A Blockchain-Based Framework for Secure and Tamper-Resistant File Integrity Verification

DrP Veeresh[1], Nagaluti Hemalatha[2], Sagri Varsha[3], Yaganti Pravallika[4], Gonapadu Rokkam Prasad[5], Malekar Akhilesh[6]

[1,2,3,4,5,6]*Dept. Of Computer Science and Engineering, St. Johns College of Engineering and Technology, Yemmiganur, 518360, India*

***Abstract*—With the rapid growth of digital data, ensuring the integrity and authenticity of files has become a critical cybersecurity challenge. Traditional centralized file verification mechanisms are vulnerable to unauthorized modifications, single points of failure, and lack transparency in integrity validation. To address these limitations, this paper proposes a blockchain-based framework for secure and tamper resistant file integrity verification. The proposed system leverages cryptographic hashing techniques, specifically SHA-256, to generate unique digital fingerprints for files. Instead of storing files directly, their corresponding hash values are recorded on a decentralized blockchain ledger, ensuring immutability and resistance to tampering. During the verification process, the hash of a file is recomputed and compared with the blockchain-stored hash to detect any unauthorized modifications. Any mismatch between the hashes indicates file tampering, thereby ensuring reliable integrity validation. In addition, the framework supports comparative file analysis to identify identical or modified versions of files, further enhancing verification accuracy. The system is implemented using Python with a user-friendly graphical interface, making it accessible for practical use by individuals and organizations. By combining decentralization, cryptographic security, and transparency, the proposed framework provides a robust and reliable solution for protecting digital file integrity against tampering and unauthorized alterations.**

Keywords: **Blockchain; Document Verification; Digital Credentials; Smart Contracts; Cryptography; Immutable Ledger; Fraud Prevention.**

I. INTRODUCTION

Credential verification has become increasingly important for governments and employers and educational institutions because the number of fraudulent cases especially involving degrees and certifications has increased. The traditional systems which depend on centralized databases and paper credentials face two major problems because they allow unauthorized access and create extended processing times and result in expensive operational expenses. The distribution of blockchain and distributed ledger technology (DLT) enables users to access secure systems through decentralized networks which maintain permanent records to safeguard their systems against unauthorized changes and fraudulent activities. The blockchain system records information permanently which makes it an appropriate method for protecting confidential credentials.

The system uses smart contracts to process automatic verification which enables users to verify credentials through an instant system without needing any manual work. The system creates trust between parties while decreasing administrative tasks and reducing the chances of fraud. The verification system establishes trustworthy verification services for government documents and professional licenses and medical certifications through educational verification by blockchain technology. Blockchain technology provides a solution to certification and employment and educational assessment because it creates secure systems which maintain full operational transparency while preventing any form of tampering.

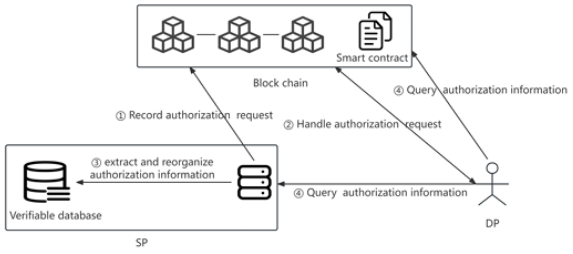

Fig: Overview of blockchain-based personal data authorisation management scheme

People depend on data for various tasks because they treat data as their most important resource according to current information standards. The rising need for data protection results in greater difficulties which organizations must handle to maintain their data security and integrity. The system face serious threats because cyber attackers and unauthorized users can access protected information which results in a loss of user confidence. Data integrity stands as a vital security principle which organizations require because its breach results in loss of trust and hampers their decision-making processes.

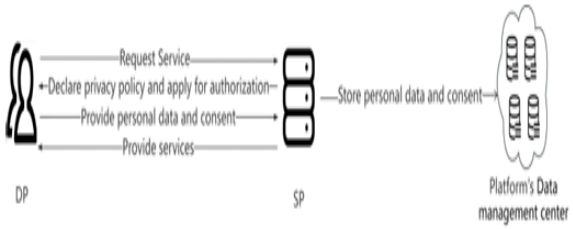

Fig: Authorization and use process of personal data

II. RELATED WORK

Document verification methods have developed from their original manual processes and centralized database systems which required high resource consumption and suffered from both human mistakes and security vulnerabilities to their present advanced systems that use digital signatures and cryptographic methods [1]. The methods establish better integrity protection through their implementation but they depend on trusted authorities which create security risks when those authorities experience security breaches [2]. The adoption of cloud-based systems has enabled organizations to store documents at scale but this practice creates two main problems which are data leaks and centralized authority control [3]. Blockchain technology has developed as a powerful solution because it provides permanent document verification through its decentralized system which prevents document hash alterations until all parties reach agreement [4].

The world now experiences a fundamental change in its approach to handling confidential documents. MIT and other universities use blockchain technology to create digital diplomas which graduates can share directly with others [5]. Estonia and India now use this technology as their main method for national record maintenance which academic institutions previously used. Companies in the private sector develop verification systems that let employers check a candidate's work history within seconds instead of needing weeks [6].

The "DNA" of a blockchain network defines its core strength because it operates as a decentralized system that maintains transparent operations while securing permanent data records [7]. The essential characteristics of blockchain technology provide strong evidence that it should be used to verify important governmental documents. The existing systems require extensive development work because they have not reached their final operational stage which can support worldwide business activities [8].

We need to develop new privacy protection tools which will make these systems suitable for government use in healthcare and education. The combination of InterPlanetary File System (IPFS) storage together with Zero-Knowledge Proofs enables us to verify data while keeping sensitive information completely hidden from view [9].

Researchers are exploring solutions to reduce computational overhead in blockchain technology. The researchers study sidechains and permissioned blockchains and Proof of Stake consensus mechanisms to solve this problem [10]. The blockchain system needs to resolve all accessibility problems together with all performance problems that impact users in developing nations before it can become a widely accepted system for document verification [11]. The combination of blockchain technology with artificial intelligence produces better fraud detection capabilities through its ability to examine metadata for unusual patterns which enables immediate verification. The process of credential exchange depends on the ability of different blockchain systems to work together including Hyperledger Fabric and Ethereum. User trust and acceptance remain significant [12]. The organization needs to establish user benefits together with legal protections while creating easy-to-use platforms and conducting educational programs to achieve wider user acceptance [13]. The future of decentralized ecosystems will combine blockchain technology with edge computing and IoT and digital identity systems to create improved verification systems which will operate more efficiently and depend less on

centralized control in sectors including healthcare and supply chains [14].

III METHODOLOGY

The proposed system aims to develop a web application that ensures file integrity and secures data stored on cloud storage solutions by utilizing blockchain technology and cryptographic techniques such as encryption. The main goal of this system is to offer users a secure method that allows them to confirm data integrity through a system which protects their information from unauthorized access. The research implementation uses immutable blockchain storage together with cryptographic methods and cloud storage to solve the security problems which exist in centralized storage systems through data breaches and inefficient data integrity verification methods which lack transparency.

User Interface and Web Application

The web application uses the Angular frontend framework to provide users with secure registration and login functions through blockchain wallet authentication that uses Ethereum smart contracts for identity verification. The application allows authenticated users to upload files which it processes through local data fragmentation. The system encrypts divided file segments and sends the protected segments to cloud storage while hash calculation occurs for each file chunk which combines to create a Merkle tree that shows the complete file integrity. The system stores this root information along with file metadata in an unchangeable format on a smart contract. The web application protects file content confidentiality while enabling users to verify file integrity through testing. The application supports users in both manual testing and automated testing of file integrity while providing users the choice between two verification methods which include automated verification. The system supports users in establishing regular hash verification procedures which test file integrity at set intervals.

Data Encryption and Integrity Verification

Your data remains private and secure through our system which uses "defense-in-depth" cryptographic methods. The system processes your uploaded file by dividing it into multiple smaller pieces instead of saving it as a single complete unit. We wrap every piece of the file with AES-256 encryption protection. The standard uses a 256-bit key which provides military-level protection to keep your data safe while maintaining its original state.The system gives you complete control of your data because we designed it to let you maintain exclusive data access. You use your blockchain wallet to sign a cryptographic message when you want to start processing a file. The system uses your unique signature to create your encryption key through a Key Derivation Function (KDF) process. The system provides major security advantages through its "just-in-time" method because it eliminates the need to store your encryption key and "salt" information on public blockchain networks.

The system protects your data from any unauthorized changes during the entire storage period. Every file segment undergoes processing through a SHA-256 hashing algorithm which produces hash outputs that we arrange in a Merkle Tree structure. The blockchain smart contract stores only the "Merkle Root" and the file metadata which establishes an unchangeable record of your file's original condition. The Merkle Tree will detect changes if any part of a chunk is modified which allows you to verify that your data remains unchanged and genuine.

Fig: Cryptographic Process

Blockchain Integration

The combination of our blockchain technology with decentralized storage systems we created delivers organizations a complete decentralized solution for their permanent recordkeeping needs. The system stores essential metadata on the Ethereum network which operates as a digital fingerprint to establish file authenticity. The operation uses smart contracts which function as self-running software programs to handle data relationships without requiring human control. The contracts function as a secure storage facility which contains both the processed Merkle Roots and the encrypted URLs for all individual file segments.

The system architecture alerts users to any attempts at unauthorized modification because it detects all changes that occur within the system. The original data

all changes to the file will produce a new Merkle Root which will differ from the one that exists in the on-chain storage system. The developers have executed the logic through a Solidity-based smart contract which they launched on the ETH Sepolia test network. The contract manages user registration connections to particular wallet addresses together with file metadata storage and real-time integrity checking functions. The system provides users complete data monitoring control through its customizable hash verification settings which verify data integrity at specified intervals.

Cloud Storage and Chunk Management

The system uses Amazon S3 service to store the Encrypted chunks of data which are stored in the cloud. I selected Amazon s3 because it's highly reliability and scalable and manages data in a manner that provides durability, high availability, minimal latency, and enhanced durability (M et al., 2022). The system processes user uploaded files by splitting them into smaller chunks which compute Merkle root and enable parallel data processing. S3 provides the uploaded chunks URLs after we encrypt all chunks and upload them to S3 and we store these URLs in smart contract for integrity verification purposes.
Users retrieve their encrypted chunks from the cloud when they want to check their file integrity or access their file. The system uses the users unique key to decrypt the encrypted chunks which are combined to reconstruct the original file uploaded by the user.

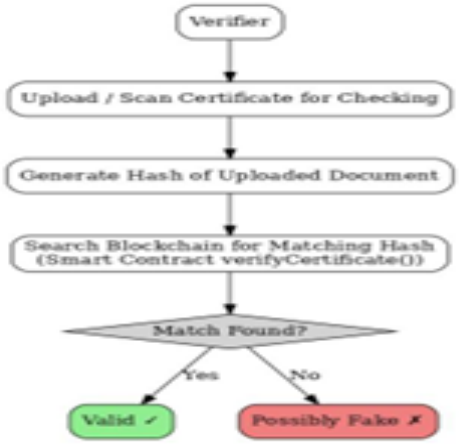

Fig: Deployment and Maintenance

Periodic Integrity Verification

The system uses periodic integrity verification to maintain automated monitoring of file integrity which operates through blockchain technology and various cloud computing solutions, while users can control the system through their ability to turn periodic integrity checks for their files on or off. The smart contract stores their preferences, while the Lambda function at scheduled times uses the SHA-256 hash algorithm to calculate their Merkle root from uploaded chunks and verifies it against the metadata stored in the smart contract. The web application performs integrity verification through manual checks, but the system handles all tasks automatically, which allows users to monitor their files for damage or contamination at all times without needing to touch the system.

IV EXISTING SYSTEM

The present systems for file storage and verification management function through their centralized system, which designates a single organization to handle all data protection measures and operational records. The existing systems require users to give complete trust to the provider's system, which they use to protect their data from unauthorized access and data loss. The platforms use standard encryption methods, but their implementation relies on server-side key management practices. This security practice exposes sensitive information because all system encryption keys become vulnerable to unauthorized access when its main authority is compromised. The systems become vulnerable to hacking because they lack a permanent independent record, which allows high-level administrators and advanced attackers to change protected files while they erasing all proof of the security breach.

The existing system creates serious security problems because it establishes a critical vulnerability point that enables complete data loss and service interruption through hardware failures and DDoS attacks. The internal verification processes for file integrity work as a "black box" system that prevents users from understanding how their files are checked. The user lacks any means to independently verify the system's file authentication claim because the centralized system provides only the provider's word for confirmation. The proposed blockchain framework from your research project establishes a verification system that enables users to validate their work

through its permanent and transparent decentralized ledger system.

V PROPOSED SYSTEM

The proposed approach utilizes blockchain technology to create a secure and transparent document verification platform. Unlike traditional systems, which are often centralized or manually verified, this decentralized ledger maintains the legitimacy and trust of the documents. Documents are not directly stored on the blockchain; instead, they are represented as cryptographic hash values, serving as unique digital fingerprints. Any modification to a document alters its hash, signalingpotential tampering.

The verification system consists of five key modules:

1.UserModule: Allows individuals to upload documents for validation, generating a unique hash stored on the blockchain.

2.AuthorityModule: Official entities create and verify credentials using electronic signatures before submitting the documents to the blockchain.

3.Verification Module: Users can verify documents by uploading them to the platform for hash comparison with the blockchain record.

4.Blockchain Layer: Ensures decentralized, immutable record-keeping for document transactions.

5. Smart Contracts: Automates the issuance and verification processes, ensuring only authorized parties can register documents.

VI SYSTEM ARCHITECTURE

The system architecture proposed for secure file storage and retrieval consists of three main components: a Web Application, Blockchain (Smart Contract), and Cloud Storage (AWS S3). The Web Application which uses Angular framework enables users to log in through MetaMask wallet and allows them to upload files after the system processes their data by dividing it into chunks and applying AES 256 encryption and creating a Merkle tree which will be uploaded to AWS S3. Users can download their files through a secure process and verify file integrity by comparing the recalculated Merkle root with the stored value. The Blockchain component provides decentralized storage which keeps unchangeable records together with user authentication data and integrity verification information. AWS S3 serves as the cloud storage for the encrypted file chunks which provides users with secure storage and retrieval capabilities when they need to download files or conduct integrity checks.

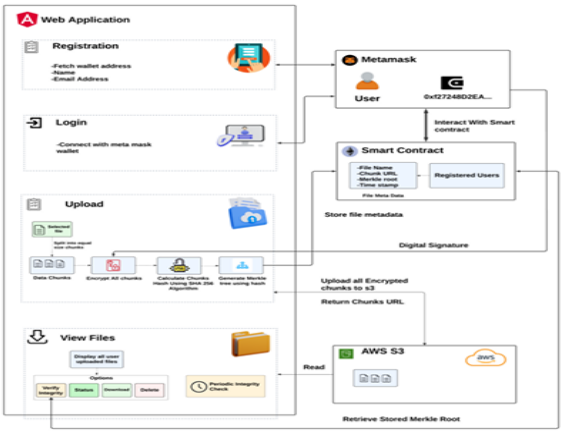

Fig: System Architecture

VII RESULTS

The document verification system developed through blockchain technology achieves two main objectives. It establishes document authenticity through rapid transparent methods while preventing document fraud. The system can handle multiple document types which include educational qualifications and institutional records through cryptographic hashing that stores data on the blockchain. The system reduces verification times from multiple hours or days to only seconds which improves hiring and admissions processes. The system builds user trust through its decentralized structure which prevents any single point of failure while creating an audit trail that shows document validation processes. The design of the user-friendly interface system provides accessibility to all users because it allows easier operation for users who lack technical knowledge. Users who require operational control can access the system through role-based access control which prevents them from making unauthorized changes. The traditional methods show better results through this system which needs further research to solve its existing scalability issues and network dependency problems. The document verification method serves as a secure framework which delivers effective results through its identified strengths and limitations.

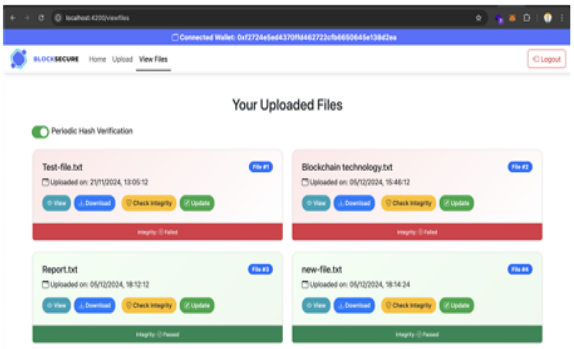

Fig: Page to view uploaded file integrity

VIII DISCUSION

The study results demonstrate that cloud storage systems achieve data protection and system integrity through the implementation of blockchain technology and cryptographic methods and cloud storage which provides various security solutions. The main experiments used Merkle Tree based integrity verification to detect data tampering by recalculating Merkle roots and comparing them with stored Merkle roots to detect tampering through the unchangeable nature of blockchain data.The second experiment tested wallet-based encryption and decryption capabilities through AES-256 encryption which confirmed results from earlier research. The decryption process took less time than the decryption process because encryption required more computational resources as file sizes increased which created possible delays for data processing tasks. The third experiment used continuous integrity monitoring to detect security breaches through real-time monitoring which enhanced user trust and system performance yet the system required scheduled data checks which resulted in potential delays.

The research identified multiple limitations which included the high transaction costs of Ethereum that restricted system growth and the dependency on AWS S3 for centralized storage which conflicted with decentralized storage systems and the performance degradation of large file operations. Organizations should explore various blockchain solutions which will enable them to decrease costs while their systems will enhance capacity through the proposed improvements. The system requires decentralized storage methods which will protect against centralization risks through real-time integrity checks with AWS Event Bridge integration. The user experience improves through the system use.

IX CONCLUSION

The research examines ways to connect user-friendly cloud storage solutions with the security requirements needed to combat persistent data breaching issues. Our team built a data protection framework through blockchain technology which we applied to centralized systems to create a solution that protects essential data content. Our system uses AES-256 encryption together with SHA-256 hashing to build a digital defense which keeps data secret while preventing unauthorized access.

The main result of our research introduced a fresh approach to establishing trustworthiness. Our system uses Merkle Trees to validate data authenticity while it stores Merkle Roots on the Ethereum blockchain to create a system which provides continuous, unchangeable inspection records. Users now have the ability to verify their "truth" stored files without depending on any outside service provider for authentication.

The research shows positive outcomes yet the study discovered multiple actual challenges which must be resolved before full implementation. The main Ethereum network requires high transaction fees which create major obstacles that prevent daily operations from continuing. Our team intends to test various blockchain systems which charge lower transaction fees while we create quicker methods to divide and secure large data volumes. Our objective requires us to achieve advanced security protection which operates at the same efficiency and cost as our current cloud storage solutions.

REFERENCES

[1] Hao, L.; Min, Z.; Dengguo, F.; Zhu, H. Research on Access Control of Big Data. Chin. J. Comput. 2017, 40, 72–91.

[2] Papageorgiou, A.; Strigkos, M.; Politou, E.; Alepis, E.; Solanas, A.; Patsakis, C. Security and privacy analysis of mobile health applications: The alarming state of practice. IEEE Access 2018, 6, 9390–9403.

[3] Nguyen, T.T.; Backes, M.; Marnau, N.; Stock, B. Share First, Ask Later (or Never?) Studying

Violations of GDPR's Explicit Consent in Android Apps. In Proceedings of the 30th USENIX Security Symposium (USENIX Security 21), Vancouver, BC, Canada, 11–13 August 2021; pp. 3667–3684.

[4] Neisse, R.; Steri, G.; Nai-Fovino, I. A blockchain-based approach for data accountability and provenance tracking. In Proceedings of the 12th International Conference on Availability, Reliability and Security, Reggio Calabria, Italy, 29 August–1 September 2017; pp. 1–10.

[5] Calani, M.; Denaro, G.; Leporati, A. Exploiting the Blockchain to Guarantee GDPR Compliance while Consents Evolve under Data Owners' Control. In Proceedings of the ITASEC, Online, 7–9 April 2021; pp. 331–343.

[6] Mahindrakar, A.; Joshi, K.P. Automating GDPR compliance using policy integrated blockchain. In Proceedings of the 2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), Baltimore, MD, USA, 25–27 May 2020; pp. 86–93.

[7] Garay, J.; Kiayias, A.; Leonardos, N. The bitcoin backbone protocol: Analysis and applications. In Proceedings of the Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, 26–30 April 2015; pp. 281–310.

[8] Peregrina-Pérez, M.J.; Lagares-Galán, J.; Boubeta-Puig, J. Hyperledger Fabric blockchain platform. In Distributed Computing to Blockchain; Elsevier: Amsterdam, The Netherlands, 2023; pp. 283–295.

[9] Peng,Z.; Wu,H.; Xiao, B.; Guo, S. VQL: Providing query efficiency and data authenticity in blockchain systems. In Proceedings of the 2019 IEEE 35th International Conference on Data Engineering Workshops (ICDEW), Macao, China, 8–12 April 2019; pp. 1–6.

[10] Peregrina-Pérez, M.J.; Lagares-Galán, J.; Boubeta-Puig, J. Hyperledger Fabric blockchain platform. In Distributed Computing to Blockchain; Elsevier: Amsterdam, The Netherlands, 2023; pp. 283–295.