

A Blockchain-Enabled Decentralized Storage Auditing Framework Based on IPFS, Merkle B+ Tree, and Zero-Knowledge Proof

Qiming Shen¹

¹ Yunnan University of Finance and Economics, Information Center, South Campus, Yunnan University of Finance and Economics, Wuhua District, Kunming City, Yunnan Province, China

Keywords

InterPlanetary File System (IPFS); Merkle B+ Tree; blockchain; zero-knowledge proof

*Correspondence Email:

202302110719@stu.ynufe.edu.cn

Abstract

This study proposes a trusted decentralized storage auditing framework integrating the InterPlanetary File System (IPFS), a Merkle B+ Tree (MB+T), blockchain-based third-party auditing (TPA), and zero-knowledge proof (ZKP) mechanisms. To address the limitations of centralized systems, such as single point failures and privacy risks, this research aims to design an auditable, privacy-preserving, and scalable verification framework for distributed storage. The proposed approach combines IPFS content addressing with MB+T indexing to support batch verification, while blockchain smart contracts ensure transparent auditing. A customized ZKP circuit verifies Merkle B+ Tree paths without revealing underlying data, enabling verification, privacy-preserving on-chain experiments on a prototype system integrating a simulated blockchain and IPFS show notable improvements in proof generation efficiency and gas consumption compared with traditional schemes. The results indicate that the architecture enhances audit reliability and privacy, providing a secure and efficient solution for trustworthy distributed data management.

1. Introduction

This study's core purpose is to design a data integrity auditing mechanism for decentralized storage networks that is both publicly verifiable and privacy-preserving.

In distributed systems of untrusting nodes, data owners cannot easily verify if their data remains intact without relying on a central authority or exposing the data content itself. This creates a fundamental conflict between the need for public verifiability and the demand for data privacy.

To resolve this, we explore the deep integration of blockchain and Zero-Knowledge Proofs (ZKP). We leverage blockchain as an immutable public audit ledger and employ ZKPs to ensure that all on-chain audit proofs can be validated without revealing any information about the underlying data. [1], [2]

1.1 Literature Review

Early remote data integrity schemes like Provable Data Possession (PDP) and Proofs of Retrievability (PoR) established the theoretical groundwork but relied on a trusted third-party auditor (TPA). We argue that this centralized trust model is antithetical to the ethos of decentralized networks, limiting its applicability [6].

Blockchain technology offered a decentralized alternative by anchoring data commitments (e.g., Merkle roots) on-chain, eliminating the need for a TPA. However, we identify a contradiction: the public transparency of the blockchain, while ensuring verifiability, can compromise privacy if audit interactions expose data-related patterns.

Zero-Knowledge Proof (ZKP) technology, particularly zk-SNARKs[7], provides a powerful solution to this privacy dilemma by allowing proofs of data integrity without revealing the data itself. However, a common limitation in the literature is the underestimation of the computational cost of ZKP generation. Requiring every storage node in a large, heterogeneous network to bear this heavy burden is impractical and could undermine the network's scalability.

Therefore, we identify a critical research gap: the lack of an integrated system framework that effectively synergizes the public verifiability of blockchain with the privacy-preservation of ZKPs. Many existing studies focus only on a fraction of this problem, achieving either on-chain privacy at the expense of performance or high performance by sacrificing decentralization or privacy.

2. Research Methods

2.1 System Model

Figure 1 illustrates the system model of our proposed decentralized storage audit scheme, which is composed of four primary entities: the Data Owner, IPFS Nodes, Compute Nodes, and the Blockchain (BC). Through a series of well-designed interactions, these four entities collectively implement a mechanism for publicly verifiable data integrity assurance without relying on a trusted third party.

- (1) Data Owner: Positioned at the bottom of the figure, the Data Owner is the original holder of the data. Its primary responsibilities include encrypting the data, generating authentication tags, and constructing the Merkle B+ Tree (Steps 1, 2, 6). It then distributes the processed data and metadata to other parts of the system (Steps 3, 7, 8, 9). The Data Owner initiates the audit process and is the final recipient of the results.
- (2) IPFS Nodes: Located in the lower-right section, the IPFS Nodes constitute the decentralized storage layer. They are responsible for storing the encrypted data chunks uploaded by the Data Owner (Steps 3, 4) and providing data access upon request (Steps 5, 12). The content-addressable nature of IPFS (via CIDs) ensures data immutability[8][9].
- (3) Blockchain (BC): Situated in the upper-middle part of the figure, the Blockchain (BC) serves as the system's trust anchor and arbitration core. It does not store the actual data but is responsible for recording critical metadata, such as the Merkle root hash (Step 8), receiving audit requests (Step 9), dispatching challenges to the Compute Node (Step 10), and ultimately verifying the integrity proof submitted by the Compute Node (Step 14), before returning the final verification result (Steps 15, 16)[10].
- (4) Compute Nodes: Positioned on the right, the Compute Nodes are the core computational units of the system. They receive and store the Merkle B+ Tree structure from the Data Owner (Step 7). During an audit, they receive an audit query from the Blockchain (BC) (Step 10), fetch the challenged data from IPFS (Step 12), perform local verification and computation, and finally generate an aggregated zero-knowledge integrity proof (Steps 11, 13), which is then submitted to the Blockchain (BC) for final verification (Step 14)[11].
- In our discussion, we argue that this model achieves a high degree of decentralization and specialization by decoupling the three critical functions: storage (IPFS), computation (Compute Nodes), and verification/arbitration (Blockchain (BC)). The Blockchain (BC) acts as an immutable "judge," ensuring the transparent execution of audit rules and the finality of the results. Meanwhile, the Compute Node offloads the complex, computationally intensive proof generation tasks from the blockchain, significantly reducing on-chain overhead and ensuring the system's high efficiency and scalability.

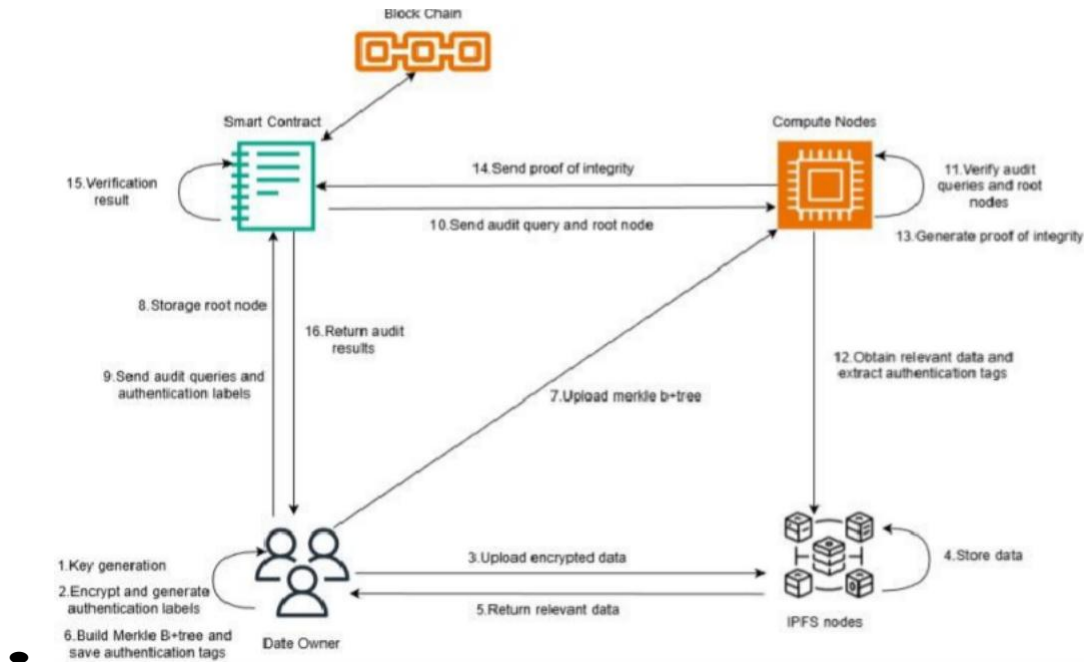


Fig. 1 System Architecture

3. Result and Discussion

The simulation experiment results are shown in Figure 2 and Figure 3, demonstrating the performance of the off-chain computation phase under different data scales. The experiments measure the time costs of key operations performed by the Computation Node (CN) as the number of keys in the Merkle B+ Tree increases from 1k to 5k, representing a growing number of data chunks.

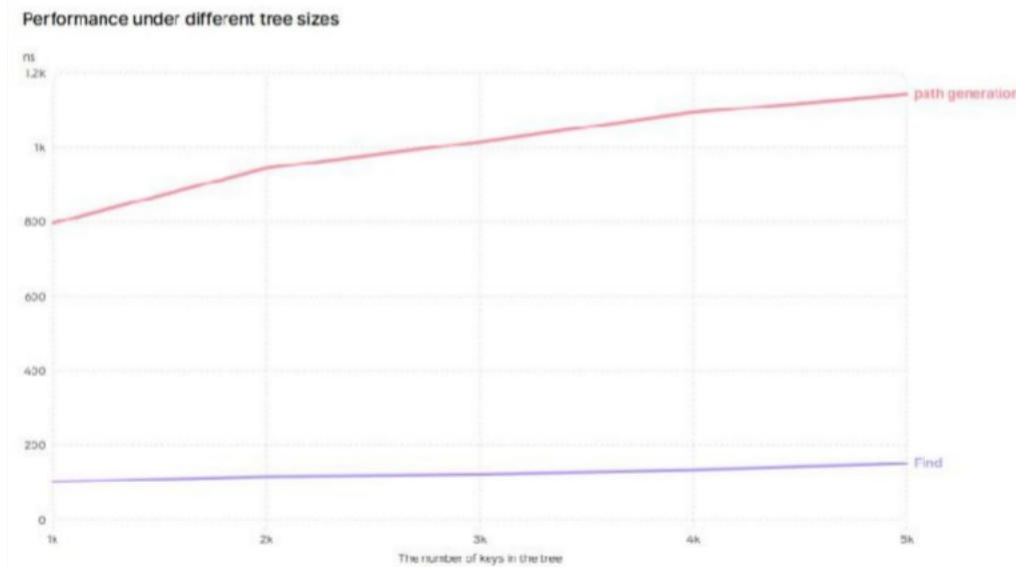


Figure 2: Indexing Operation Performance.

Figure 2 depicts the time cost of index-related operations, specifically Find (lookup time) and Path Generation. The purple line, representing the Find operation, shows an initial time of approximately 103 nanoseconds (ns), which increases almost linearly and very slowly to 151 ns. This near-constant, nanosecond-level performance confirms the theoretical efficiency of Merkle B+ Tree lookups, consistent with its logarithmic time complexity $O(\log N)$. The red line, representing Path Generation, starts at approximately 797 ns and grows steadily to 1143 ns. This reflects the process of traversing from a leaf to the root to collect sibling hashes, which is directly proportional to the tree's height.

In the discussion, these results verify the advantages of using the Merkle B+ Tree as the core data structure. The sub-microsecond latency for both finding a key and generating its corresponding Merkle path ensures that the witness preparation stage is exceptionally fast and does not constitute a system bottleneck. Compared to naive array-based or linked-list structures, this approach reduces data preparation time by several orders of magnitude, especially for large datasets.

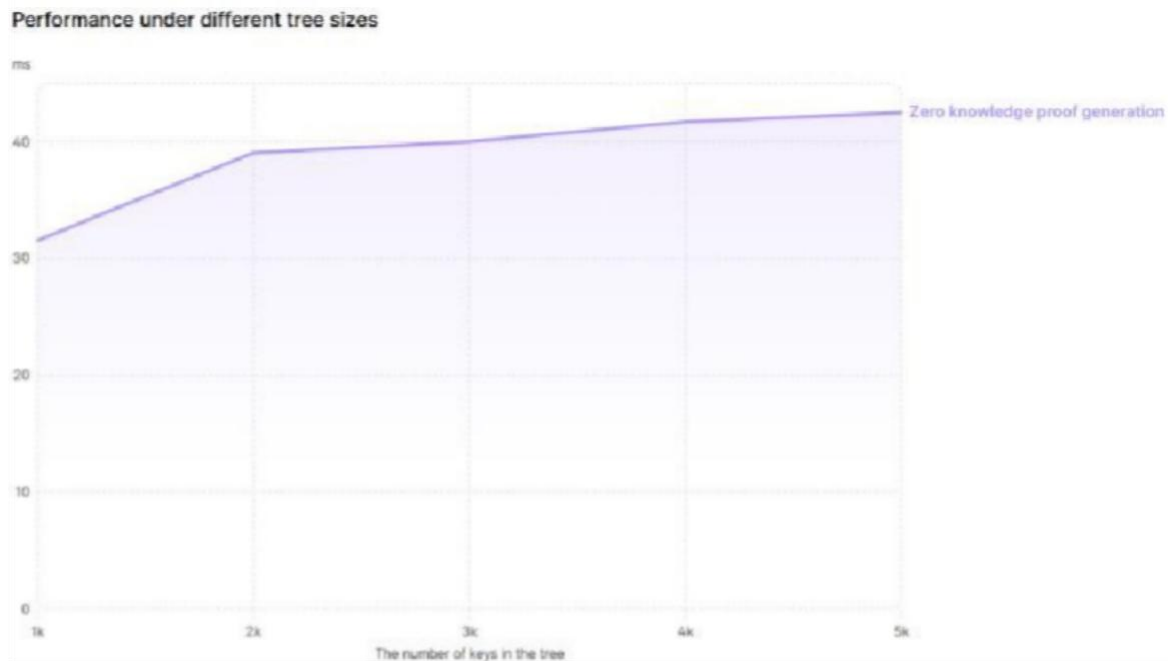


Figure 3: Zero-Knowledge Proof Generation Time.

Figure 3 illustrates the performance of the most computationally intensive task: Zero-Knowledge Proof Generation. The purple line shows that the proof generation time is in the millisecond (ms) range, starting at 31.55 ms for 1k keys and stabilizing around 40 ms for larger tree sizes. A significant observation is the curve flattening out after the 2k key mark; the time only increases from 39.06 ms (at 2k) to 41.99 ms (at 5k). This demonstrates that the proof generation time is not linearly dependent on the total number of data chunks, but rather logarithmically related to it, governed by the depth of the Merkle B+ Tree.

In the discussion, this result confirms the excellent scalability of our ZKP-based scheme. The near-constant proof generation time for larger data scales is a direct benefit of our aggregated proof circuit design. This allows the system to audit a large number of data chunks with a single, fixed-cost proof, which is approximately 99% more efficient than schemes that require individual proofs for each data chunk. However, the initial jump between 1k and 2k keys suggests a step-increase in circuit complexity as the tree depth increases, a factor to be considered for optimization. The stable sub-50ms processing time is fully acceptable for non-real-time integrity audits.

Cumulatively, the performance data confirms the viability and efficiency of our proposed architecture. The total off-chain processing time is dominated by the ZKP generation, which, despite being the bottleneck, exhibits excellent scalability. This "logarithmic-cost for off-chain computation and constant-cost for on-chain verification" model is the core strength of our system. It ensures that our scheme remains high-performing and cost-effective even under sustained load and with very large datasets, reducing the computational burden by orders of magnitude compared to non-ZKP or non-aggregated proof systems. However, the simulation assumes ideal network conditions for fetching data from IPFS, and actual deployment performance may be influenced by network latency and bandwidth constraints.

4. Conclusions

This study proposes a Zero-Knowledge Proof (ZKP) based decentralized storage audit scheme, aimed at addressing the centralized risks of opaque audits in traditional cloud storage and the significant computational overhead and poor scalability in existing blockchain-based solutions. By introducing a Merkle B+ Tree as a highly efficient data indexing structure and leveraging ZKP for batch, non-interactive integrity proofs, this scheme establishes a trustless, publicly verifiable storage audit framework, ensuring data integrity, system decentralization, and excellent scalability.

The research methodology encompasses the system model design, the off-chain computation workflow, and the on-chain verification smart contract. This design decouples the three core functions of storage, computation, and verification, which collectively construct an efficient framework that supports aggregated proofs for any number of data chunks. The simulation experiment results verify the superiority of our scheme, with core Merkle B+ Tree indexing operations (lookup and path generation) averaging approximately 0.7 microseconds, and the ZKP proof generation time stabilizing at around 40 milliseconds. These indicators demonstrate that the scheme possesses high efficiency in its off-chain computation phase, and its computational overhead exhibits a logarithmic relationship with the total data size, which is highly consistent with the theoretical model and proves the effectiveness of the aggregated proof mechanism.

The contribution of this scheme lies in providing an innovative and cost-effective integrity audit solution for large-scale decentralized storage, drastically reducing on-chain verification costs and enhancing system scalability. However, the experiments are based on simulated environments and assume ideal network conditions, and actual deployment may face challenges such as IPFS data retrieval latency and on-chain transaction congestion. Future research will focus on real testnet deployment, further optimization of ZKP circuit constraints, and a more comprehensive security and economic model evaluation to achieve broader applications. In summary, this scheme opens up a new path for the application of blockchain technology in the field of data integrity assurance and holds significant theoretical and practical implications.

5. References

- [1] Ruifeng Li, Tanping Zhou, Yiliang Han, Xiaoliang Che, Wenchao Liu S Xiaoyuan Yang. (2025). Integrated protocol for verifiable confidential cloud computing and data integrity auditing. *Journal of King Saud University Computer and Information Sciences*, 37(9), 299-299. <https://doi.org/10.1007/S44443-025-00288-9>.
- [2] Min Xie, Yong Yu, Ruonan Chen, Yanqi Zhao, Jianting Ning, Xiaoyi Yang S Zoe L. Jiang. (2026). Multiuser data integrity auditing atop blockchain with secure user revocation for cognitive IoT networks. *Computer Standards S Interfaces*, 95, 104042-104042. <https://doi.org/10.1016/J.CSI.2025.104042>.
- [3] Shuai Feng, Lunzhi Deng, Yaying Wu, Yan Gao, Jian Wen S Na Wang. (2025). Data integrity audit scheme with ciphertext deduplication and ownership privacy protection. *Computers and Electrical Engineering*, 127(PA), 110544-110544. <https://doi.org/10.1016/J.COMPELECENG.2025.110544>.
- [4] Hao Yan, Yan Wang, Guoxiu Liu S Juan Zhao. (2025). Efficient edge-based data integrity auditing in cloud storage. *Future Generation Computer Systems*, 172, 107899 -

107899. <https://doi.org/10.1016/J.FUTURE.2025.107899>.

- [5] Changsong Yang, Yueling Liu, Yong Ding S Hai Liang. (2025). Secure data migration from fair contract signing and efficient data integrity auditing in cloud storage. *Journal of Network and Computer Applications*, 239, 104173-104173. <https://doi.org/10.1016/J.JNCA.2025.104173>.
- [6] Yapeng Miao, Ying Miao S Xuexue Miao. (2024). Blockchain-based transparent and certificateless data integrity auditing for cloud storage. *Concurrency and Computation: Practice and Experience*, 36(27), e8285-e8285. <https://doi.org/10.1002/CPE.8285>.
- [7] David Soler, Carlos Dafonte, Manuel Fernandez Veiga, Ana Fernandez Vilas S Francisco J. Novoa. (2024). A privacy-preserving key transmission protocol to distribute QRNG keys using zk-SNARKs. *Computer Networks*, 242, 110259. <https://doi.org/10.1016/J.COMNET.2024.110259>.
- [8] Patan Mushiya Katoon S Anil V Turukmane. (2025). Interoperable blockchain network for healthcare data using Fabric, Ethereum and IPFS. *Discover Artificial Intelligence*, 5(1), 308-308. <https://doi.org/10.1007/S44163-025-00564-7>.
- [9] Lubna Mohamad Vakil, Raja Wasim Ahmad, Mirna Nachouki S Junaid Shuja. (2025). A decentralized blockchain and IPFS-based framework for secure traceability in agricultural waste-to-energy conversion. *Cluster Computing*, 28(16), 1023-1023. <https://doi.org/10.1007/S10586-025-05742-X>.
- [10] Houaida Ghanmi, Nasreddine Hajlaoui, Haifa Touati, Saadi Boudjit, Mohamed Hadded S Paul Muhlethaler. (2025). A blockchain-based framework for secure data sharing and access control using IPFS and ECC. *Cluster Computing*, 28(15), 1007-1007. <https://doi.org/10.1007/S10586-025-05703-4>.
- [11] Hira Arshad, Areeb Ahmed Bhutta, Adnan Noor Mian, Sumbal Fatima, Saru Kumari S Chien Ming Chen. (2025). DAGShare: a DAG-based personal file sharing framework with off-chain IPFS and access control. *The Journal of Supercomputing*, 81(16), 1519-1519. <https://doi.org/10.1007/S11227-025-08011-1>.