BLOCKCHAIN BASED DATA SHARING

Mr. Vikas Narkhede[1] and Dr. Satpalsing Rajput[2]

[1]*SSBT's COET,Bambhori PG Student*
[2]*SSBT's COET,Bambhori Department of Computer Engineering*

Abstract: In traditional cloud storage systems, all data is stored in the cloud server, which could lead to major issues including key misuse and privacy data leaking. However, because the traditional cloud storage system depends on centralized storage, a single point of failure could cause the system to fail. With the advancement of blockchain technology, decentralized storage has become more well-known. Decentralized storage can overcome the problem of a single point of failure in typical cloud storage systems while also providing a variety of benefits over centralized storage, including low cost and fast throughput. In this work, focus on data storage and sharing in decentralized storage systems and propose a framework that includes IPFS, an encryption technique, and blockchain technologies. The data holder can distribute secret keys to data users and encrypt shared data location by Blockchain network. In this research contribute to data integrity verification in the blockchain network.

Keywords: Blockchain, IPFS, Encryption

1. INTRODUCTION

The Blockchain is a singly Linked List of the block, with each block containing some transactions. It provides a decentralized, immutable data store that can be used across a network of users, creates property, and serves as a shared ledger for all transactions. Each transaction can be easily queried, according greater transparency and trust to all parties involved.

.

1.1 Overview of Blockchain

Blockchains are one form of distributed ledger technology. Necessarily, not all distributed ledgers employ a chain of blocks to provide a secure and valid transaction. A blockchain is distributed across the system and regulated by peer-to-peer networks. It can exist without a centralized power or server controlling it because it is a distributed ledger, and its data quality can be maintained using data replication and computational security. The structure of the blockchain, on the other hand, distinguishes it from other types of distributed ledgers. Blocks are used to gather and organize data on a blockchain. After that, the blocks are linked together and encrypted using cryptography. [1].
Blockchain is a ledger of records arranged in data batches called blocks that uses crypto-graphic validation to link themselves together. Put simply each block reference and Identify the hash of the previous block forming an unbroken chain [1].

1.1.1 Types of Blockchain

1. Public Blockchain: Anyone who wishes to participate in the public blockchain can do so by using open-source software. This has the characteristics of a decentralized network. The chain can be extended by any of the participants. An open network, similar to the bitcoin architecture, that anybody can access. Users can add to the chain by adding new blocks, but they do not have the right or authorization to modify or change existing blocks, as this could lead to incorrect hash derivation.

 Creative Commons Attribution 4.0 International

Example: bitcoin, ethereum and factom etc [1].

2. Private Blockchain:
The private blockchain incorporates permission to review and modify the contents of locks. So this behaves as a centralized system. The most preferred option of the banks is that it is a closed, con_dential, and non-transparent system which checks all the details and the accessing capacities are being controlled at a central level.
Example: quorum, hyperledger
(corda, fabrix and sawtooth) etc. [1].

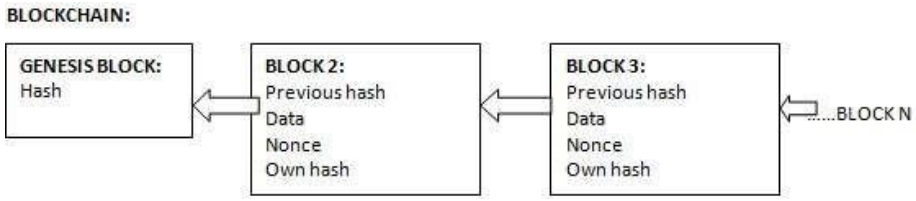


Figure 1.1: Structure of BlockChain

Fig shows 1.1 each block is a typical blockchain has at least the following components:

Details of blockchain components are
Index(Block 1) - The position of the First block has index 0 or Genesis block.
Previous Hash- Check it is valid or not previous block.
Time Stamp - When the new block has added the information in the store at that time.
Data - The information is stored on the block.
Nonce - Number generation of the new block.
Hash - hash is generated for the current block. Merkle Root - the root of the Merkle tree that contains the block's transactions.

All the above information that is Index, previous Hash, Timestamp, Data, and Nonce are used for the creation of the present blocks Hash. Each transaction is recorded and stored in a block. Each block is connected to the one before and after it. All the transactions are blocked together. Modifying/Changing a block will change its Hash. All subsequent blocks in the chain will be invalidated as a result. The reason for this is the hash of the present block is derived from the previous blocks hash.

1.1.2 Merkle Trees
Each block stores a summary of all the transactions in the block in a multi-level data structure called Merkle Tree. In Figure, a Merkle tree example is shown. Merkle Trees are binary trees containing cryptographic hashes built by hashing recursively the children nodes, using a bottom-top approach. This allows summarizing the contents of large data sets and provides a secure and efficient form of verification of the data set integrity. Checking if a data element is included in a tree with N hashed elements takes at most calculations. This data structure is an essential component in the Bitcoin protocol since it prevents tampering with the transactions by malicious users. To successfully swap fake transactions into the bottom of the Merkle tree, it would be necessary to recalculate all the hashes of the nodes that are in the same sub-tree, up to the Merkle root. Because that each block may include hundreds of transactions, this might be a highly costly procedure, and by the time it's _nished, fresh blocks pointing to the preceding actual block may have already been mined.

 Creative Commons Attribution 4.0 International

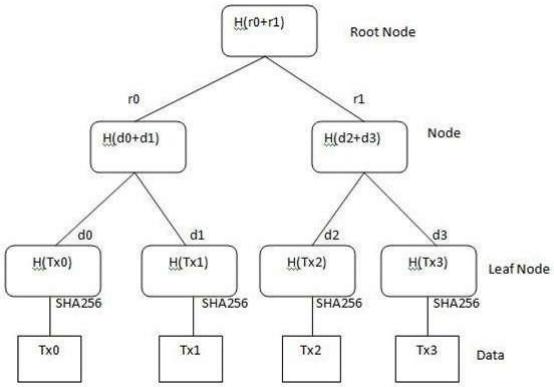


Figure 1.2: Merkle Tree

Fig shows 1.2 The Binary Merkle Tree, as seen above, is the most frequent and basic version of a Merkle tree. There are four transactions in a block: TX0, TX1, TX2, and TX3. As you can see, there is a root hash, also known as the Merkle Root, which is the hash of the entire tree. Each of these is hashed many times and saved in each leaf node., resulting in Hash (r0+r1). Consecutive pairs of leaf nodes are then summarized in a parent node by hashing H(Tx0) and H(Tx1), resulting in H(d0+d1), and separately hashing HTx2 and HTx3, resulting in H(d2+d3). The two hashes (H(d0+d1) and H(d1+d2) are then hashed again to produce the Root Hash (H(r0+r1) or the Merkle Root. Merkle Root is stored in the header. The block header is the part of the block which gets hash in the process of mining. In a Merkle Tree, it holds the hash of the previous block, a Nonce, and the Root Hash of all the transactions in the current block. As a result, putting the Merkle root in the block header secures the transaction. Because this Root Hash contains the hashes of all transactions within the block, these transactions may result in disc space savings. Merkle trees have their advantages It allows for data integrity and validity to be preserved. It also helps to save memory and disc space, and it is computationally simple and rapid. The transfer of very small bits of data over networks is required for their proof and management.

2.1 Background

Traditional centralized systems are primarily built on the client-server model, in which a client can save entries in a central server and retrieve an updated copy of the information each time the server is accessed. In conventional centralized systems, Every authorization must be verified by a central trusted party, resulting in substantial costs and network-level at the central servers. Direrently,a transaction in the blockchain network can be conducted between any two peers (P2P) without authentication by the central party. In this manner, blockchain has the potential to greatly cut server costs, including development and operating expenditures, as well as optimize network level at the central server. In contrast to this Blockchain are one form of distributed ledger technology. A blockchain is distributed across the system and regulated by peer-to-peer networks. It can exist without centralization or server controlling it because it is a distributed ledger, and its data quality can be maintained using database replication and computational security. The structure of the blockchain, on the other hand, distinguishes it from other types of distributed ledgers. Blocks are used to gather and organize the data on a blockchain. After that, the blocks are linked together and secured using cryptography. [1].

 Creative Commons Attribution 4.0 International

Blockchain is a ledger of records arranged in data batches called blocks, that uses cryptographic validation to link themselves together. Put simply each block reference and Identify the previous block hash forming an unbroken chain. Each block includes the previous block hash, connecting the two links. Thus, the linked blocks form a chain. Blockchain eliminates the risk of data being held centrally. IPFS [2] has a special property of content addressing at the HTTP layer for the identification of files. IPFS represents a file by the hash on it, instead of representing it by which server it is stored on InterPlanetary File System (IPFS) is a protocol and network designed to create a content-addressable, peer-to-peer method of storing and sharing hypermedia in a distributed file system. IPFS was initially designed by Juan Benet and is now an open-source project developed with help from the community. IPFS is a distributed _le system that aims to connect all computing devices to the same file system. In some ways, IPFS is similar to the World Wide Web, but IPFS could be seen as a single Bit Torrent swarm, exchanging objects within one Git repository. In other words, IPFS is a content-addressed block storage format with content-addressed hyperlinks that deliver high throughput. This results in a Merkle-directed acyclic network that is more generic (DAG) . IPFS is a self-certifying namespace that incorporates a distributed hash table, an incentive block exchange, and an identity hash table. IPFS doesn't have a single point of failure, and nodes don't have to trust one another not to tamper with data in transit. HTTP problems with DDoS attacks, therefore Distributed File Sharing saves.

2.2 RelatedWord

A blockchain-based decentralized storage system used in many domains like Cloud IOT ,Big data.

2.2.1 Blockchain in big data

The rise of the big data age on the Internet has led to the explosive growth of data size. However, trust issue has become the biggest problem of big data, leading to di_culty in data safe circulation and industry development. In blockchain technology to solve these issues, LI Yue et al. In [4] The article describes a credible massive data-sharing strategy based on blockchain technology and smart contracts to assure the secure circulation of data resources. Elena Kara_loski et al. In [3] The Blockchain technology contributes significantly to the storage, organization, and processing of Big Data. Its proposed solutions for decentralized private data management, digital property resolution, and reforms are having a big impact on how Big Data will emerge. Shilpa K. Sangode et al. In [10]] Metadata is a collection of basic information concerning data block distribution, block size, and replica management. The newer HDFS architecture for big data storage does away with the concept of a Name Node, with the metadata being stored utilizing blockchain technology instead. With the help of BlockChain Miners, mathematical cryptography techniques such as hashcode can be used.

2.2.2 Blockchain in Cloud

With the growing popularity of cloud storage, data security has become a must. Cloud user authentication is mostly required for safe data exchange. With centralized techniques, data storage, communication, and processing all have increased overheads. Shangping Wang et al. In [8] proposes a framework that combines the decentralized storage system IPFS, the Ethereum blockchain, and attribute-based encryption (ABE) technology to solve, the keyword search function on the ciphertext of the decentralized storage systems is implemented, that also eliminates the single point of failure in standard cloud storage solutions while also providing various advantages over the centralized system. Priyanka Maharu Salunke et al. In [9] In a decentralized access control system, it concludes by assisting in the concealment of user identification from the cloud. Instead of recognizing user identity, the cloud simply identifies user credentials. User revocation is also supported by this system, which removes users from shared data and prevents replay attacks. This proposed systemis both decentralized and robust in comparison to a

 Creative Commons Attribution 4.0 International

centralized approach. It also solves the issue of data retrieval. Shangping Wang et al. In [11] Blockchain provides a new solution to protect the personal health records sharing system as a distributed architecture with decentralized and tamper-proof characteristics. Bao-Kun Zheng et al. In [14] cryptosystem is used to create a trusted data sharing scheme that prevents the shared data from being tampered with and uses the Paillier cryptosystem to ensure the confidentiality of the shared data. The transaction information is secure, and the shared data can be traded. Qi Xia et al. In [16] As a result of this architecture, more accountability is ensured because all users are already known and the blockchain keeps track of their actions. After their identities and cryptographic keys have been validated, users can request data from the common pool. Our approach is lightweight, scalable, and effcient, according to the results of the system evaluation. Xueping Liang et al. In [18] Using blockchain technology, performance study show that ProvChain delivers security characteristics such as tamper-proof provenance, user privacy, and reliability for cloud storage applications while having low overhead.

2.2.3 Blockchain in Healthcare

Bingqing Shen et al. In [5] A session-based healthcare data-sharing strategy is created based on MedChain, which provides exibility in data sharing. The findings of the evaluation suggest that MedChain can improve productivity while also meeting security criteria in data exchange.

Ayesha Shahnaz et al. In [13] the framework by de_ning granular access controls. Furthermore, this framework addresses the scalability issue that blockchain technology has in general by utilizing off-chain record storing. This framework gives the EHR system the advantages of a blockchain-based solution that is scalable, secure, and integrated.

2.2.4 Blockchain in data sharing

Dagang Li et al. In [6] Our model's security study reveals that, due to the Meta-unique key's architecture, the proxy re-encryption used in our system is inherently free of collusion- attack. Saqib Ali, Guojun Wang et al. In [12] The PingER framework will get decentralized storage, distributed processing, and efficient lookup capabilities as a result of this.

2.2.5 Blockchain Privacy and security

Mangesh Gosavi et al. In [7] the method allows data owners to specify their access controls over user attributes and enforce those constraints in the data to be delivered. By establishing an escrow-free key issuing protocol that is developed using secure multiple computing between the key generating and data management facilities, this method also removes the Key Escrow Problem. This method is efficient in securely managing data dispersed in the data-sharing system, according to the performance and security study.

2.2.6 Blockchain in IoT

The Internet of Things (IoT) plays a critical role in automating our daily lives. Electronic gadgets communicate with one another via the internet and frequently share and exchange information. Given the privacy and security considerations, a system must be developed o assure data integrity and digital device authentication. Hossein Shafagh et al. In [15] Our system is unconcerned with physical storage nodes, and it also enables the use of cloud storage resources as storage nodes. Yun Zhang et al. In [17] For various levels/types of attacks, extensive simulation results show that the suggested scheme can provide a greater security level and stronger resilience to attack than a standard database-based data sharing method.

3.1 Proposed Solution

 Creative Commons Attribution 4.0 International

In the proposed approach In practical applications, blockchain are not suitable for storing large files (up to max size.) due to block bloated. In our scheme, blockchain based data sharing use file hash (IPFS CID) of the files are stored on the blockchain whereas the actual files are stored in IPFS use the SHA-256 algorithm to create a unique hash of the entire block that is utilized to form the chain by the related blocks. IPFS as well uses this algorithm to generate the hash of the shared file and AES Encryption is a form of symmetric, cryptographic encryption that depends on a secrete key between the sender and receiver to access any file.

Algorithm 1: Algorithm to uploading File

Input: *File type* F_t, *Encrypt File Fenc, Secrete Key K, Block B*

Output: *File Hash Fh*

```
begin
while n number of file do
    If Ft = true then
        Generate Fenc using AES
        convert Fenc into buffer
        call function to upload Fenc in to IPFS.
        store the Fenc in to Upload folder.
        return Fh
        update B with Fh
    else
        return error
    end if
end while
end
```

As seen in Algorithm 1, using function on Submit to store the file on the IPFS when uploading the file. Before uploading the file to the IPFS network, the Secret key entered here will be used to encrypt it with AES encryption. The sender and receiver of a shared file share a unique key to strengthen the security of the file(s) on the blockchain network..In this scheme, the storage nodes in IPFS are not considered here. The files are stored on the IPFS network, The IPFS network returns a hash of the file. Then it start the mining process to put the transaction in a block while the blockchain stores only the file hash of the file. Because IPFS uses the SHA-256 hashing technique, each file will have a unique hash. As either a result, the file is stored in a secure decentralized network and accessible through the blockchain. It is simple to get the hash generated by the file. As a result, IPFS eliminates the constraint of keeping entire files on the blockchain.Using AES

 Creative Commons Attribution 4.0 International

encryption and the uploader's file key, encrypt the file. As a result, if a user attempts to download the file straight via IPFS, they will receive an unreadable file. As a result, only users with a valid secret key can access the viewable file contents, strengthening the blockchain's and file content's security. Learn how to use IPFS's decentralized storage power to improve the blockchain's security and efficiency.

Algorithm 2: Algorithm to downloading File

Input: *File hash* F_h, *Decrypt File* F_{dec}, *encrypt File* F_{enc} *Secrete Key K,*

Output: *File type* F_t

```
begin
while n number of file do
    If F_h = true and K = true then
        call function to download F_enc from IPFS.
        Decrypt F_enc using AES method
        store F_dec in to download folder.
    else
        return error
    end if
end while
end
```

As seen in Algorithm 2, using function on Submit to download the file from the IPFS network. Before downloading the file from IPFS network, the Secret key entered here will be used to decrypt it with AES encryption.after downloading the file,file hash store in block of blockchain network.Even every any size of file return file hash has constant length.

$$f(x)=K$$

where x is the number of any size of files.
K is the constant file hash length.
This feature can prove the scaling solution of various blockchain.

4.1 Experimental Setup

A simulation conducted using propose Data sharing system and decentralized storage systems, namely ipfs ,windows 10 pc system,i5 processor,4gb ram,500gb harddisk. These environment using different file size with respective to upload and download times are recorded at similar speeds The average transfer speeds were calculated as:
i) 222 kilobytes/s for upload, and ii) 706 kilobytes/s for download Figure 4.1 and 4.2 shows the upload and download times for the files of the same size over the network. Table I and II summarize the obtained values

 Creative Commons Attribution 4.0 International

Table 1. Upload Times

File Size(inMB)	Upload Time (in seconds)	File Hash
5	25	QmUMXeGFoW7fjKvgTSjVx3RpEJMBd4v31DNgV1S1TRW6HT
10	43	QmXbMg5f5s161fjpPb5QDarNu4f9wrxgv5rzg857W1ynCE
20	93	QmeYhjgay7o9TCzxANdyiHLiNsj8MgR75VZ8shKu4e3jKe
30	145	QmNYV9SkHGWwNcDkwYygoNrBWhe6FDJuJZoMtuUGu3sSdi
40	188	QmYxr3vQ1w3ufYfJKprXzADQa7HLDrhxrzD2oJzuE4KX3A
50	238	QmRUJpuYiPfbxQLY5ayBDyNXJECYwPTwU2T9dafFkBrLKq

Table 1 shows that file size ,upload time in sec and file hash return from ipfs

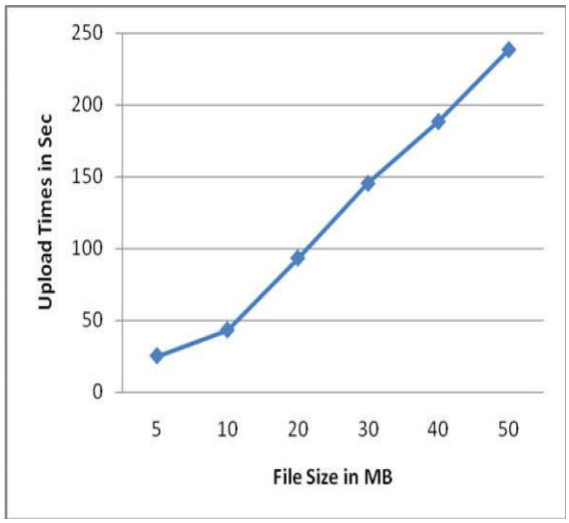


Figure 4.1. Upload Times in Decentralized Storage Systems

Figure 4.1 shows that file size in MB is directly proportional to Upload time in sec

Table 2. Download Times

File Size(inMB)	Download Time (in seconds)	File Hash
5	7	QmUMXeGFoW7fjKvgTSjVx3RpEJMBd4v31DNgV1S1TRW6HT
10	12	QmXbMg5f5s161fjpPb5QDarNu4f9wrxgv5rzg857W1ynCE
20	29	QmeYhjgay7o9TCzxANdyiHLiNsj8MgR75VZ8shKu4e3jKe
30	41	QmNYV9SkHGWwNcDkwYygoNrBWhe6FDJuJZoMtuUGu3sSdi
40	52	QmYxr3vQ1w3ufYfJKprXzADQa7HLDrhxrzD2oJzuE4KX3A
50	61	QmRUJpuYiPfbxQLY5ayBDyNXJECYwPTwU2T9dafFkBrLKq

 Creative Commons Attribution 4.0 International

Table 2 shows that file size ,download time in sec and file hash return from ipfs

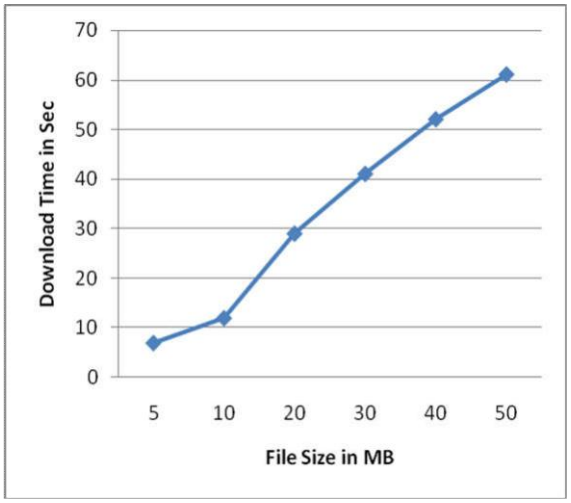
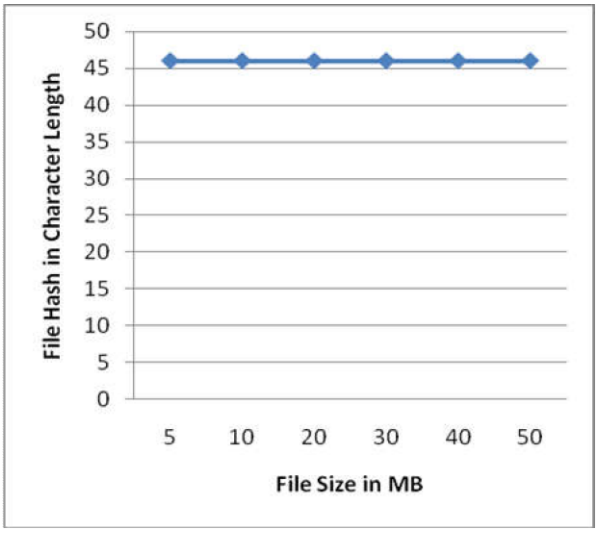


Figure 4.2. Download Times in Decentralized Storage Systems

Figure 4.2 shows that file size in MB is directly proportional to download time in sec . From Figure 4.1 and 4.2, it can be inferred that the transfer speeds of files in the systems are similar.

Figure 4.3. File Hash



This ensures that large files are stored on IPFS and only file hash which are relatively very small get recorded on the Blockchain. Figure 4.3 shows how the size of the file grows in relation to files added to the IPFS network. These must point out that the average size of file hash recorded on the ledger is around 46 character string length for each activity performed on the overlaying IPFS network. Consequently, the ledger size grows

 Creative Commons Attribution 4.0 International

but at a much slower rate and remains very stable even when the size of files added to the IPFS network drastically increases.

5.1 Discussion

5.1.1 Data integrity

When a user uploads a file to the IPFS network, the hash of the file is returned. The hash will link the user to the IPFS gateway at http://ipfs.io/ipfs/, where they can download the file without having to use the IPFS command line. To identify the content of a file, IPFS generates a hash. Even if files have similar names and sizes, IPFS's integrity function can establish each file's integrity. It can assist users in determining whether or not files have been tampered with.

5.1.2 Security and privacy analysis

For security and privacy purpose Files are chunked and stored on different storage nodes in IPFS. Firstly files are encrypted and stored on IPFS storage nodes using the AES method. The storage nodes can only see a portion of the ciphertexts and cannot access any file information. The File Hash (file location hash ciphertext) get from IPFS and saved on the blockchain using the AES algorithm. Although everyone may read the ciphertext content after encryption using the AES algorithms, Any user fulfil the access but cannot decrypt the encrypted file from IPFS. As a result, we were able to achieve data access control. The suggested system can be regarded to be safe as long as the blockchain network scheme are secure.

CONCLUSION

In comparison to a centralized system, a decentralized system is more secure, scalable, and has higher throughput. Decentralized systems perform similarly to centralized systems. The failure of the central node/server in centralized systems takes the entire network down, whereas decentralized systems have more fault tolerance. Blockchain are not suitable for storing large files (up to max size.) due to block bloated. In our scheme, blockchain based data sharing use file hash (IPFS CID) of the files are stored on the blockchain whereas the actual files are stored in IPFS The use of CIDs to address content increases data integrity in the proposed data sharing system. Decentralized systems offer more storage space for less money. As a result, in today's environment, decentralized systems have more to offer than centralized systems.

11.1. Journal Article

[1] *S. S. N. L. Priyanka and A. Nagaratnam, "Blockchain evolution a survey paper", IJSRSET, vol. 4, no. 8, 2018.*

[2] *V. Narkhede and S. D. Rajput, "Analysis of blockchain algorithms", IJIRSET, vol. 9,no. 3, March 2020.*

[3] *E. Kara_loski and A. Mishev, "Blockchain solutions for big data challenges", IEEE EUROCON, July 2017.*

[4] *L. Yue, H. Junqin, Q. Shengzhi, and W. Ruijin, "Big data model of security sharing*

based on blockchain", IEEE, 2017.

[5] *B. Shen, J. Guo, and Y. Yang, "Medchain efficient healthcare data sharing via blockchain", Appl. Sci., 2019.*

[6] *D. Li, R. Du, Y. Fu, and M. H. Auz, "Meta-key a secure data-sharing protocol underblockchain-based decentralised storage architecture", IEEE, 2019.*

[7] *M. Gosavi and T. Maktum, "Security e_ciency enhancement in attribute based distributed data sharing", IJSTE, vol. 3, no. 11, May 2017.*

[8] *S. WANG, Y. ZHANG, and Y. ZHANG, "A blockchain-based framework for data sharing with _ne-grained access control in decentralized storage systems", IEEE Access,2018.*

[9] *P. M. Salunke and V. V. Mahale, "Secure data sharing in distributed cloud environment", IEEE Xplore, 2018.*

[10] *S. K. Sangode1 and H. K. Barapatre, "Generate distributed metadata using blockchain technology within hdfs environment", IRJET, 2017.*

[11] *S. WANG1, D. ZHANG1, and Y. ZHANG, "Blockchain-based personal health records sharing scheme with data integrity verifiable", IEEE Access, Mar 2018.*

[12] *S. Ali, G. Wang, B. White, and R. L. Cottrell,"A blockchain-based decentralized data storage and access framework for pinger", IEEE, 2018.*

[13] *A. SHAHNAZ, U. QAMAR, and A. KHALID, "Using blockchain for electronic healthrecords", IEEE Access, October 2019.*

[14] *B.-K. Zheng, L.-H. Zhu, M. Shen1, F. Gao1, C. Zhang, Y.-D. Li, and J. Yang, "Scalable and privacy-preserving data sharing based on blockchain", J Comput Sci Technol, March2018.*

[15] *H. Shafagh, L. Burkhalter, and A. H. andSimon Duquennoy, "Towards blockchain-based auditable storage and sharing of iot data", CCSW 17, November 2017.*

[16] *Q. Xia, E. B. Sifah, A. Smahi, S. Amofa, and X. Zhang, "Bbds blockchain-based data sharing for electronic medical records in cloud environments", Information, 2017.*

[17] *Y. Zhang, D. He, , and K.-K. R. Choo, "Bads blockchain-based architecture for data sharing with abs and cp-abe in iot", Wireless Communications and Mobile Computing, 2018*

[18] *X. Liang, S. Shetty, D. Tosh, C. Kamhoua, K. Kwiat, and L. Njilla, "Provchain a blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability", IEEE, 2017.*

[19] *Ashok Aravind S N, Divyesh P K, Prateek Joshi, Srinivasan R, Rekha P "Analysis Of Ipfs Based Decentralized Storage" JCR vol 7, issue 14, 2020.*