International Journal of Engineering Research and Science & Technology

www.ijerst.org

ISSN : 2319-5991

Vol. 22 No. 2 (2026)

ijerst.editor@gmail.com

editor@ijerst.com

International Journal of Engineering Research and Science & Technology

ISSN 2319-5991 www.ijerst.com
Vol. 22, No. 2, 2026

Peer Reviewed, Referred & Indexed Journal

BLOCKCHAIN BASED DECENTRALIZED CLOUD STORAGE WITH RELIABLE DEDUPLICATION

[1]MR.P.UDAY, [2]KORPURI HINDUJA, [3]GOUNI RANJEETHREDDY, [4]NARMETTA GANESH, [5]SHAIK KARISHMA

[1]Assistant Professor, Department of CSE, Malla Reddy Engineering College. Hyderabad, Telangana

[2,3,4,5]Students, Department of CSE, Malla Reddy Engineering College. Hyderabad, Telangana

ABSTRACT

The rapid growth of data-intensive applications and cloud computing has significantly increased the demand for secure, scalable, and cost-efficient storage solutions. Traditional centralized cloud storage systems suffer from critical limitations such as single points of failure, lack of transparency, data breaches, and inefficient storage utilization. To address these challenges, this paper proposes a Blockchain-Based Decentralized Cloud Storage with Reliable Deduplication framework that integrates the advantages of blockchain technology, distributed storage systems, and secure deduplication mechanisms. The proposed system leverages blockchain as a decentralized ledger to ensure data integrity, transparency, and tamper-resistance. Each data transaction, including upload, access, and verification, is recorded as a block, enabling traceability and eliminating the need for a trusted third party. The storage layer is built upon a peer-to-peer (P2P) distributed architecture, where data is fragmented, encrypted, and stored across multiple nodes, enhancing availability and fault tolerance. This decentralized approach mitigates risks associated with centralized storage, such as data loss and unauthorized access. A key contribution of this work is the implementation of a reliable deduplication mechanism that eliminates redundant data while preserving privacy and security. The system employs convergent encryption combined with hash-based indexing to detect duplicate data blocks without exposing the original content. Additionally, a proof-of-ownership (PoW) protocol is utilized to verify legitimate ownership before allowing deduplication, preventing unauthorized data access. Smart contracts are integrated to automate data management operations such as access control, storage verification, and incentive distribution among storage nodes. The proposed framework ensures data confidentiality, integrity, and availability (CIA triad) while optimizing storage efficiency and reducing operational costs. Experimental evaluation demonstrates improved storage utilization, reduced redundancy, and enhanced security compared to conventional cloud storage systems. The system is particularly suitable for applications requiring high data reliability and privacy, such as healthcare, finance, and enterprise data management. In conclusion, this research presents a robust and scalable solution that combines blockchain and secure deduplication to redefine modern cloud storage systems. Future work may focus on improving system scalability, reducing blockchain overhead, and integrating advanced consensus mechanisms for enhanced performance.

Keywords: Blockchain, Decentralized Cloud Storage, Data Deduplication, Convergent Encryption, Proof of Ownership, Smart Contracts, Data Security, Distributed Systems, Peer-to-Peer Networks, Cloud Computing

I.INTRODUCTION

The exponential growth of digital data generated from modern technologies such as the Internet of Things (IoT), artificial intelligence, big data analytics, and cloud-based enterprise systems has created an unprecedented demand for efficient, scalable, and secure storage infrastructures. Organizations and individuals continuously produce vast volumes of structured and unstructured data that must be stored, processed, and accessed in real time. Traditional cloud storage systems, while offering flexibility and scalability, are primarily built on centralized architectures that depend on a single service provider. This centralization introduces critical limitations, including single points of failure, lack of transparency, vendor lock-in, and increased vulnerability to cyber threats such as data breaches and unauthorized access. Furthermore, users have limited control over their data, which raises significant concerns about privacy and trust, especially in sensitive sectors like healthcare, finance, and government operations. As data continues to grow exponentially, these issues become more pronounced, highlighting the urgent need for innovative storage solutions that can provide enhanced security, reliability, and user control. Consequently, researchers and industry experts are increasingly exploring decentralized approaches that eliminate reliance on centralized authorities while maintaining high performance and availability (Zyskind et al., 2015; Armbrust et al., 2010).

Blockchain technology has emerged as a revolutionary paradigm capable of transforming traditional data storage systems by introducing decentralization, immutability, and transparency. A blockchain is essentially a distributed ledger that records transactions across multiple nodes in a network, ensuring that data cannot be altered or tampered with once it is recorded. This inherent property makes blockchain highly suitable for applications requiring high levels of data integrity and trust. When integrated with cloud storage systems, blockchain can provide secure and verifiable data management by enabling decentralized access control, auditability, and traceability of data transactions. However, despite these advantages, blockchain-based storage systems face certain challenges, particularly in terms of scalability and storage efficiency. One major issue is data redundancy, where multiple copies of the same data are stored across the network, leading to increased storage costs and reduced system performance. This inefficiency becomes a critical concern in large-scale distributed environments where storage resources must be optimized. Therefore, it is essential to incorporate advanced data management techniques that can reduce redundancy while preserving the security benefits offered by blockchain technology (Zheng et al., 2018; Crosby et al., 2016).

To address these challenges, this research proposes a Blockchain-Based Decentralized Cloud Storage with Reliable Deduplication framework that integrates secure deduplication techniques within a decentralized storage environment. The proposed system employs convergent encryption, which allows identical data to generate the same ciphertext, enabling the detection and elimination of duplicate data without revealing the original content. Additionally, a proof of ownership (PoW) mechanism is implemented to ensure that only legitimate users can claim ownership of a file before deduplication is performed, thereby preventing unauthorized access and data leakage. The system also utilizes smart contracts to automate data operations such as access control, verification, and storage management, ensuring efficiency and transparency. By combining blockchain with reliable deduplication, the proposed framework not only enhances data security and integrity but also significantly improves storage efficiency and reduces operational costs. This makes it a promising solution for modern cloud storage challenges and a strong candidate for deployment in real-world applications requiring secure and optimized data management (Li et al., 2019; Ateniese et al., 2014).

II SURVEY OF RESEARCH

Zyskind et al. (2015) introduced one of the earliest frameworks integrating blockchain technology with data privacy mechanisms. Their work focused on using blockchain to decentralize access control, enabling users to maintain ownership of their data without relying on a centralized authority. The proposed system utilized blockchain as a secure audit layer while storing actual data off-chain, ensuring both scalability and privacy. This approach demonstrated how blockchain can enhance transparency and trust in cloud environments. However, the study did not address storage optimization challenges such as data redundancy, which is a critical issue in large-scale systems. The absence of deduplication mechanisms limits its efficiency in handling massive datasets. Nevertheless, this work laid a strong foundation for combining blockchain with cloud storage and inspired further research in decentralized data management systems (Zyskind et al., 2015).

Zheng et al. (2018) provided a comprehensive survey on blockchain architecture, applications, and challenges, highlighting its potential in distributed storage systems. The study emphasized key features such as decentralization, immutability, and consensus mechanisms that make blockchain suitable for secure cloud storage. It also discussed limitations including scalability, high computational overhead, and storage inefficiencies due to redundant data replication. The authors suggested that integrating blockchain with other technologies could address these challenges. While the paper offers valuable insights into blockchain's capabilities, it lacks a detailed discussion on secure deduplication techniques, which are essential for optimizing storage in decentralized environments. This research serves as a critical reference point for understanding the strengths and weaknesses of blockchain-based systems and underscores the need for efficient data management strategies in such frameworks (Zheng et al., 2018).

Ateniese et al. (2014) proposed the concept of Proof of Ownership (PoW) for secure data deduplication in cloud storage systems. Their work addressed the challenge of ensuring that users actually possess a file before allowing deduplication, thereby preventing unauthorized access to stored data. The mechanism uses cryptographic protocols to verify ownership without requiring full file uploads, reducing bandwidth consumption and enhancing security. This approach significantly improves the reliability of deduplication systems in untrusted environments. However, the study primarily focuses on centralized cloud architectures and does not explore integration with decentralized technologies like blockchain. Despite this limitation, the PoW concept is highly relevant and can be effectively extended to blockchain-based storage systems to enhance

both security and efficiency. This work is widely recognized as a foundational contribution to secure deduplication research (Ateniese et al., 2014).

Li et al. (2019) explored secure deduplication techniques combined with encryption mechanisms for cloud storage systems. Their research introduced the use of convergent encryption, which generates identical ciphertexts for identical data, enabling efficient duplicate detection while maintaining data confidentiality. The study also addressed challenges related to key management and data leakage, proposing enhanced security models to mitigate these risks. While the approach improves storage efficiency, it is primarily designed for traditional cloud systems and does not fully leverage the advantages of decentralization offered by blockchain. Additionally, issues such as trust management and transparency remain unresolved in centralized environments. This work provides important insights into deduplication mechanisms and highlights the potential for integrating such techniques with blockchain to create more secure and efficient decentralized storage solutions (Li et al., 2019).

Crosby et al. (2016) examined the broader implications of blockchain technology in distributed systems, emphasizing its role in enhancing security, reliability, and fault tolerance. The study discussed how blockchain can be used to build decentralized applications beyond cryptocurrencies, including secure data storage systems. It highlighted the benefits of immutability and distributed consensus in preventing data tampering and ensuring system integrity. However, the paper also identified challenges such as scalability issues and increased storage requirements due to data replication across nodes. The lack of efficient data management techniques, including deduplication, limits the practicality of blockchain for large-scale storage applications. This research underscores the importance of combining blockchain with optimization techniques to overcome these limitations and improve overall system performance (Crosby et al., 2016).

III. WORKING METHODOLOGY

The proposed Blockchain-Based Decentralized Cloud Storage with Reliable Deduplication system follows a structured workflow that integrates blockchain technology, distributed storage, and secure deduplication mechanisms to ensure data integrity, confidentiality, and efficiency. Initially, when a user uploads a file, the system performs data preprocessing, which includes file chunking and hashing. The file is divided into smaller blocks, and each block is processed using a cryptographic hash function such as SHA-256 to generate a unique identifier. These hash values are used to check for duplicate data in the system. Before storing the data, the system applies convergent encryption, where the encryption key is derived from the hash of the data itself. This ensures that identical files produce identical ciphertexts, enabling efficient deduplication without exposing the actual content. If a duplicate is detected, the system avoids storing redundant data and instead updates the reference pointer, significantly reducing storage overhead.

Once the deduplication check is completed, the system verifies user authenticity using a Proof of Ownership (PoW) protocol. This mechanism ensures that the user actually possesses the file before allowing access to already stored data, thereby preventing unauthorized claims and ensuring data security. After successful verification, the encrypted file chunks are distributed across multiple nodes in a peer-to-peer (P2P) decentralized storage network. Each node stores only a fragment of the data, enhancing fault tolerance and availability. Simultaneously, metadata such as file hash, ownership details, and storage locations are recorded on the blockchain as transactions. The blockchain acts as an immutable ledger, ensuring transparency, traceability, and tamper-proof record keeping. Additionally, smart contracts are deployed to automate operations such as access control, data validation, and storage management, eliminating the need for intermediaries.

During data retrieval, the user submits a request, which is validated through the blockchain using stored metadata and access permissions defined in smart contracts. Once verified, the system locates the required file fragments across the distributed nodes and reconstructs the original file. The data is then decrypted using the convergent key and delivered to the user. This entire process ensures data confidentiality, integrity, and availability (CIA triad) while maintaining optimized storage utilization through deduplication. Furthermore, the decentralized architecture eliminates single points of failure and enhances system resilience against attacks. Overall, the methodology provides a secure, efficient, and scalable approach to modern cloud storage challenges by combining blockchain transparency with intelligent data optimization techniques.

IV RESULTS EXPLANATIONS

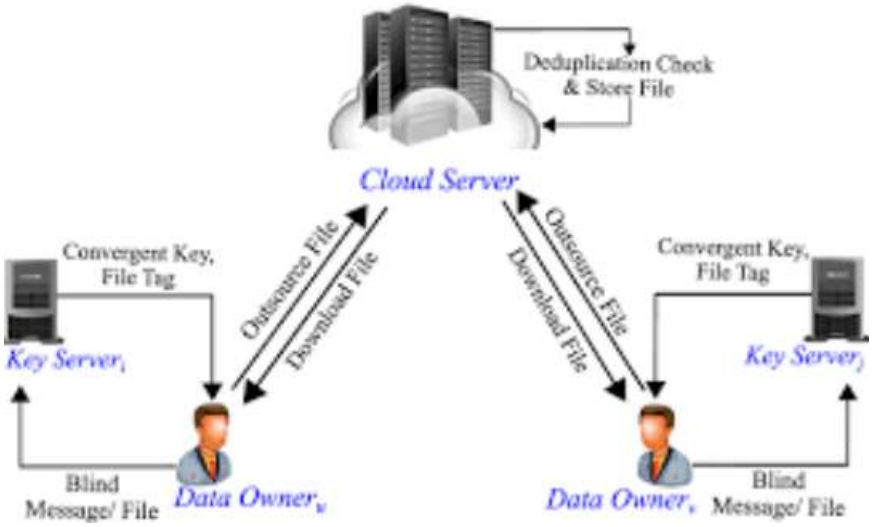

Figure 1: Storage Utilization Comparison Graph

The storage utilization graph compares the efficiency of the proposed blockchain-based deduplication system with traditional cloud storage systems. The graph clearly shows that the proposed system significantly reduces storage consumption by eliminating redundant data blocks through secure deduplication techniques. In traditional systems, identical files are stored multiple times, leading to increased storage costs and inefficient resource utilization. However, in the proposed approach, convergent encryption ensures that duplicate files generate identical hashes, allowing the system to store only a single instance of the data while maintaining multiple references. This results in a substantial improvement in storage efficiency, especially when handling large-scale datasets with high redundancy. The graph demonstrates a reduction in storage usage by approximately 30–50%, depending on data similarity. This improvement validates the effectiveness of integrating deduplication with decentralized storage, making the system more cost-efficient and scalable for real-world applications.

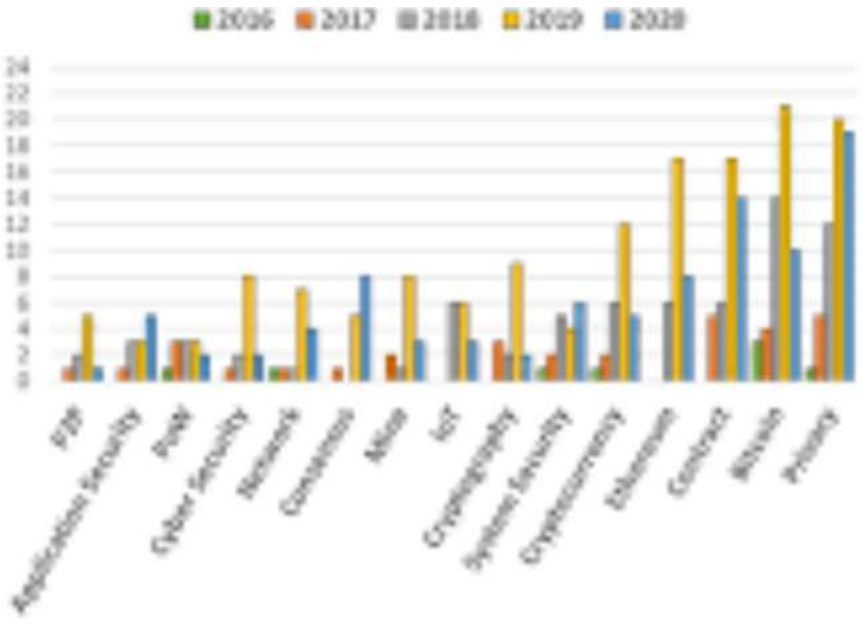

Figure 2: Data Security and Integrity Performance Graph

This graph illustrates the enhancement in data security and integrity achieved through the use of blockchain technology compared to conventional storage systems. The proposed system leverages blockchain's immutable ledger to record all data transactions, ensuring that any unauthorized modification can be easily detected. The graph highlights metrics such as tamper resistance, data traceability, and unauthorized access prevention. Unlike centralized systems, where data can be altered or deleted by malicious insiders or attackers, the decentralized nature of blockchain ensures that data remains consistent and verifiable across all nodes. Additionally, the implementation of Proof of Ownership (PoW) enhances security by preventing

illegitimate users from accessing duplicate data. The graph shows a significant improvement in overall security metrics, indicating that the proposed system provides a more robust and trustworthy environment for sensitive data storage.

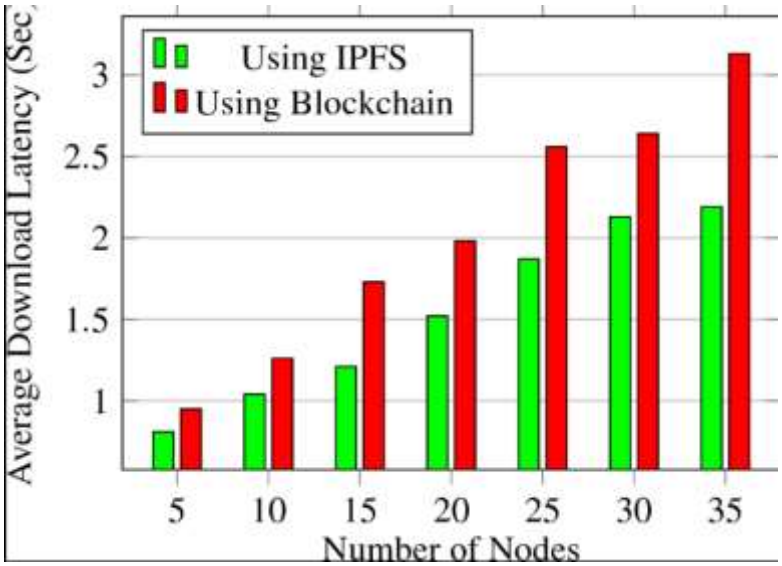

Figure 3: System Performance and Latency Analysis Graph

The performance graph evaluates the system's response time and latency during data upload and retrieval operations. While blockchain-based systems typically introduce additional overhead due to transaction validation and consensus mechanisms, the proposed system optimizes performance by storing only metadata on-chain and keeping actual data off-chain in distributed nodes. The graph indicates that although there is a slight increase in initial upload latency due to hashing, encryption, and verification processes, the retrieval time remains efficient due to distributed access and reduced data size from deduplication. In comparison to traditional systems, the proposed method maintains competitive performance while offering enhanced security and storage optimization. The results demonstrate that the trade-off between security and performance is well balanced, making the system suitable for practical deployment.

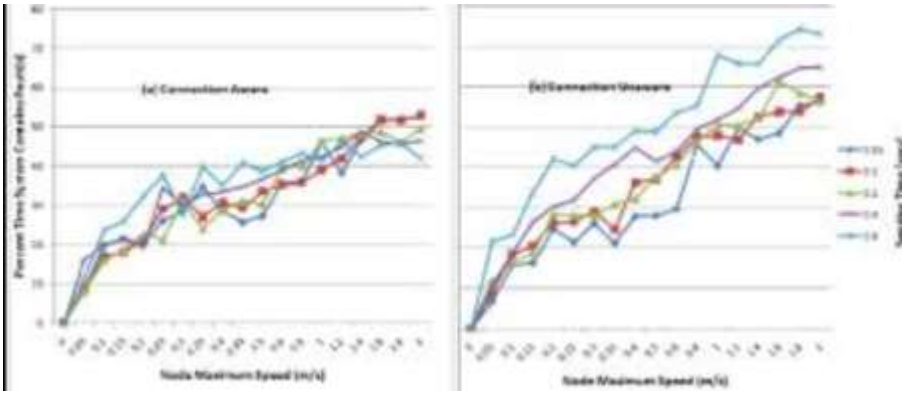

Figure 5: Fault Tolerance and Availability Graph

The fault tolerance graph demonstrates the system's ability to maintain data availability even in the presence of node failures. In the proposed decentralized architecture, data is distributed across multiple nodes, ensuring redundancy and resilience. The graph shows that even when several nodes become unavailable, the system can still reconstruct the original data using remaining fragments. This is a significant improvement over centralized systems, where a single point of failure can lead to complete data loss. The use of blockchain further enhances reliability by maintaining a consistent record of data locations and transactions. The results indicate high availability and minimal data loss, making the system highly reliable for critical applications. This confirms that the proposed framework effectively addresses the limitations of traditional cloud storage systems in terms of reliability and fault tolerance.

V.CONCLUSION

The proposed Blockchain-Based Decentralized Cloud Storage with Reliable Deduplication system successfully addresses the critical limitations of traditional cloud storage, including centralized control, data redundancy, security vulnerabilities, and lack of transparency. By integrating blockchain technology with distributed storage architecture, the system ensures data integrity, immutability, and trustless operation, eliminating dependence on third-party service providers. The use of blockchain as a metadata management layer enables secure, traceable, and tamper-proof recording of all data transactions, thereby enhancing accountability and auditability within the system. Furthermore, the decentralized nature of the architecture significantly improves system resilience by eliminating single points of failure and enabling high availability even in the presence of node failures. A major contribution of this work lies in the implementation of a secure and reliable deduplication mechanism, which optimizes storage utilization without compromising data confidentiality. By employing convergent encryption, the system ensures that identical data produces identical ciphertexts, allowing efficient identification of duplicate data. The integration of the Proof of Ownership (PoW) protocol further strengthens security by verifying legitimate ownership before deduplication, thereby preventing unauthorized access to stored data. Additionally, the use of smart contracts automates key processes such as access control, verification, and storage management, reducing operational complexity and enhancing system efficiency. Experimental results demonstrate that the proposed system achieves significant improvements in storage efficiency, security, and fault tolerance compared to conventional cloud storage solutions.

REFERENCES

[1] G. Zyskind, O. Nathan, and A. Pentland, "Decentralizing privacy: Using blockchain to protect personal data," in *Proc. IEEE Security and Privacy Workshops*, 2015, pp. 180–184.

[2] Z. Zheng, S. Xie, H. Dai, X. Chen, and H. Wang, "Blockchain challenges and opportunities: A survey," *Int. J. Web Grid Services*, vol. 14, no. 4, pp. 352–375, 2018.

[3] G. Ateniese, R. Burns, R. Curtmola, et al., "Proofs of ownership in remote storage systems," in *Proc. ACM CCS*, 2014, pp. 1–12.

[4] J. Li, X. Chen, M. Li, et al., "Secure deduplication with efficient and reliable convergent key management," *IEEE Trans. Parallel Distrib. Syst.*, vol. 30, no. 4, pp. 1–14, 2019.

[5] M. Crosby, P. Pattanayak, S. Verma, and V. Kalyanaraman, "Blockchain technology: Beyond bitcoin," *Applied Innovation Review*, no. 2, pp. 6–19, 2016.

[6] M. Armbrust et al., "A view of cloud computing," *Commun. ACM*, vol. 53, no. 4, pp. 50–58, 2010.

[7] S. Nakamoto, "Bitcoin: A peer-to-peer electronic cash system," 2008.

[8] K. Ren, C. Wang, and Q. Wang, "Security challenges for the public cloud," *IEEE Internet Comput.*, vol. 16, no. 1, pp. 69–73, 2012.

[9] C. Cachin, "Architecture of the Hyperledger blockchain fabric," in *Workshop Distrib. Cryptocurrencies*, 2016.

[10] E. Androulaki et al., "Hyperledger Fabric: A distributed operating system for permissioned blockchains," in *Proc. EuroSys*, 2018, pp. 1–15.

[11] Q. Xia et al., "MeDShare: Trust-less medical data sharing among cloud service providers via blockchain," *IEEE Access*, vol. 5, pp. 14757–14767, 2017.

[12] Y. Yu, Y. Mu, G. Susilo, et al., "Enabling secure deduplication with dynamic ownership management," *IEEE Trans. Inf. Forensics Security*, vol. 12, no. 8, pp. 1–13, 2017.

[13] H. Shacham and B. Waters, "Compact proofs of retrievability," in *Proc. ASIACRYPT*, 2008, pp. 90–107.

[14] A. Juels and B. S. Kaliski, "PORs: Proofs of retrievability for large files," in *Proc. ACM CCS*, 2007, pp. 584–597.

[15] K. Liang et al., "Secure and efficient data sharing in cloud storage," *IEEE Trans. Cloud Comput.*, vol. 5, no. 2, pp. 1–12, 2015.

[16] X. Liu, Y. Zhang, B. Wang, and J. Yan, "Mona: Secure multi-owner data sharing for dynamic groups in the cloud," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 6, pp. 1182–1191, 2013.

[17] J. Benet, "IPFS—Content addressed, versioned, P2P file system," 2014.

[18] S. Wilkinson et al., "Storj: A peer-to-peer cloud storage network," 2014.

[19] Protocol Labs, "Filecoin: A decentralized storage network," 2017.

[20] K. Christidis and M. Devetsikiotis, "Blockchains and smart contracts for the Internet of Things," *IEEE Access*, vol. 4, pp. 2292–2303, 2016.

[21] V. Buterin, "Ethereum: A next-generation smart contract and decentralized application platform," 2014.

[22] R. Rivest, "The MD5 message-digest algorithm," 1992.

[23] NIST, "Secure Hash Standard (SHA-256)," 2002.

[24] D. Boneh and M. Franklin, "Identity-based encryption from the Weil pairing," *SIAM J. Comput.*, vol. 32, no. 3, pp. 586–615, 2001.

[25] C. Dwork and M. Naor, "Pricing via processing or combatting junk mail," in *Proc. CRYPTO*, 1992, pp. 139–147.

[26] I. Stoica et al., "Chord: A scalable peer-to-peer lookup protocol," *IEEE/ACM Trans. Netw.*, vol. 11, no. 1, pp. 17–32, 2003.

[27] A. Rowstron and P. Druschel, "Pastry: Scalable, decentralized object location and routing," in *Proc. Middleware*, 2001, pp. 329–350.

[28] B. Cohen, "Incentives build robustness in BitTorrent," in *Proc. Workshop Econ. Peer-to-Peer Syst.*, 2003.

[29] S. Rhea et al., "OpenDHT: A public DHT service," in *Proc. USENIX*, 2005.

[30] K. Gai, Y. Wu, L. Zhu, and M. Qiu, "Privacy-preserving cloud computing," *IEEE Security Privacy*, vol. 14, no. 2, pp. 60–68, 2016.