IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

COMPARISON OF INTERPLANETARY FILE SYSTEM WITH BLOCKCHAIN BASED FILE SYSTEM

[1]Jatin Pahuja, [2]Dr. Neha Agrawal, [3]Dr. Meenu Garg

[1]BTech Scholar, [2]Assistant Professor, [3] Assistant Professor

[1]Information Technology

[1]Maharaja Agrasen Institute of Technology, Delhi, India

Abstract: In the following paper we have compared a blockchain based file storage system with InterPlanetary File system to gain some insights and establish some common ground. The debate over decentralized file systems has been gaining traction due to the privacy concerns regarding the centralized data storage servers controlled by a handful companies. It's obvious as to why we need decentralized file systems but choosing which system to use is a tough task since there are lot of alternatives each having its own advantages and disadvantages.

IPFS is a peer-to-peer hypermedia protocol and a distributed file system using content based addressing and cryptographic hashes. The Blockchain based file system has been made on the underlying principles of a distributed ledger and implement by us using Hyperledger fabric, docker and smart contracts

Index Terms **- IPFS, Blockchain, decentralized, file system.**

I. INTRODUCTION TO IPFS

Today, the Internet is such an important tool in our everyday life, we use it to consume media, to communicate with friends and colleagues, to learn to handle our finances and much more. But the web, as we know it, has a problem. The information on it is centralized. It's all stored on big server farms like this one. And these are usually controlled by a single company. I mean, have you ever wondered what would happen if sites like YouTube or Wikipedia would go offline?

How would you watch cat videos or spend hours reading one Wikipedia page after the other? The centralization brings another problem with it, and that is censorship, because content is hosted on just a few servers, it's easy for governments to block access to them. In 2017, Turkey ordered Internet providers to block access to Wikipedia because the administration called it a threat to national security. So, you get the idea. Centralization of the web isn't a good thing. But then why do we keep using such a model?

Well, that's because we have high expectations when it comes to the Web. We want pages, images and videos to load instantly, and we want them in high quality. Centralizing servers allows companies to have complete control over how fast it can deliver all of this content. Another reason we use this model is that there just isn't a good and fast alternative. But that might be about to change. Meet IPFS, the interplanetary file system. That's a fancy name and they have very ambitious goals as well.

They want to make the web completely distributed by running it on top of a peer-to-peer network that works similarly to how BitTorrent works. Let's take a look at how IP can accomplish these goals. But first, you have to understand how we access content on the Web. Right now, let's say you want to download a photo from the Internet. When you do that, you tell the computer exactly where to find the photo. In this case, the location of the photo is the IP address or the domain name.

This is called location-based addressing. You tell the computer where to get the information, but if that location isn't accessible, in other words, the server is down, you won't get that photo. However, when that happens, there is a high chance that someone else has downloaded that picture before and still has a copy of it. But yet your computer won't be able to grab a copy from that person to fix this IP address moves from location based addressing to content-based addressing.

Instead of saying where to find a resource, you just say what it is you want. But how does that work? Well, every file has a unique cache, which can be compared to a fingerprint when you want to download a certain file. You just asked the network who has the file with this hash and someone on the first network will provide it to you. Now, you might think hold on a minute. How do I know that that person hasn't tampered with the file?

Well, because you use a hash function to request the file, you can verify what you have received. You request a file with a certain hash. So, when you receive the file, you check of the hash matches, what you have received, security built in. Another nice feature of using hashes to address content is deprecation. When multiple people publish the same file, an IP address, it will only be created once and that makes the network very efficient.

Enough with this high-level overview. Let's take a look at how app stores file and makes them accessible to others. Files are stored inside IPFS objects, and these objects can store up to two hundred and fifty-six kilobytes worth of data. They can also contain links to other IPFS objects. A simple hello world text file, which is very small, can be stored in a single Ipfs object. But what about files that are larger than 256 kilobytes like an image or a video?

For instance, while those are split up into multiple IP office objects that are all 256 kilobytes in size and afterwards the system will create an empty IP office object that links to all the other pieces of the file IP addresses. Data architecture is very simple, but yet it can be very powerful. This architecture allows us to really use it as a file system. Here's a simple directory structure with some files in it. We can translate this into ABFS objects as well, creating one object for each file and each directory.

But that is not all you see because IP address users' content-based addressing. Once something is added, it cannot be changed anymore. It's an immutable data store, much like a block chain. But then how do we change stuff on it? While IP supports versioning of your files, let's say you're working on an important document that you want to share with everyone over IP effects. When you do, that IP address will create a new comet object for you. This object is really basic. It just tells Apophis which comet went before it and it links to the ABFS object of your file. Now let's imagine that after a while you want to update this file while you just add your updated file to Apophis and then the software will create a new comet object for your file. This common object now links to the previous comet, and this process can be repeated endlessly. IPv6 will make sure that your file, plus all its entire history, is accessible to other notes on the network. This all sounds great, but it's not without its limitations or drawbacks. The biggest problem that IPFS has faces is keeping files available. Every node on the network keeps a cache of the files that it has downloaded and helps to share them if other people need them. But if a specific file is hosted by, let's say, four nodes and those nodes go offline, then that file becomes unavailable and no one can grab a copy of it.

A bit like BitTorrent swarms without seeders, there are two possible solutions for this problem. Either incentivize people to store files and make them available, or we can proactively distribute files and make sure that there is always a certain number of copies available on the network. And that's exactly what File coin intends to do. File coin is created by the same group of people that have created IPFS. It's basically a block chain built on top of IPFS that wants to create a decentralized market for storage.

If you have some free space on your hard drive, you can rent it out to others and make money of it. In the process, Phalcon creates a strong incentive for nodes to keep the files online for as long as possible because otherwise they won't get rewarded. The system also makes sure that files replicated on many nodes so they cannot become unavailable. That's just a quick summary of pholcodine and how it intends to build on top of IP efforts to solve some of its issues. Another nice application is DTube, which is basically sites like YouTube, but entirely distributed and hosted on IP effects. Anyone can publish videos and anyone can help to support the network. Pretty clever. But by now you must be wondering why is IP office called the interplanetary file system? Is it suited to run across multiple planets? Well, let's assume that we have a base on Mars. Communicating from Mars to Earth is quite difficult. Depending on the position of the two planets, a signal can take somewhere between four and twenty-four minutes to travel between them.

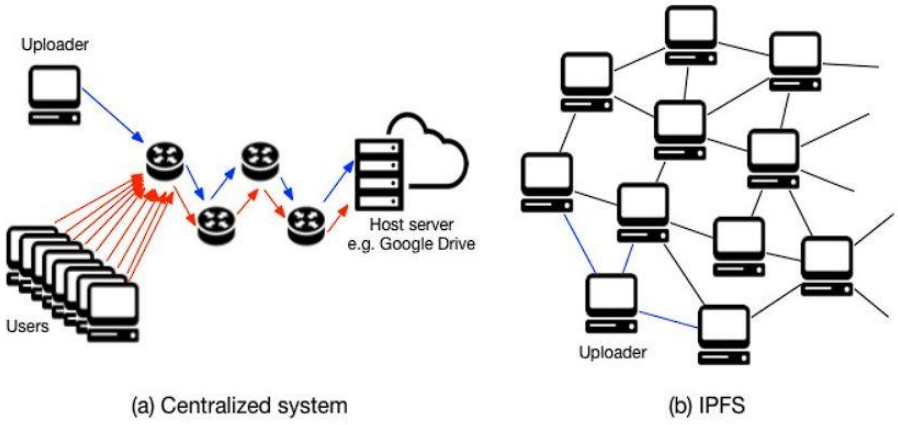

Fig.1 Difference between centralized system and IPFS

II. RELATED WORK

Traditional centralized databases are often based on the client-server architecture, in which the client can store entries in a central server and view a modified copy of the records any time the server is accessed. Blockchain, on the other hand, is an increasing collection of blocks that are connected and encrypted using cryptographic algorithms. Satoshi Nakamoto created this invention in 2008 with the intention of using it in his cryptocurrency Bitcoin [1]. Each block in the blockchain contains a list of transactions, a hash of the previous block, and a hash of the current block. The genesis block is the first block of the blockchain. Blockchain is a distributed ledger technology that is managed by a peer-to-peer network of nodes. To update the distributed ledger, the network's

participating nodes must reach a collective consensus. The consensus protocol is at the heart of a blockchain, determining how it operates.

The authors of Ref. [2] explained their version of record storage system in blockchain that can store records on IBM based blockchain using smart contracts and Hyperledger. They are yet to deploy this system although the tech behind it looks promising but the fact that IBM as an intermediary of the blockchain does defeat the core principle of decentralization.

IPFS [3] is a distributed and versioned file system that can link several processing nodes to the same file system and control it by monitoring its versions over time. For the recognition of files, IPFS has the special property of content addressing at the HTTP layer. Instead of defining a file by the server on which it is stored, IPFS represents it by its hash. In IPFS, the hash of a file typically starts with "Qm," and the hash is simply a multihash. Since the name of the files in IPFS is not a part of the IPFS object, two files with different names and the same content would have the same hash values.

The Merkle Patricia tree framework [5] of the Ethereum blockchain can also be emulated as IPFS properties. Larger pieces of data need a higher cost to be stored on the Ethereum network, but only hashes of files are stored on it rather than the whole dataset. Furthermore, this hash of the file can be connected with the file on IPFS to gain access to it [4]. It proposes a novel zig-zag-oriented storage paradigm based on IPFS and blockchain to solve the problem of high-throughput for individual IPFS users. Smart contracts make it simple to gain access to the Ethereum network. Ethereum smart contracts are written in Solidity [13], a high-level scripting language inspired by coding languages such as C++, JavaScript, and Python. Remix IDE [7], a browser-based IDE, can be used to create Ethereum smart contracts.

The Truffle system [6] is another choice, as it includes built-in smart contract compilation, linking, execution, and binary management. It can be used to implement networks in both public and private contexts. Ganache, an internal JavaScript version of the Ethereum blockchain, is a one-click blockchain support mechanism in the truffle system. It also supports front-end libraries via Drizzle. MetaMask [8] can be used to run Ethereum decentralised applications in the browser without running a complete node. The methods mentioned above can be used in tandem to create an efficient Ethereum decentralised application development environment.

C. Udokwu [9] discussed the limitations of blockchain technologies in the context of education system. The main issue with blockchain technology is that every peer must have the same chain (ledger), and the transaction must be verified by almost every peer in the network; however, this approach creates a problem for certain peers who do not want to see the different information or set of information of the ledger, or the peers who do not want to use in the verification.

H. Shen [10] used the consortium blockchain solution (a mix of public and private blockchains) to store the candidate response in the subchain (private chain). These answers are compared to the prime chain (public chain), which already holds the list of quiz question answers. The authors used this model to assess the course credit system through quizzes. However, the writers make no note of how consensus is reached or how transactions are distributed using this method.

The authors of Ref. [11] used a blockchain-based solution to store the university's graduation prerequisite table. This index would be saved by the evaluator for various forms of student assessment. This methodology also includes the ability to assess the university's course outcomes. However, the learning block is divided into two parts: the student skill chain and the course chain. The student ability chain holds the student's information, while the course chain keeps track of the course's outcomes depending on student results.

B. Wu, Y. Li proposed blockchain-based storage to store the digital certificate of the candidate to improve the digital education system [12]. The blockchain provides the decentralized platform, where the record can be traced easily. The blockchain is used to store the operational skill data set that enables the various matrices, which ensure the trust in digital education system. The work in presented Hyperledger composer-based permissioned blockchain storage for the storage of credit store of student which will be visible to employer industries. The information of students can be easily traced out on the chain. Hence, this approach becomes suitable for the industry person to identify the students. The authors of Ref. [13] have used private and public blockchain to store the big data storage, where the public chain stores the storage address of the person and the private chain stores the data of the person. There are two parallel chains that are being used in the storage scheme. The authors have used this scheme to evaluate the learning behavior of the student and teaching behavior of the teacher. The private chain is stored using distributed storage techniques, whereas the public chain uses storage of centralized storage. Distributed chain stores the achievement, credits, and rewards, whereas the central stores the record achievement, credits, and rewards which include videos, audios, and image data.

The work in Ref. [14] presented the blockchain approach b-learning management system to store the data of students, where different smart contracts are being applied for managing the QUIZ, credits management, content sharing, and instant rewarding. The storage scheme manages the different certification data, knowledge data, and shared data

In Ref. [15], the authors have focused on the higher education system, where the same university has different campuses in the country. To manage the secure transaction that includes student profile and certification, the blockchain-based structure has been proposed. The authors of Ref. [16] have addressed the data science skill gap between industry and data science education. To link the data science education, industry blockchain-based approach has been proposed by the authors. It is feasible for the industry to search the candidate of data science through this platform. Sankar LS et al., in [17], analyses and research different consensus protocols in blockchain, as well as the viability and utility they offer in various platforms. Blockchain may be envisioned as a trustworthy record-keeping mechanism focused on archival science – an ancient discipline aimed at record preservation.

III. METHODOLGY

So, we divided the work into the following steps:

1. The first step was to install both the systems to start testing them. Firstly, we installed the IPFS Python client. To access the IPFS api, we had to build our own gui to patch with this python client. IPFS's dashboard has five main options: status, file, explore, peers, and settings. The status is used to provide information about the peers, such as the size of shared files by the peers, the number of peers linked, and the bandwidth used by the peer. There are two kinds of bandwidth used: IN and OUT. The IN bandwidth is used to share files on IPFS, while the OUT bandwidth is used to view files from IPFS.
2. Then we installed the Blockchain based file system and ran it using Visual studio and docker. The implementation has been explained in the previous paper [2].
3. After getting both the systems we will then simultaneously try out the upload, deletion and various other features on files. We will then tabulate the comparisons based on a number of factors in trying to find out new insights.

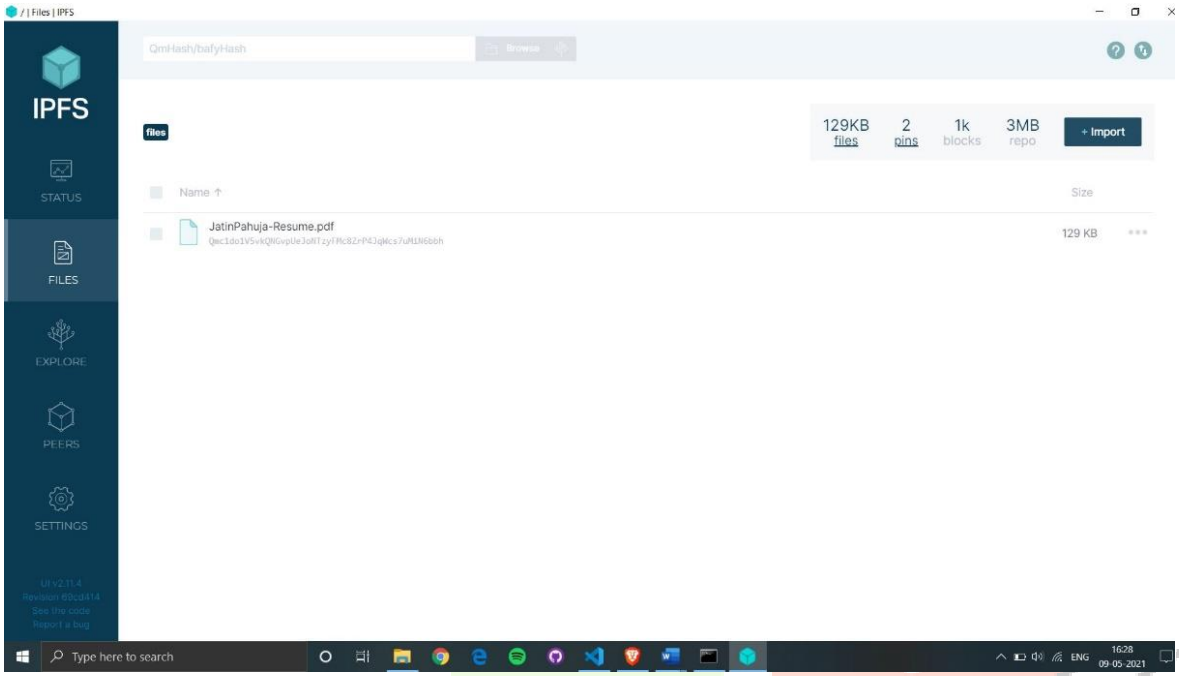

Fig.2 Image showing the files tab in IPFS

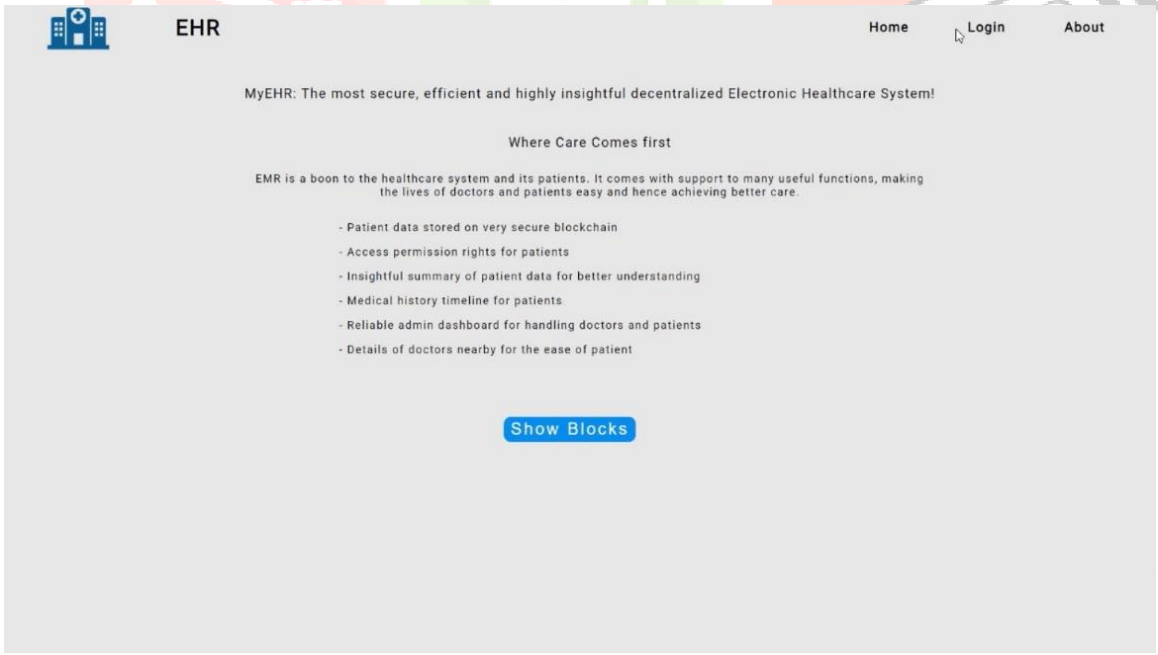

Fig.3 The homepage has been shown above

IV. COMPARISON

We have evaluated the two systems on a number of factors as shown below. We will then write the advantages and disadvantages of these systems in detail.

Advantages of Blockchain based file system over IPFS

1. The first thing that counts is the user experience and obviously this system was made keeping that in mind and for a beginner user this system is extremely handy and easy to use and can be operated by anyone while the IPFS as good as it is it still requires you to have some understanding of how to use it and installation can be quite time consuming.
2. The data that we uploaded is easy to share with others as it has built in permissions mode where anyone requesting data can be given permission while in IPFS it's a little more complicated than this obviously you can share files in IPFS and it doesn't allow duplicates but sharing isn't as easy.
3. The data can never go offline. It's possible that the file you uploaded in IPFS can't be found if your node goes offline and nobody pinned your file then it's likely that your file may be lost but this isn't the case in this system because the files are stored in blockchain rather than in the local node in IPFS

Limitations of Blockchain based file system

1. This system is using IBM's blockchain technology which defeats the whole point of decentralization. There should not be any dependence on any large multinational to support the system so as to when it crashes or servers go down this system goes down with it. This is a serious disadvantage.
2. Blockchains are prone to attacks 51% attacks, forking, mining pool attacks while the consensus in this is quite different yet the security isn't at par with IPFS on scale that the IPFS has fundamental principles protecting it. The entire architecture of IPFS has been made to be prone to any attacks which isn't quite here.
3. In blockchain schemes, the user's private key is known as a protection and authentication credential provided by the user, and no third party is involved in this phase. When a user builds a wallet for a cryptocurrency, he or she must also import the private key into the wallet. This private key is loaded into the ledger to ensure the cryptocurrencies' security and authentication. If the private key is missing or broken, it cannot be retrieved, which ensures that the customer cannot enter the wallet by any other means and that any of his cryptocurrencies in the wallet are inaccessible.
   Since blockchain networks are not regulated by third-party agencies, data may be changed by untraceable criminals in the event of a missing or stolen private key.

Advantages of IPFS over blockchain based system

1. Decentralization — The files are distributed through a network of nodes using hashes as identifiers. File coin provides an incentive for nodes to host archives.
2. Fault Tolerance — If one node crashes, the file remains accessible as long as there are other nodes hosting it. There isn't a single point of weakness in this system.
3. Persistent Storage — The key point of IPFS is data storage: the entire file history can be accessed as well as items relating to the original data, as well as all new copies, are available. IPFS objects can be stored permanently without being changed since data blocks are stored locally across the network and can be cached forever.
4. Censorship Resistance — Since content is spread through an entire network, no single authority can delete it until it has been submitted to IPFS. The file is not fully deleted if it is just removed from one node. It indicates that there are already versions of it on other nodes.
5. Scalability — The more nodes hosting files, the better and more accessible they become to network users.

Limitations of IPFS

1. Running an IPFS node currently necessitates the use of considerable bandwidth, which can be prohibitively expensive for many users, including those in developing countries. Excessive bandwidth consumption can stymie IPFS adoption in many parts of the world. Although there are several approaches to dealing with this issue, financial incentives may be the most effective. Getting paid to host content on IPFS will help offset the costs of operating nodes and promote adoption.
2. The latest IPFS implementation cannot guarantee data availability as needed. Content pinning, which involves actively storing copies of written content on an IPFS node, is one method for ensuring availability. To meet the availability guarantees, this IPFS node must be online at all times. File coin and other projects are developing an opportunity layer to allow nodes to save content in exchange for rewards.

By definition, content released to IPFS is open to the public. Anyone who has the content hash will access the content. Currently, IPFS lacks a built-in workaround for saving private data. In IPFS, encryption can be used to store and/or share private data. Perhaps a more sophisticated approach is to use the IPFS protocol to build a private network in which nodes can only link to a given list of nodes in the private network.

	Blockchain system	IPFS
Speed of file upload (100kb)	5s	10ms
File retrieval time	10s	5s
Become a node	✗	✓
File can't be found	Never	If node is offline
Analytics	✗	✓
Incentives	✗	✗

Fig.4 Comparison Table

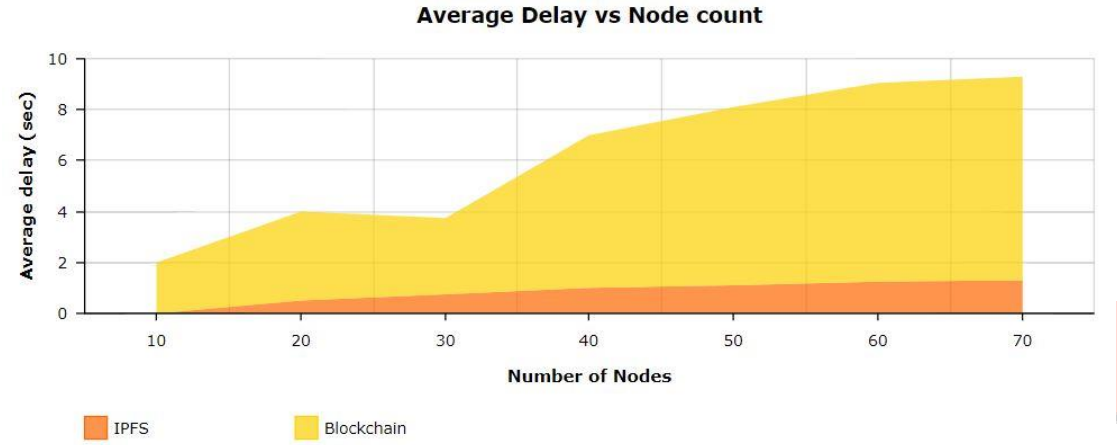

Fig.5 The average delay of both these systems was calculated by trial running the two simultaneously and then charting it on scale.

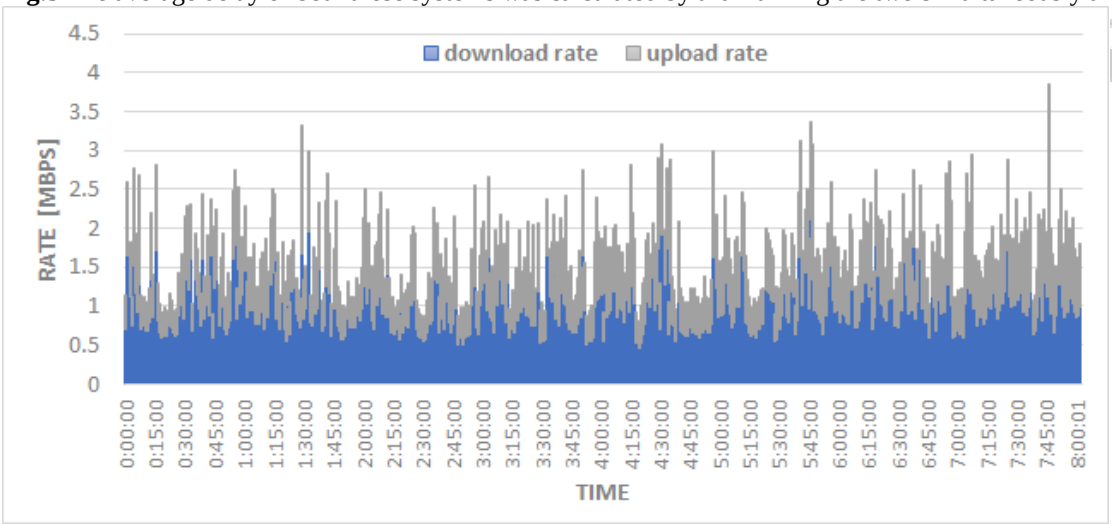

Fig.6 IPFS download and upload rate

V. CONCLUSION AND FUTURE SCOPE

In this paper we examined the advantages and disadvantages of both these systems. The blockchain based system though it provides the basic use case but it's still in beta case along with limited features and hidden centralized problems. In blockchain schemes, the user's private key is known as a protection and authentication credential provided by the user, and no third party is involved in this phase. IPFS is much more decentralized, established, secure and provides much more use cases. It's evident from the results that IPFS is the preferred choice over a purely blockchain based system as explained in the paper.

A hybrid blockchain based system based on IPFS that incentivizes the nodes along with bigger block size could be revolutionary like File coin but with much more efficiency in mining and making the incentives available on a decentralized platform rather than having to go on to a centralized exchange to buy it and then using it to access the storage system which kind of defeats the purpose

of all it. Having the currency of the storage system as a reward along with not having unrealistic requirements could help to incentivize nodes and keep the systems running which could turn into something very promising in the future.

REFERENCES

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

[2] Pahuja, Jatin, and Dr. Neha Agrawal. "Record Storage and Management System Using Blockchain." (2020).

[3] Benet, J. (2014). IPFS-content addressed, versioned, P2P file system. arXiv preprint arXiv:1407.3561.

[4] Dr. Christian Lundkvist and John Lilic, An Introduction to IPFS – ConsenSys – Medium. (2018). Medium. [Online]. Available:https://medium.com/@ConsenSys/anintroduction-to-ipfs-9bba4860abd0.

[5] ethereum/wiki. (2018). GitHub. [Online]. Available: https://github.com/ethereum/wiki/wiki/Patricia-Tree.

[6] Truffle Suite - Your Ethereum Swiss Army Knife. (2018). Truffle Suite. [Online]. Available: http://truffleframework.com/.

[7] Remix - Solidity IDE. (2018). Remix.ethereum.org. [Online]. Available: https://remix.ethereum.org/.

[8] MetaMask. (2018). Metamask.io. [Online]. Available: https://metamask.io/.

[9] C. Udokwu, A. Kormiltsyn, K. Thangalimodzi, A. Norta, The state of the art for blockchain-enabled smart-contract applications in the organization, 2018 Ivannikov Ispras Open Conference (ISPRAS), IEEE, 2018, pp. 137144.

[10] H. Shen, Y. Xiao, Research on online quiz scheme based on double-layer consortium blockchain, 2018 9th International Conference on Information Technology in Medicine and Education (ITME), IEEE, 2018, pp. 956960.

[11] A. Alammary, S. Alhazmi, M. Almasri, S. Gillani, Blockchain-based applications in education: A systematic review, Appl. Sci. 9 (12) (2019) 2400.

[12] B. Wu, Y. Li, Design of evaluation system for digital education operational skill competition based on blockchain, 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE), IEEE, 2018, pp. 102109.

[13] X. Gong, X. Liu, S. Jing, G. Xiong, J. Zhou, Parallel-education-blockchain driven smart education: Challenges and issues, Chinese Automation Congress (CAC), IEEE, 2018, pp. 23902395.

[14] G. Sahonero-Alvarez, Blockchain and peace engineering and its relationship to engineering education, 2018 World Engineering Education Forum-Global Engineering Deans Council (WEEF-GEDC), IEEE, 2018, pp. 16.

[15] K. Al Harthy, F. Al Shuhaimi, K.K.J. Al Ismaily, The upcoming blockchain adoption in higher-education: requirements and process, 2019 4th MEC International Conference on Big Data and Smart City (ICBDSC), IEEE, 2019, pp. 15.

[16] A. Mikroyannidis, J. Domingue, M. Bachler, K. Quick, Smart blockchain badges for data science education, 2018 IEEE Frontiers in Education Conference (FIE), IEEE, 2018, pp. 1-5.

[17] Sankar, L. S., Sindhu, M., &Sethumadhavan, M. (2017, January). Survey of consensus protocols on blockchain applications. In Advanced Computing and Communication Systems (ICACCS), 2017 4th International Conference on (pp. 1-5). IEEE.

[18] Huang, H., Lin, J., Zheng, B., Zheng, Z., & Bian, J. (2020). When blockchain meets distributed file systems: an overview, challenges, and open issues. *IEEE Access*, *8*, 50574-50586.

[19] Naz, M.; Al-zahrani, F.A.; Khalid, R.; Javaid, N.; Qamar, A.M.; Afzal, M.K.; Shafiq, M. A Secure Data Sharing Platform Using Blockchain and Interplanetary File System. *Sustainability* 2019, *11*, 7054. https://doi.org/10.3390/su11247054

[20] Jonas Bostoen, A Hands-on Introduction to IPFS https://medium.com/coinmonks/a-hands-on-introduction-to-ipfs-ee65b594937#:~:text=IPFS%20works%20by%20connecting%20all,versions%20of%20content%20on%20IPFS).

[21] R. Kumar and R. Tripathi, "Implementation of Distributed File Storage and Access Framework using IPFS and Blockchain," 2019 Fifth International Conference on Image Information Processing (ICIIP), 2019, pp. 246-251, doi: 10.1109/ICIIP47207.2019.8985677.

[22] Kumar, R., & Tripathi, R. (2020). 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). Academic Press.

[23] Vishwa Vidyapeetham, A. (2018). A blockchain and ipfs based framework for secure research record keeping. *International Journal of Pure and Applied Mathematics*, *119*(15), 1437-1442.

[24]Vincent Tabora (2020), Using IPFS For Distributed File Storage Systemshttps://medium.com/0xcode/using-ipfs-for-distributed-file-storage-systems-61226e07a6f