Data Storage and Retrieval Using IPFS and Blockchain

Subjects: Others

Contributor: Muhammad Bin Saif, Sara Migliorini, Fausto Spoto

Blockchain technology has been successfully applied in recent years to promote the immutability, traceability, and authenticity of previously collected and stored data. However, the amount of data stored in the blockchain is usually limited for economic and technological issues. Namely, the blockchain usually stores only a fingerprint of data, such as the hash of data, while full, raw information is stored off-chain. This is generally enough to guarantee immutability and traceability, but misses to support another important property, that is, data availability. This is particularly true when a traditional, centralized database is chosen for off-chain storage. For this reason, many proposals try to properly combine blockchain with decentralized IPFS storage. However, the storage of data on IPFS could pose some privacy problems. This entry proposes a solution that properly combines blockchain, IPFS, and encryption techniques to guarantee immutability, traceability, availability, and data privacy.

Keywords: blockchain ; decentralized storage ; IPFS ; data traceability

1. Introduction

In the digital information age, the exponential growth in data volume has posed significant challenges in terms of efficiency and scalability. Indeed, traditional centralized data storage and retrieval systems act as a single computational node, which consequently become a trust bottleneck and a single point of failure. To overcome these challenges, data storage is shifting to distributed systems, where data spread across multiple public nodes. In distributed systems, the decentralized peer-to-peer nodes can synchronize with each other without the need for a central authority. For instance, a torrent file-sharing protocol [1] uses a peer-to-peer network to connect all participating computers, allowing them to share files in a decentralized manner.

More recently, blockchain technology has emerged as an innovative solution for reaching a consensus about a global state in a decentralized manner, without the need for a central authority. It allows a decentralized network of untrustworthy nodes to agree on the content and state of the blockchain, independently from each other. Furthermore, the information stored in blockchain is also immutable, since the content of a block cannot be modified without invalidating the content of all the subsequent blocks. In blockchain, smart contracts play a vital role in developing decentralized applications on top of the blockchain. In this context, a smart contract is a piece of code deployed in blockchain to execute predefined actions, enabling automated execution by ensuring the immutability and transparency of agreements in a trustless environment [2]. Therefore, smart contracts can significantly contribute to the spread of blockchain technology in several application domains, such as healthcare [3], tourism [4], energy and water management [5], identity management [6], supply chains [7], and so on.

In contrast to traditional centralized solutions that lack data traceability and transparency, blockchain and smart contract technology could help to overcome these challenges. Indeed, blockchain technology inherently provides traceability, immutability, and trustworthiness [8]. These properties can secure and tamper-proof critical information in many domains. The final aim is to increase the trust of end users by issuing non-repudiable certificates. On the other hand, storing large amounts of data in blockchain is not feasible, due to scalability and transaction cost challenges. Indeed, increasing data volume increases the transaction cost and decreases the blockchain throughput (scalability). Therefore, as the volume of data grows, blockchain solutions become increasingly inefficient and economically unviable. However, a solution, called off-chain data storage [9], has been proposed to address the challenges of efficient data storage strategies, minimizing on-chain data while ensuring integrity. The idea is to develop proper solutions that guarantee the consistency and immutability of off-chain data from those of the on-chain data.

The Interplanetary File System (IPFS) is a peer-to-peer distributed system for storing, accessing, and sharing files, websites, applications, and data. First introduced in 2015, IPFS developed upon a decentralized environment and incorporates distributed and bandwidth-saving techniques from torrent [10]. Blockchain and IPFS function as decentralized technologies but serve different purposes and have distinctive characteristics. IPFS offers an efficient, peer-to-peer

decentralized public network for large distributed data storage and access. It aims to improve the efficiency and resilience of traditional web protocols by allowing files to be stored in multiple locations, making them resistant to censorship and ensuring availability, even if some nodes go offline. On the other hand, blockchain serves primarily as a decentralized ledger, recording transactions or data transparently and in a tamper-proof way. Integrating these technologies represents an efficient solution to the challenges mentioned above related to scalability, efficiency, immutability, and data availability.

2. Blockchain and IPFS for Data Storage

Blockchain technology is receiving a lot of attention due to its potential for secure and decentralized data storage. The studies on blockchain mainly explored the blockchain's immutable ledger system for data storage; however, they frequently emphasized the limits associated with the high costs and inefficiencies of keeping large volumes of raw data directly in the blockchain. In [11], the authors propose an in-depth analysis of these challenges, emphasizing the need for more effective data storage solutions within the blockchain framework. In this regard, IPFS is becoming increasingly common as a large data storage solution, thanks to its decentralized and effective design. In [12], the authors employ blockchain technology to transmit data in a peer-to-peer network and IPFS as data-sharing infrastructure to share pre-trained deep learning models to stakeholders. Meanwhile, in [13], the authors propose a blockchain and cloud-based decentralized secure storage for data availability, privacy, and efficient resource utilization. The work in [14] proposes the integration of blockchain and IPFS, showing the potential for secure data retrieval and storage. The proposed approach leverages different privacy modes for data privacy and security by storing nonsensitive data on IPFS without encryption and sensitive data with two-layered encryption. This approach mitigates the risk of a single point of failure and data tampering. However, data retrieval and query require the decryption of data stored on IPFS, which may introduce additional security and privacy risks associated with the exposure of private keys.

3. Data Security and Privacy Methods

Data privacy has been a significant challenge in decentralized systems. Encryption and hashing have been pivotal in securing data within the blockchain and IPFS frameworks. The literature on various encryption algorithms and hashing techniques studied by [15] shows the advancement in these methods in maintaining data integrity and security. Furthermore, ref. [16] focuses on the role of these techniques in preserving data privacy in decentralized systems and on the importance of a balance between accessibility and security. In [17], the authors delve into privacy-preserving approaches such as zero-knowledge proofs and homomorphic encryption. These studies emphasize the development of techniques that secure data without compromising privacy. However, these techniques can provide solutions for data security and privacy but often fall short in terms of balancing data security with system efficiency, such as computation overhead, scalability, and dynamic data handling. In addition, relying on encryption inadvertently exposes private keys during decryption processes, making it a substantial security risk. The work in [18] proposes a blockchain-based framework for privacy preservation and secure access control in cloud storage. The proposed scheme combines the Ethereum blockchain and ciphertext-policy attribute-based encryption (CP-ABE) for user data security and privacy. However, due to privacy issues and data leakage, the cloud might not be a trustworthy source for data storage.

4. Query Optimization Techniques

Leveraging blockchain and IPFS in decentralized storage systems has transformed data storage and retrieval, particularly in sectors demanding high integrity and efficiency. These studies illustrate how blockchain and IPFS can improve query optimization, security, and scalability in distributed storage systems. In [19], the authors focus on agricultural product traceability, using IPFS to store multiple data types and blockchain for securing IPFS hash addresses to improve query efficiency and data authenticity. However, the sensor data are collected and processed on a centralized private server before storing data on IPFS. Therefore, this approach may pose security risks related to a single point of failure, data loss, or compromised server. In [20], the authors integrate blockchain, IPFS, and Elasticsearch to address big data storage challenges in distributed systems, resulting in reduced storage overhead, search latency, and access control. The proposed approach relies on blockchain's inherent security features. Moreover, the elastic search approach focuses primarily on search latency and precision, without considering the security implications of query handling mechanisms. The work in [21] proposes a secure and decentralized framework for managing cloud data provenance by integrating blockchain technology with the IPFS. The proposed solution ensures provenance data availability, security, and integrity by utilizing decentralized storage and blockchain for immutability. The method for verifying data integrity is secure; however, it might be computationally intensive, particularly in scenarios where the frequent validation of large data sets is necessary.

References

1. Pouwelse, J.; Garbacki, P.; Epema, D.; Sips, H. The Bittorrent P2P File-Sharing System: Measurements and Analysis. In Proceedings of the Peer-to-Peer Systems IV; Castro, M., van Renesse, R., Eds.; Springer: Berlin/Heidelberg, Germany, 2005; pp. 205–216.
2. Mohanta, B.K.; Panda, S.S.; Jena, D. An Overview of Smart Contract and Use Cases in Blockchain Technology. In Proceedings of the 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Bengaluru, India, 10–12 July 2018; pp. 1–4.
3. Pinto, R.P.; Silva, B.M.C.; Inácio, P.R.M. A System for the Promotion of Traceability and Ownership of Health Data Using Blockchain. IEEE Access 2022, 10, 92760–92773.
4. Rana, R.L.; Adamashvili, N.; Tricase, C. The Impact of Blockchain Technology Adoption on Tourism Industry: A Systematic Literature Review. Sustainability 2022, 14, 7383.
5. Xia, W.; Chen, X.; Song, C. A Framework of Blockchain Technology in Intelligent Water Management. Front. Environ. Sci. 2022, 10.
6. Stockburger, L.; Kokosioulis, G.; Mukkamala, A.; Mukkamala, R.R.; Avital, M. Blockchain-enabled decentralized identity management: The case of self-sovereign identity in public transportation. Blockchain Res. Appl. 2021, 2, 100014.
7. Agrawal, T.K.; Kumar, V.; Pal, R.; Wang, L.; Chen, Y. Blockchain-based framework for supply chain traceability: A case example of textile and clothing industry. Comput. Ind. Eng. 2021, 154, 107130.
8. Zhu, H.; Zhou, Z.Z. Analysis and outlook of applications of blockchain technology to equity crowdfunding in China. Financ. Innov. 2016, 2, 29.
9. Rajasekar, V.; Sondhi, S.; Saad, S.; Mohammed, S. Emerging Design Patterns for Blockchain Applications. In Proceedings of the ICSOFT, Online Event, 7–9 July 2020; pp. 242–249.
10. Bauer, D.P. InterPlanetary File System. In Getting Started with Ethereum: A Step-by-Step Guide to Becoming a Blockchain Developer; Apress: Totowa, NJ, USA, 2022; pp. 83–96.
11. Monrat, A.A.; Schelén, O.; Andersson, K. A Survey of Blockchain From the Perspectives of Applications, Challenges, and Opportunities. IEEE Access 2019, 7, 117134–117151.
12. ul Haque, A.; Ghani, M.S.; Mahmood, T. Decentralized Transfer Learning using Blockchain & IPFS for Deep Learning. In Proceedings of the 2020 International Conference on Information Networking (ICOIN), Barcelona, Spain, 7–10 January 2020; pp. 170–177.
13. Shah, M.; Shaikh, M.; Mishra, V.; Tuscano, G. Decentralized Cloud Storage Using Blockchain. In Proceedings of the 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) (48184), Tirunelveli, India, 15–17 June 2020; pp. 384–389.
14. Zheng, X.; Lu, J.; Sun, S.; Kiritsis, D. Decentralized industrial IoT data management based on blockchain and IPFS. In Proceedings of the IFIP International Conference on Advances in Production Management Systems, Novi Sad, Serbia, 30 August–3 September 2020; Springer: Cham, Switzerland, 2020; pp. 222–229.
15. Huang, H.; Lin, J.; Zheng, B.; Zheng, Z.; Bian, J. When blockchain meets distributed file systems: An overview, challenges, and open issues. IEEE Access 2020, 8, 50574–50586.
16. Hassan, M.U.; Rehmani, M.H.; Chen, J. Privacy preservation in blockchain based IoT systems: Integration issues, prospects, challenges, and future research directions. Future Gener. Comput. Syst. 2019, 97, 512–529.
17. Satybaldy, A.; Nowostawski, M. Review of techniques for privacy-preserving blockchain systems. In Proceedings of the 2nd ACM International Symposium on Blockchain and Secure Critical Infrastructure, Taipei, Taiwan, 6 October 2020; pp. 1–9.
18. Wang, S.; Wang, X.; Zhang, Y. A Secure Cloud Storage Framework With Access Control Based on Blockchain. IEEE Access 2019, 7, 112713–112725.
19. Hao, J.; Sun, Y.; Luo, H. A safe and efficient storage scheme based on blockchain and IPFS for agricultural products tracking. J. Comput 2018, 29, 158–167.
20. Arer, M.M.; Dhulavvagol, P.M.; Totad, S. Efficient big data storage and retrieval in distributed architecture using blockchain and ipfs. In Proceedings of the 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 7–9 April 2022; IEEE: Toulouse, France, 2022; pp. 1–6.
21. Hasan, S.S.; Sultan, N.H.; Barbhuiya, F.A. Cloud data provenance using IPFS and blockchain technology. In Proceedings of the Seventh International Workshop on Security in Cloud Computing, Auckland, New Zealand, 8 July 2019; pp. 5–12.