Implementation of Distributed File Storage and Access Framework using IPFS and Blockchain

Randhir Kumar
Department of Information Technology
National Institute of Technology,Raipur
CG, India-492010
rkumar.phd2018.it@nitrr.ac.in

Rakesh Tripathi
Department of Information Technology
National Institute of Technology,Raipur
CG, India-492010
rtripathi.it@nitrr.ac.in

***Abstract*—Many critical applications are designed on the distributed structure using the blockchain technology to ensure the availability, immutability, and security. However, these applications are facing the storage problem owing to the data volume growth of transaction. The number of transactions and its size in a block is growing in the blockchain day by day because of the feature of immutability and append-only. The growing nature of transactions in a block is not only making the problem for storage but also in access to the block transactions. In this paper, we propose an IPFS based blockchain storage model to solve the storage problem of transaction in a block along with access of transaction of a particular block. In the propose storage model, the miners stores transaction on IPFS distributed file system storage and get the returned IPFS hash of transaction into the block of the blockchain. The feature of the IPFS network and its resultant hash reduce the size of transactions in a block. To secure access of transaction for a particular block content-addressed (IPFS hash) storage technique has been proposed. We have applied this scheme on a transaction which includes image storage on IPFS and hash storage into the blockchain. In this paper, we have also proposed the content-addressed technique in contrast to the location addressed for the access of transaction. To implement the framework we have used anaconda python, python flask, and IPFS.**

***Index Terms*—Blockchain, InterPlanetary File System(IPFS), Distributed storage, Content-addressed technique.**

I. Introduction

Recently, the blockchain technology is emerging as a decentralized database storage technology [1]. Moreover, the blockchain provides features such as immutable, shared, and decentralized database that stores history of assets and transactions across a distributed peer-to-peer network. The blockchain is successfully utilized in bitcoin developed by Satoshi Nakamoto but now it is gaining popularity in the various fields including healthcare, supplychain, and financial sector [2].

There are many distributed file storage applications currently running all over the world. However, these applications are having problem of storage space and access with other peers in the network. The BitTorrent is the distributed file storage and sharing application which provides the facilities to coordinate with untrusted peers of the distributed network [4]. The more important and notable thing of the BitTorrent is that , about 170 million users uses BitTorrent every month [5]. Moreover, BitTorrent protocols takes 40% of the internet traffic on daily basis [5]. However, these application does not secure and efficient when compared with HTTP. The HTTP protocol is one of the most considerable protocol in terms of development for distributed applications worldwide. However, the HTTP protocols work on location-addressed based protocols which is not much efficient in terms distributed access of file. To overcome the drawback of BitTorrent and HTTP , the Interplanetary File System(IPFS) [3], distributed file storage network is used for building a better web for us. The IPFS works with the content-addressed technique to store and access of file.

The distributed storage provides structure where data is disseminated on different nodes, which refers the same distributed storage system. The each node is peers with one another in the distributed network. The all peers keeps same capabilities to share and access the files i.e., no one is important and superior to others. Moreover, the distributed network also replicates the information which ensures the availability of information in the system.

IPFS(Interplanetary File System) provides distributed file storage system which facilitates connection with P2P network. The IPFS calculates the unique Hash of a file which is accessible to all the peers in the network. The hash get modified every time when file gets updated [6].

The IPFS is considered as the backbone of web 3.0 which provides P2P decentralized file storage system and content-addressable technique to access the stored file [3]. IPFS facilitate the distributed hash table (DHT) which is much efficient than BitTorrent, and Git file storage system. IPFS is also known as version controlled system which ensure the security, reliability, and scalability faced by the existing file storage and sharing system. The same transaction has the same hash in the IPFS that ensures the originality of the content. The IPFS hash gets disseminated to all the peers which also ensures the consistency among the peers. Interplanetary File system provides high throughput content-addressed block storage model which ensures security of transactions.

IPFS provides a high throughput in terms of access using content-addressed block storage model, we also notice that

978-1-7281-0899-5/19/$31.00 ©2019 IEEE

storage is a new challenge in blockchain technology owing to the append only features [4]. In this paper, we try to leverage the concept of decentralized P2P system in IPFS to eliminate the limitations of blockhain storage and ensure data availability along with reliability using IPFS and blockchain. In this paper, we have considered the blockchain transaction as a image File.

AIM: We propose a blockchain data storage model that illustrates how to use IPFS networks to reduce the storage of transaction size of block in blockchain. Moreover, we have also used the content-addressed based access of transactions.

Organization: This paper is organized as follows Section-II describe related work of blockchain in healthcare, Section-III describe the proposed structure, Section-IV describe the Implementation and Result of the proposed framework, Section-V concludes the paper.

II. Related Work

The authors in [7]–[9] has reported that currently a complete ledger of Bitcoin network occupies up to 200GB, and this number constantly increasing at about 0.1 GB per day, causing great pressure on data storage. Moreover, the work in [8] proposed network coding-based distributed storage (NC-DS) framework, which divides the transactions in different block and again block into sub-block.The each sub-block is encoded and distributed to all the peers(nodes) in network. This scheme reduced the transaction size into some extends but at the same time it also increased the complexity using encoding and decoding process.

The author in [10] adopted the method of recording block techniques which does not record the expired transaction history. The header information of expired transaction is stored in the recording block technique. This approach reduces the size of block in blockchain but at the same time traceability of the complete transaction is not possible owing to the header part storage only.

The authors in [12] adopted the summary block techniques which includes the active transaction records in blockchain. These records are kept in external file and the remaining old transactions are deleted from the blockchain. This technique reduces the size of block in blockchain but the drawback of this technique is that it can not retain the history of blockchain.
The authors in [13] has discussed the account-tree techniques which only includes transactions of the peers who is part of the networks. In this approach the account of each peers are recorded in off-chain(Database)storage and matched during the transactions. However, this approach does not maintain record of old transactions. the following approach is suitable to reduce the size of block owing to the does not include the history of old transactions. The account-tree techniques does not suitable for complete traceability of blockchain.

The above techniques has some limitations and drawback, and the size of the transactions is also not reduced effectively. In this paper, we are using blockchain and IPFS to manage the block transactions. We are using image file as block transaction data in blockchain. The transaction data (image file) is stored into the IPFS. Only the IPFS hash data is stored as a transaction in the block of blockchain [14].

Moreover, we provide the content-addressed techniques which creates permanent storage while access of transaction. The provided hash by the IPFS point the same transaction exactly owing to the content-addressed technique. Moreover, the content-addressed has many implication from a access perspective like transaction storage, and transaction access on IPFS. Once the hash of transaction gets created then it can be accessed by other peers in the network efficiently without revealing of location of the transactions [3].

Motivation: - The work in this paper is motivated by the following observation from the Literature [10], [12], [13], we have notice that storage is a challenge in blockchain network owing to the append only feature [4]. In this paper, we have proposed IPFS-based blockchain storage system which eliminates the limitations of existing blockchain storage. We have also proposed the content-addressed access of transaction rather than simply location based access of transactions.

III. Proposed Structure

A. Work flow synchronization of Miners

There are two major work flow is performed by miners i.e, participating in network process and mining process.
Process of new peers synchronizations:
The newly peer(node) join the network must have synchronized with the existing distributed structure. The transaction made by the peer must be synchronized to maintain the consistency in network. To maintain the synchronization in network all the newly join peers must have to follow the consensus of blockchain.

Process of Mining
The mining is process in which all the transactions is verified before submissions to the block of blockchain by miners. The verification of transaction involves format of transactions, input of transactions, and the transactions meet the requirement of consensus. once the transactions are verified then it gets deposited to the mining pool. To add the transactions in a new block, the miners verify the previous block. Furthermore, the previous block hash is added to the new block before packed into a new block in blockchain. To successful addition of a new block, the miners solve a crypto-puzzle by calculating the block hash [15]. The miners who first calculates the block hash disseminates the block to the other peers(nodes) in a blockchain network. The other nodes in the network mutually verify the correctness of block. Moreover, the verified block by the peers get added by the other miners as a new block in blockchain network [16].

As shown in Algorithm 1, the files are uploaded by the peers of the network. to store the files into the IPFS distributed file storage ipfsapi has been used. The connection of ipfs is established on the port number 5001. The hash value (Hv) is calculated in binary format and then hash of

Algorithm 1: Algorithm for content addressed hash

```
Input: Image Files
Output: Content-addressed hash of the File
//the file uploaded by the peers in the network
$file \leftarrow request.files['upload']$
//initiated the IPFS distributed file storage
$api \leftarrow ipfsapi.connect('localhost', 5001)$
//original file addition into IPFS storage
$res \leftarrow api.add(file)$
// file to binary conversion
$Hv \leftarrow convert(file, binary)$
// created the hash message digest of the file (create
  hash − > ch)
$digest \leftarrow ch('sha256').update(file).digest()$
// created the message digest using salting methods
$Mds \leftarrow (digest.bytelength.toString(16), 'hex')$
// combining binary conversion and salting message
  digest for content-addressed hash
$content\_addressed \leftarrow combine(Hv, Mds, digest)$
// returning content addressed hash to store in
  blockchain
Return $content\_addressed\_hash$
```

the file gets created using sha-256 method [11], [17]. The salting method approach is applied to convert the message digest into the hex format. Finally, the message digest (Mds) and hash value are combined together to generate content-addressed hash of the file. The content-addressed hash is stored into the blockchain network in order to reduce the size of the distributed ledger.

B. IPFS based blockchain storage data model

Introduction to IPFS

IPFS is a distributed data storage which provides content addressed and allocates the unique hash for stored file. IPFS facilitate the high-throughput and efficient storage model with concurrent access. The IPFS creates hash which is 46 bytes long. Therefore, storing transaction data in IPFS and storing the hash returned by IPFS into the block of blockchain provides large reduction in storage space.

Blochchain Storage Model

To provide the efficient blockchain storage model and the feature of IPFS, the IPFS-based data storage model has been proposed for blockchain, as shown in Fig.1.

In this model the miners collect the transactions and puts the valid transactions into the mining pools and stores them into the IPFS distributed network. The IPFS provides the unique hash for the stored transactions during the mining process which is use to create a new block. The stored transaction can be accessed by the unique hash value provided by the IPFS which is known as content addressable access. The proposed technique does not disclose the actual access name of the transaction.

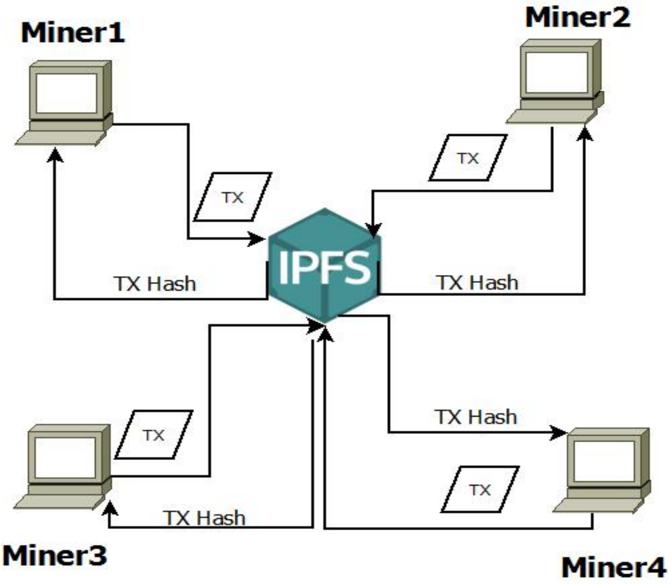


Fig. 1. Working Model of Blockchain and IPFS

C. Implication of Content-addressed Model of IPFS

1) To increases the integrity of Transaction: The IPFS content-addressed increases the integrity of transaction owing to the fingerprint(hash). The obtained hash from IPFS makes transaction reliable and resilient. This types of fingerprint validation becomes difficult with location-addressed.

2) Easy to Access Transaction: In the IPFS distributed network, as long as one node has copy of the transaction then other peers will be able to access it. The attractive feature of the IPFS is the hash of the transaction get changed when transaction get updated without changing the peers connected to the IPFS hash. The content-addressed techniques of IPFS signify the importance of transaction and provides belief to the peers that they are accessing exact content which was originally stored.

3) Easy to Distribute Transaction: In the IPFS peers can access the content and distribute the content to others peers. The "PIN" command is used to distribute the content to others peer. To get the details of "PIN" command must visit the URL "$https : //docs.ipfs.io/reference/api/cli/\#ipfs - pin$"

IV. Implementation and Results

The implementation of our proposed IPFS-based blockchain storage model is carried out in the IPFS distributed file sharing system where each transaction stored on IPFS and calculated unique hash is stored in blockchain transaction. The experimental setup consists of python anaconda, python flask. The setup is performed on Intel(R) Xeon(R) W-2175 CPU @ 2.50GHZ running Window x64- based processor with 128 GB of RAM and 1 TB of local storage.

A. File upload to IPFS

As shown in Fig.2 and Fig.3, the transaction(Image file) is uploaded onto the IPFS. The transaction which is large in size should not be stored in the block of blockchain directly. To achieve the efficient storage of the transaction we are storing the transaction on IPFS.

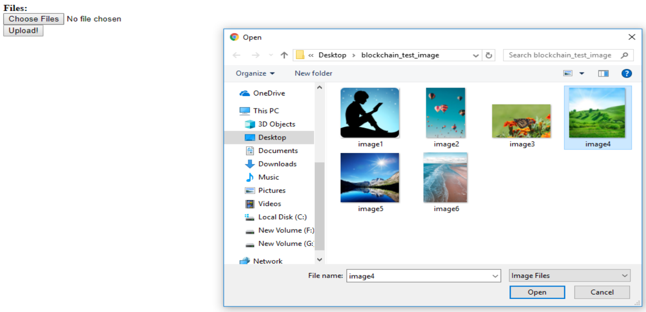


Fig. 2. Image as transaction on Blockchain using IPFS

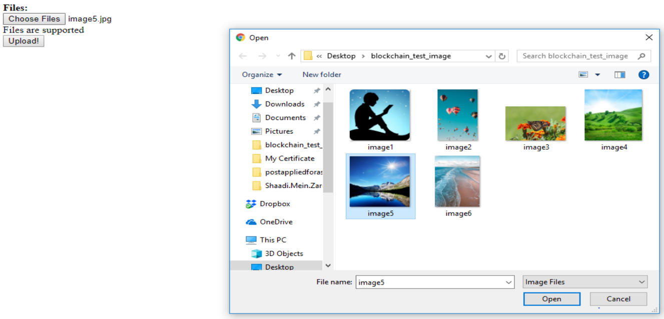


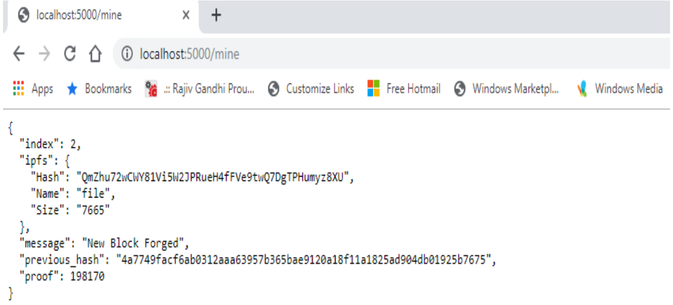
Fig. 3. Image as transaction on Blockchain using IPFS

B. Process of Mining to retrieve the IPFS Hash

As discussed in the proposed structure, the transaction is stored to the IPFS. The IPFS provides unique hash of the transaction during mining process as shown in Fig.4 and Fig.5. We have mined the both of the transaction performed in the Fig2. and Fig.3, and retrieved the unique hash for both the transaction along with their size. The mining process includes the index number which is block number in blokchain and the previous hash maintain the chain of block. Moreover, the list of transactions can be mined during the mining process to create a single block in blockchain. The generated IPFS hash value is 46 bytes which provides efficient and less storage size than original size of the file i.e., 4KB as shown in Fig.5. In the mining process we are generating the IPFS hash value rather than to store the original size of the transactions.

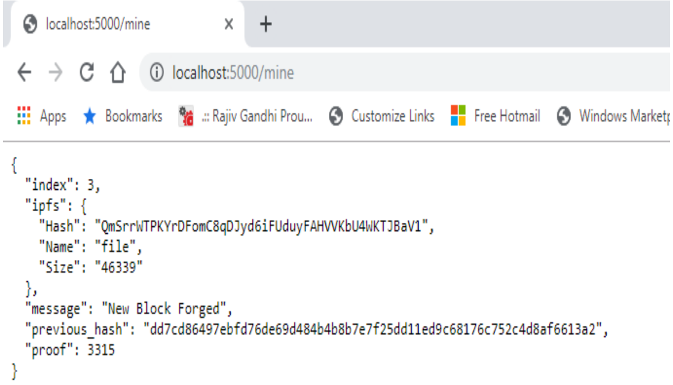
C. List of transactions in blockchain

The list of transactions are shown in Fig.6, which contains all the IPFS hash value with their block number(index number). The first block is genesis block which does not contains any IPFS hash value. Every block contains timestamp which denotes the creation time of block to ensure the reliability. The prrof-of-work maintains consistency of the block over the peers of blockchain. AS shown in Fig.6, the length denotes the chain size of blockchain. The resultant blockchain facilitate availability, reliability and immutability and reduced size of the transactions.



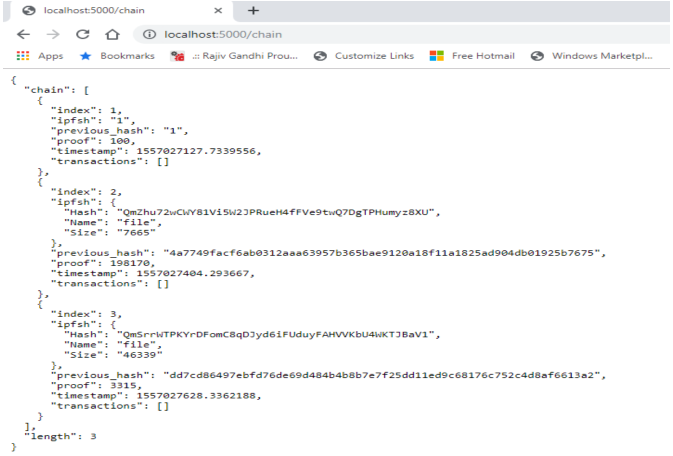
```
{
  "index": 2,
  "ipfs": {
    "Hash": "QmZhu72wCWY81Vi5W2JPRueH4fFVe9twQ7DgTPHumyz8XU",
    "Name": "file",
    "Size": "7665"
  },
  "message": "New Block Forged",
  "previous_hash": "4a7749facf6ab0312aaa63957b365bae9120a18f11a1825ad904db01925b7675",
  "proof": 198170
}
```

Fig. 4. To retrieve the unique hash of the transaction



```
{
  "index": 3,
  "ipfs": {
    "Hash": "QmSrrWTPKYrDFomC8qDJyd6iFUduyFAHVVKbU4WKTJBaV1",
    "Name": "file",
    "Size": "46339"
  },
  "message": "New Block Forged",
  "previous_hash": "dd7cd86497ebfd76de69d484b4b8b7e7f25dd11ed9c68176c752c4d8af6613a2",
  "proof": 3315
}
```

Fig. 5. To retrieve the unique hash of the transaction

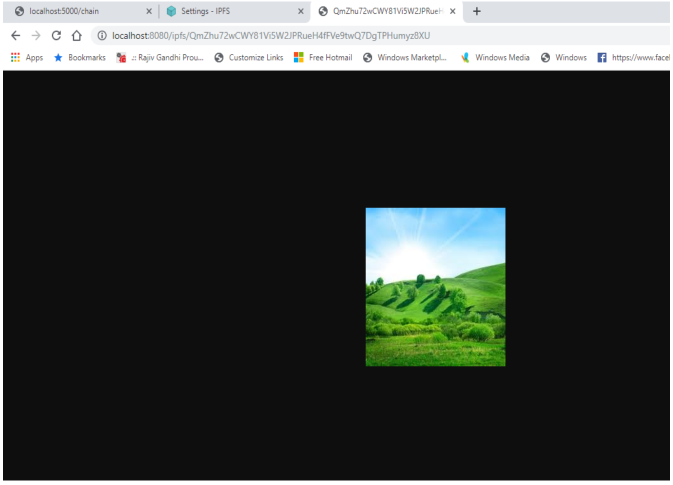


```
{
  "chain": [
    {
      "index": 1,
      "ipfsh": "1",
      "previous_hash": "1",
      "proof": 100,
      "timestamp": 1557027127.7339556,
      "transactions": []
    },
    {
      "index": 2,
      "ipfsh": {
        "Hash": "QmZhu72wCWY81Vi5W2JPRueH4fFVe9twQ7DgTPHumyz8XU",
        "Name": "file",
        "Size": "7665"
      },
      "previous_hash": "4a7749facf6ab0312aaa63957b365bae9120a18f11a1825ad904db01925b7675",
      "proof": 198170,
      "timestamp": 1557027404.293667,
      "transactions": []
    },
    {
      "index": 3,
      "ipfsh": {
        "Hash": "QmSrrWTPKYrDFomC8qDJyd6iFUduyFAHVVKbU4WKTJBaV1",
        "Name": "file",
        "Size": "46339"
      },
      "previous_hash": "dd7cd86497ebfd76de69d484b4b8b7e7f25dd11ed9c68176c752c4d8af6613a2",
      "proof": 3315,
      "timestamp": 1557027628.3362188,
      "transactions": []
    }
  ],
  "length": 3
}
```

Fig. 6. Blockchain Transactions List

D. Content-Addressed based access of transaction

As shown in Fig.7, the transaction is accessed by their hash value which is provided by the IPFS at the time of mining process. The IPFS follow content-addressed storage rather than location-addressed. In location addressed the storage keeps the record of list, or directory, of the locations. However, the location-addressed reveal the path of the transaction when request is made for a particular transaction. The content-addressed storage (IPFS) keeps record of hash value of the transaction rather than keeping the record of location of transaction. The content-addressed storage does not reveal the location of the transaction , hence keeping the transaction on IPFS distributed storage is much efficient in-contrast with location-addressed storage. As shown in Fig.7, we are accessing the transaction using their content-address. To see the transaction on IPFS using content-addressed access the url must be followed ”localhost:8080/ipfs/<hash of file >”


localhost:5000/chain

Settings - IPFS

QmZhu72wCWY81Vi5W2JPRueH

localhost:8080/ipfs/QmZhu72wCWY81Vi5W2JPRueH4fFVe9twQ7DgTPHumyz8XU

Fig. 7. Transaction access based on the IPFS hash

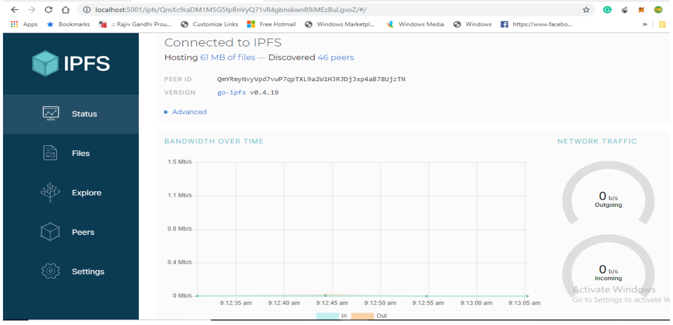
E. IPFS Connection

Fig.8, shows the connection of IPFS to see the details of status, files and peers on distributed environment. The status describes bandwidth occupied by the IPFS during storage and access of transaction. As shown in Fig.9 and Fig.10, the details of all peers can be verified in IPFS such as how many peers are connected at a time and how many are accessing the resource on IPFS. The all peers are verified by the ”PEER ID” and their ipfs address. Moreover we can also see the peers location on IPFS.

To establish the connection with IPFS the url must be followed ”127.0.0.1:5001/webui”.

F. Access of Bandwidth by IPFS

The process of data storage and access on IPFS distributed network takes bandwidth which is shown in Fig. 11, There are two different types of bandwidth is occupied by the IPFS such as “IN” and “OUT” bandwidth. The “IN” bandwidth is utilized when transactions gets stored into IPFS and “OUT” bandwidth is utilized during transactions accessed from IPFS.


localhost:5001/ipfs/QmXc9raDM1M5G5fpBnVyQ71vR4gbnskwnB9iMEzBuLgvoZ/#/

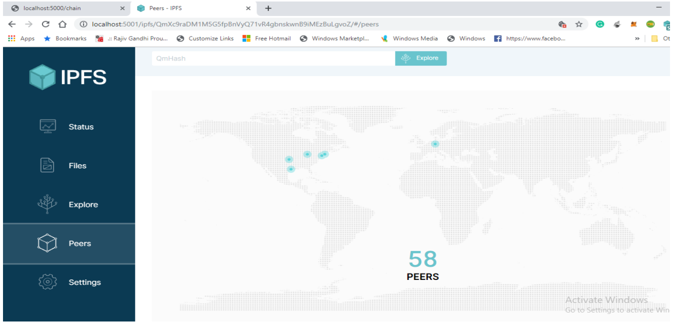
Connected to IPFS

Hosting 61 MB of files — Discovered 46 peers

PEER ID QmYRmyNvyVpd7vwP7qpTXL9a2W1HJRJDjJxp4aB78UjzTN

VERSION go-ipfs v0.4.19

Fig. 8. Connection to IPFS distributed File storage system


localhost:5000/chain

Peers - IPFS

localhost:5001/ipfs/QmXc9raDM1M5G5fpBnVyQ71vR4gbnskwnB9iMEzBuLgvoZ/#/peers

Fig. 9. Connection to IPFS distributed File storage system

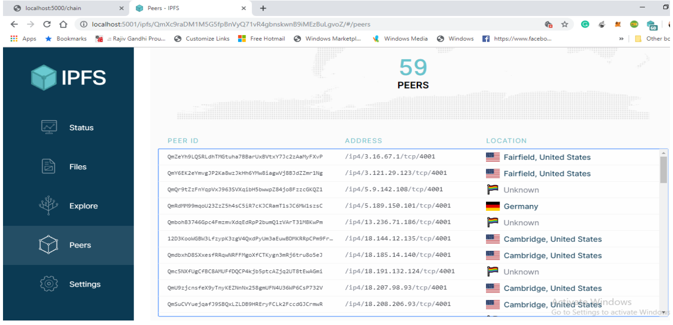
V. Conclusion

In this paper, we have designed the IPFS-based storage model for blockchain to mitigate the drawback of storage and transaction access of a block in blockchain network. The result of our IPFS storage model provides efficient storage space owing to the content-addressed hash of the transactions. Moreover, the content-addressed scheme is applied to access the transactions of block using the hash value provided by the IPFS distributed storage. In the proposed model, we are storing the hash of the transaction rather than storing the complete transaction or original transaction to ensure the efficient storage scheme for blockchain network. our proposed structure of storage model can be utilized with the other types of transaction files such as video and audio on blockchain. The existing storage model like bitcoin, ethereum, hyperledger suffers from the storage of bulky data in the distributed ledger of blockchain network. Thus, our model can also be applied on existing storage model to reduce the size of each block.

References

[1] Andreas, M., and Masteing Bitcoin. ”Unlocking Digital Cryptocurrencies.” (2015): 49-68.

[2] Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H. (2017, June). An overview of blockchain technology: Architecture, consensus, and future trends. In 2017 IEEE International Congress on Big Data (BigData Congress) (pp. 557-564). IEEE



59
PEERS

PEER ID	ADDRESS	LOCATION
QmZeYh9LQSRLdhTMGtuha7BBarUxBVtxY7Jc2zAaMyFXvP	/ip4/3.16.67.1/tcp/4001	Fairfield, United States
QmY6EK2eYmvgJP2Ka8wzJkHh6YMw8iagwVj8BJdZZmr1Ng	/ip4/3.121.29.123/tcp/4001	Fairfield, United States
QmQr9tZzFnYqpVxJ963SVXqibH5bwwpZ84jo8FzzcGKQZ1	/ip4/5.9.142.108/tcp/4001	Unknown
QmRdMM99mqoU23ZzZ5h4sC5iR7cKJCRamT1sJC6MW1szsC	/ip4/5.189.150.101/tcp/4001	Germany
Qmboh83746Gpc4FmzmvXdqEdRpP2bumQ1zVArT31M8KwPm	/ip4/13.236.71.186/tcp/4001	Unknown
12D3KooWGBW3LfzypK3zgV4QxdPyUm3aEuwBDMKRRpCPm9Fr…	/ip4/18.144.12.135/tcp/4001	Cambridge, United States
QmdbxhD8SXxesfRRqwNRFFMgoXfCTKygn3mRj6tru8o5eJ	/ip4/18.185.14.140/tcp/4001	Cambridge, United States
Qmc5NXfUgCfBC8AMUFfDQCP4kjb5ptcAZjq2UT8tEwAGmi	/ip4/18.191.132.124/tcp/4001	Unknown
QmU9zjcnsfeX9yTnyKEZNnNx258gmUFN4U36WP6CsP732V	/ip4/18.207.98.93/tcp/4001	Cambridge, United States
QmSuCVYuejqafJ9SBQxLZLDB9HREryFCLk2FccdGJCrmwR	/ip4/18.208.206.93/tcp/4001	Cambridge, United States

Fig. 10. Information of IPFS peers

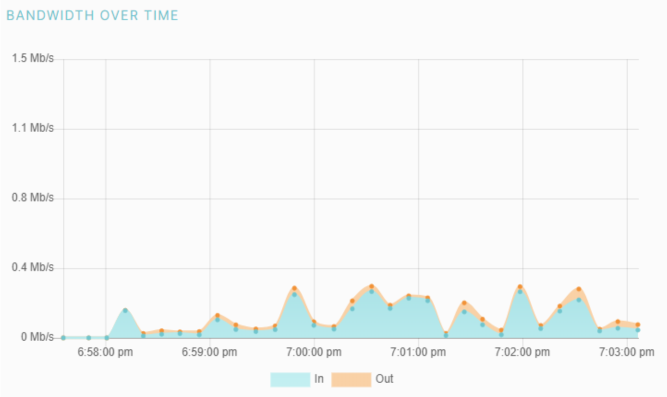


Fig. 11. Transaction submission and Access Bandwidth by IPFS

[3] Benet, J. (2014). Ipfs-content addressed, versioned, p2p file system. arXiv preprint arXiv:1407.3561.

[4] Chen, Y., Li, H., Li, K., & Zhang, J. (2017, December). An improved P2P file system scheme based on IPFS and Blockchain. In 2017 IEEE International Conference on Big Data (Big Data) (pp. 2652-2657). IEEE.

[5] Napoli, C., Pappalardo, G., & Tramontana, E. (2014, June). Improving files availability for bittorrent using a diffusion model. In 2014 IEEE 23rd International WETICE Conference (pp. 191-196). IEEE.

[6] Alessi, M., Camillo, A., Giangreco, E., Matera, M., Pino, S., & Storelli, D. (2018, June). Make Users Own Their Data: A Decentralized Personal Data Store Prototype Based on Ethereum and IPFS. In 2018 3rd International Conference on Smart and Sustainable Technologies (SpliTech) (pp. 1-7). IEEE.

[7] Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system.

[8] Dai, M., Zhang, S., Wang, H., & Jin, S. (2018). A low storage room requirement framework for distributed ledger in blockchain. IEEE Access, 6, 22970-22975.

[9] Clark, J. B. A. M. J., Edward, A. N. J. A. K., & Felten, W. (2015). Research perspectives and challenges for bitcoin and cryptocurrencies. url: https://eprint. iacr. org/2015/261. pdf.

[10] Dorri, A., Kanhere, S. S., & Jurdak, R. (2017, April). Towards an optimized blockchain for IoT. In Proceedings of the second international conference on Internet-of-Things design and implementation (pp. 173-178). ACM.

[11] Kumar, A., Ghrera, S. P., & Tyagi, V. (2015). A comparison of buyer-seller watermarking protocol (BSWP) based on discrete cosine transform (DCT) and discrete wavelet transform (DWT). In Emerging ICT for Bridging the Future-Proceedings of the 49th Annual Convention of the Computer Society of India (CSI) Volume 1 (pp. 401-408). Springer, Cham.

[12] Palai, A., Vora, M., & Shah, A. (2018, February). Empowering light nodes in blockchains with block summarization. In 2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS) (pp. 1-5). IEEE.

[13] Zheng, Q., Li, Y., Chen, P., & Dong, X. (2018, December). An Innovative IPFS-Based Storage Model for Blockchain. In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI) (pp. 704-708). IEEE.

[14] Benet, J. (2014). Ipfs-content addressed, versioned, p2p file system. arXiv preprint arXiv:1407.3561.

[15] Tosh, D. K., Shetty, S., Liang, X., Kamhoua, C. A., Kwiat, K. A., & Njilla, L. (2017, May). Security implications of blockchain cloud with analysis of block withholding attack. In Proceedings of the 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (pp. 458-467). IEEE Press.

[16] Hao, Y., Li, Y., Dong, X., Fang, L., & Chen, P. (2018, June). Performance Analysis of Consensus Algorithm in Private Blockchain. In 2018 IEEE Intelligent Vehicles Symposium (IV) (pp. 280-285). IEEE.

[17] Kumar, A., Ghrera, S. P., & Tyagi, V. (2015). Modified buyer seller watermarking protocol based on discrete wavelet transform and principal component analysis. Indian Journal of Science and Technology, 8(35), 1-9.