Decentralized Storage System to Store Data using Blockchain Technology

Prof. Sunil Khachane, [1]Akshar Mehta, [2]Sakshi Pandey, [3]Karan Parmar, [4]Nilesh Sahu
[1]*Head of Department, Computer Engineering, MCT's Rajiv Gandhi Institute of Technology, Mumbai, India*
[2,3,4] *BE Student, Department of Computer Engineering, MCT's Rajiv Gandhi Institute of Technology, Mumbai, India*

***Abstract*—In today's society, data storage is becoming more and more crucial. In the lives of customers, data reliability, secrecy, security, and accessibility have all been crucial factors. All web2.0 vulnerabilities are removed by blockchain in general by introducing new technology and structure. With the power of the blockchain, data storage problems might be resolved and customer privacy and security of their data could be guaranteed. This reduces the possibility of a single point of failure and encourages ideas like tokenization and incentivization to reward customers and storage providers for making data storage compatibility possible. As a result, we offer a decentralised storage system that use a blockchain network and smart contractsto store user data.**

***Index Terms*—Blockchain, vulnerability, data storage.**

I. INTRODUCTION

Storage is a mechanism that allows a computer to keepdata indefinitely or temporarily. Storage devices, such as flash drives and hard discs, are essential components of most digitaldevices because they allow users to save various types of data,such as images, files, videos, and unprocessed data.

Cloud storage systems have become increasingly popularin recent years to meet the data storage and sharing needs of businesses, organisations, and individuals. Users who investin these services must fully trust the companies to keep their important and private data secure for an extended period of time. However, centralised systems are more vulnerable to being attacked or having their services severely disrupted. Furthermore, data leakage cases such as Facebook-Cambridge Analytica have resulted in a strong shift away from centralised to decentralised data storage systems at the time.

Cloud computing is the delivery of various services via the Internet. Data storage, servers, databases, networking, and software are examples of these resources. Cloud-based storage enables files to be stored to a remote database as opposed to a proprietary hard drive or local storage device. As long as an electronic device has internet access, it has access to data and the software programs needed to run it.

Because the information being accessed is situated virtually,or in the cloud, cloud computing derives its name. Users can access data via the Internet by storing files and apps on remote servers thanks to cloud service providers. This enables remote working since the user does not need to be in a certain placeto access it.

The Internet becomes the cloud, and your data, work, and applications become accessible from any device that can connect to the Internet, wherever in the globe. The cloud can be both public and private. For a fee, public cloud services offer their services over the Internet. Private cloud services,on the other hand, only serve a limited number of people. A network infrastructure called these services offers hosted services. There is also a hybrid option that includes aspects of both public and commercial services.

With cloud computing, all of the labor-intensive data processing is offloaded from the device you carry or sit on while working. Additionally, it offloads all of that work to enormous computer clusters in the internet. Cloud computing is becoming increasingly popular among individuals and businessesfor a variety of reasons, including cost savings, increased productivity, speed and efficiency, performance, and security.Google Cloud,

Microsoft Azure, Amazon Web Services, andmany other platforms already exist as data storage solutions. The fact that all of this data is kept on one central server increases the possibility of a single point of failure or anassault meant to bring the server down. Large tech businesses make up the existing structure, and a single organization purchases and offers all services.

When data is kept by a third party, there are significant concerns about data security and confidentiality. Due to asingle point of failure, other storage systems are prey to cyber attacks. Large IT corporations might also stop supplyingservices to the client because they have control and authority over the client's data.

Blockchain storage is a method of storing data in a decentralised network that uses unused hard disk space from users all over the world to store files. Decentralized cloud storage is an alternative to centralised cloud storage and can solve many of the problems associated with centralised systems.

II.LITERATURE REVIEW

Decentralized storage systems have grown in popularityin recent years as a result of its capacity to offer a more secure, effective, and dependable method of data storage and exchange. One of the most well-known decentralised storage systems that has gained widespread acceptance in both the research community and industry is the InterPlanetary File System (IPFS). We will talk about IPFS's primary features, aswell as its possible advantages and difficulties in creating a decentralised storage system, in this literature review.

The decentralised storing and sharing of files is made possible by the IPFS protocol. In contrast to conventional centralised storage systems, IPFS distributes and stores data through a peer-to-peer network. Since IPFS is built on content addressing, each file is given a distinct hash, which is then usedto locate the file on the network. Since data is duplicated over numerous nodes and may be retrieved even if some nodes are unavailable, this method makes IPFS more resilient to failures.

One of IPFS's key benefits is its capacity to offer a more private and secure method of data sharing and storing. Datais spread among several nodes, making it more challenging for attackers to access it unlawfully. Moreover, IPFS provides end-to-end encryption, enabling data to be encrypted prior to being stored on the network and ensuring that only authorised users may access it.

Another benefit of IPFS is its ability to make data shar- ing and storing less expensive and complicated. IPFS does not require a centralised server to store and distribute files because it utilises a peer-to-peer network. As a result, it is less expensive to operate a central server and simpler to grow the storage system as necessary. Moreover, IPFS offers incremental updates, which eliminates the need to store and disseminate the full file by only storing and distributing the changes to a file.

Despite the potential advantages of IPFS, there are numerous issues that must be resolved if a dependable and effective decentralised storage system is to be created. The network's scalability is one of the major issues. It gets harder to maintainconsistency and make sure that every node has access to the same data as the network grows in size. Moreover, network latency and bandwidth restrictions may have an impact on IPFS performance.

Another concern is the issue of data-availability. It is possible for some nodes to go offline or become inaccessible since data is spread over numerous nodes, making it challenging to retrieve data. Since at least one node in the network holdsa copy of the data, IPFS offers content-addressed storage to solve this problem.

An effective technology for creating a decentralised storagesystem is IPFS, in conclusion. It offers a more dependable, efficient, and secure means to store and share data and may beless expensive and complex than conventional storage systems.To create a dependable and scalable decentralised storage system using IPFS, there are various additional issues that must be resolved.

III.RELATED WORK

A. Blockchain Technology

Decentralized cryptocurrencies (like Bitcoin, Ethereum, Litecoin, etc.) have gained popularity in recent years, andthe blockchain technology that powers cryptocurrencies is receiving increasing attention. The blockchain has recentlybecome very important in the world of finance. It has also been

discovered to be helpful in numerous other non-financial do- mains. Examples include decentralised storage, identity-based PKI, decentralised document existence proof, decentralisedIOT, and decentralised supply chains.

A personal data management system based on blockchain technology has been proposed in order to enable data owners to own and control their own data. This system can improve the privacy of user data. A blockchain architectural system forIOT was developed to address the issue of data privacy in IOT systems, and advanced encryption standard (AES) technology was employed to safeguard sensitive data. A blockchain-basedaccess control architecture for increasing the security of big data platforms was presented to address the security and privacy challenges that are impeding the growth of big data. The use of conventional database techniques to store financial transactions has a number of drawbacks. Take the saleof a piece of real estate as an example. The buyer acquires ownership of the property following the exchange of funds. Both the buyer and the seller are capable of maintaining their own records of financial transactions, but neither can be trusted. It is possible for both the buyer and the seller to claim that money has been made even though it hasn't, and both parties can simply contest this.

Transactions must be monitored and confirmed by a reliable third party to prevent potential legal problems. In addition to making the transaction more challenging, the presence of this centralised authority creates a weak point. Both parties may suffer if the main database is compromised.

Blockchain resolves these problems by developing a decentralised, unchangeable mechanism for transaction recording. Blockchain develops separate ledgers for the buyer and seller in real estate transactions. All transactions are subject to approval by both parties and are automatically updated inboth of their ledgers in real time. Any change to previous transactions will reflect poorly on the entire ledger. Due to these features, blockchain technology has proven valuable ina variety of fields, including the creation of digital currencies like Bitcoin.

B. Advanced Encryption Standard Technology

The United States government selected the symmetric block cypher known as the Advanced Encryption Standard (AES) to safeguard sensitive data. To encrypt sensitive data, AES is used in hardware and software across the globe. Cybersecurity and government computer security are both dependent on it for theprotection of electronic data. The Data Encryption Standard (DES), which was becoming increasingly susceptible to brute-force attacks, prompted the National Institute of Standards and Technology (NIST) to identify the need for an alternative in 1997.

The more modern and sophisticated encryption method, according to NIST, must be declassified and able to "secure sensitive government information well into the [21st] century."It was designed to have simple hardware and software im- plementation, work well in constrained contexts like a smart card, and provide strong protections against a variety of attackvectors.

AES is an iterative cypher as opposed to a Feistel one. Its foundation is a "substitution-permutation network." It consistsof a number of interconnected operations, some of which sub- stitute certain outputs for inputs (substitutions), while others require shifting bits about (permutations).

It's interesting to note that AES uses bytes rather than bitsfor all of its calculations. As a result, AES considers a plaintextblock's 128 bits to be 16 bytes. For processing as a matrix, these 16 bytes are organised into four columns and four rows.

In contrast to DES, the number of rounds in AES varies andis based on the size of the key. For 128-bit keys, AES employs10 rounds; for 192-bit keys, 12 rounds; and for 256-bit keys, 14 rounds.Each of these rounds uses a different 128-bit round key, which is calculated from the original AES key.

IV.PRELIMINARIES

The background information and specific aspects that will be utilized in this article are reviewed in this section.

A. IPFS

Computers all around the world can store and serve files asa part of a sizable peer-to-peer network thanks to a distributed file storage protocol known as the Interplanetary File System (IPFS). Downloading the IPFS software allows any machineto start hosting and serving data from any location in theworld. A file uploaded to the IPFS network can be viewedand downloaded by anybody who has IPFS installed on

their machine.

With IPFS, there will be just one network in existence. If two users publish a block of data with the same hash, thepeers downloading the content from ”user 1” and the peers downloading it from ”user 2” will exchange data. IPFS aimsto displace the technologies now in use for providing static online material by utilising HTTP-accessible gateways.

The way we currently utilise the Web and IPFS are quite similar. A unique cryptographic hash string is created when a file is posted to the IPFS system and can be used to retrievethe content. Similar to a Web Uniform Resource Locator is thehash string (URL). Moving future, the file location will simplybe referred to as the hash string. Blockchains are not practical for storing large data due to block bloat and transaction fees (video, audio, etc.).

As a result, we store the encrypted file in IPFS accordingto the plan. In the Ethereum blockchain, a small amount of metadata is stored.Before their attribute set satisfies the accesspolicy defined by the data owner, users won't be permitted to receive data from the Ethereum blockchain, decrypt the file

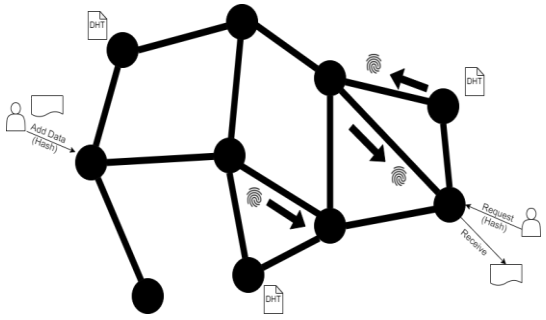

Fig. 1. A simplified IPFS network and its components.

location, download the encrypted file from the IPFS via thefile location, and then decrypt it.

B. Transaction life cycle

A transaction needs to be authorised and validated before it can be posted to the blockchain. A transaction must go throughseveral crucial phases before it can be put to the blockchain. We'll focus on cryptographic key authentication, proof of work authorization, mining's role in blockchain networks, and the more recent use of proof of stake protocols. This lifetime tracks a single transaction as it passes through each step of the blockchain integration process. Simply said,a transaction is the act of transmitting money by one partyto another and having that other party accept it. Although conducted digitally, the blockchain transaction is also very comparable. The life cycle of a blockchain transaction includesthe following stages:

If there are two Bitcoin users, A and B. A wishes to sendB 1 bitcoin.

- A obtains B's wallet address first (a wallet in the blockchain is a digital wallet that allows users to manage their transactions). With this knowledge, he creates a newtransaction from his wallet for 1 bitcoins with a 0.003 bitcoin transaction fee.
- He then confirms the data before sending the transaction. Every transaction that is launched has a digital signature from the sender, which is essentially the sender's private key. This is done to increase the transaction's securityand guard against fraud.
- The transaction signing mechanism is then launched by A's wallet, signing his transaction with his private key.
- The transaction has now been published to the network's memory pool.
- The miners eventually agree to this transaction. This transaction is grouped into a block by the miners, who also determine the Proof of Work and give the block a hash value before mapping it to the blockchain.
- The Blockchain now has this block.
- This block is recognised by the network as a legitimatetransaction as it receives confirmation.
- The moment this transaction is approved,

B receives hisbitcoin.

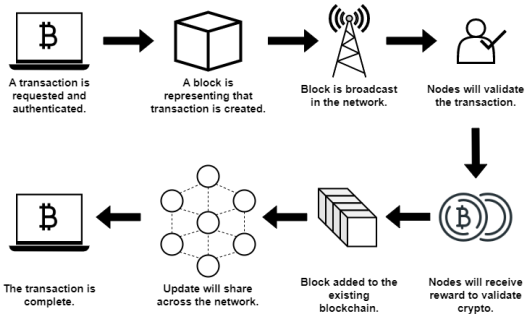

Fig. 2. Blockchain transaction process.

C. Merkle Tree

The Merkle tree serves as the essential foundation for blockchain technology. It is a form of mathematical data structure made up of hashes of different data blocks that is used to summarise all the transactions in a block of data. It alsomakes it possible to quickly and securely verify content acrossa vast body of data. It

also helps to verify the data's correctness and comprehensiveness. Both Ethereum and Bitcoin use the Merkle Tree structure. Merkle Tree is also known as "Hash Tree."

A Merkle tree records all of the transactions in a blockby generating a digital fingerprint of the full collection of transactions. It allows the user to verify if a transaction canbe included in a block or not.

Merkle trees are created by continuously hashing node pairsuntil there is just one hash left. This hash is the Merkle Root, sometimes referred to as the Root Hash. The Merkle Trees are constructed from the bottom up.

Each leaf node is a hash of transactional data, while the non-leaf node is a hash of its prior hashes. An even numberof leaf nodes are required since Merkle trees are in a binary tree. If the number of transactions is odd, the final hash willbe copied once to create an even number of leaf nodes. The Merkle Root information can be found in block headers. The block header is the area of a bitcoin block that is hashed duringmining. It contains the Root Hash of the most recent block's transactions in a Merkle Tree as well as the Root Hash ofthe block before that, which is a Nonce. Consequently, by including the Merkle root in the block header, the transaction isrendered impregnable. Each leaf node is a hash of transactional data, while the non-leaf node is a hash of its prior hashes.

An even number of leaf nodes are required since Merkle trees are in a binary tree. If the number of transactions isodd, the final hash will be copied once to create an evennumber of leaf nodes. The Merkle Root information can be found in block headers. The block header is the area of a bitcoin block that is hashed during mining. It contains the Root Hash of the most recent block's transactions in a Merkle Tree as well as the Root Hash of the block before that, which is a Nonce. Consequently, by including the Merkle root in the block header, the transaction is rendered impregnable.

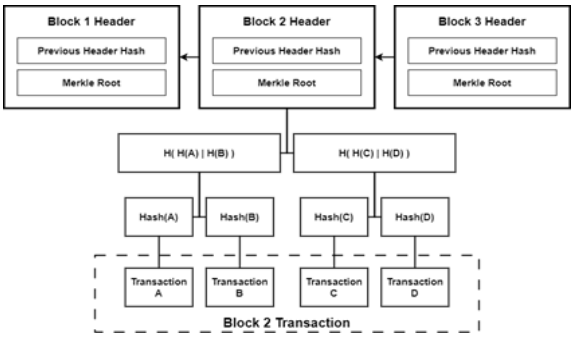

Fig. 3. Blockchain transaction process.

V.METHODOLOGY

The suggested system runs through six phases:

A. Upload file

The user uploads a file using the file picker. The system verifies the file size and availability of network storage. The file is uploaded once there is enough storage space. The systemthen moves on to the following phase. Users are informed to try again when adequate storage is not available.

B. File Encryption

AES 256 bit encryption is applied to the submitted file. Theuser's wallet address is combined with a random salt value to create the encryption key. Data pertaining to users is encryptedusing this encryption key and an IV. This guarantees that the user's information is kept secret.

C. File Storage across Multiple Peers

The IPFS protocol is then used to distribute the encrypted file throughout the network in 256KB blocks. To enable registered peers to store the file on the network, the suggested method makes use of a private IPFS network. The file block isduplicated across several peer storages using the IPFS cluster to increase availability. IPFS provides a hash value that showsthe path of the file.

D. Reimbursing Peers for File Storage

The total amount of cryptocurrency is calculated and re- moved from the user's wallet after the file has been disseminated across peers. The user's wallet sends this cryptocurrencyfirst to the smart contract. This sum is distributed by the smartcontract to the peers who have saved the user's file.

E. Storing IPFS Hash value

The user's wallet address, metadata, and the hash

value are all stored in the blockchain using a smart contract. Similar to agreements, smart contracts function without the involvement of a third party. They have some degree of control over assets transferred between parties or transactions between nodes. Code snippets are kept on a blockchain network and run automatically when certain criteria are satisfied. For the smart contract to function under our proposed system, there must be sufficient network storage space for files and the user must have enough funds in their wallet to pay their peers.

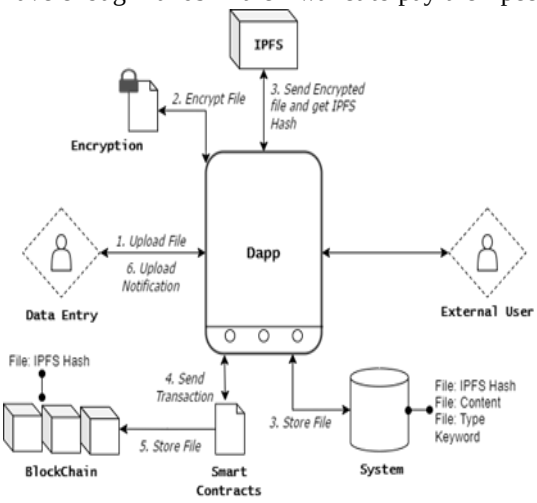

Fig. 4. System Architecture

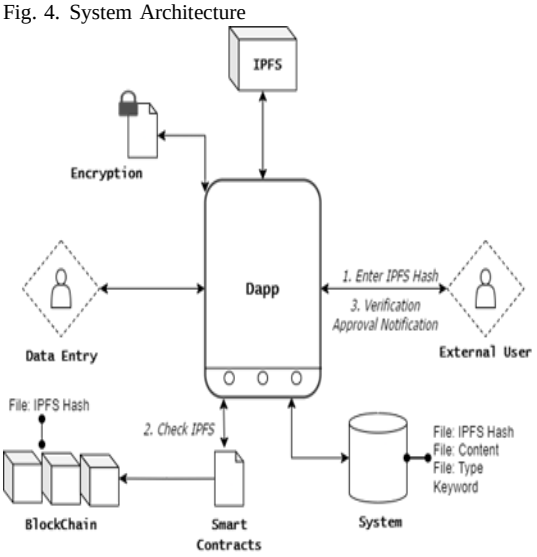

Fig. 5. Verification of the File Upload

F. File Upload Notification

The user is then told if the file upload was approved or denied by the blockchain system.

RESULT

Data storage using decentralised IPFS (InterPlanetary File System) technologies is safe and effective. Rather of storing data based on its location, IPFS stores data using a content- addressed method. Data is kept on several network nodes rather than a single central server, which facilitates data storageand retrieval and increases security.

Here are some advantages of using IPFS for decentralized data storage:

- Decentralization: Because IPFS is a decentralised net- work, data is kept on several network nodes rather than a central server. As a result, it is more difficult for attackersto penetrate the network by focusing on a single point of failure, increasing security.
- Data integrity: Data is saved based on its content rather than its location thanks to IPFS's usage of content- addressed storage. This makes the data tamper-proof since any alteration to the data's content will producea new hash.
- Faster data access: Since data is cached on network nodesto make it easier to obtain the data from the network, IPFS enables faster data access.
- Lower costs: Since IPFS doesn't need expensive servers or data centres, it may result in lower costs for datastorage and retrieval. Instead, data can be kept on the network's nodes, which can be managed by people or businesses with extra computer power.
- Data privacy: Given that data is encrypted and stored across a number of network nodes, IPFS offers great dataprivacy. Unauthorized individuals find it challenging to access the data as a result.

Overall, IPFS-based decentralised storage systems offer a safe, effective, and economical solution to store data. It hasthe potential to completely change how data is accessible and kept, particularly in sectors that demand the highest standards of data security and privacy. Like any new technology, it does, however, have significant difficulties that must be overcome, including scalability, interoperability, and user adoption.

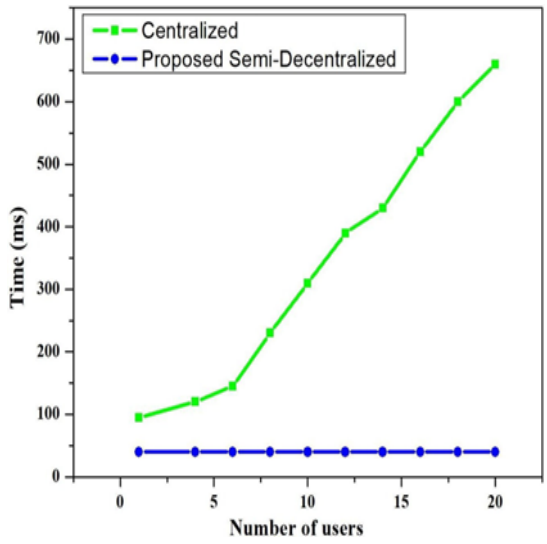

Fig. 6. Result

Fig. 6 demonstrates the computational burden of user-side decryption. From the user's perspective, the suggested method delivers better decryption efficiency because blockchain handles the majority of the computational processes. Using the user-associated decryption engine, the domain authority retrieves the ciphertext supplied by the data owner from the block chain during the decryption phase. The final decryption just requires the user to execute simple addition and multiplication operations once they have the decryption keys, so the number of users in the domain has no impact on how long it takes to complete. Decryption time for the user is therefore almost constant.

CONCLUSION

By encrypting the data and distributing it among a large number of peers in the system, the suggested solution improves data security. To ensure the privacy of the user's data, the implemented system encrypts data using the AES 256-bit encryption algorithm. Afterwards, encrypted data is distributed and stored among network peers via the IPFS protocol. Our method not only addresses the privacy and security issues related to centralised cloud storage, but it also gives peersa platform to rent out idle storage and receive cryptocurrency in return, making the best use of the capacity that is available.

ACKNOWLEDGMENT

We wish to express our sincere gratitude to Dr. Sanjay U. Bokade, Principal, and Prof. S. P. Khachane, Head of Department of Computer Engineering at MCT's Rajiv Gandhi Institute of Technology, for providing us with the opportunity to work on our project," Decentralised Storage System to Store Data using Blockchain Technology." This project would not have been possible without the guidance and encouragement of our project guide, Prof. S. P. Khachane, and the valuable insights of our project expert, Prof. Bhushan Patil We would also like to thank our colleagues and friends who helped us complete this project successfully.

REFERENCES

[1] IPFS. "What Is IPFS? — IPFS Docs." IPFS Documentation — IPFS Docs, https://docs.ipfs.tech/concepts/what-is-ipfs/. Accessed 29 Oct. 2022.

[2] Karanth, Harsha, and Shuwam Rana. "Decentralized Storage. Introduc- tion — by Harshakaranth — Coinmonks — Medium." Medium, Coin- monks, 21 Apr. 2022, https://medium.com/coinmonks/decentralized-storage-701d53c4aa76.

[3] Y. He, H. Li, X. Cheng, Y. Liu, C. Yang and L. Sun,"A Blockchain Based Truthful Incentive Mechanism for Distributed P2P Applications,"in IEEE Access, vol. 6, pp. 27324-27335, 2018, doi: 10.1109/AC- CESS.2018.2821705.

[4] Halpin and M. Piekarska,"Introduction to Security and Privacy on the Blockchain," 2017 IEEE European Symposium on Security and Privacy Workshops (EuroSPW), 2017, pp. 1-3, doi: 10.1109/EuroSPW.2017.43

[5] Pham, Van-Duy, et al. "B-Box - A Decentralized Storage System Using IPFS, Attribute-Based Encryption, and Blockchain." 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), IEEE, Oct. 2020. Crossref, doi:10.1109/ rivf48685.2020.9140747.

[6] Shah, Meet, et al. "Decentralized Cloud Storage Using Blockchain." 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), IEEE, June 2020. Crossref, doi:10.1109/icoei48184.2020.9143004.

[7] Wang, Shangping, et al. "A Blockchain-Based Framework for Data Sharing With Fine-Grained

Access Control in Decentralized Storage Systems." IEEE Access, Institute of Electrical and Electronics Engineers (IEEE), 2018, pp. 38437–50. Crossref, doi:10.1109/access.2018.2851611.