference on Blockchain*, Atlanta, GA, USA, July 14-17, 2019, pp. 18-25.
[22] Xu J, Xue K, Li S, Tian H, Hong J, Hong P, Yu N. Healthchain: A blockchain-based privacy preserving scheme for large-scale health data. *IEEE Internet Things J* 2019, 6(5): 8770-8781.
[23] Kumar R, Marchang N, Tripathi R. Distributed off-chain storage of patient diagnostic reports in healthcare system using IPFS and blockchain. In *Proceedings of 2020 International Conference on COMmunication Systems & NETworkS*, Bengaluru, India, January 7-11, 2020, pp. 1-5.
[24] Hao JT, Sun Y, Luo H. A Safe and Efficient Storage Scheme based on BlockChain and IPFS for Agricultural Products Tracking. *J Comput* 2018, 29(6): 158-167.
[25] Ali M S, Dolui K, Antonelli F. Iot data privacy via blockchains and IPFS. In *Proceedings of the 7th International Conference on the Internet of Things*, Linz, Austria, October 22-25, 2017, pp. 1- 7.
[26] Trón V. The book of swarm. 2020. Available: https://www.ethswarm.org/The-Book-of-Swarm.pdf (accessed on 10 January 2023).
[27] Ogden M, McKelvey K, Madsen M B. Dat - distributed dataset synchronization and versioning. 2017. Available: http://terrymarine.com/wp-content/uploads/2017/07/724d7267d90052778b0530807512474b.pdf.
[28] Lambert N, Bollen B. The safe network a new, decentralised internet. 2014. Available: https://docs.maidsafe.net/Whitepapers/pdf/TheSafeNetwork.pdf.
[29] Storj Labs Inc. Storj: A decentralized cloud storage network framework. 2018. Available: https://www.storj.io/storj.pdf (accessed on 10 January 2023).
[30] Williams S, Diordiiev V, Berman L, Raybould I, Uemlianin I. Arweave: A protocol for economically sustainable information permanence. *Comput. Sci.* 2019.
[31] Libp2p. Available: https://github.com/libp2p (accessed on 10 January 2023).
[32] Maymounkov P, Mazières D. Kademlia: A peer-to-peer information system based on the XOR metric. In *Lecture Notes in Computer Science* Proceedings of the 1st International Workshop on Peer-to-Peer Systems, Cambridge, MA, USA, March, 2002, pp. 53–65.
[33] Henningsen S, Florian M, Rust S, Scheuermann B. Mapping the interplanetary filesystem. In *Proceedings of the 19th IFIP Networking Conference*, Paris, France, June 22-26, 2020, pp. 289–297.
[34] Protocol Labs. IPFS. Available: https://github.com/ipfs (accessed on 10 January 2023).

[35] Protocol Labs. Filecoin: A decentralized storage network. Available: https://filecoin.io/filecoin.pdf (accessed on 10 January 2023).
[36] BitTorrent Foundation. Bittorrent (btt) white paper. Available: https://www.allcryptowhitepapers.com/bittorrent-whitepaper-btt/ (accessed on 10 January 2023).
[37] Ethersphere. Available: https://github.com/ethersphere (accessed on 10 January 2023).
[38] Hypercore Protocol developers. Hypercore protocol. Available: https://github.com/hypercore-protocol (accessed on 10 January 2023).
[39] MaidSafe. Safe network. Available: https://github.com/safenetwork (accessed on 10 January 2023).
[40] Storj Labs. Available: https://github.com/Storj (accessed on 10 January 2023).
[41] ArweaveTeam. Arweave. Available: https://github.com/ArweaveTeam (accessed on 10 January 2023).
[42] De la Rocha A, Dias D, Psaras Y. Accelerating content routing with bitswap: A multi-path file tra nsfer protocol in IPFS and filecoin. 2021. Available: https://research.protocol.ai/publications/accelerating-content-routing-with-bitswap-a-multi-path-file-transfer-protocol-in-ipfs-and-filecoin/.
[43] Protocol Labs. FAQ | IPFS Docs. Available: https://docs.ipfs.tech/concepts/faq/ (accessed on 9 May 2023).
[44] Trautwein D, Raman A, Tyson G, Castro I, Scott W, *et al.* Design and evaluation of IPFS: a storage layer for the decentralized web. In *Proceedings of the ACM SIGCOMM 2022 Conference*, Amsterdam, Netherlands, August 22-26, 2022, pp. 739-752.
[45] Pouwelse J, Garbacki P, Epema D, Sips H. The bittorrent P2P file-sharing system: Measurements and analysis. In *Lecture Notes in Computer Science* Proceedings of the 4th International Workshop on Peer-To-Peer Systems, Ithaca, NY, USA, February, 2005, pp. 205-216.
[46] Bharambe A R, Herley C, Padmanabhan V N. Analyzing and improving a bittorrent networks performance mechanisms. In *Proceedings of the 25th IEEE International Conference on Computer Communications*, Barcelona, Catalunya, Spain, April 23-29, 2006, pp. 1-12.
[47] Xia R L, Muppala J K. A survey of bittorrent performance. *IEEE Commun. Surv. Tutor.* 2010, 12(2): 140-158.
[48] Ramanathan S, Hossain A, Mirkovic J, Yu M, Afroz S. Quantifying the impact of blocklisting in the age of address reuse. in IMC '20: Proceedings of the ACM Internet Measurement Conference, Virtual Event, USA, Oct. 2020. New York: Association for Computing Machinery, 2020, pp. 360-369.
[49] Huang H, Lin J, Zheng B, Zheng Z, Bian J. When blockchain meets distributed file systems: An overview, challenges, and open issues. *IEEE Access* 2020, 8: 50574-50586.
[50] Web3 Foundation. Available: https://web3.foundation/about/ (accessed on 10 January 2023).
[51] Johnson N. ERC-137: Ethereum Domain Name Service – Specification. Available: https://eips.ethereum.org/EIPS/eip-137 (accessed on 10 January 2023).
[52] Xu S, Ford B, Estrada-Galiñanes V. Dissecting IPFS and Swarm to demystify distributed decentralized storage networks. 2023. Available: https://www.epfl.ch/labs/dedis/wp-content/uploads/2023/01/report_2022-3-SixiaoXu-DissectingIPFSandSwarm.pdf.
[53] Keall D. How dat works. Available: https://dat-ecosystem-archive.github.io/how-dat-works/ (accessed on 31 January 2021).
[54] Nabben K. Decentralized Technology in Practice: Social and technical resilience in IPFS. In *Proceedings of the 42nd International Conference on Distributed Computing Systems Workshops (ICDCSW)*, Bologna, Italy, July 10, 2022, pp. 66-72.
[55] MaidSafe. The safe network primer. Available: https://primer.safenetwork.org (accessed on 10 January 2023).
[56] Irvine D. Self-authentication. Available: https://docs.maidsafe.net/Whitepapers/pdf/SelfAuthentication.pdf (accessed on 10 January 2023).
[57] Irvine D. Self encrypting data. Available: https://docs.maidsafe.net/Whitepapers/pdf/SelfEncryptingData.pdf (accessed on 10 January 2023).
[58] Paul G, Hutchison F, Irvine J. Security of the maidsafe vault network. *Wireless World Research Forum Meeting 32*, Marrakech, Morocco, May 20-22, 2014.

[59] Jacob F, Mittag J, Hartenstein H. A security analysis of the emerging p2p-based personal cloud platform maidsafe. In *Proceedings of the 2015 IEEE Trustcom/BigDataSE/ISPA*, Helsinki, Finland, August 20-22, 2015, pp. 1403-1410.
[60]Protocol Labs. IPFS. Available: https://github.com/ipfs/ipfs (accessed on 9 May 2023).
[61]Network Research and Infrastructure Group. SAFE. Available: https://github.com/RENCI-NRIG (accessed on 9 May 2023).
[62] Reed IS, Solomon G. Polynomial codes over certain finite fields. *J. Appl. Ind. Math.* 1960, 8(2): 300-304.
[63] De Figueiredo S, Madhusudan A, Reniers V, Nikova S, Preneel B. Exploring the storj network: A security analysis. In *Proceedings of the 36th ACM/SIGAPP Symposium On Applied Computing*, Gwangju, Korea, March 22-26, 2021, pp. 257-264.
[64] Zhang X, Grannis J, Baggili I, Beebe N L. Frameup: An incriminatory attack on storj: A peer to peer blockchain enabled distributed storage system. *Digit Investig* 2019, 29: 28–42.
[65] Corallo M . Compact block relay. Available: https://github.com/bitcoin/bips/blob/master/bip-0152.mediawiki (accessed on 10 January 2023).
[66] Sun J, Yao XM, Wang SP, Wu Y. Blockchain-based secure storage and access scheme for electronic medical records in IPFS. *IEEE Access* 2020, 8: 59389-59401.
[67] Sun J, Yao XM, Wang SP, Wu Y. Non-repudiation storage and access control scheme of insurance data based on blockchain in IPFS. *IEEE Access* 2020, 8: 155145-155155.
[68] Wei X, Li X. Anti-collusion data auction mechanism based on smart contract. *Inf. Sci.* 2021, 555: 386-409.
[69] Wei X, Li X. Smart contract based data trading mode using blockchain and machine learning. *IEEE Access* 2019, 7: 102331-102344.
[70] Bethencourt J, Sahai A, Waters B. Ciphertext-policy attribute-based encryption. In *Proceedings of Symposium on security and privacy (SP'07)*, Berkeley, CA, USA, May 20-23, 2007, pp. 321-334.
[71] Szabo N. Formalizing and securing relationships on public networks. *First Monday* 1997, 2(9).
[72] Tikhomirov S, Voskresenskaya E, Ivanitskiy I, *et al.* Smartcheck: Static analysis of ethereum smart contracts. In *Proceedings of the 1st International Workshop on Emerging Trends in Software Engineering for Blockchain*, Gothenburg, Sweden, May 27, 2018, pp. 9-16.
[73] Meng N, Zhang S. University education resource sharing based on blockchain and IPFS. In *Big Data Analytics for Cyber-Physical System in Smart City* Proceedings of the Big Data Analytics for Cyber-Physical-Systems, Shanghai, China, December 28-29, 2020, pp. 1808-1813.
[74] Huang Z, Zhang X, Zhao J, Zou H. Design of knowledge sharing mechanism based on blockchain (in Chinese). *J Chongqing Univ Technol* 2021, 35(09): 143-151.
[75] Li L, Yue Z, Wu G. Electronic medical record sharing system based on hyperledger fabric and interplanetary file system. In *Proceedings of the 5th International Conference on Compute and Data Analysis*, Sanya, China, February 2-4, 2021, pp. 149-154.
[76] Androulaki E, Barger A, Bortnikov V, Cachin C, Christidis K, *et al.* Hyperledger fabric: a distributed operating system for permissioned blockchains. In *Proceedings of the thirteenth EuroSys conference*, Porto, Portugal, April 23-26, 2018, pp. 1-15.
[77] Ding Y, Xiang H, Luo D, Zou X, Liang H. Scheme for electronic certificate storage by combining Fabric technology (in Chinese). *J Xidian Univ* 2020, 47(05): 113-121+158.
[78] Daemen J, Rijmen V. The design of rijndael: aes-the advanced encryption standard. Berlin, Germany: Springer, 2013.
[79] Li H, Pei L, Liao D, Chen S, Zhang M, *et al.* FADB: A fine-grained access control scheme for VANET data based on blockchain. *IEEE Access* 2000, 8(1): 85190-85203.
[80] Li Z. Design and implementation of high-quality educational resource platform in colleges and universities based on blockchain technology (in Chinese). *Electro Technol & Softw Eng* 2001, (07): 160-161.
[81] Wang G, Ding H. Blockchain based scheme for secure sharing of business collaboration data (in Chinese). *J. Inf. Secur. Res.* 2021, 7(07): 606-614.

[82] Li X, Peng C, Li M, Li G, Tao J. Statistical data processing based on homomorphic encryption (in Chinese). *Cyberspace Secur* 2021, 6(07): 22-25+28.
[83] Liu W, Li Y, Tian Z, Peng Y, She W. IDDS: double-chain structure infectious disease data sharing blockchain model. *Appl Res Comput* 2021, 38(03): 675-679.
[84] Li X, He Q, Jiang B, Qin X, Qin K, *et al*. Bts-pd: a blockchian based traceability system for p2p distribution. In *Blockchain and Trustworthy Systems* Proceedings of the International Conference on Blockchain and Trustworthy Systems, Guangzhou, China, December 7-8, 2019, pp. 607-620.
[85] Khatal S, Rane J, Patel D, Patel P, Busnel Y. Fileshare: A blockchain and ipfs framework for secure file sharing and data provenance. In *Advances in Machine Learning and Computational Intelligence* Proceedings of the MoSICom '20: International Conference on Modelling, Simulation & Intelligent Computing, Dubai, United Arab Emirates, January 29-31, 2020, pp. 82-833.
[86] Hoang VH, Lehtihet E, Ghamri-Doudane Y. Privacy-preserving blockchain-based data sharing platform for decentralized storage systems. In *Proceedings of the 19th IFIP Networking Conference*, Paris, France, June 22-26, 2020, pp. 280-288.
[87] Sahai A, Waters B. Fuzzy identity-based encryption. Annual international conference on the theory and applications of cryptographic techniques. In *Advances in Cryptology – EUROCRYPT 2005* Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques, Aarhus, Denmark, May 22-26, 2005, pp. 457-473.
[88] Goyal V, Pandey O, Sahai A, Waters B. Attribute-based encryption for fine-grained access control of encrypted data. In *Proceedings of the 13th ACM conference on Computer and communications security*, Alexandria, VA, USA, October 30- November 3, 2006, pp. 89-98.
[89] Blaze M, Bleumer G, Strauss M. Divertible protocols and atomic proxy cryptography. In *Advances in Cryptology-EUROCRYPT'98* Proceedings of the Theory and Applications of Cryptographic Techniques, Espoo, Finland, May 31-June 4, 1998, pp. 127-144.
[90] Xie Y. Research on block chain and furcations. Master Degree, South China University of Technology, China, 2018.
[91] Hasan H R, Salah K. Proof of delivery of digital assets using blockchain and smart contracts. *IEEE Access* 2018,6: 65439-65448.
[92] Benet J, Dalrymple D, Greco N. Proof of replication. 2017. Available: https://filecoin.io/proof-of-replication.pdf (accessed on 10 January 2023).
[93] Gao P. Discussion on data backup mode based on cloud computing (in Chinese). *China Inf.* 2021,6: 53-54.
[94] Toka L, Dell'Amico M, Michiardi P. Online data backup: a peer-assisted approach. In *Proceedings of the Tenth International Conference on Peer-to-Peer Computing (P2P)*, Delft, Netherlands, August 25-27, 2010, pp. 1-10.
[95] Defrance S, Kermarrec A M, Merrer E L, Scouarnec N L, Straub G, *et al*. Efficient peer-to-peer backup services through buffering at the edge. In *Proceedings of the IEEE International Conference on Peer-to-peer Computing*, Kyoto, Japan, August 31- September 2, 2011, pp. 142-151.
[96] Nguyen D N, Tran X H, Nguyen H S. A cluster-based file replication scheme for DHT-based file backup systems. In *Proceedings of the 2016 International Conference on Advanced Technologies for Communications*, Hanoi, Vietnam, October 12-14, 2016, pp. 204-209.
[97] Skowron P, Rzadca K. Exploring heterogeneity of unreliable machines for p2p backup. In *Proceedings of the High Performance Computing and Simulation (HPCS)*, Helsinki, Finland, July 1-5, 2013, pp. 91-98.
[98] Bernard S, Fessant F L. Optimizing peer-to-peer backup using lifetime estimations. In *Proceedings of the 2009 EDBT/ICDT '09 joint conference*, Saint-Petersburg, Russia, March 22, 2009, pp. 26-33.
[99] Cox L P, Noble B D. Samsara: honor among thieves in peer-to-peer storage. *ACM SIGOPS Oper Syst Review* 2003, 37(5): 120-132.
[100] Skowron P, Rzadca K. Flexible replica placement for optimized P2P backup on heterogeneous, unreliable machines. *Concurr Comput* 2016, 28(7): 2166-2186.

[101]Rzadca K, Datta A, Buchegger S. Replica placement in P2P storage: complexity and game theoretic analyses. In *Proceedings of the 30th International Conference on Distributed Computing Systems*, Genoa, Italy, June 21-25, 2010, pp. 599-609.