Decentralized File Storage System Using IPFS

Rahul Mane
Computer Engineering.
A P Shah Institute of Technology
Thane, India
rahulmane9403@gmail.com

Simran Singh
Computer Engineering.
A P Shah Institute of Technology
Thane, India
annusingh2894@gmail.com

Vidhi Jain
Computer Engineering.
A P Shah Institute of Technology
Mumbai, India
vidhijain031101@gmail.com

Yashovardhan Sarda
Computer Engineering.
A P Shah Institute of Technology
Thane, India
yashsarda003@gmail.com

Dr. Rahul Ambekar
Computer Department.
A P Shah Institute of Technology
Thane, India
rahulambekar0103@gmail.com

Abstract—**The purpose of storing and exchanging data, centralized storage systems has been widely embraced by several businesses, organizations, and people in recent years. However, these solutions raise issues for users regarding a single point of failure and the participation of a centralized organization or third party (server). Every system aims to achieve key principles of security i.e., availability, integrity, and confidentiality (CIA). Smart contracts are secure platforms to store data more efficiently through the process of decentralization; to enhance the security of data decentralized storage is used along with Blockchain. Blockchain uses the hashing mechanism which further helps to protect the file from being tempered or deleted. But storing files on Blockchain is very expensive both in terms of time and money. So, the solution to this is a combination of an Interplanetary File System (IPFS) and Ethereum blockchain to store users' data in a distributed manner. Distributed user data storage is one of the benefits of the IPFS network. In addition, the document that the user must upload is encrypted using SHA-256. For the benefit of all users using our system, the hash that the IPFS network returned will be kept in the Ethereum blockchain network. Storage system using IPFS, SHA-256 for encryption, and blockchain technologies together to ensure decentralized, secure, and transparent characteristics for storing and sharing data.**

Keywords—SHA-256, IPFS, Blockchain, CIA, encryption, security, smart contract, storage, decentralization.

Introduction

Information technology has grown significantly in recent years; there has been a rapid increase in demand for computation and resource storage. People are investigating novel computing techniques to satisfy their desire for greater computing power and more storage space. People then began putting their reliance on reliable third parties to store their information. Centralized systems typically have a higher risk of assault or severe service interruption. Data Transparency is a major concern for centralized servers which questions their reliability, because if an attacker attacks the server there may be a high possibility of data loss, misplacement, and data leaks due to which it becomes a very easy target for the attackers. Also, the data is sometimes used by a third party for analytical and marketing purposes without any consent of the respective owner. But when it comes to blockchain the data cannot be accessed due to its decentralized nature and no one can access the data without the concern of the owner. Blockchain also maintains a copy of the ledger with each node across its network which makes it highly fault tolerant. According to a Forbes article (released in 2018), There are 2.5 quintillion bytes of data created each day at our current pace and this will keep on increasing. The cost to accommodate data in these centralized storages keeps increasing with data generation. The scalability of data is very difficult to meet existing demand, compared to blockchain. Blockchain relies on distributed ledger technology (DLT). The DLT acts as a decentralized database of information about transactions between various parties. With decentralized storage, data is encrypted and stored across multiple locations, or nodes, that are run by individuals or organizations that share their extra disk space for a fee. Only the data's owner holds the private encryption key; storage providers cannot access the data. In many cases, the files are also shared and spread across multiple locations, providing yet another layer of storage security. Decentralized data storage products often use blockchain to track storage transactions. Blockchain is a decentralized ledger whereby it is used to record transactions in a different node in the network. The data stored in a blockchain is immutable i.e., once a block is formed it cannot be changed or tampered with.

I. Problem Statement

Under third-party data storage, the confidentiality and integrity of the data are still questionable. Keep extremely sensitive data, traditional storage methods are not particularly secure. So, to address this issue we tried to implement a decentralized system to tackle this problem with the proposed system.

II. Objective

To build an online file storage platform that acts as a content hosting and sharing, as well as an online drive for keeping and retrieving data like files and media on torrents The platform connects to the IPFS network to store the data securely, with encryption applied to the data prior to uploading to ensure protection. This means that even if someone gains access to the file hash, the data remains secure, as it can only be unlocked with the password known only to the user.

A. Abbreviations and Acronyms

DLT: - Distributed Ledger Technology
DHT: - Distributed Hash Table
IPFS: - InterPlanetary File System
CID: - Content Identifier

III. Literature review

In this proposed system a smart contract is implemented to control the access privilege and the modified version of IPFS software is utilized to enforce the predefine access-control list. This work proposed a file-sharing environment based on blockchain by clouting the Interplanetary file system (IPFS) and Public Key Infrastructure (PKI). The outcome generated shows transparency, security, access, and quality of data

[1] Nguyen et al. propose a decentralized storage system that combines an InterPlanetary File System (IPFS), Attribute-based Encryption (ABE), Multi-Authority ABE (MA-ABE), and Ethereum blockchain to overcome the drawbacks of centralized storage systems. The authors use IPFS to store user data in a distributed manner, while MA-ABE is used to encrypt documents that need to be shared among multiple organizations. The hash returned by the IPFS network is stored in the Ethereum blockchain to provide trustworthiness for all users participating in the system. The authors claim that their proposed system is the first to use IPFS, ABE, MA-ABE, and blockchain technologies together to ensure decentralized, secure, and transparent characteristics for storing and sharing data. Overall, Nguyen et al.'s proposed system is a promising solution to the issues of centralized storage systems. However, further research is needed to evaluate the scalability and performance of the system, as well as its potential vulnerabilities to attacks. [2] The article titled "Enhanced Security Mechanism for Decentralized Cloud Storage using Blockchain Technology" discusses the need for secure storage of data on the cloud and proposes a decentralized cloud storage approach using blockchain technology to enhance data security. The authors highlight the importance of confidentiality, integrity, and availability (CIA) of data, and suggest that existing centralized cloud storage solutions do not adequately provide these properties. The article goes on to describe the use of blockchain technology to enhance data security and prevent unauthorized tampering or deletion of data. The data stored in the blockchain are linked to each other by a chain of blocks, and each block has its hash value, which is stored in the next block. The authors suggest that this reduces the chances of data altering. The SHA-512 Hashing algorithm is used for this purpose. The authors also discuss the use of symmetric and asymmetric encryption methods to convert plain data into cipher text and vice versa. The Advanced Encryption Standard (AES) algorithm is suggested for encrypting and decrypting data due to its significant features.[3] The paper introduces a system that utilizes blockchain technology to provide secure distributed data storage with a keyword search service, allowing clients to upload encrypted data, distribute it to cloud nodes, ensure data availability, and grant permission for others to search the data. The authors argue that traditional cloud storage has relied on large storage providers as trusted third parties, which poses issues such as data availability, high operational cost, and data security. The proposed system addresses these issues by leveraging blockchain technology to provide secure and efficient data storage and retrieval. The article provides a detailed description of the proposed system architecture and its components, including the blockchain-based storage layer, the distributed cloud nodes, and the keyword search module. The authors also discuss the security and privacy aspects of the system, highlighting the use of cryptographic techniques to ensure data confidentiality, integrity, and availability. The paper concludes with a discussion of the potential applications and future research directions of the proposed system. Overall, the article presents a novel approach to secure and efficient distributed data storage with keyword search, leveraging blockchain technology. The proposed system addresses some of the shortcomings of traditional cloud storage and provides a promising solution for secure and efficient data storage and retrieval. However, further research is needed to evaluate the performance, scalability, and usability of the system in real-world scenarios. [4] The article discusses the challenges of storing large files on blockchains and the inefficiency of doing so. It proposes a modified version of the InterPlanetary Filesystem (IPFS) that leverages Ethereum smart contracts to provide access-controlled file sharing. The smart contract is used to maintain the access control list, while the modified IPFS software enforces it. The article analyzes and discusses the impact of the access-controlled IPFS using an experimental setup. The article provides a detailed explanation of the proposed system and its implementation, including the smart contract code and the modified IPFS software. It also discusses the security implications of the system and how it addresses the limitations of traditional IPFS. Overall, the article presents a novel solution to the problem of access-controlled file sharing on IPFS and provides a thorough analysis of its effectiveness.[5] The article proposes a decentralized platform for storing and sharing patient medical records using blockchain technology, specifically Ethereum and IPFS. The proposed framework aims to address the scalability issues associated with blockchain technology by utilizing the off-chain scaling mechanisms of IPFS. The article also evaluates the performance of the proposed framework through experimental setups and simulations. The literature review of the article reveals that the authors have provided a comprehensive overview of the proposed framework, including its design, architecture, and performance. The authors have also discussed the related work in the domain of healthcare being implemented using blockchain technology, highlighting the solutions proposed for solving the prevalent problems in blockchain technology, such as scalability and data sharing through blockchain. The authors have proposed a solution to the scalability problem by using an underlying database, about some ONC requirements and any other defined standards to solve them. Overall, the article provides a detailed and well-researched proposal for a decentralized electronic health record-sharing system based on the Ethereum blockchain and IPFS, with a thorough evaluation of its performance.

IV. Methodology

A. Why we are using Blockchain-based data storage

Blockchains are immutable. Once anything is saved on the blockchain, it can't be erased or modified. It's a database that can only be added to, not updated or removed. Conventional transactional databases are intended to be constantly updated. From away, this makes blockchains excellent for certain but not all use cases.

Blockchains have many administrators, rather than just one. This eliminates the need to rely on any one blockchain administrator or person. The blockchain serves as proof of validity as well as a barrier against fraud or distrust.

B. How Blockchain Works

Blockchain technology is based on distributed ledger technology (DLT). The DLT functions as a decentralized database of transaction information between numerous parties. Operations fill the DLT in chronological sequence and are stored as a series of blocks in the ledger. A blockchain is created when a linked chain is constructed between blocks, with each one referring to the one before it.

In blockchain storage, data are initially partitioned in a process known as sharding. Each shard is replicated to prevent data loss in the case of a transmission mistake. The files are additionally encrypted with a private key, making it difficult for other nodes in the network to see them. The replicated shards are distributed among decentralized nodes all over the world. The interactions are recorded in the blockchain ledger, which allows the system to confirm and synchronize transactions across blockchain nodes. Blockchain storage is intended to record these exchanges in perpetuity, and the data can never be modified.

Blockchains aren’t efficient for storing large file sizes. Storage of data “on-chain” can be very expensive. This isn’t a very scalable or efficient route for more than core ledger data and related hashes. Costs can rack up per terabyte on the chain per transaction, with fees each time you want to read that data. Because the cost is too expensive, we are using IPFS to upload the file or data.

C. How IPFS Work

IPFS is a decentralized storage and delivery network which is built on fundamental principles of P2P networking and content-based addressing

Every file uploaded to IPFS is assigned a unique address generated from the file's hash. This address is referred to as a content identifier (CID). IPFS currently uses SHA-256 by default, which generates a 256-bit (32-byte) output that is Base58 encoded. Base58 is a binary-to-text encoding that has the advantage of not including letters that could be confused for each other in certain fonts (such as zero and the capital letter O).

IPFS is fundamentally a Distributed Hash Table (DHT) that maps CIDs to people who have content addressed by that CID. Because no single node in the network holds the complete table, the hash table is spread. Instead, each node keeps a chunk of the hash table and information about which nodes are holding other relevant sections. The main innovation of IPFS is the use of Distributed Hash Tables (DHT) for file system storage and retrieval. This stores files as key-value pairs on a blockchain. The information is divided into 256 KB chunks and distributed over a network of nodes or computers. It is well-coordinated to allow for quick access and lookup amongst nodes. There are no duplicates on IPFS because the hash will always refer to the file or a chunk of the file when it was uploaded. When someone talks about 'uploading' material to IPFS, what they generally mean is that they are notifying the network that they have content by adding an entry to the DHT that maps from CID to their IP address. Someone else who wishes to get their data would search up the CID in the DHT, discover the person's IP address, and download the data from them directly. I PFS's speed and reliability benefits stem from the fact that several persons can upload the same data, and subsequently, downloads are distributed among all of them. If one of them falls or decides to cease hosting the data, the others can step in.

Files are not uploaded to IPFS in the same way that they are uploaded to the cloud. The hash ID is used to identify all material on IPFS. When someone wants that data, they are asking it by its hash ID rather than the actual file itself. Thus, IPFS provides an abstraction to the file's actual location, thus the application is unconcerned about the actual physical location.

D. How Pinata Connects to IPFS

Pinata is an IPFS pinning service that simplifies IPFS for developers. Creators access their original material on IPFS using our simple web interface. Once uploaded with Pinata, it is instantly pinned to IPFS and assigned a unique CID. The next step, of course, is to share it! One method is to use Dedicated Gateways, which act as hyperspeed bridges between IPFS and conventional websites. With a Dedicated Gateway, you can instantly distribute material, utilize a custom domain to fit your brand, stream movies, and more.

E. Architecture

An architecture diagram for a system where files are stored in IPFS from a React browser and the IPFS file URLs are later stored on the Ethereum blockchain might look like this:

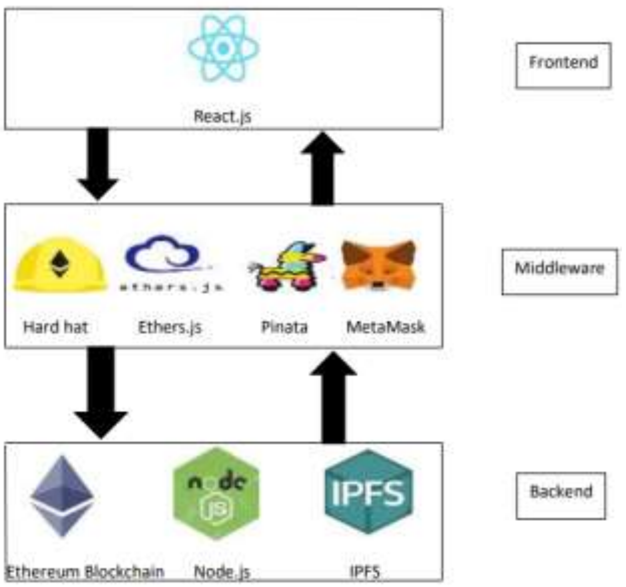

Fig1. Layer Diagram

If you use Ethers.js in place of a web3 provider, the architecture of the system would be similar, but with a few changes.

1. IPFS Node: This component is responsible for storing and retrieving files from the IPFS network.
2. React App: The front end of the application is built using the React JavaScript library. It communicates with the API to upload and retrieve files.
3. API: This component acts as a bridge between the React app and the IPFS node. It makes requests to the IPFS node on behalf of the React app, such as uploading files and returning file URLs.
4. Smart Contract: This component is deployed on the Ethereum blockchain and is responsible for storing the IPFS file URLs. It can also handle additional functionality such as authentication and access control.
5. Ethers.js: This JavaScript library is used by the React app to interact with the Ethereum blockchain instead of a web3 provider. Ethers.js provides a simpler and more user-friendly interface to interact with the Ethereum blockchain

The fig2 illustrates that the sender is authenticated by a smart contract. A new user block is added to the blockchain by smart contract after successful user authentication. The sender uploads a file to Interplanetary File System (IPFS). IPFS returns the cryptography hash on successful storage of the file. If the hash key returned by IPFS is trusted, it is stored on the blockchain. This authentication is done by Smart Contract. Here similarly smart contracts are leveraged to directly store file parts, the data can more easily be accessed and the reassembly information could be stored as well. However, sending and storing large files, even partially, using smart contracts is expensive (for example regarding gas costs) and needs to be executed at every mining or verifying node. On the other hand, operating the mining nodes becomes more expensive. More data needs to be propagated through the network, processed, and stored by the node. Mining nodes would thus require connections with higher bandwidths and more storage space to store the blockchain, even partially, thus leading to increased costs. The report comes to the conclusion that huge files should not be shared or stored on a blockchain. The uploaded file's cryptographic hash key can now be accessed by the sender using blockchain. The user initiates file sending operation by entering the receiver's ether account address (public key). Authentication of a particular receiver is checked by a smart contract. Smart Contract stores the cryptographic hash key on the receiver's block. Authenticated receivers receive the hash key sent by the sender. The receiver requests a particular file to IPFS. The receiver gets access to the file, if and only their private key pair matches with the public key which was used for the encryption of a file while sending it

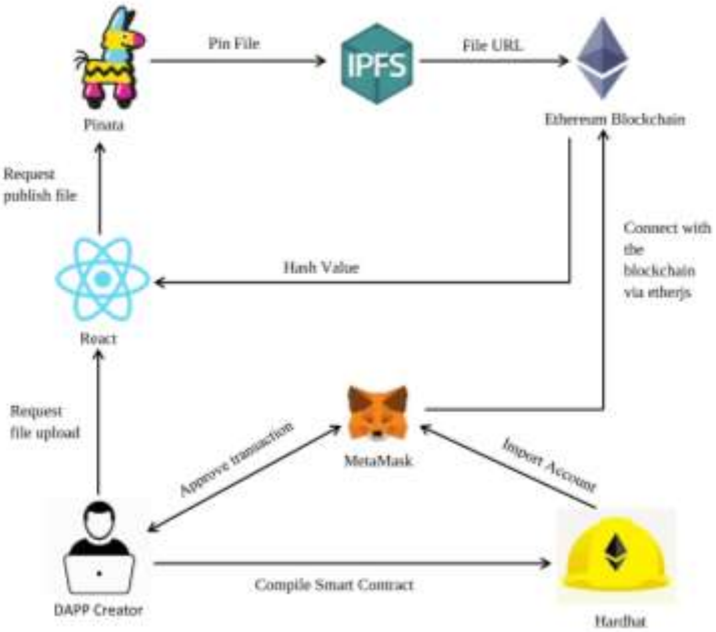

Fig2 Architecture Diagram

COMPARATIVE PARAMETER

Feature	Austria, Phillipe, “Analysis of Blockchain-based Storage Systems”	Our System
Storage Platform	Sia	IPFS
Security	Threefish block cipher	SHA 256-bit encryption
Privacy	Sia has fewer data access to provide more privacy	Distributed storage across multiple peers’ high privacy
Performance	Slower upload	Little faster

		than Sia
Cost	High redundancy factors will incur high costs	Maximizes storage resource utilization and lower cost
Technology	Skynet, P2P cloud storage	Pinata, Hardhat, Ethers.js, remix, Metamask

RESULT AND DISCUSSION

The system is able to share and store files in a secure way. A combination of the Ethereum blockchain and InterPlanetary File System (IPFS) works together efficiently. Blockchain and InterPlanetary File System (IPFS) ensures high data security for the system.

- Scalability- Blockchain has the scalability issue as the size of file increase. Our purpose system uses offchain mechanism to store file i.e. IPFS. This solves the scalability issue mentioned. The data size being stored on the blockchain has now decreased the transactions could also be performed faster. As mentioned earlier, IPFS uses cryptographic hash which is stored in the decentralized manner using peer-to-peer network. This also ensures that while solving the scalability problem the security of the framework is not compromised.
- Content address storage- Content-addressable storage refers to the off-chain storage mechanism of IPFS used in the proposed framework. The sensitive data of user is stored on the IPFS, which ensures that a hash of the stored record is generated. That hash is now stored in the blockchain and is accessed when needed by the user. The IPFS generates the cryptographically secure hash which ensures the security of the data being stored on it. And this also ensures security in our proposed framework.
- Integrity-Blockchain does not compromise with this feature. It ensure that the data store is temper proof and reliable. The information store in this network cannot be changed by any unauthorised user. This is done by using the access rules which ensure that the private data of user are not accessible and remain temper-proof.

CONCLUSION

The major advantage of the project is a large reduction in operational expenses. The cost of storing 1TB of data on Amazon, for example, is around $20. Using the above technologies, it is possible to bring down the cost of storage approximately ten times! Data redundancy indicates that the data is safe and secure until 51% of the network goes down or is attacked, which is exceedingly rare and extremely difficult to achieve, implying 100% safety and security. Data stored cannot be viewed by anybody other than the person for whom it is intended since it is encrypted before posting.

REFERENCES

[1] V.-D. Pham *et al.*, “B-Box - A Decentralized Storage System Using IPFS, Attributed-based Encryption, and Blockchain,” 2020. [Online]. Available: https://v-chain.vn/solutions/b-box.

[2] R. Pise and S. Patil, “Enhancing security of data in cloud storage using decentralized blockchain,” in *Proceedings of the 3rd International Conference on Intelligent Communication Technologies and Virtual Mobile Networks, ICICV 2021*, Institute of Electrical and Electronics Engineers Inc., Feb. 2021, pp. 161–167. doi: 10.1109/ICICV50876.2021.9388521.

[3] H. G. Do and W. K. Ng, “Blockchain-Based System for Secure Data Storage with Private Keyword Search,” in *Proceedings - 2017 IEEE 13th World Congress on Services, SERVICES 2017*, Institute of Electrical and Electronics Engineers Inc., Sep. 2017, pp. 90–93. doi: 10.1109/SERVICES.2017.23.

[4] M. Steichen, B. Fiz, R. Norvill, W. Shbair, and R. State, “Blockchain-Based, Decentralized Access Control for IPFS,” 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, Jul. 2018, pp. 1499–1506. doi: 10.1109/Cybermatics_2018.2018.00253.

[5] A. Shahnaz, U. Qamar, and A. Khalid, “Using Blockchain for Electronic Health Records,” *IEEE Access*, vol. 7, pp. 147782–147795, 2019, doi: 10.1109/ACCESS.2019.2946373.

[6] Vaigai College of Engineering and Institute of Electrical and Electronics Engineers, *Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020) : 13-15 May, 2020.*

[7] 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. 38437–38450, Jun. 2018, doi: 10.1109/ACCESS.2018.2851611.

[8] Wei-Meng Lee, “Using the MetaMask Chrome Extension”.

[9] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, “An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends,” in *Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017*, Institute of Electrical and Electronics Engineers Inc., Sep. 2017, pp. 557–564. doi: 10.1109/BigDataCongress.2017.85.

[10] S. Nakamoto, “Bitcoin: A Peer-to-Peer Electronic Cash System.” [Online]. Available: www.bitcoin.org

[11] A. Nguyen, “Build a blockchain application using React and Solidity Title: Build a blockchain application using React and Solidity Number of Pages: 54 pages,” 2022.

[12] A. Cavalli, E. Montes De Oca, and M. Núñez, “TestNet: Let’s test together!,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 2644, pp. 258–264, 2003, doi: 10.1007/3-540-44830-6_19.

[13] “Storj: A Decentralized Cloud Storage Network Framework,” 2018. [Online]. Available: https://github.com/storj/whitepaper

[14] “Grayscale Building Blocks | An Introduction to Filecoin,” 2021. [Online]. Available: https://www.nytimes.com/2017/08/01/business/amazon-china-internet-censors-apple.html.

[15] M. Alizadeh, K. Andersson, and O. Schelén, “Efficient Decentralized Data Storage Based on Public Blockchain and IPFS.”

[16] H. Thakur and A. Chattopadhayay, “DISTRIBUTED DECENTRALIZED DATA STORAGE USING IPFS,” *International Research Journal of Engineering and Technology*, p. 1641, 2008, [Online]. Available: www.irjet.net