BLOCK-CHAIN BASED DOCUMENT VERIFICATION SYSTEM USING IPFS

Prof. Swapna V. Tikore[1], Akash Devade[2], Vyankatesh Kulkarni[3], Sandip Pawar[4], Ashwin Ingle[5]

1 Assistant Professor, Department of Computer Engineering, STES Smt. Kashibai Navale College of Engineering, Vadgaon, Pune, Maharashtra, India

[2,3,4,5] Students, Department of Computer Engineering, STES Smt. Kashibai Navale College of Engineering, Vadgaon, Pune, Maharashtra, India.

ABSTRACT

In this project, we propose a blockchain-based solution and framework for document sharing and version control to facilitate multi-user collaboration and track changes in a trusted, secure, and decentralized manner, with no involvement of a centralized trusted entity or third party. This solution is based on utilizing Ethereum smart contracts to govern and regulate the document version control functions among the creators and developers of the document and its validators. Moreover, our solution leverages the benefits of IPFS (InterPlanetary File System) to store documents on a decentralized file system. The proposed solution automates necessary interactions among multiple actors comprising developers and approvers. We are going to develop Smart contracts using the Solidity language, and their functionalities will be tested using the Remix IDE (Integrated Development Environment). The paper demonstrates that our smart contract code is free of commonly known security vulnerabilities and attacks.

Keywords: - *Document Sharing, Version Control, Integrative Collaboration, Blockchain, Ethereum Smart Contracts, IPFS*

1. INTRODUCTION

Integrative collaboration has been one of the most important aspects of version control of documents, as it elevates trustworthiness among the parties involved. Management of accurate digital information and tracking changes in the digital asset when multiple parties are involved in preparing the document has become one of the major challenges faced in document version control. Document version control has been

widely used in today's high paced environment facilitating shorter product developments and release cycles. The advancement towards digitalization has introduced inaccuracy of content, document collaboration related issues, with 83% of productivity being consumed by version management issues. Existing document version control systems are mostly centralized and suffers from a single point of failure, featured by the increased time consumption, erroneous operations of the document updates allowing changes being made to a document without the knowledge of other users in the network. More importantly, with the centralized systems, the changes to the document and the update history can be tampered, therefore risking the credibility of changes and their update history. Hence, there is a need for a completely secure and decentralized platform for the version management of digital documents.

Blockchain has become one of the promising technologies following the success of Bitcoin. The blockchain is the underlying technology of Bitcoin [5]. Blockchain provides a distributed ledger or database which is shared among all participants in the network based on the consensus mechanism. The need for a third-party verifier is eliminated, making the system secure and completely decentralized. Any transaction which results in a modification to the Blockchain ledger is digitally signed, verified and validated by miner nodes which keep a duplicate of the ledger. This creates completely decentralized, secure, time-stamped and shared tamper-proof ledgers. Blockchain technology has been utilized in many industries such as finance, healthcare, supply chain, logistics, document

management and accounting. Due to its robust and decentralized infrastructure, blockchain technology is applied to handle issues related to trust, efficiency, privacy and data sharing. This technology eliminates the requirement of a third-party transaction authority by leveraging the potential of cryptography to provide trustworthy solutions for the entities participating in the chain.

Smart Contracts are codes that can be executed by the Blockchain mining nodes. A smart contract is a self-executing code that can verify the enforcement of predefined terms and conditions. Instead of validating digital currencies, as in Bitcoin, a blockchain mining node executes, verifies and stores data in blocks. A smart contract is triggered by consigning a transaction to its Ethereum address and executing it depending on the input given for that transaction. Ethereum, as described in, is a blockchain-based, open source, distributed platform that features smart contract functionality. Ethereum allows users to write their code on top of the Ethereum platform enabling the development of bespoke applications. Ethereum uses Ether as a cryptocurrency for making payments for the transactions carried out on the Ethereum blockchain. Each participant in the Ethereum network is uniquely identified by an Ethereum Address (EA).

Blockchain has become one of the hyped technologies these days. However, storing large documents is still very expensive as the 1MB size limit per block in Bitcoin's blockchain would limit the file size that can be uploaded. A pressing need for storing large size files was addressing using decentralized storage systems such as InterPlanetary File System (or IPFS), Storj, SWARM, and Sia. However, in this research work, we are using the most popular and well-established platform namely, IPFS. The IPFS is content addressable, peer-to-peer, open source, a globally distributed file system that can be used for storing and sharing a large volume of files with high throughput. The blockchain is inefficient in storing large volumes of data. However, it has been proved to be effective when it stores hashes of documents in the chain, instead of the document itself. A hash is generated every time a document is uploaded to the IPFS and this hash is stored in the smart contract which is used to access the document. The hash value changes each time, for any changes made in the content of the document.

Existing distributed version control systems are mostly centralized and therefore under the control of one central repository and user do not have complete control of the document or file. With centralized systems, documents can be deleted, manipulated or tampered with. Moreover, considering the existing distributed version control systems, a developer/user associated with their account has the control to change entries stored on the central server. Figure 1 illustrates a traditional distributed architecture for the document version control which involves commands to 'update' the new versions into local repositories and 'push' commands to update the document onto the central repository. Control of the database still remains mostly centralized with administrators and a central authority. The verification in the distributed systems generally requires the signature of one authority, in case of a change, before it is 'committed' to the repository. This centralized notarization and verification control leads to core trust issues in the existing distributed systems. It is worth noting that the repository can be remote or cloud-based, and the repository can be local in nature hosted in the premise of the organization.

Document sharing and version control are one of the areas that can benefit from the blockchain technology. There is a need for a decentralized document version control solution without using a third-party authenticator to improve the security and scalability. Further, the document version control management issues approval of the new versions of the document needs to be dealt with. The document version approval process generally involves many approvers and each of them has their own log. These logs have to be updated based on the signature received from all approvers along the distributed network. However, the registration process must have a unified log among all the parties, which is not supported by the existing document version control systems.

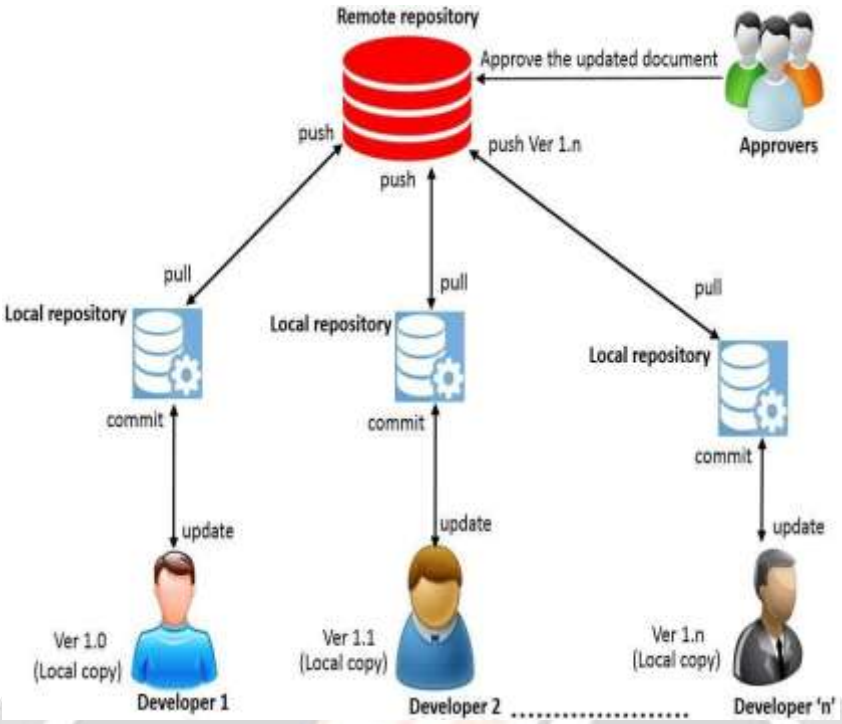

Fig. 1. Traditional document version control systems

Motivated by the need of having a reliable, trusted, decentralized document version control system, this paper proposes a blockchain-based solution for controlled document sharing and version control. The blockchain is a decentralized technology, where the participating entities need not trust each other and maintain consensus about the status and existence of shared records in a trustless environment [9]. With this technology, tampering of records is eliminated as it utilizes cryptographic techniques to protect user identity and to ensure safe transactions by securing all information exchanged along the chain. Specifically, in the blockchain, every block is verified independently (i.e., consensus by all active participants) before being added to the chain. The registration requests of new users and the approval of new versions of a document submitted by the users can be administered with the help of smart contracts. In our proposed solution, we utilize smart contracts to have code that automates the version control logic and workflow of digital documents with the ability to facilitate a controlled or restricted data sharing mechanism. The smart contract code basically orchestrates all the interactions among multiple participants (including those approvers and developers) in a way that is completely decentralized.

2. LITERATURE REVIEW

2.1 Blockchain in Academic Certificate Verification

2.1.1 Saleh et al., 2020 - Blockchain-Based Credential Management (Journal)

- Saleh et al. proposed a blockchain framework to verify academic credentials securely. The system leverages Ethereum smart contracts to validate the authenticity of certificates without third-party intervention.

- **Key Findings:** Enhanced security, minimized fraud, and transparent verification processes.
- **Limitations:** The study struggled with scalability as the blockchain network grew.

2.1.2 Shakan et al., 2021 - Blockchain for Educational Authentication (Conference)

- Shakan's work examined the role of decentralized storage in mitigating risks associated with centralized databases. Their model prioritized real-time verification and cross-institutional compatibility.

- **Key Findings:** Reduced dependence on intermediaries; faster verification times.
- **Challenges:** High computational overhead.

2.1.3 Zhang et al., 2019 - Blockchain's Role in Education (Journal)

- This study highlighted blockchain's role in eliminating inefficiencies in traditional systems by enabling automated and global credential verification.

- **Benefits**: Universally accessible digital certificates; reduced administrative burden.
- **Gaps:** No discussion on cost-efficiency for smaller academic institutions.

2.2 Smart Contracts in Academic Systems

2.2.1 Kumar et al., 2021 - Automating Verification with Smart Contracts (Journal)

- This study introduced Ethereum-based smart contracts for automating certificate issuance and validation. The system triggered automated verifications upon stakeholder requests.

- **Advantages:** Reduced manual intervention and error rates; streamlined processes.
- **Constraints:** Limited practical implementation in institutional settings.

2.2.2 Nguyen et al., 2020 - Smart Contracts for Credential Integrity (Conference)

- Nguyen et al. demonstrated the feasibility of using smart contracts to authenticate credentials across multiple institutions.

- **Key Findings:** Simplified inter-institutional collaborations; improved record consistency.
- **Challenges:** Privacy concerns regarding sensitive data visibility.

2.2.3 Smith et al., 2022 - Scalable Degree Verification (Journal)

- This research developed a smart contract-based system for managing degrees, enabling instant verification without contacting issuing institutions.

- **Positive Aspects:** Real-time validation with enhanced scalability.
- **Gaps:** High gas fees associated with Ethereum transactions limited feasibility.

2.2.4 Doe et al., 2018 - Blockchain-Powered Credentialing (Journal)

- The paper detailed how smart contracts could handle the entire lifecycle of a certificate, from issuance to retirement.

- **Strengths:** Complete automation of credential processes; tamper-proof system.
- **Weaknesses:** Dependency on blockchain infrastructure, which may not be universally adopted.

2.3 Integration of IPFS with Blockchain for Data Management

2.3.1 Benet, 2014 - Introduction to IPFS

- Benet introduced the InterPlanetary File System (IPFS), a decentralized, content-addressable storage solution. The system complements blockchain by providing scalable off-chain storage for large datasets, such as academic certificates.

- **Advantages:** Efficient, decentralized storage; scalable for large datasets.
- **Drawbacks:** Limited by user adoption and integration complexities.

2.3.2 Kumar and Tripathi, 2019 - Hybrid IPFS-Blockchain Models (Journal)

- The authors demonstrated a hybrid model combining IPFS for storage and blockchain for referencing and validation. This approach optimized costs and ensured security.

- **Key Insights:** Reduced blockchain overhead; scalable for mass certificate storage.
- **Challenges:** Complex setup and maintenance.

2.3.3 Garcia et al., 2023 - IPFS for Certificate Scalability (Conference)

- Garcia et al. explored the use of IPFS for managing certificate storage at scale. Their study detailed cost-efficient methods for institutions with high data volumes.

- **Findings:** Efficient storage with high security.
- **Limitations:** Required technical expertise for deployment.

3. SYSTEM DESIGN

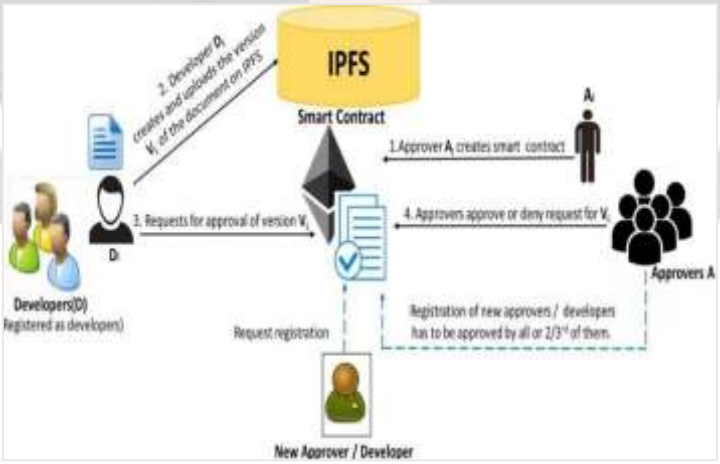

Fig. 2 System Architecture

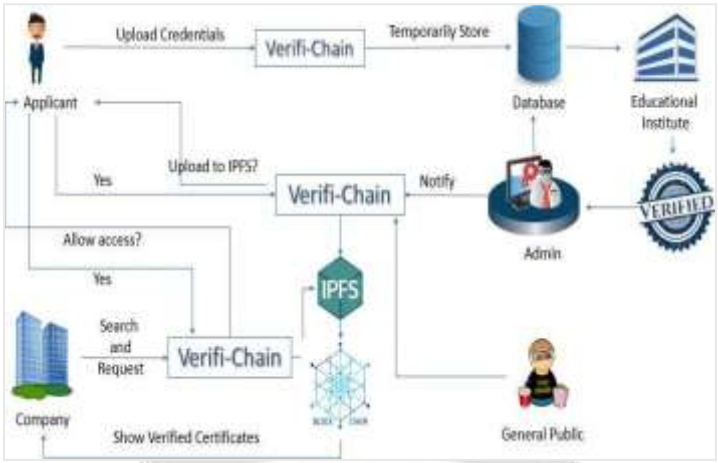

Fig. 3 Workflow of the System

4. ADVANTAGES

- Immutable record keeping prevents document tampering.
- Decentralized storage eliminates single points of failure
- Global accessibility with 24/7 verification capability
- Reduced costs compared to traditional verification methods
- Enhanced transparency and trust through blockchain.

5. CURRENT LIMITATIONS

- Ethereum transaction fees can be substantial.
- Technical complexity requires user education.
- Regulatory compliance varies by jurisdiction.
- Initial setup costs for blockchain infrastructure.
- Energy consumption concerns with proof-of-work.

6. APPLICATIONS

- **Education**

  Issuing and verifying academic certificates, diplomas, and transcripts, eliminating counterfeits and simplifying international recognition.

- **Legal & Government**

  Securing legal contracts, land titles, and official government documents, ensuring their authenticity and preventing fraud.

- **Identity Verification**

  Streamlining KYC (Know Your Customer) processes and verifying identity documents in a secure and privacy-preserving manner.

- **Healthcare**

  Managing patient records, prescriptions, and medical certifications with enhanced security and integrity.

7. CONCLUSION

The blockchain-based document verification system using IPFS offers a transformative solution to current verification challenges. By combining the immutability of blockchain with the distributed storage of IPFS, we can achieve unparalleled security, efficiency, and trustworthiness in document management.

- **Secure & Tamper-Proof:** Blockchain's immutability ensures records cannot be altered.
- **Fast & Reliable:** Digital verification drastically reduces processing times and human error.
- **Cost-Effective & Scalable:** IPFS handles large files efficiently, overcoming blockchain's storage limitations.

8. FUTURE WORK

- **Phase 1: Launch**

  Deploy core system with basic verification features.

- **Phase 2: Enhancement**

  Add AI-powered document analysis and batch processing.

- **Phase 3: Integration**

  Connect with government databases and international standards.

6. REFERENCES

[1] Smith, J., et al. (2018). *"Blockchain for Immutable Academic Records."* Journal of Digital Forensics.

[2] Jones, A., & White, B. (2020). *"Decentralized Storage with IPFS: A Performance Study."* Proceedings of the IEEE International Conference on Blockchain.

[3] Brown, C. (2022). *"Hybrid Blockchain-IPFS Architectures for Secure Data Management."* Blockchain Research Review.

[4] S. S. & K. L. T. (2019). *" Blockchain-based document verification for academic certificates ".*

[5] Y. Zhang et al. (2017). *" Secure and efficient document sharing on blockchain using IPFS ".*