Privacy-Preserving Blockchain Systems for Data Sharing

Abhilasha Sharma
B.Tech Student, Department of CSE, Global Institute of Technology, Jaipur, Rajasthan, India
23egjcs006@gitjaipur.com

Aaditya Agrawal
B.Tech Student, Department of CSE, Global Institute of Technology, Jaipur, Rajasthan, India
23egjad001@gitjaipur.com

Dharmveer Jangid
Assistant Professor, Department of CSE, Global Institute of Technology, Jaipur, Rajasthan, India
dharmveer.jangid@gitjaipur.com

ABSTRACT: The exponential growth of data across sectors such as healthcare, the Internet of Things (IoT), and finance has necessitated robust frameworks for secure data sharing. Traditional centralized architectures often suffer from single points of failure, lack of transparency, and privacy vulnerabilities. Blockchain technology, with its decentralized and immutable ledger, offers a promising solution but inherently struggles with the tension between transparency and confidentiality. This paper reviews state-of-the-art privacy-preserving blockchain systems designed for data sharing. It critically analyzes methodologies including Zero-Knowledge Proofs (ZKPs), Homomorphic Encryption (HE), and Proxy Re-Encryption (PRE), alongside architectural patterns like consortium and hybrid chains. We evaluate implementations in real-world scenarios, highlighting performance metrics such as transaction latency and scalability. The review concludes that while cryptographic integration enhances privacy, significant trade-offs in computational overhead and system maturity (Technology Readiness Levels) remain barriers to widespread adoption.

KEYWORDS: Blochchain, Data Security, IoT, ZKPs, Encryption, Data Sharing, Zero-Knowledge Proofs (ZKPs).

1. INTRODUCTION

In the rapidly evolving digital economy, data has emerged as one of the most valuable assets. However, secure and seamless data sharing continues to pose significant challenges. Traditional centralized storage architectures such as cloud-based solutions are vulnerable to unauthorized access, data breaches, and misuse, as users often lose full control over their sensitive information. Additionally, data silos within organizations limit interoperability and restrict the efficient utilization of data across multiple systems or stakeholders [1]-[3].

Blockchain technology provides an innovative foundation for trustless and decentralized data sharing, enabling peer-to-peer interactions without reliance on third-party authorities. Its transparent, immutable, and distributed architecture enhances data integrity and reduces single-point-failure risks. Despite these strengths, public blockchain frameworks face major privacy limitations. Since every transaction recorded on the blockchain is visible to all participants by design, this transparency conflicts with stringent data protection laws and confidentiality requirements, including frameworks such as GDPR (General Data Protection Regulation) and HIPAA (Health Insurance Portability and Accountability Act) [4]-[6].

To address this conflict between auditability and privacy, researchers and industry practitioners have focused on developing privacy-preserving blockchain solutions. These advanced systems aim to ensure that data validity can be verified without exposing raw or sensitive content, ultimately supporting secure, compliant digital interactions [7]-[11].

Privacy-preserving blockchain architectures are typically designed around three essential objectives:

- **Confidentiality:** Ensuring that private data remains hidden from unauthorized participants, even when referenced or validated on a public ledger. Cryptographic techniques such as encryption, secure multi-party computation, and zero-knowledge proofs are widely employed to safeguard data confidentiality.
- **Auditability and Integrity Verification:** Enabling transparent validation of data access, updates, or sharing events without disclosing the underlying information. Immutable records stored on the blockchain allow administrators and regulators to perform traceable compliance checks.
- **User Sovereignty and Access Control:** Empowering users with full ownership of their data and granular control over who can access it, when, and under what conditions. Decentralized identity (DID), tokenization, and smart contracts are core technologies supporting this principle.

This study provides a thorough review of recent advancements in privacy-preserving blockchain mechanisms used for secure data sharing. The survey categorizes cryptographic primitives—such as homomorphic encryption, zk-SNARKs, ring signatures, and attribute-based encryption—and analyzes how they enable data confidentiality, identity protection, and secure validation in blockchain environments. Additionally, the paper explores architectural approaches including off-chain storage, consortium blockchain models, and hybrid blockchain-cloud systems that improve scalability while retaining privacy guarantees. Overall, privacy-enhanced blockchain systems represent a promising pathway toward a secure, interoperable, and user-centric data-sharing ecosystem. These innovations are expected to play a crucial role in future digital infrastructure, particularly within industries that handle highly sensitive information, such as healthcare, finance, governance, and smart cities.

2. CRYPTOGRAPHIC AND ARCHITECTURAL FRAMEWORKS

The development of privacy-preserving data sharing systems relies on a synergy between advanced cryptographic techniques and specific blockchain network topologies [12]-[14].

A. Cryptographic Primitives

- **Zero-Knowledge Proofs (ZKPs):** ZKPs (e.g., zk-SNARKs) allow a prover to demonstrate the validity of a transaction (e.g., "I have the authority to share this file") without revealing the underlying data or the prover's identity. This is crucial for validating permissions without exposing access policies on the public ledger.
- **Homomorphic Encryption (HE):** HE permits computations to be performed directly on encrypted data. In data sharing scenarios, this allows third parties (like AI models) to analyze shared datasets and generate insights without ever decrypting or seeing the raw information.
- **Proxy Re-Encryption (PRE):** PRE is a key management protocol where a semi-trusted proxy transforms ciphertext encrypted for one user (the data owner) into ciphertext decryptable by another (the data requester). This enables secure sharing without the data owner needing to be online to manually encrypt files for every new requester.
- **Attribute-Based Encryption (ABE):** Specifically Ciphertext-Policy ABE (CP-ABE), where data is encrypted with an access policy (e.g., "Doctor AND Cardiology"). Only users whose private keys possess attributes satisfying this policy can decrypt the data, embedding access control directly into the cryptography.

B. System Architectures

- **On-Chain vs. Off-Chain Storage:** Due to the high cost and low privacy of storing large datasets on-chain, most effective systems utilize a hybrid model. Encrypted data payloads are stored in decentralized off-chain storage (e.g., IPFS), while the blockchain stores only the metadata (hashes, access pointers, and audit logs).
- **Consortium Blockchains:** Unlike public chains (Bitcoin/Ethereum), consortium chains (e.g., Hyperledger Fabric) restrict node participation to known entities (e.g., hospitals or banks). They employ "Channels" to create private sub-ledgers between specific parties, isolating sensitive transaction data from the broader network.
- **Trusted Execution Environments (TEEs):** Some systems integrate hardware-based TEEs (like Intel SGX) to perform secure computations off-chain, with the blockchain acting merely as a verification layer for the TEE's integrity proofs.

3. IMPLEMENTATIONS

Privacy-preserving blockchain systems have seen diverse implementations across high-stakes domains.

A. Healthcare and Electronic Health Records (EHR)

Healthcare is the primary driver for these innovations due to strict compliance needs. Systems typically grant patients full sovereignty over their records.

- **MedBlock & MedRec:** Implementations often use a "summary contract" on the blockchain that points to records held in provider databases. Patients grant access via smart contracts that trigger re-encryption keys for researchers or specialists.
- **Genomic Data Sharing:** Specialized platforms use secure multiparty computation (SMPC) combined with blockchain to allow researchers to query genomic databases for matches (e.g., finding a donor) without any single institution revealing its full patient DNA database.

B. Internet of Things (IoT) and Smart Cities

IoT devices generate massive streams of data that require automated, machine-to-machine sharing.

- **Edge Computing Privacy:** Implementations deploy lightweight blockchain nodes on edge gateways. For example, a smart grid system might use ZKPs to prove that a household's energy usage is within a certain tier for billing purposes, without revealing the granular, minute-by-minute usage pattern that could imply user lifestyle habits [1].

- **Industrial IoT (IIoT):** Supply chain implementations use consortium chains where competitors share logistics data (e.g., location, temperature) to ensure product quality. Privacy is maintained by hashing commercially sensitive details (like supplier pricing) while keeping the "state" (e.g., "item arrived") visible.

4. EVALUATION

Evaluating these systems requires analysing the trade-off between privacy guarantees and system performance.

Performance Metrics

Transaction Latency: The addition of privacy layers significantly impacts throughput. Standard Ethereum transactions might take 15-30 seconds, but privacy-preserving transactions involving ZKP generation can take significantly longer (minutes) depending on hardware. However, optimized consortium implementations (e.g., Hyperledger Fabric with private channels) have achieved latencies as low as 15.3 ms for specific transaction types.

Computational Overhead: Techniques like Homomorphic Encryption are computationally intensive. Studies show that fully homomorphic encryption can increase processing time by a factor of 10^3 to 10^6 compared to plaintext operations, making it currently viable only for specific, low-volume use cases.

Storage Efficiency: Hybrid on/off-chain models demonstrate high efficiency. By storing only 32-byte hashes on-chain, systems avoid the "blockchain bloat" problem while maintaining 100% data integrity verifiability.

Table 1: Comparative Analysis of Techniques

Feature	Zero-Knowledge Proofs (ZKP)	Proxy Re-Encryption (PRE)	Trusted Execution (TEE)
Privacy Level	Very High(No data revealed)	High(Data encrypted in transit)	Medium-High(Hardware dependent)
Computation Cost	High (Proof generation is slow)	Low to Medium	Low (Near native speed)
Trust Model	Trustless (Math-based)	Semi-Trusted (Proxy required)	Hardware Trust (Intel/AMD)
Best Use Case	Authentication & Verification	File/Document Sharing	High-speed Data Processing

5. CONCLUSION

Privacy-preserving blockchain systems have matured from theoretical concepts to viable prototypes, particularly in the domains of healthcare and IoT. The integration of off-chain storage with on-chain access control offers a pragmatic solution to the "privacy vs. transparency" paradox. However, significant challenges remain. Scalability is the primary bottleneck; cryptographic proofs (like ZKPs) are computationally heavy, limiting transaction throughput. Usability is another hurdle, as managing private keys and complex access policies remains difficult for average users. Finally, Regulatory uncertainty persists regarding whether immutable blockchain hashes of personal data constitute "personal data" under laws like GDPR (specifically the "Right to be Forgotten").

REFERENCES

[1] G. Sharma, N. Hemrajani, S. Sharma, A. Upadhyay, Y. Bhardwaj, and A. Kumar, “Data management framework for IoT edge-cloud architecture for resource-constrained IoT application,” Journal of Discrete Mathematical Sciences and Cryptography, vol. 25, no. 4, pp. 1093–1103, 2022.

[2] D. Shekhawat and R. Ajmera, “Docker: A review and comparison with virtualization,” International Journal of Scientific Research in Computer Science and Management Studies, vol. 8, no. 1, 2019.

[3] S. Pathak, S. Tiwari, K. Gautam, and J. Joshi, “A review on democratization of machine learning in cloud,” International Journal of Engineering Research and Generic Science, vol. 4, no. 6, pp. 62–67, 2018.

[4] A. Rathour, A. Shahi, A. Tiwari, B. Maurya, and M. Jha, "Decentralized file system (storage and sharing) using blockchain," International Journal of Advance Research and Innovative Ideas in Education, vol. 10, no. 3, pp. 4333–4338, 2024.

[5] K. K. Gautam, S. Prakash, and R. K. Dwivedi, "Patients medical record monitoring using IoT based biometrics blockchain security system," in Proc. Int. Conf. IoT, Communication and Automation Technology (ICICAT), pp. 1–6, 2023.

[6] A. Agarwal, R. Joshi, H. Arora, and R. Kaushik, "Privacy and security of healthcare data in cloud based on the blockchain technology," in Proc. 7th Int. Conf. Computing Methodologies and Communication (ICCMC), pp. 87–92, 2023.

[7] S. Mishra, H. Arora, G. Parakh, and J. Khandelwal, "Contribution of blockchain in development of metaverse," in Proc. 7th Int. Conf. Communication and Electronics Systems (ICCES), pp. 845–850, 2022.

[8] D. Shekhawat and R. Ajmera, "Survey on security implication for the downtime of VM in cloud," in Proc. IEEE 2nd World Conf. Smart Trends in Systems, Security and Sustainability (WorldS4), 2018.

[9] H. Arora, G. K. Soni, R. K. Kushwaha, and P. Prasoon, "Digital image security based on the hybrid model of image hiding and encryption," in Proc. 6th Int. Conf. Communication and Electronics Systems (ICCES), pp. 1153–1157, 2021.

[10] S. P. Chaturvedi, A. Yadav, A. Kumar, and R. Mukherjee, "Unlocking IoT security: Enabling the future with lightweight cryptographic ciphers," in Intelligent Computing Techniques for Smart Energy Systems (ICTSES 2023), Lecture Notes in Electrical Engineering, vol. 1277, pp. 189–199, 2025.

[11] S. Singhal, R. Misra, "A Review on Blockchain and Applications", International Conference on Recent Trends in Engineering & Technology (ICRTET-2023), 2023.

[12] G. K. Soni, A. Rawat, S. Jain, and S. K. Sharma, "A pixel-based digital medical images protection using genetic algorithm with LSB watermark technique," Smart Systems and IoT: Innovations in Computing, Smart Innovation, Systems and Technologies, vol. 141, pp. 483–492, 2020.

[13] H. Kaushik, H. Arora, R. Joshi, K. Sharma, M. Mehra, and P. K. Sharma, "Digital image security using hybrid model of steganography and cryptography," in Proc. Int. Conf. Electronics and Renewable Systems (ICEARS), pp. 1009–1012, 2025.

[14] P. Jha, K. K. Sharma, B. Jain, and V. Sharma, "Digital image encryption using AES algorithm," EIJO Journal of Engineering, Technology and Innovative Research, vol. 4, no. 2, 2019.