Off-chain Data Storage Blockchain using Interplanetary File System

T.S. Vasughi[1] , P.Muthulakshmi[2] D. Gritto[3]

[1] Research Scholar, Department of Computer Science, CSH, SRM Institute of Science and Technology,Kattankulathur, India.
[2]Associate Professor, Department of Computer Science, CSH, SRM Institute of Science and Technology, Kattankulathur, India
[3]Assistant Professor, SRM Arts and Science College,Kattankulathur, India.

Abstract

The data stored on-chain must be immutable and unmodifiable that has unique signifiers in the blockchain network. The primary challenge for storing enormous data in the blockchain network will face the scalability and computational overhead issue. Off-chaining enhances the blockchain scalability Solution to store and access the data through a hash pointer. Interplanetary File System (IPFS) is content-based addressing that creates a new hash for each file and stores the hashes in the blockchain instead of storing a huge file. IPFS network reduces the storage of transaction size of the block in the blockchain.

Keywords: Blockchain, Off-chain, Interplanetary File System, Content-based addressing.

I. INTRODUCTION

Many applications are designed on the distributed file storage using blockchain technology to ensure immutability and security, a decentralized database that stores the transaction in a peer-to-peer network. Trusted a third party to store the personal data and sensitive data may lead to attacks and misuse of the user data without compromising the security. Developing legal laws and regulatory decisions about collecting, sharing, and storing sensitive data in the blockchain with a decentralized platform.[1] A data management system ensures the users own and control the data. The blockchain protocol recognizes the user's data to grant permission to access by the access control manager.The traditional peer-to-peer systems include Distributed Hash Table(DHT), BitTorrent, Git, and SFS. [2]BitTorrent is a peer-to-peer file-sharing system to distribute the file among each untrusted peer using a quasi tit-for-tat strategy to reward the honest user and punish the malicious user that leeches other's resources.The DHT and BitSwap allow IPFS to link the cryptographic hashes using Merkle Directed Acyclic Graph (DAG) with these properties content addressing, Tamper resistance, and Deduplication.[3]The integration of 5G and industry 4.0 and Blockchain data storage systems has ensured security and increased the amount of data transfer.5G mobile networks increase channel capacity and improve network efficiency compared to other generations.However, the storage space and access the information from other peers is an issue for many applications.

Interplanetary File System and the blockchain will solve the problem of storage space and access among the peers of transaction of the block.The consensus algorithm of the blockchain validates the storage and retrieval process to avoid users injecting fake transactions into the ledger.

The remaining paper is organized as follows: Section II describes related work, Section III describes the Blockchain Techonology and IPFS, and Section IV concludes the paper.

II. RELATED WORK

Traditional data storage involves a third party for data storage and is untrustworthy, and insecure, non-transparent, unreliable. The proposed system exploits the benefit of an Interplanetary File System.IPFS eliminates the limitation of storage space in the blockchain network and ensures data availability and reliability of the network.The high demand for storage space and bandwidth to synchronize data of many nodes in the network.[4]The miner's store all the transaction data in the IPFS. IPFS produces a hash of transactions into the block. A new node is attached to the network, and complete data synchronization is necessary for the blockchain network. For the mining process, build an unspent output of all transactions pool for the transaction verification process.IPFS network reduces the storage of transaction size of the block in the blockchain.the transaction data deposited in the IPFS by miners and return the hash of the transactions and store it in the block. The evaluation of the synchronization speed of all nodes in terms of good performance. The compression ratio can reach 0.0817. [5]Content-based addressing in IPFS implications it easy to access transactions, distributes the content to other peers in the network, and increases the integrity of transactions.[6]IPFS stores the multimedia objects such as images and video as transactions to prevent copyright infringement. The copyright information circulates to all the peers in the blockchain network, and the copyright infringement verifies in the blockchain distributed ledger.The data storage management and privacy and security of information in a centralized manner for cloud-based storage may lead to a single point of failure.[7]The cluster node of IPFS ensures the authentication of the healthcare devices. The cluster layer is proposed for device-generated data in hash-based storage and securely transmitting over the consortium blockchain.[8]IPFS is used to store the data using the Secure hashing algorithm SHA-256 for the service verification scheme. The miner node uses the smart contract as a service dispute between the client and service provider to maintain the distributed ledger of the blockchain network.[9]The combination of IPFS and Blockchain addresses the problem of high throughput for individual users of content service providers.

III. BLOCKCHAIN TECHNOLOGY

The core concept of blockchain technology of digital cryptocurrencies such as Bitcoin was first introduced by Nakamoto in 2009. [10]Blockchain is a digital distributed ledger technology, decentralized, tamper-resistant, immutable based on cryptographic algorithms. Digital information recorded in a chain structure called blocks. Each block keeping a set of transactions validated through consensus algorithm to establish data provenance. A chain of data blocks connected by hashes that allows peer to peer who may not necessarily trust each other to share information.Core components of blockchain are:

Node: Any computer connected to the blockchain network can validate the transaction. A node can contain an entire history of all transactions of a specific blockchain.

Block: A data structure keeps a set of valid transactions in a distributed ledger.

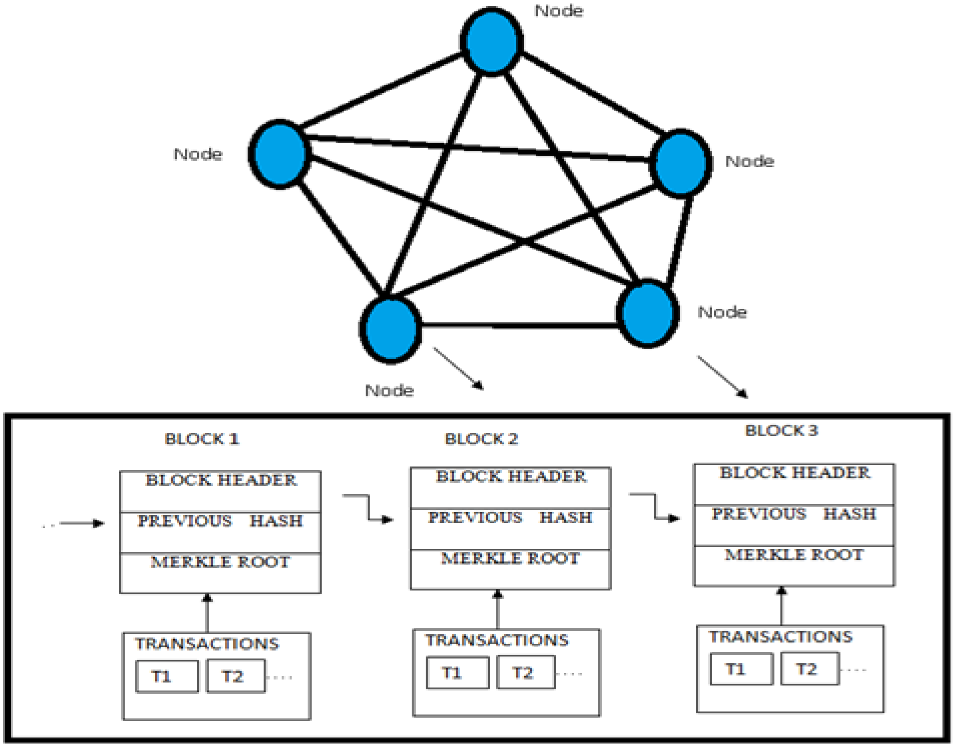
Chain: A sequence of blocks arranged on the blockchain in chronological order.

Transaction: Each transaction on the blockchain is verified and added to the block. Once a transaction is stored in a Block. It cannot be altered or modified.

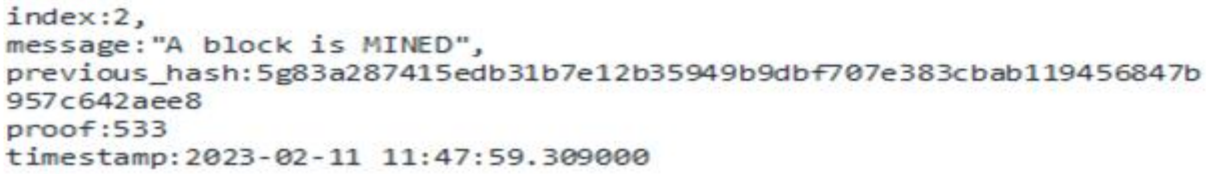
Consensus: Creating a new block to validate the transactions through consensus protocols. Some of the most common consensus protocols are Proof of Work (PoW), Proof of Stake (PoS), Practical Byzantine Fault Tolerance (PBFT), etc.

Merkle Tree: The hash pointer of all transactions in a block.

Public Key Infrastructure: Uniquely identify the blocks and transactions of the participating nodes of the blockchain network using the cryptographic keys. The content of each block validates by the hash function.



Blockchain Architecture



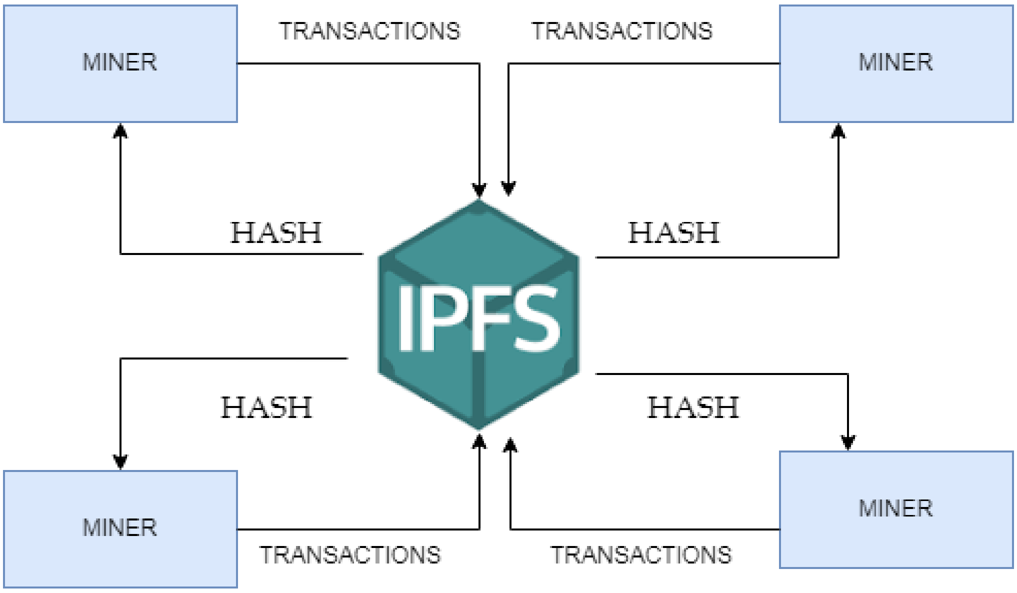
```
index:2,
message:"A block is MINED",
previous_hash:5g83a287415edb31b7e12b35949b9dbf707e383cbab119456847b
957c642aee8
proof:533
timestamp:2023-02-11 11:47:59.309000
```

IV. OFF-CHAIN DATA STORAGE

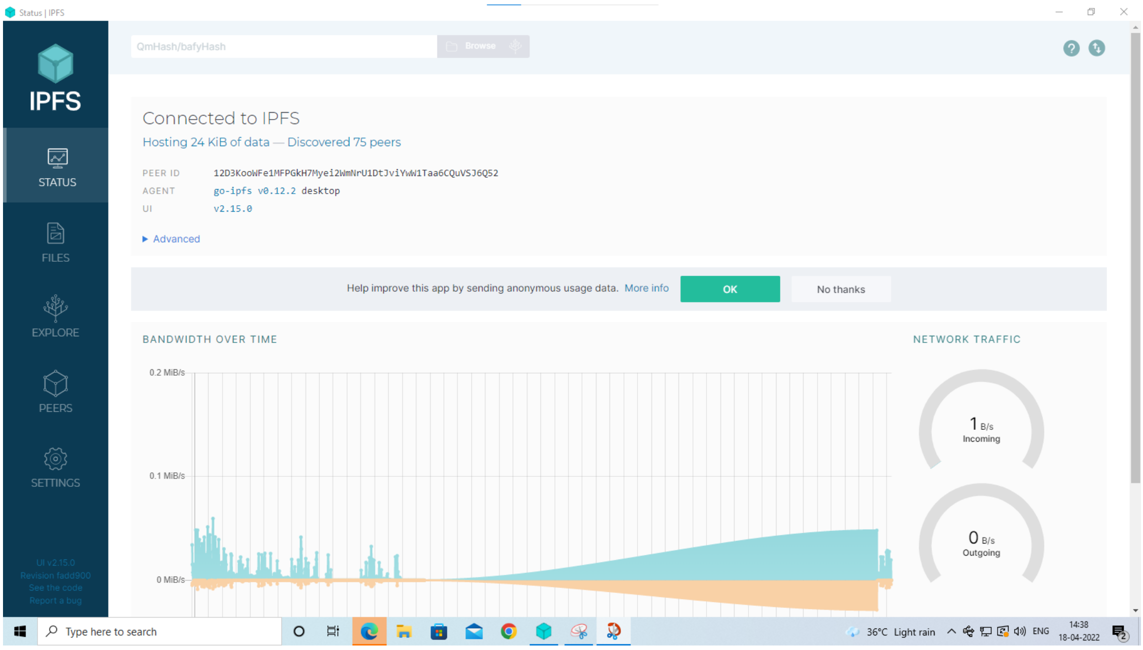
The data stored on-chain must be immutable and unmodifiable that has unique signifiers in the blockchain network. The primary challenging for storing a large datas in the blockchain network which face the scalability and computational overhead issue. Off-chaining enhance the blockchain scalability Solution to store and access the data through a hash pointer. The different security credentials such as cryptographic public-private key pair protecting off-chain data.Off-chain systems required proper back up to avoid single point of failure must maintain certain functionality with high accessibility.The most popular and well-established platforms for decentralized storage systems such as Interplanetary File System, Stroj, SWARM, and Sia.

Interplanetary File System (IPFS)

Interplanetary File System (IPFS) is one of the important off-chain storage approaches. IPFS is a content based addressed data storage protocol, allocating a unique hash for each stored file. The hash will modify every time when it is updated which is 46 bytes long.All transaction data is stored in the IPFS by the miners. IPFS inserts a hash of transactions into each block. The new node is added to the network, and the blockchain network requires complete data synchronization. Create an unspent output pool of all transactions for the transaction verification procedure for the mining process. The IPFS network minimizes the amount of the block's transaction storage on the blockchain. Miners put transaction data in IPFS, return the hash of the transactions, and store it in the blocks.



The IPFS distributed bandwidth is classified the network traffic into incoming and outgoing. The IN bandwidth specifies the transactions stored into IPFS and OUT bandwidth specfies the access from IPFS.



Connected to IPFS

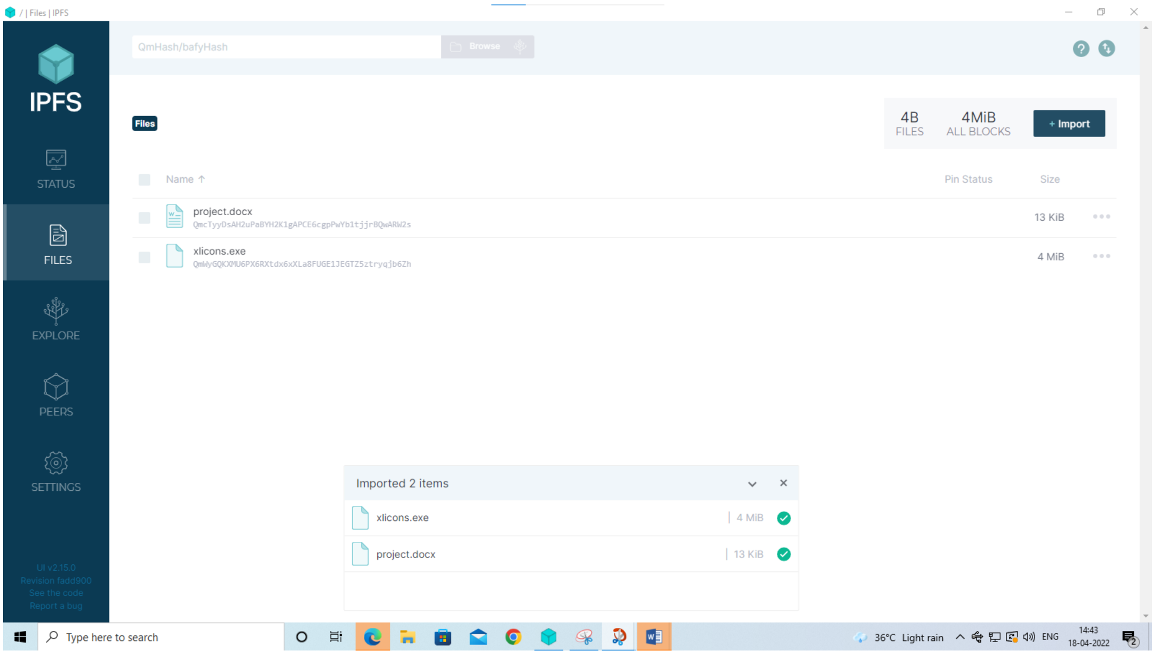
Hosting 24 KiB of data — Discovered 75 peers

```
PEER ID   12D3KooWFe1MFPGkH7Myei2WmNrU1DtJviYwW1Taa6CQuVSJ6Q52
AGENT     go-ipfs v0.12.2 desktop
UI        v2.15.0
```

▸ Advanced

Help improve this app by sending anonymous usage data. More info

The content-based address does not reveal the location of the transaction in contrast with the location-based address keeping the transactions on IPFS distributed storage. At the time of the mining process, the transactions on IPFS access the URL “localhost:8080/ipfs/hash of file”.

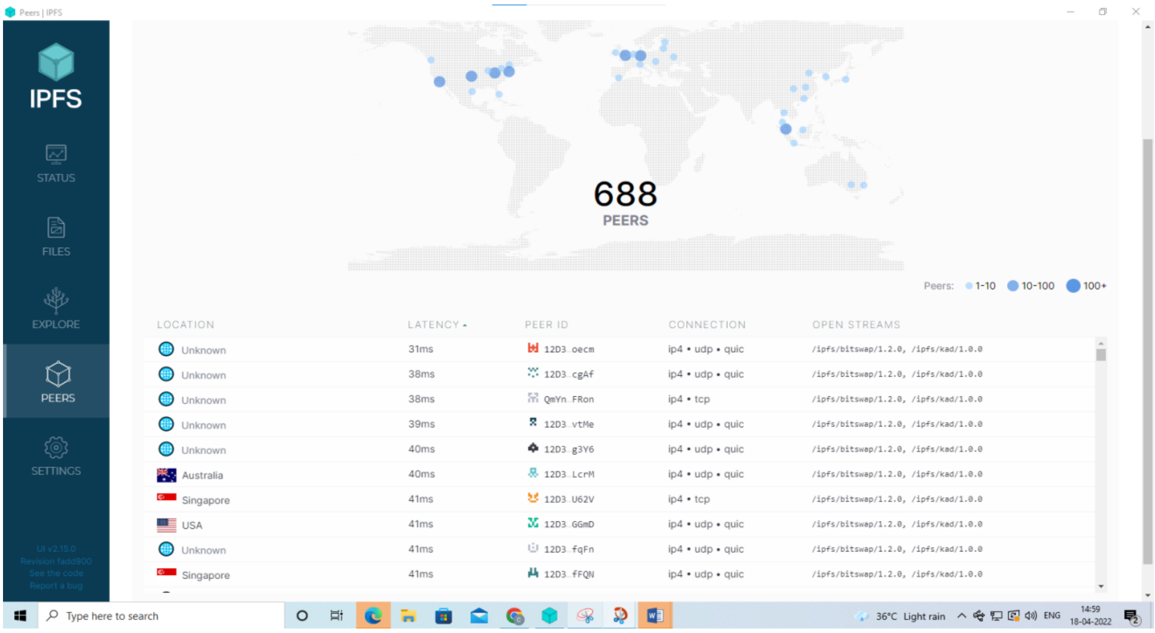


Name ↑	Pin Status	Size
project.docx QmcTyyDsAH2uPaBYH2K1gAPCE6cgpPwYb1tjjrBQwARW2s		13 KiB
xlicons.exe QmWyGQKXMU6PX6RXtdx6xXLa8FUGE1JEGTZ5ztryqjb6Zh		4 MiB

Imported 2 items

xlicons.exe	4 MiB ✓
project.docx	13 KiB ✓

The connection of IPFS shows the detail of all peers connected at a time and resources access in a distributed environment. All the peers must have the Peer ID, Location, Latency, Connection, and Open streams.



LOCATION	LATENCY ▴	PEER ID	CONNECTION	OPEN STREAMS
Unknown	31ms	12D3…oecm	ip4 • udp • quic	/ipfs/bitswap/1.2.0, /ipfs/kad/1.0.0
Unknown	38ms	12D3…cgAf	ip4 • udp • quic	/ipfs/bitswap/1.2.0, /ipfs/kad/1.0.0
Unknown	38ms	QmYn…FRon	ip4 • tcp	/ipfs/bitswap/1.2.0, /ipfs/kad/1.0.0
Unknown	39ms	12D3…vtMe	ip4 • udp • quic	/ipfs/bitswap/1.2.0, /ipfs/kad/1.0.0
Unknown	40ms	12D3…g3Y6	ip4 • udp • quic	/ipfs/bitswap/1.2.0, /ipfs/kad/1.0.0
Australia	40ms	12D3…LcrM	ip4 • udp • quic	/ipfs/bitswap/1.2.0, /ipfs/kad/1.0.0
Singapore	41ms	12D3…U62V	ip4 • tcp	/ipfs/bitswap/1.2.0, /ipfs/kad/1.0.0
USA	41ms	12D3…GGmD	ip4 • udp • quic	/ipfs/bitswap/1.2.0, /ipfs/kad/1.0.0
Unknown	41ms	12D3…fqFn	ip4 • udp • quic	/ipfs/bitswap/1.2.0, /ipfs/kad/1.0.0
Singapore	41ms	12D3…fFQN	ip4 • udp • quic	/ipfs/bitswap/1.2.0, /ipfs/kad/1.0.0

V. CONCLUSION

The existing storage models like Bitcoin, Ethereum, and Hyperledger distributed ledger of the blockchain suffer enormous data storage will face the scalability and computational overhead issue. In this paper, the proposed Blockchain Data Storage Model based on Interplanetary File System provides the efficient storage of the hash of the transactions in the network instead of storing the entire file. IPFS network reduces the storage of transaction size of the block in the blockchain.

References

[1] G. Zyskind, O. Nathan, and A. S. Pentland, "Decentralizing privacy: Using blockchain to protect personal data," *Proc. - 2015 IEEE Secur. Priv. Work. SPW 2015*, pp. 180–184, 2015, doi: 10.1109/SPW.2015.27.

[2] J. Benet, "IPFS - Content Addressed, Versioned, P2P File System," no. Draft 3, 2014, [Online]. Available: http://arxiv.org/abs/1407.3561

[3] I. Jovović, S. Husnjak, I. Forenbacher, and S. Maček, "5G, Blockchain and IPFS: A General Survey with Possible Innovative Applications in Industry 4.0," 2018, doi: 10.4108/eai.6-11-2018.2279695.

[4] Q. Zheng, Y. Li, P. Chen, and X. Dong, "An Innovative IPFS-Based Storage Model for Blockchain," *Proc. - 2018 IEEE/WIC/ACM Int. Conf. Web Intell. WI 2018*, pp. 704–708, 2019, doi: 10.1109/WI.2018.000-8.

[5] R. Kumar and R. Tripathi, "Implementation of Distributed File Storage and Access Framework using IPFS and Blockchain," *Proc. IEEE Int. Conf. Image Inf. Process.*, vol. 2019-November, pp. 246–251, 2019, doi:

10.1109/ICIIP47207.2019.8985677.

[6] R. Kumar, R. Tripathi, N. Marchang, G. Srivastava, T. R. Gadekallu, and N. N. Xiong, “A secured distributed detection system based on IPFS and blockchain for industrial image and video data security,” *J. Parallel Distrib. Comput.*, vol. 152, pp. 128–143, 2021, doi: 10.1016/j.jpdc.2021.02.022.

[7] R. Kumar and R. Tripathi, Towards design and implementation of security and privacy framework for Internet of Medical Things (IoMT) by leveraging blockchain and IPFS technology, vol. 77, no. 8. Springer US, 2021. doi: 10.1007/s11227-020-03570-x.

[8] H. Zareen, S. Awan, M. B. E Sajid, S. M. Baig, M. Faisal, and N. Javaid, “Blockchain and IPFS Based Service Model for the Internet of Things,” *Lect. Notes Networks Syst.*, vol. 278, pp. 259–270, 2021, doi: 10.1007/978-3-030-79725-6_25.

[9] Y. Chen, H. Li, K. Li, and J. Zhang, “An improved P2P file system scheme based on IPFS and Blockchain,” *Proc. - 2017 IEEE Int. Conf. Big Data, Big Data 2017*, vol. 2018-January, pp. 2652–2657, 2017, doi: 10.1109/BigData.2017.8258226.

[10] C. S. Wright, “Bitcoin: A Peer-to-Peer Electronic Cash System,” *SSRN Electron. J.*, pp. 1–9, 2019, doi: 10.2139/ssrn.3440802.