IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Systematic research on Decentralised file storage system using block chain technology

[1]Nilu Bhagwan Puri, [2]Prof. Pallavi P. Rane, [3]Prof. Nilesh N. Shingne, [4]Prof. Harshal S. Deshpande

[2,4]Assistant professor, Rajarshi Shahu College of Engineering, Buldhana, Maharashtra

[3]Assistant professor, Sanmati engineering College, Washim, Maharashtra

Abstract: Traditional centralized storage systems face limitations like single points of failure, security breaches, and limited scalability. Decentralized file storage systems using blockchain offer a compelling alternative. This paper presents such a system, focusing on data integrity, privacy, and accessibility without a central authority. Blockchain's distributed ledger secures data transactions, ensuring transparency and immutability. Cryptographic hashing, data encryption, and consensus algorithms enhance security and prevent unauthorized access. Files are sharded and distributed across a node network, with blockchain storing metadata for retrieval. This architecture improves fault tolerance and enables cost-effective scaling by leveraging unused storage. Token-based incentives encourage node participation. Potential applications include secure data sharing, archival storage, and sensitive information management in sectors like healthcare, finance, and IoT. This blockchain-based storage system offers a secure, scalable, and transparent platform, addressing the vulnerabilities of traditional systems.

Keywords: Scalability, Decentralized, Blockchain, Hashing, Data files

1. Introduction

With growing concerns about privacy, security, and user control, data sovereignty has become crucial for digital platforms. Traditionally, it's been understood as data being subject to the laws of the country where it's collected or processed, often leading to data localization policies. However, cross-border digital interactions highlight the shortcomings of this model. Individuals and organizations now demand greater data autonomy. Centralized storage systems also face scrutiny due to vulnerabilities to breaches, censorship, and unauthorized access. These issues necessitate solutions that empower users with data control, leading to the concept of data self-sovereignty (DSS).

DSS expands on data sovereignty by giving individuals and organizations complete control over their data, regardless of storage or processing location. It emphasizes user autonomy, allowing individuals to decide how their data is stored, accessed, and shared, without relying on central authorities. This aligns with the trend toward decentralized digital infrastructures. Decentralized storage systems, frequently using blockchain, are essential for DSS. Blockchain's decentralized nature, combined with transparency, immutability, and cryptography, ensures data security and user control. Smart contracts, self-executing agreements on the blockchain, automate and enforce data rules without intermediaries, preserving data ownership. Decentralized storage, by distributing data across multiple nodes, minimizes single points of failure and improves security, privacy, reliability, and user autonomy.

Centralized systems are vulnerable to outages if the central server or data center fails. Decentralized systems distribute data across multiple nodes, ensuring data availability even if some nodes go offline. Centralized storage is a prime target for hackers. A successful attack can compromise vast amounts of data. Decentralized systems, with their distributed nature and cryptographic security, offer enhanced protection against cyberattacks. Centralized authorities can censor or restrict access to data. Decentralized systems, by distributing data and control, make censorship more difficult. Users often have limited visibility into how their data is being handled by centralized providers. Blockchain's transparent ledger provides an auditable trail of data transactions. Centralized storage providers can impose high fees, and costs can escalate with increasing storage needs. Decentralized storage can offer more competitive pricing by leveraging unused storage capacity across a network.

Blockchain-based systems often employ strong encryption methods to protect data from unauthorized access. Blockchain's immutability ensures that data cannot be tampered with once it's recorded on the ledger. Decentralized systems empower users with greater control over their data, including who can access it and how it's used.

2. Project Objectives:

- To ensure that data is securely stored and protected from unauthorized access and breaches.
- To Provide immutability through blockchain, ensuring that stored data cannot be tampered with or altered without detection.
- To Empower users with full control over their data, allowing them to decide how it is stored, accessed, and shared without reliance on centralized entities.
- To eliminate single points of failure and ensure consistent availability even if some nodes go offline.
- To safeguard sensitive information while preventing unauthorized entities from accessing the data.

3. Literature Review

Csirmaz et. al. states that synchronizing diverged copies of some data stored on a variety of devices and/or at different locations is an ubiquitous task. The last two decades saw a proliferation of practical and theoretical works addressing this problem. File synchronization is a feature usually included with backup software in order to make is easier to manage and recover data as and when required. File synchronization usually delivered through cloud services. Dedicated file synchronizing solutions frequently come with additional tools not just for managing the saved data, but also to allow for file sharing and collaboration with stored files and documents. These cloud storage services are easily accessible for the end-user because the service front-ends are very well integrated into web clients as well as desktop and mobile environments. Simple user interfaces hide the complex and sophisticated service back-ends. Collaboration services are frequently integrated into the “cloud storage” environment. For example, Google Docs is an application layer integrated into Google Drive storage, Office 365 is integrated with One Drive storage and Dropbox Paper service is an extension of [1]

Dürsch et. al. states that an application that enable digitally conducting QDA are grouped as Computer assisted qualitative data analysis software. One representative of Computer assisted qualitative data analysis software (CAQDAS) is called QDAcity1. QDAcity is a cloud-based web application, developed and operated by the Professorship for Open Source Software at the Friedrich-Alexander-Universität Erlangen- Nürnberg. QDAcity provides an environment for multiple analysts or researchers to collaboratively conduct QDA. Since QDA deals with big amounts of fuzzy and subjective data, enabling researchers to share and discuss different interpretations, ideas, and conclusions can be very beneficial for the process of QDA. The approach of enhancing a process by promoting close collaboration and "shrinking the feedback loop" can also be found in other fields. Agile approaches of software development like Extreme Programming (Beck, 2000) serve as examples of this. However, currently QDAcity only allows the simultaneous collaboration of multiple researchers in a shared project, but not on a more granular level in a shared document. Real-time collaborative editing of a shared document is a classic form of digitally enabled, close collaboration. [2]

Martins et. al. states that in recent years the cloud has become ubiquitous. Many apps and services with users spread across the world resort to these solutions. Cloud applications with global scale user base like social media tend to resort to distributed databases that prioritize lower latency over strong consistency. Such solutions don’t require coordination which would require reads and writes to contact a majority of replicas in a communication process that can cross continents, penalizing performance. This kind of of applications along with the database replicas usually run in multiple datacenters. Instances are usually geo-replicated to accommodate users from different parts of the globe with fast response time. When the amount of replicas grows it also becomes important to partition data in a way that doesn’t break the fault tolerance guarantees of replication, since having every piece of data in every replica of the database might not be necessary and can definitely become very expensive. In such a setting it is not enough to have database replicas close to the users. The coordination between replicas performing reads and writes also needs to be minimized in order to achieve the desired low latency. Consider that you have a local server close to a client. In order for a client database operation to complete, it would need to contact a majority of database instances. This would completely break the desired low latency. For this reason weak consistency models have been rising in popularity recently. [3]

4. Motivation

Centralized storage systems are susceptible to vulnerabilities like single points of failure, data breaches, censorship, and high costs. Decentralized storage offers a more resilient and secure alternative. Growing concerns about data privacy, ownership, and regulatory compliance necessitate empowering users with greater data control. Decentralized storage supports data sovereignty by prioritizing user autonomy. Blockchain's immutability, transparency, and cryptographic security can revolutionize data management by ensuring integrity and trust without intermediaries. The increasing sophistication of cyberattacks and data breaches makes secure and private storage systems critical. Decentralized systems, by distributing data, enhance security. Advances in peer-to-peer networking and blockchain provide a solid base for developing decentralized storage, and further research promises innovative solutions.

5. Proposed Methodology:

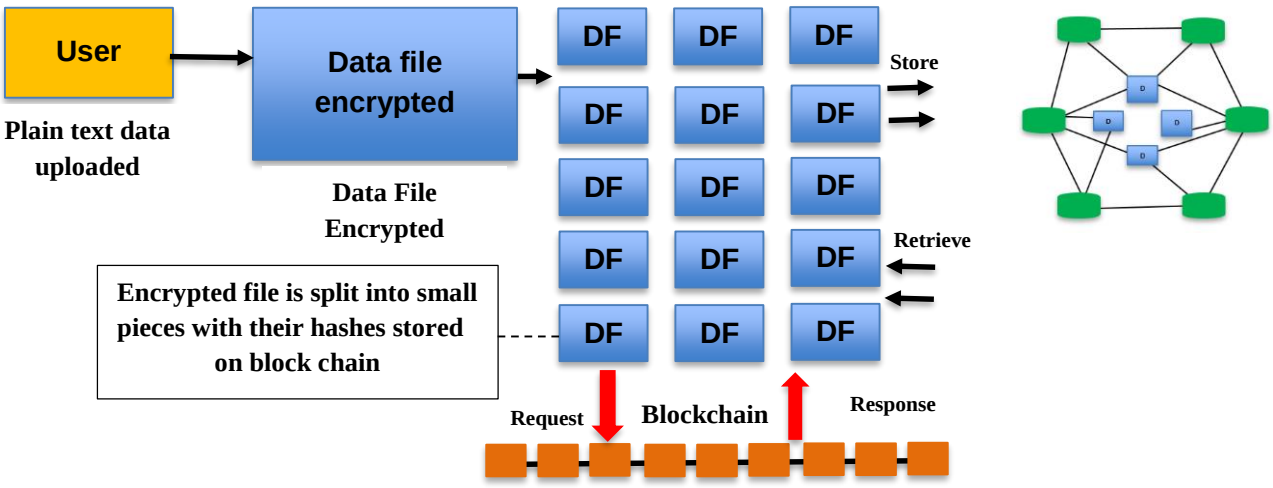

Fig. 1 Architecture of Block Chain

Decentralized storage systems operate on a peer-to-peer (P2P) network framework where participants contribute unused storage space in exchange for incentives like tokens or cryptocurrencies. Blockchain technology underpins this model by facilitating the issuance and management of digital tokens that reward network contributors. This incentivized approach encourages active participation, ensuring the sustainability, scalability, and efficiency of the storage ecosystem. By aligning participant incentives with the system's overall health, decentralized storage systems maintain continuous data availability, reliability, and security.

A typical blockchain-based decentralized storage system functioning on a P2P network involves four key steps:

1. Data Uploading: Users upload plaintext files to the decentralized storage system.
2. Data Encryption: Uploaded files are encrypted using advanced cryptographic algorithms. This process converts plaintext into ciphertext, ensuring data privacy and confidentiality. Only users with the correct decryption keys can access the original data.
3. Data Fragmentation (Sharding): Encrypted data files are divided into smaller fragments, or shards. This process enhances system scalability and performance by enabling secure distribution of data fragments. Sharding also improves retrieval speed, as individual fragments can be accessed independently.
4. Data Chunk Distribution: Encrypted fragments are distributed across multiple nodes in the P2P network. Each node, contributing storage and participating in data operations, ensures redundancy and high availability. Even if some nodes fail or go offline, the system maintains data integrity and accessibility by storing fragments on other nodes.

This architecture leverages blockchain's transparency and cryptographic security to create a resilient, efficient, and user-driven storage environment.

6. Applications:

Decentralized file storage systems using blockchain technology have a wide range of potential applications, offering improvements in security, privacy, and accessibility compared to traditional centralized systems. Here are some key applications:

Secure Data Storage and Backup:

- Personal Data: Individuals can securely store and back up personal files, photos, and documents, maintaining control over their data and reducing the risk of data loss or theft.
- Sensitive Information: Organizations can store sensitive information, such as financial records or medical data, with enhanced security and privacy, complying with data protection regulations.
- Archival Storage: Decentralized systems can provide long-term archival storage for important documents and records, ensuring data integrity and accessibility over time.

Data Sharing and Collaboration:

- Secure File Sharing: Users can securely share files with others, controlling access permissions and ensuring data privacy.
- Collaborative Workspaces: Decentralized platforms can facilitate collaborative work on documents and projects, with version control and transparent audit trails.
- Content Distribution: Creators can distribute content (e.g., music, videos, software) directly to audiences, bypassing intermediaries and maintaining control over their work.

Digital Identity and Access Management:

- Self-Sovereign Identity: Individuals can manage their digital identities and control access to their personal data, reducing reliance on centralized identity providers.
- Secure Authentication: Decentralized storage can be used to store and manage authentication credentials, enhancing security and reducing the risk of identity theft.

Supply Chain Management:

- Product Tracking: Blockchain-based storage can be used to track products throughout the supply chain, ensuring transparency and authenticity.
- Supply Chain Data: Securely store and share supply chain data among stakeholders, improving efficiency and collaboration.

Healthcare:

- Electronic Health Records: Patients can securely store and control access to their electronic health records, enabling seamless sharing with healthcare providers.
- Medical Research: Decentralized platforms can facilitate secure sharing of medical data for research purposes, while protecting patient privacy.

Internet of Things (IoT):

- IoT Data Management: Decentralized storage can provide a scalable and secure solution for managing the massive amounts of data generated by IoT devices.
- Secure Device Communication: Blockchain can be used to secure communication between IoT devices, preventing unauthorized access and data breaches.

7. Conclusion:

This paper examines decentralized storage systems, focusing on their characteristics, performance, and role in sustainable data self-sovereignty (DSS). Blockchain-based decentralized storage offers a promising path to realizing DSS by giving users greater control, privacy, and security. The paper highlights the need to match these systems to user needs for effective platform selection. While blockchain storage excels in security and integrity, systems vary considerably in complexity, cost, and performance. Users must therefore carefully consider these factors to choose the best option. These systems are crucial for self-sovereign data management, providing secure, resilient, and user-centric solutions. With growing global attention to data ownership, privacy, and security, decentralized storage platforms are increasingly important for achieving data sovereignty in the digital age. This paper is a valuable resource for users, developers, and researchers, supporting informed decisions about selecting and implementing decentralized storage that meets their sovereignty and operational needs.

References:

[1] Elod P. Csirmaz and Laszlo Csirmaz, “Synchronizing Many Filesystems in Near Linear Time”, arXiv:2302.09666v2 [cs.IT] 17 May 2023

[2] Martin Dürsch, “Scaling Real-time Collaborative Editing in a Cloud-based Web App”, Erlangen, 19 April 2023

[3] João Gonçalves Martins, “Query Processing in Cloud Databases with Partial Replication” NOVA University Lisbon March, 2023

[4] Masoumeh Hajvali, Sahar Adabi, Ali Rezaee and Mehdi Hosseinzadeh, “Decentralized and scalable hybrid scheduling-clustering method for real-time applications in volatile and dynamic Fog-Cloud Environments” (2023) 12:66, https://doi.org/10.1186/s13677-023-00428-4, Journal of Cloud Computing: Advances, Systems and Applications

[5] Elod P. Csirmaz 1,_ and Laszlo Csirmaz, “Data Synchronization: A Complete Theoretical Solution for Filesystems” Future Internet 2022, 14, 344. https://doi.org/10.3390/fi14110344

[6] Novak Boˇskov, Ari Trachtenberg, and David Starobinski. Enabling costbenefit analysis of data sync protocols, 2023.

[7] Elod P. Csirmaz and Laszlo Csirmaz. Data synchronization: A complete theoretical solution for filesystems. *Future Internet*, 14(11), 2022.

[8] Elod Pal Csirmaz. Algebraic file synchronization: Adequacy and completeness. *CoRR*, abs/1601.01736, 2016.

[9] John Day-Richter. What’s different about the new Google Docs: Making collaboration fast, 2010.

[10] Shimon Even. *Graph Algorithms*. Cambridge University Press, USA, 2nd edition, 2011.

[11] JiuLing Feng, XiuQuan Qiao, and Yong Li. The research of synchronization and consistency of data in mobile environment. In *2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems*, volume 02, pages 869–874, 2012.

[12] Rusty Klophaus. Riak core: Building distributed applications without shared state. In *ACM SIGPLAN Commercial Users of Functional Programming*, CUFP ’10, New York, NY, USA, 2010. Association for Computing Machinery.

[13] Zhenhua Li, Christo Wilson, Zhefu Jiang, Yao Liu, Ben Y. Zhao, Cheng Jin, Zhi-Li Zhang, and Yafei Dai. Efficient batched synchronization in dropbox-like cloud storage services. In David Eyers and Karsten Schwan, editors, *Middleware 2013*, pages 307–327, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.