A Secure and Distributed Framework for sharing COVID-19 patient Reports using Consortium Blockchain and IPFS

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

2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC) | 978-1-7281-7132-6/20/$31.00 ©2020 IEEE | DOI: 10.1109/PDGC50313.2020.9315755

***Abstract*—Today healthcare industries are maintaining COVID-19 patients' information electronically which includes patients' diagnostic reports, patients' private information, and doctor prescriptions. However, the COVID-19, patient sensitive information is currently stored in centralized or third-party storage model. One of the key challenge of centralized storage model is the preserving privacy of patient information and transparency in the system. The privacy risk include illegitimate access to sensitive information of patient such as identification details access and misutilization of patient information and their clinical records. To overcome this challenge, we proposed a distributed on-chain and off-chain storage model using consortium blockchain and interplanetary file systems (IPFS). The proposed framework though maintaining patient privacy makes it easier for legitimate entities like healthcare providers (e.g., physicians and clinical staffs) to access clinical data of COVID-19 patients'.**

***Index Terms*—Blockchain, off-chain storage, Interplanetary File Systems (IPFS), Local Mining, peer-to-peer (P2P) storage and sharing model.**

I. Introduction

Today healthcare industry is generating large volume of COVID-19 patients' information that must be preserved, publicize, and accessed accordingly. For example, each COVID-19 patient goes under the various clinical test viz., X-Ray, CT-SCAN, and tomography, to identify the corona positive patient. This ailment must be disseminated among other healthcare provider to identify the symptoms of COVID-19 patient's. These data must be preserved so that it can be easily accessible from the other healthcare physician, if need arises. These COVID-19 information must be kept private and tempered-proof [1], [2]. These requirements demands the immutable, transparent, and decentralized process for storing COVID-19 clinical records.

The traditional infrastructure of patient reports sharing and storage are built on the centralized (cloud services) infrastructure. However, the underlying storage layer is prone to single point of failure, mutability, privacy, and security concerns. The peers (hospitals or doctors) have to trust and rely on this third-party storage (centralized storage layer) for sharing and access the patient's records. Thus, the current storage layer must be decentralized where each peers (hospitals) have control over their records without dependency of third-party.

The blockchain technology fulfil the requirement of transparent information sharing. It provides decentralized storage, wherein a COVID-19 patient's clinical records could be disseminated easily with the peers (doctors or hospitals) in healthcare system [3], [4]. The underlying technology guarantees diverse features such as immutability (tempered-proof), privacy, and consistency which are key requirement of healthcare system that is working on COVID-19 patients' clinical record collections. The blockchain contains list of subsequent blocks linking with a cryptographic hash which ensures immutability of patient's clinical records [5].

The details of COVID-19 clinical records including all ailments details of the patients' could be processed efficiently and stored into the blockchain network [6], [7], this records are transparent to all the healthcare providers (hospitals or doctors) working within the distributed network. Each information is recorded on the block-chain network can be easily accessed by the registered peers (hospitals or doctors) in the network. However, to maintain the large volume of the COVID-19 patients' records, distributed off-chain storage is required that supports peer-to-peer (P2P) working model. Moreover, these off-chain storage need to be secure and must facilitate persistent storage of COVID-19 patient's records.

Interplanetary File systems (IPFS) facilitate peer-to-peer distributed file storage model where large volume of the COVID-19 patients records (patient private information, doctor prescriptions, and diagnostic reports) can be stored easily. The underlying storage model stores content-addressed hash of the files and removes the duplicate files using distributed hash table (DHT) and version-control techniques [8]. The IPFS generates content-addressed hash of the files locally by high frequency request in the system, to provide availability of the hash in the network as per needs. In addition, the IPFS facilitates high throughput, secure transaction hash mapping, and transparent access of records in the network. To integrate the IPFS with blockchain facilitate better scalability of the network, as the IPFS storage layer is distributed.

Rather than storing the actual clinical reports of COVID-19 patient's in blockchain (on-chain storage) network, the content-address hash must be stored for the patient's clinical records. To access these COVID-19 clinical records, the peers (hospitals or doctors) can access corresponding hash (content-addressed) of the clinical reports. The hash of the patient's clinical records is created and added to the blockchain network, to facilitate the scalability in the network [9], [10].

Motivation: From the various literature [1]–[3], [5]–[7], [9], [10], we observed that most of the work is performed to share the patient reports with small sizes in KB. However, to maintain the privacy certificate authority (CA) based services has been implemented that executes on centralized server. Thus, we have used the consortium, proof-of-identity (PoI), and wavelet-based hash approach in the proposed framework to avail the privacy and security services.

Novelty: In this study, we designed and implement a consortium block-chain network, to store and manage the details of COVID-19 patient's records. In a consortium block-chain network only authorized (registered) peers can be a part of the network. The consortium block-chain network ensures better privacy of the patients' clinical reports. In the consortium network, each peers (hospitals or doctors) are assigned with a unique identity (Proof-of-Identity), to access the information in the network. we also used wavelet hash to identify the similar COVID-19 reports.

The rest of the paper is organized as follows: Section 2 discusses the related study of blockchain- IPFS storage layer. Section 3, describes the proposed model and their working, Section 4, discusses about the implementation of the model, and Section 5, concludes the paper.

II. Related Study

The state-of-the-art in healthcare measuring and comparison work has a relatively very long history. However, there is currently very little work on recommending and evaluating privacy and security of COVID-19 applications, due to newness of the work. Authors in [1]–[3], [5]–[7], [9], [10] have proposed the blockchain based decentralized storage layer for the COVID-19 patient reports, as the blockchain storage layer facilitiate the traceability, privacy, security, integrity of the data. The authors have also mentioned that, the underlying technology can address the various challenges of the centralized storage systems. However, the authors have not yet given the adequate implementation of the blockchain decentralized model in case of COVID-19 patient reports sharing framework. The work in [4] presented the blockchain based storage and sharing for medical records of patient, but didn't mentioned the off-chain storage layer, how the actual data is stored. The work in [13] presented the IPFS based distributed storage layer for data sharing and improved the scalability of the blockchain network by preserving content-addressed hash. However, authors didn't mentioned any perceptual hash based techniques, to find the similarity of the shared document.

III. Proposed Model

In this section, we present an integration of the on-chain (blockchain) and off-chain (IPFS) storage model using blockchain and IPFS for COVID-19 patient clinical (diagnostic) reports. The underlying on-chain storage model provides immutability and keeps privacy of the patient's records where as off-chain storage model stores actual diagnostic reports by providing content-addressable hash. The key purpose of this process is to preserve the privacy of clinical records. Fig. 1 illustrate the framework. We've split our framework into three separate modules to maintain privacy i.e., *reports upload*, *process of mining*, and *reports storage*. The health care providers (hospitals or doctors) could upload each patient's information via an interface of web users. The mining process is carried out to verify the transaction (clinical reports) and maintaining consistency in the network of consortium blockchain.Finally, the uploaded records' content-addressable hash is stored in the consortium blockchain network, to provide the patients' clinical records privacy (COVID-19 diagnostic report). These are the steps which follow the working process of proposed framework. The working of the model is pictorially illustrated in Fig. 3:

1) To receive the Proof of Identity (PoI), the healthcare provider must participate (register) in a consortium network.
2) The health-care provider (doctors) uses a web user interface to upload COVID-19 patient reports.
3) Using proof of work (PoW) technique, the submitted patient report gets validated by miners. The underlying approach also maintains the consistency in network of consortium blockchain.
4) The miners distribute the transactions to the valid peers in network of consortium blockchain to verify the transaction by seeing their respective local copy, to create and add a new block in the network of consortium blockchain.
5) Once the transaction gets verified then it is stored into the distributed off-chain storage (IPFS). Additionally, the IPFS generates the hash(content-addressed) which get stored in the on-chain storage (blockchain network).

Only those peers who have signed into the network can access the list of transactions. To become a member of the consortium network a peer (healthcare provider) must have to register into the network.

A. Peers Authorization in Consortium Network

Algorithm 1 shows the verification of the peers (doctors or hsopitals) in the consortium network for reports uploading and sharing. The purpose of this algorithm is to prevent the malicious peers and their access to the shared reports in consortium network. The Proof-of-Identity (PoI) based peer identification ensures the security in the proposed model. Once a peer signed into the network of consortium blockahin, they are facilitated by a unique identifica-

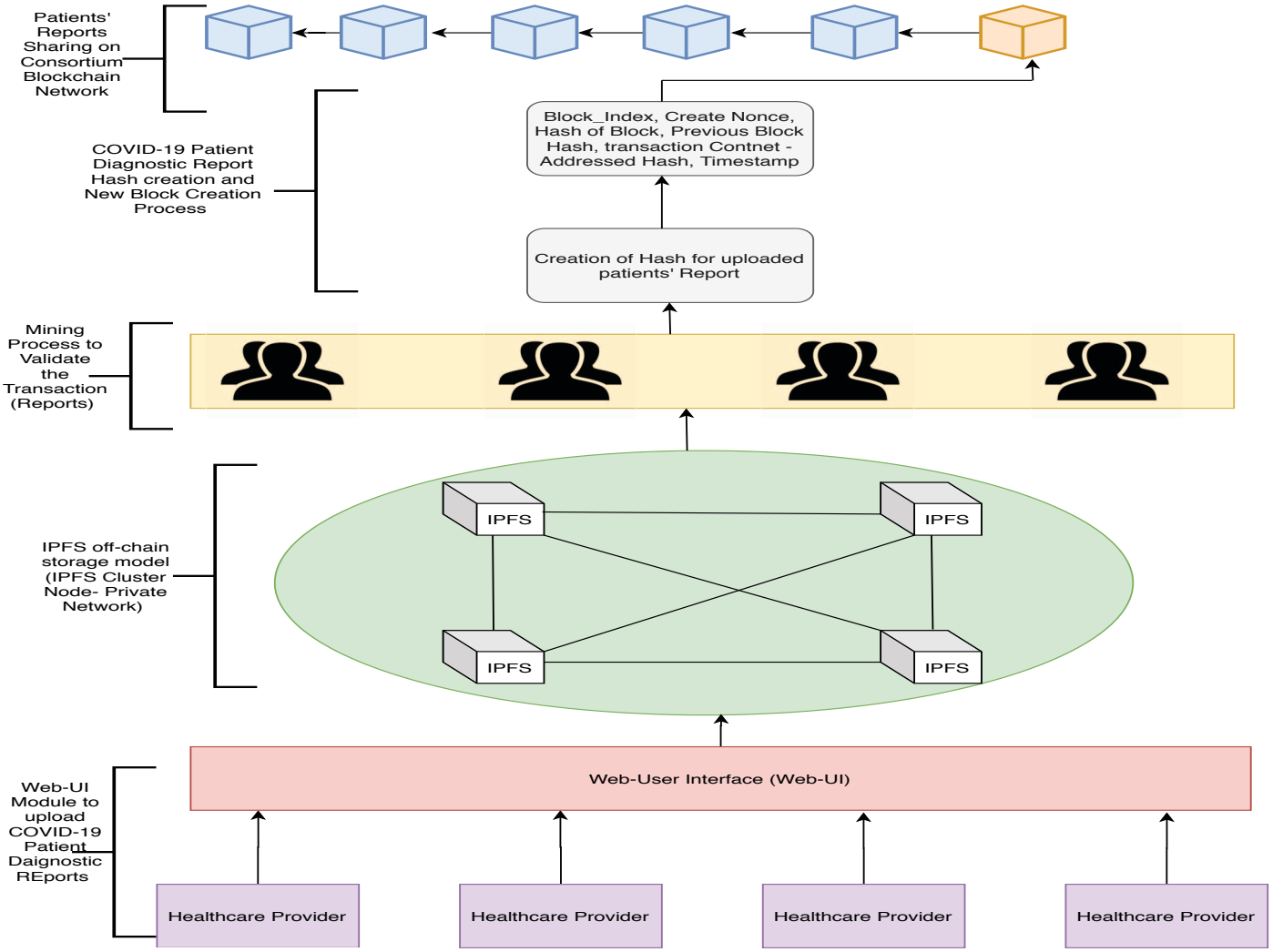

Fig. 1: The Proposed on-chain and off-chain storage Framework for sharing COVID-19 patient Reports

Algorithm 1: Algorithm for peers verification in consortium network

Input: Proof-of-Identity (PoI)
Output: verification of peers
// Verifying the authorization of peers//
if *(msg.sender is not valid)* **then**
  return "Invalid Peer";
else
  Peer can upload/share the patient's reports to/from off-chain (IPFS) storage
  return "Valid Peer";
end

Algorithm 2: Algorithm for the Integration of storage layer on-chain and off-chain

Input: PoI, report_document
Output: on-chain and off-chain integration while report_document upload
// Adding the reports to off-chain storage & receiving it's Hash value//
Hash_value = report_document | ipfs add
// Adding the Hash of reports into on-chain storage (consortium network)//
Block_Index = append_Hash(Hash_value)
//Integrity Mapping of Reports by adding PoI, Block_Index, and Hash_value of shared reports //
map_peer(PoI, Block_Index, Hash_value, PoW, Previous_Block_Hash, timestamp)

tion i.e., PoI **(http://<url:port>/nodes/register)**. The PoI of peers(hospital) gets verified while uploading of patient reports in the consortium network. The peers who upload the patient reports are responsible to manage the records of corresponding patient's in the network, as the network structure is decentralized.

B. Data Storage in Consortium Network

once the reports is uploaded by the peers (doctor or hospitals), the reports will be added to the off-chain stoage model (IPFS) and the content-addressed hash will be stored into the on-chain storage (blockchain network). The mapping is performed to maintain the integrity of storage layer (on-chain and off-chain) by proof-of-Identity (PoI) and Block_index, and Hash_value which is shown in Algorithm 2.

*1) **Validation of CoVID-19 Reports**:* The validation process of the CoVID-19 reports are illustrated in the Fig. 2. To validate the patients' diagnostic reports, the perceptual hash of the CoVID-19 patients' diagnostic reports (X-ray and CT-SCAN) are stored in the consortium network with genesis block (first block). As new reports are uploaded by the healthcare providers, the similarity of the perceptual hash is calculated with existing reports perceptual hash and majors are taken accordingly. This process is applied in the framework for sharing only CoVID-19 reports. By using the underlying approach other reports are easily gets discarded. The wavelet hash (WHash) is applied to compute the perceptual hash of the patients' diagnostic reports by using the Eq.(1). The wavelet transform discompose the daignostic reports into two different parts such as high pass and low pass. The high pass is computed using the Eq. (2), to evaluated the pixel intensity, and low pass is computed using the Eq. (1), to find the approximation in the pixels intensity. The 64 bit perceptual hash is computed on the basis of approximation of pixel intensity. The hamming distance(H_d) is evaluated using Eq.(4), to find the similarity of uploaded reports perceptual hash with existing perceptual hash in the blockchain network.

$$Z^m = \sum_{t=0}^{M} x[t]g[M-t], \{\forall M \in [0,64]\} \tag{1}$$

$$Z^m_{low} = \sum_{t=0}^{64} x[t]g[2M-t], \{\forall M \in [0,64]\} \tag{2}$$

$$Z^m_{high} = \sum_{t=0}^{64} x[t]h[2M-t], \{\forall M \in [0,64]\} \tag{3}$$

$$H_c = \sum_{r=1}^{8}\sum_{s=1}^{8} |h^1_{r,s} - h^2_{r,s}| \tag{4}$$

In Fig. 2, $<= 50\%$ similarity is taken as the threshold value. The reason behind this is that for evaluation of text similarity such a threshold value is considered in an existing work [12], [15].

Algorithm 3: Algorithm for transaction verification during transaction access

```
Input: Proof-of-Identity (PoI), Block_Index
Output: Valid or Invalid Transaction
if (PoI, Block_Index)==valid then
    M = get_hash_of_Block_Id (Block_Index)
    N = get_value_from_IPFS(Hash_Value)
    if (M ! = N) then
        return "Invalid Transaction"
    else
        return "Valid Transaction";
    end
else
    return "Invalid Transaction"
end
```

*2) **Verification of Report Hash in Consortium network**:* As shown in Algorithm 3, the hash of transactions (reports) is checked, to verify whether peers are accessing correct reports hash or not.

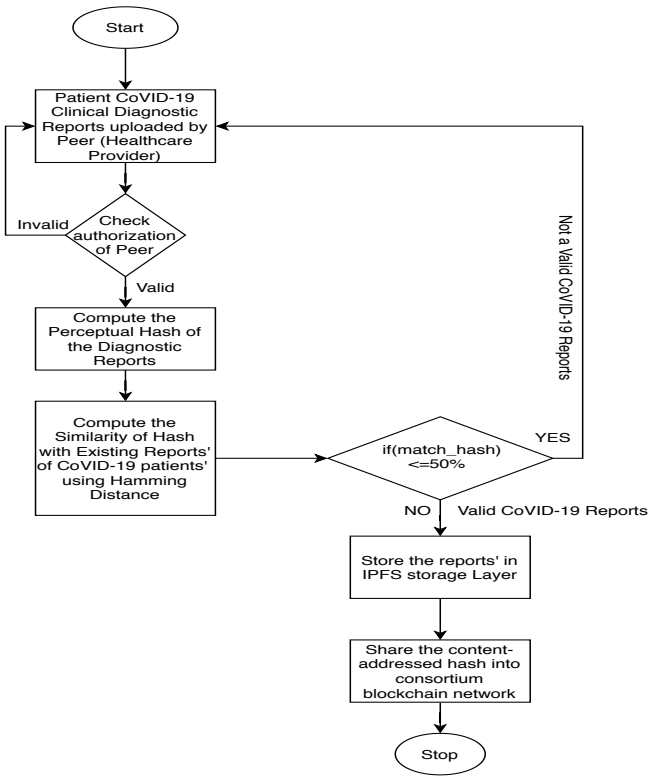

Fig. 2: CoVID-19 Reports' Validation

As stated earlier, in the proposed solution only hash of the individual reports are stored in the network of consortium blockchain because of the following reasons. Firstly, If all clinical reports and their corresponding hashes are inserted in the network of consortium blockchain, as more patients' reports are uploaded, the blockchain's size would rapidly increase. Secondly, when a new peer join a network , they have to copy the whole chain that includes patient information of all the results of COVID-19 clinical records. This method is certainly not scalable, particularly as the sizes of these clinical records (patients' reports) increases in multiple gigabytes.

For shared off-chain access and storage of the patients' diagnostic reports and their corresponding hashes, IPFS usages the distributed P2P hypermedia protocol. IPFS generates unique cryptographic content addressed hash for each file and maintain these hashes in distributed hash table [11], [14]. These hashes removes the redundancy in the network. Moreover, they are used for retrieval of actual patients' diagnostic reports. By including the hashes in the consortium blockchain network, we're making big size savings. This also helps to protect the privacy of reports (COVID-19 clinical reports).

IV. Implementation Results And Analysis

The implementation of our proposed model is divided in four different modules: *healthcare provider(doctor or hospitals)*, *data validation (mining process)*, *blockchain storage (on-*

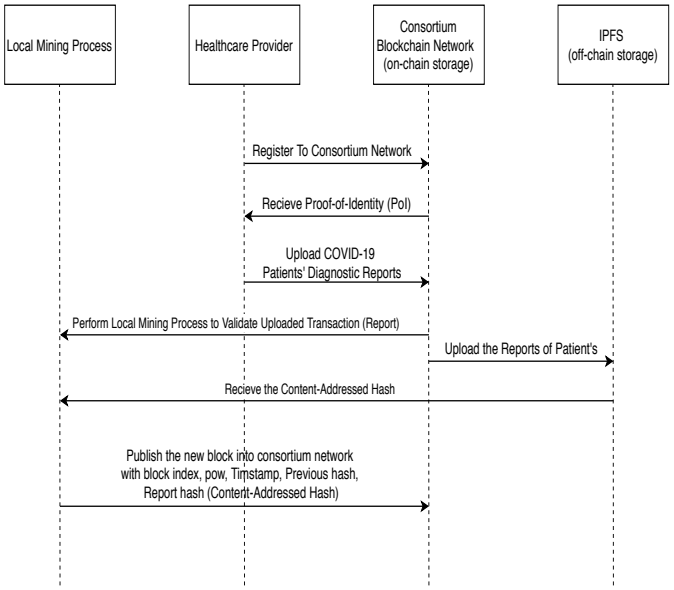

Fig. 3: Sequence Diagram for COVID-19 patient Reports upload/sharing

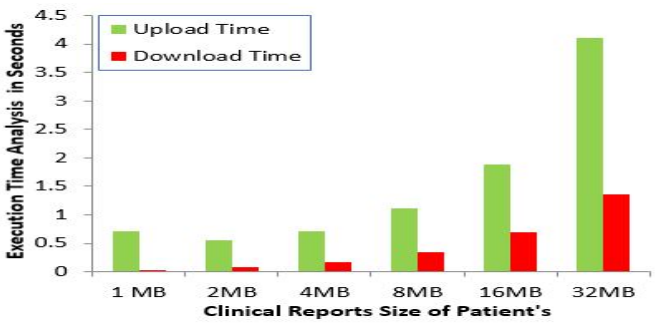

Fig. 4: Transaction Upload and Download Time Analysis

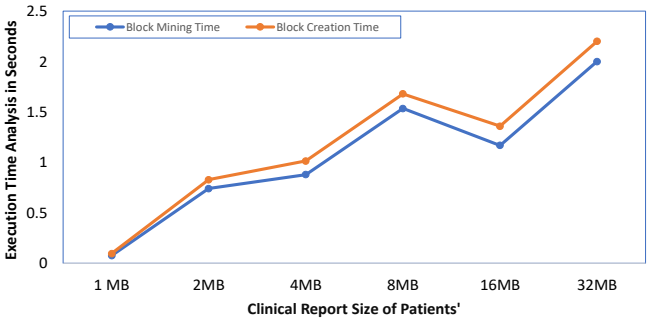

Fig. 5: Block Mining and Creation Time Analysis

chain storage), and *IPFS distributed storage (off-chain storage)*. The functioning of these modules are independent to each other.

The implementation of our proposed framework has been simulated on *Ethereum platform* using solidity programming and the implementation of blockchain is simulated using *Ropstent EtherScan*. The hash of the patient's diagnostic reports are stored into the EtherScan blockchain network (on-chain storage) whereas the data storage layer is provided by IPFS (off-chain storage).

The framework is configured with IPFS version 0.4.21 for off-chain storage along with the following system configuration: Tyrone PC run by Intel(R) Xeon(R) Silver 4114 CPU @ 2.20GHz (2 processors), 2 TB hard disk and 128 GB of RAM. Explanation of the working of modules and how they are related as follows:

1) **Healthcare Provider:** This modules are responsible to collect patient diagnostic reports. These reports are finally uploaded or shared to the network of consortium blockchain. The upload and download time of the varying size of patient's reports are shown in Fig. 4. The graph shows that reports upload time is computationally more intensive than reports download time for all different sizes of reports.
2) **Validation or Mining Process:** The module is accountable for enforcement of the consensus in the blockchain network , to validate the uploaded patient's reports as a transaction and new block creation into the network of consortium blockchain. A new block consists of patient's detailed information with content-addressable hash. Fig. 5 shows the block mining and block creation time. We can easily notice that mining of a block takes more time than creation of a block for varying sizes of patient reports. The block mining and block creation time is almost approximate for the 2MB of files.
3) **Off-chain storage:** A peer stores a patient report in IPFS, and it retrieves the unique content-addressable hash in response to that. In the future, this content-addressable hash can be used to retrieve the patient report from the off-chain storage.
4) **On-chain storage:** This content-addressed hash of the original report's gets stored in the network of consortium blockchain. In a network of consortium blockchain, each block consists of proof-value (nonce) as proof-of-existence, that ensures consistency in network, timestamp, hash of diagnostic report, current block hash , and hash of subsequent block, to maintain on-chain integrity of the block. The time taken for report access within the consortium network by varying number of peers and varying sizes of reports is shown in Fig. 6. It is noted that, as the size of reports increases, the time taken to access the reports increases. In fact, it takes more time as more peers (doctor or hospital) access the reports.

V. Conclusion

This paper illustrates the limitation of centralized storage model. and how blockchain can provide decentralized storage model to fulfill the requirements. We have designed and implemented an integration of on-chain and off-chain storage model using consortium blockchain and IPFS for the

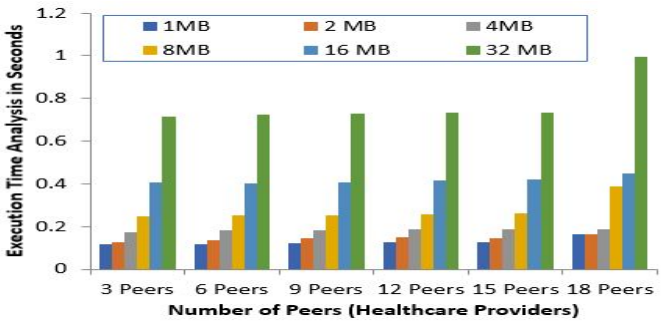

Fig. 6: Transaction Access Time by varying size of Peers

management of patient data and their privacy. For scalability, only content-addressable hashes of patients' clinical reports are stored within the consortium blockchain. Unlike the centralized data sharing system currently available among health care providers, the proposed framework is of a completely decentralized in nature. In addition, the proposed model is not dependent on a intermediary (third-party infrastructure) and facilitates fair services to authorized peers. In future, we plan to implement the model with more numbers of peers and multiple sizes of megabytes of reports sharing system.

References

[1] Chang Min Cheol, Park Donghwi, "How can blockchain help people in the event of pandemics such as the covid-19?, journal of medical systems. 44, 102," 2020. [Online; accessed 10-April-2020].

[2] Sohini Bagchi, "Could blockchain be a panacea to covid-19 pandemic?," 2020. [Online; accessed 14-April-2020].

[3] Sarah Tran, "Blockchain against corona: How the world health organization uses dlt compared to the rest of the globe," 2020.

[4] H. Jin, Y. Luo, P. Li, and J. Mathew, "A review of secure and privacy-preserving medical data sharing," *IEEE Access*, vol. 7, pp. 61656–61669, 2019.

[5] COINTELEGRAPH CONSULTING, "Blockchain as a tool to combat coronavirus," 2020. [Online; accessed 15-April-2020].

[6] MARIE HUILLET, ""new blockchain solution to fight covid-19 complies with eu data privacy regs"," 2020.

[7] VISHAL CHAWLA, "Organizations tracking covid-19 using blockchain," 2020. [Online; accessed 16-April-2020].

[8] Y. Chen, H. Li, K. Li, and J. Zhang, "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, 2017.

[9] TOSHENDRA KUMAR SHARMA, "How blockchain can solve major challenges of covid-19 faced by healthcare sectors?," 2020.

[10] SUSAN MILLERAPR, "Building a blockchain to verify covid-19 data," 2020. [Online; accessed 19-April-2020].

[11] R. Kumar and R. Tripathi, "Implementation of distributed file storage and access framework using ipfs and blockchain," in *2019 Fifth International Conference on Image Information Processing (ICIIP)*, pp. 246–251, IEEE, 2019.

[12] O. Uzuner, R. Davis, and B. Katz, "Using empirical methods for evaluating expression and content similarity," in *37th Annual Hawaii International Conference on System Sciences, 2004. Proceedings of the*, pp. 8–pp, IEEE, 2004.

[13] R. Kumar and R. Tripathi, "Blockchain-based framework for data storage in peer-to-peer scheme using interplanetary file system," in *Handbook of Research on Blockchain Technology*, pp. 35–59, Elsevier, 2020.

[14] A. Kumar and S. Srivastava, "Object detection system based on convolution neural networks using single shot multi-box detector," *Procedia Computer Science*, vol. 171, pp. 2610–2617, 2020.

[15] A. Kumar, S. S. S. Reddy, and V. Kulkarni, "An object detection technique for blind people in real-time using deep neural network," in *2019 Fifth International Conference on Image Information Processing (ICIIP)*, pp. 292–297, IEEE, 2019.