DSTORAGE FILE SYSTEM

NAMAN SINGHAL
INFORMATION TECHNOLOGY
Meerut Institute of Engineering and Technology
Meerut,India

AKASH CHAUDHARY
INFORMATION TECHNOLOGY
Meerut Institute of Engineering and Technology
Meerut,India

Abstract

As a tool for human technological advancement, the peer-review system acts as a gateway for ensuring academic paper qualities. However, the system has proven to be slow and expensive. Also, biasedness remains an unsolved problem. Such issues could become a major bottleneck, which can adversely impact research progress and dissemination of knowledge. This paper aims to propose a double-blind paper review system to preserve the authors and reviewers anonymity. This system also addresses issues concerning the reviewers payment, inconsistent review metrics, and biased reviews. The proposed solution utilizes the Hyperledger Fabric blockchain with the InterPlanetary File System (IPFS).

Keywords—*ipfs, and blockchain*

I. INTRODUCTION

Academic research record keeping is important for the research planning and management, replication of results, documentation of collaborations, publishing and peer review, and for complying with governmental and institutional rules and regulations. Good research records consist of much more than just research data. They include protocol description, data manipulation and analysis procedures, personal and group interpretation of the results, and important communications and group decisions among collaborators. So the data must be confidential, secure and tamper proof in order to avoid any discrepancies. While considering academic research, the Principal Investigator(PI) is the main actor who ensures proper research planning, management and execution of the ongoing research. Documents like proposal of funding agencies, project reports, memorandum of understanding, minutes of meetingSocket programming is a process which allows two nodes to be connected to each other for communication over a network. In this process, one node

Blockchain technology can create tamper-proof, secure record of events in a distributed, peer-to-peer network of several nodes of computers. The cryptocurrency based transaction system like bitcoin is based on this technology. Blockchain ensures anonymity and security of the users involved in the transactions. Blockchain consists of a growing list of blocks which are linked by cryptographic algorithms. It is based on the Distributed Ledger Technology (DLT) which is a system for recording digital transactions in a distributed storage with no centralized data stores. The distributed ledger technology can be used to write smart contracts or digital contracts or blockchain contracts which are self-executing contracts that can be converted to computer code with the help of certain platforms, and can be replicated, shared and supervised by network of computers that run on the blockchain. Smart contracts avoids middleman by automatically defining and enforcing rules and obligations made by the parties in the ledger. While blockchain can be used for storage of less amount of data like transaction metadata information, hash values etc., IPFS can be used as a peer-to-peer, distributed system to store hypermedia in large quantities.

InterPlanetary File System (IPFS) [3] is a peer-to-peer hypermedia protocol and distributed file system that is to replace the web of tomorrow. It has a block storage model with hyperlinks to address the contents forming a Merkle Directed Acyclic Graph (DAG). Since IPFS is distributed, it has no single point of failure.

II. LITERATURE REVIEW

Nowadays, there are various types of distributed storage systems, such as cloud storage systems, and peer-to-peer (p2p) storage systems. In all these storage systems, data can be stored, archived, and back up over distributed nodes, such as AmazonS3. Users can make use of their stored files any time anywhere; this is an outstanding advantages as to distributed storage systems. There are many researches focused on the design and construction of distributed file systems. Napster [15], Kazaa [16], and Gnutella [17] implement distributed file systems and prompt it to be an exciting and popular research area. Bit-torrent [18] is one of the most popular and successful peer-to-peer distributed file systems and has more than 100 million online users presently. It is a large-scale deployed in which millions of users log-in and log-out every day. Storage resources, as well as system clients in a distributed file system, are scattered in the network. In these systems, users act as both creators and consumers of data, therefore, to provide massive of incentives by a secure and efficient approach.program)

Traditional centralized databases are mostly based on the client-server architecture, where the client can store entries in a central server, and can access updated copy of the information on each time of accessing the server. In contrast to this, blockchain is a growing list of blocks which are linked and secured using cryptographic algorithms. This technology was invented by Satoshi Nakamoto in 2008, for the purpose of using it in his cryptocurrency Bitcoin [1]. Each block in the blockchain contains list of transactions, hash of the previous block and hash of the current block. The first block in the blockchain is called the genesis block. Blockchain is a distributed ledger technology maintained by a peer-to-peer network consisting of nodes. For updating the distributed ledger, the participating nodes in the network should derive at a common consensus. The consensus protocol is the core and it decides how a blockchain works. Sankar LS et al., in [2] provides an analysis and study of various consensus protocols in blockchain and the feasibility and efficiency they provide in various platforms. Blockchain can be visualized as a trusted record keeping system based on archival science – an ancient science aimed for preservation of records [21].

. IPFS 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. The hash of files in IPFS always begins with "Qm" and the hash is actually a multihash. Name of files is IPFS is actually not a part of the IPFS object, so two files with different names and same content will have the same hash values. Ethereum blockchain's Merkle Patricia tree structure [5] can also be emulated as IPFS objects. For larger pieces of data to be stored on the ethereum blockchain, a larger amount of fee has to be paid, so only the hashes of files are stored on the ethereum blockchain rather than storing the whole file on it. Further, this hash of the file can be linked with the file on the IPFS to access it [4]. A novel zig-zag based storage model based on IPFS and blockchain is provided in [9] to address the issue of high-throughput for individual users in IPFS.

FS. Smart contracts provide an easy way to access the ethereum blockchain. Ethereum smart contracts are written in a high-level coding language called Solidity [13] which is influenced by coding languages such as C++, javascript and Python. To develop ethereum smart contracts, Remix IDE [7] can be used, which is a browser based IDE. Another one is the Truffle framework [6], which supports built-in smart contract compilation, linking, deployment and binary management. It supports both public and private network deployment environments. The truffle framework has a oneclick blockchain support mechanism called Ganache, which is an internal javascript implementation of the ethereum block chain

III. DATA PROVENANCE REFERS TO THE TRACKING AND RECORDING OF THE ORIGINS OF DATA. IT REFERS TO THE COLLECTION OF HISTORY OF DATA SUCH AS CREATION, ATTRIBUTION AND DATA VERSIONING. PROVENANCE METADATA IS VERY IMPORTANT FOR FORENSICS PURPOSES AND AUDITING. BLOCKCHAIN CAN BE USED AS A PLATFORM FOR PROVENANCE DATA MANAGEMENT IN A TRUSTWORTHY MANNER.

IV. PROPOSED METHODOLOGY

Storage is a critical issue in blockchain especially when large amount of data needs to be stored on network nodes. Storage capacity of terminal nodes is limited, as it does not allow to store very large size data. This problem gives rise to multiple issues, such as computational power and high cost of processing large amount of data. Considering the aforementioned challenges, Stiechen et al. in proposed a decentralized storage mechanism by using IPFS and access control policies. Files are stored in the form of chunks on each node. However, a file cannot be accessed until unless proper permissions are assigned to users. This is a good approach for preserving the privacy of confidential data. The proposed schemes suffer delay while accessing files from server due to blockchain interaction. Access control and securing sensitive data is the fundamental need of recent era. Blockchain is an emerging technology which provides solution to multiple problems, i.e., security, privacy, access control, trust, traceability and many more. In this perspective, research work presents a blockchain-
based solution for securing the patient's data in medical health records. Access control rules are defined to restrict the access of data, so that privacy of patients can be preserved. A secure interoperability among heterogenous blockchains is achieved by exploiting the use of smart contracts. A set of smart contracts are designed to achieve multiple purposes. However, certain measures need to be adapted to increase adaptability and robustness. Access control reflects the system's security by restricting the access of data to a specific set of people. It is important while dealing with the sensitive data such as patient's health condition and medical record. Existing systems lack the access control policies to ensure data integrity and privacy. For this purpose, the authors of proposed a consortium blockchain with distributed ledger. Each operation performed by smart contract is stored as a transaction record in blockchain. A proof of concept is designed to validate the performance of system as compared to existing approaches.

• The proposed solution ultimately achieves integrity, scalability, and authenticity. However, privacy of patient's personal information is not considered. Online data that is stored on a third party centralized server is vulnerable to attacks.

•

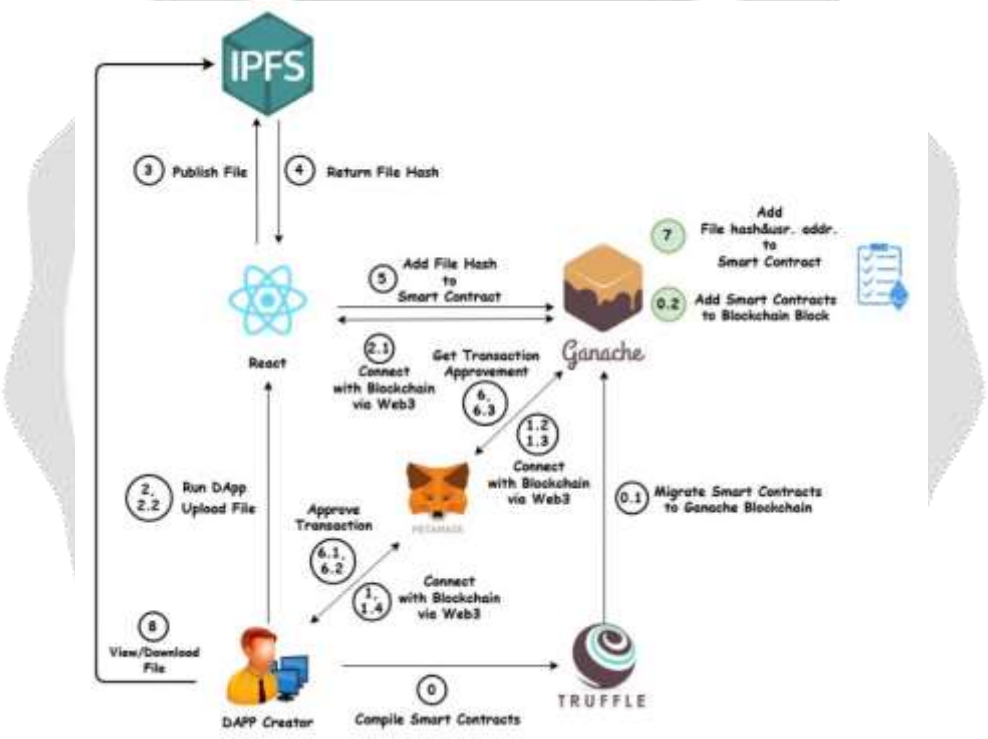
.

• The documents such as project reports, project funding details, memorandum of understanding, attendance records, and minutes of meeting are encrypted and stored in the IPFS

Recently, blockchain has become a buzzword in both industry and academia, and the combination of blockchain and distributed file system is becoming a promising solution, where blockchain is expected to provide incentives and security for the stored files in systems. Currently, the popular blockchain-
based distributed file systems include IPFS [7], Swarm [8], Storj [9], and PPIO [10]. Within those file systems, IPFS is a peer-to-
peer distributed file system for storing and accessing files, websites, applications and data; Swarm is a distributed storage platform and content distribution service based on Ethereum; Storj is another peer-to-
peer decentralized cloud storage platform that allows users to share data without relying on a third-
party data provider; and PPIO is a decentralized programmable storage network that permits users store and retrieve any data from anywhere on web. With respect to the combination with blockchains, IPFS, Swarm, and Storj file systems adopt Filecoin [11], Ethereum [12], and Metadisk [13] as their incentive mechanisms, respectively. PPIO exploits up to 4 proof algorithms, which are explained in Section III, for its incentive layer. Considering that the technologies of all distributed file systems are similar to IPFS and Swarm, we review the recent cutting-edge studies of blockchain-
based DFSs mainly focusing on IPFS and Swarm. The contribution of this survey includes the following aspects. • This paper first introduces the layered structure of blockchain-
based DFSs. We then make a comprehensive taxonomy of the cutting-

edge studies on the scalability and privacy perspectives. • We also clarified the challenges, open is sues and future directions of the blockchain-
based DFSs. • To the best of our knowledge, this is the first survey related to the blockchain-
based DFSs. Our review in this article can help subsequent researchers well understand both the cu rrent development and the future trends of the blockchain-
based DFS. The rest of this paper is organized as follows. In Section II, we explain necessary pre liminaries and basic concepts. Section III shows the layered structure of distributed file systems. Se ction IV summarizes the cutting-
edge studies. Section V discusses open issues, challenges and future directions. Finally, section VI concludes this paper.

Since the blockchain-
based distributed file systems emphasized on this article have a close correlation with the basic dat a structure of blockchains, we first introduce the preliminaries of Merkle Tree and Merkle DAG. T hen, we have an overview of BitTorrent, which can help us understand the rationale of distributed file systems such as IPFS acan filter out the remote and local tunnels using GUI as per their co nvenience.



V. CONCLUSION

The new generation of blockchain-
based distributed file systems, such as IPFS and Swarm, have shown their great potentials with the ir key characteristics: novel solutions of incentive, low-
latency data retrieval, automated auditing, and censorship-
resistant, etc. This paper first presents the rationale, layered structure and an overview of blockchai n-
based distributed file systems, particularly focusing on IPFS and Swarm systems. Then, we review t he cutting-
edge studies, and reveal a series of challenges that constrain their development. Open issues and fu ture directions are also discussed. We believe that the blockchain-
based distributed file systems will become very promising solutions for the next-
generation websites and data-
sharing platforms. We anticipate that this article can trigger blooming investigations on blockchainba sed distributed file systems …

VI.REFERNCES

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. [2] 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. [3] Benet, J. (2014). IPFS-content addressed, versioned, P2P file system. arXiv preprint arXiv:1407.3561. [4] An Introduction to IPFS – ConsenSys – Medium. (2018). Medium. [Online]. Available:https://medium.com/@ConsenSys/anintroduction-to-ipfs-9bba4860abd0.

Case, Amber (4 October 2015). "Why The Internet Needs IPFS Before It's Too Late". TechCrunch. Archived from the original on 5 February 2022. Retrieved 16 July 2019.

Palmer, Danny (11 June 2019). "This unusual Windows malware is controlled via a P2P network". ZDNet. Archived from the original on 6 September 2019. Retrieved 31 August 2019.

Abrams, Lawrence (4 October 2018). "Phishing Attacks Distributed Through Cloudflare's IPFS Gateway". Bleeping Computer. Archived from the original on 9 October 2019. Retrieved 31 August 2019.

Dale, Brady (10 May 2017). "Turkey Can't Block This Copy of Wikipedia". Observer Media. Archived from the original on 18 October 2017. Retrieved 20 December 2017.

Johnson, Steven (16 January 2018). "Beyond the Bitcoin Bubble". The New York Times. Archived from the original on 21 December 2021. Retrieved 26 September 2018.

"Public Gateway Checker | IPFS". ipfs.github.io. Archived from the original on 24 August 2020. Retrieved 29 August 2020.

About". Protocol Labs. Archived from the original on 28 April 2021. Retrieved 28 April 2021.

Porter, Jon (19 January 2021). "Brave browser takes step toward enabling a decentralized web". The Verge. Archived from the original on 26 February 2021. Retrieved 29 January 2021.