Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2024) 1–12

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering (ICMLDE 2023)
19-20 December 2024, Bhopal, India

Data Provenance in Cloud Environment using Blockchain with a Reputation-based Consensus Algorithm

Neha Chandavari[1], Rutuja D Chikkorde[a], Nirmala Raju Kanti[1], Prajwal Mutnal[1], Narayan D G[1], Maheshgouda Patil[1]

[a]*School of Computer Science and Engineering, KLE Technological University, Hubli-580031, India*

Abstract

Cloud computing has transformed digital infrastructure with its flexible, cost-effective, and scalable data access and storage solutions. However, this shift has also intensified concerns about data integrity and privacy in dynamic, distributed environments. This study addresses these challenges by integrating blockchain and InterPlanetary File System (IPFS) technologies into a Dropbox-like application for ownCloud. IPFS efficiently manages metadata storage, while blockchain ensures data provenance by storing metadata hashes immutably. This integration establishes a secure, decentralized method for managing data transactions in cloud environments, crucial for building trust and reliability. Additionally, the study introduces a novel Reputation-based Proof of Stake (rPOS) algorithm to enhance fairness and efficiency in blockchain networks. Unlike traditional Proof of Stake (PoS), rPOS integrates node reputation into miner selection, promoting equitable participation and enhancing network reliability. Experimental results validate rPOS's effectiveness, achieving an average block creation time of approximately 15 seconds with 40 nodes, compared to around 17 seconds with PoS. These results indicate that rPOS maintains lower transaction and block creation times even as the network scales, demonstrating superior scalability and efficiency. Integrating blockchain and rPOS not only strengthens data security and integrity but also enhances network performance and fairness.

Keywords: Blockchain, IPFS, Proof of Stake, rPOS (Reputation-based Proof of Stake), Fairness

1. Introduction

Cloud computing has become a fundamental technology in the digital transformation era, allowing businesses and individuals to access and store data and applications over the internet rather than on local servers or personal devices[1]. This paradigm shift offers numerous advantages, including unparalleled scalability, cost efficiency, and flexibility. Organizations can dynamically adjust their IT resources to meet fluctuating demands, significantly reduce capital expenditures, and benefit from the rapid deployment of applications and services. However, this shift also introduces significant security challenges that require robust and innovative solutions to ensure the protection of data, applications, and services from various threats and vulnerabilities within cloud environments.

Several studies have explored diverse applications of blockchain technology, each emphasizing data provenance and utilizing specific blockchain solutions. G. Haveri and B. Deewan (2023) compared Ethereum's

Proof of Work (PoW) and Proof of Authority (PoA) within OpenStack SWIFT, analyzing data such as RecordID, UserName, DateTime, FileName, and Action to conclude PoA's superior scalability in varying network and hardware conditions[2]. In IoT contexts, Ethereum Rinkeby and Ropsten networks were tested for token and input provenance, crucial for applications like vaccine supply chains, highlighting blockchain's role in ensuring data integrity amidst challenges of latency, throughput, and gas prices [3]. Almamun et al. proposed SciChain for scientific applications on HPC systems, leveraging POST and POET consensus mechanisms to manage provenance data encompassing processing time, speedup, and performance metrics, demonstrating scalability improvements over traditional ledgers [4]. Celik et al. focused on surveillance data provenance in collaborative workflows using Ethereum's Kovan network, enhancing transparency and accountability in BIM data management processes [5]. Other studies explored metadata validation in cloud environments using Ethereum, showcasing blockchain's ability to secure file operations and improve scalability across platforms like ownCloud and AWS S3, thereby advancing data management practices in distributed computing infrastructures [6, 7, 8]. These efforts collectively underscore blockchain's versatility in enhancing data provenance, scalability, and security across varied computing domains.

In the context of enhancing blockchain consensus mechanisms, several studies have explored the integration of reputational factors to improve reliability, fairness, and scalability. For instance, [9]explores enhancing Delegated Proof of Stake (DPoS) by introducing reputational factors, utilizing metrics such as trusted state value, number of votes, rank, reward, and transaction fee reward. [10] investigates augmenting Proof-of-Work (PoW) with reputation mechanisms, employing statistical measures like the arithmetic mean and standard deviation of miner reputation histories to enhance consensus reliability. [11] introduces Practical Byzantine Fault Tolerance (PBFT) and Cumulative Reputation (CuR), aiming to measure node contribution over time to ensure equitable participation in consensus groups. [12] introduces a comprehensive reputation evaluation framework (Retu) integrating historical, consensus, and transactional reputations with weighted factors across various blockchain protocols (PoS, PoW, PoA), emphasizing continuous node engagement and fairness.[13] focuses on accumulating and evaluating historical reputation records (PoR) to bolster transparency and integrity in blockchain operations. Lastly, [14]addresses security concerns in Bitcoin mining pools by defining miner lifetime and attack probabilities (PPoR) to mitigate risks associated with malicious activities. Together, these studies underscore the evolving strategies and methodologies aimed at improving blockchain technologies within cloud computing contexts.

Ensuring cloud security is critical as data is frequently distributed across multiple locations and managed by third-party providers. Data provenance, which tracks data origins, movements, and changes, is essential for maintaining integrity, accountability, and compliance. Cloud infrastructures' distributed nature complicates data tracking, necessitating innovative solutions. InterPlanetary File System (IPFS) and blockchain technologies offer solutions: IPFS decentralizes data storage, and blockchain ensures immutability. Our research integrates these into a Dropbox-like app for ownCloud. Metadata in IPFS enables efficient, secure retrieval, with the hash on blockchain ensuring immutability for enhanced data integrity. Traditional Proof of Stake (PoS) algorithms often favor high-stake nodes, posing fairness challenges. Our Reputation-based PoS (rPOS) integrates reputation into PoS, promoting fairness and system security.

In our research, we developed a Dropbox-like application for ownCloud, a popular cloud storage platform, enabling file operations like uploading, deleting, and downloading. Our project's innovation lies in integrating blockchain and IPFS technologies to enhance data provenance and security. Metadata for file operations is stored in IPFS, ensuring efficient retrieval and security, with the IPFS hash recorded on the blockchain for immutability. This method safeguards data integrity by detecting any attempt to alter the data that would mismatch with the recorded hash. Furthermore, traditional Proof of Stake (PoS) algorithms often favor nodes with higher stakes, creating a "rich-get-richer" scenario and limiting participation for nodes with lower stakes. To address this fairness issue, we developed an enhanced PoS algorithm known as Reputation-based Proof of Stake (rPOS). This algorithm integrates a reputation mechanism into the existing PoS framework used by Geth (Go Ethereum), aiming to foster a more equitable mining process while improving overall system fairness and security.

In our project, we have achieved several key objectives that contribute to the advancement of secure and efficient data management in cloud environments:

- We created a web application that integrates ownCloud for seamless file management, enabling operations such as uploading, deleting, and downloading files.
- To support data provenance, we store metadata in IPFS and validate its hash on the blockchain. This approach ensures data integrity and immutability, preventing discrepancies if data is tampered with.
- Our improvements to the Proof-of-Stake consensus algorithm include integrating a reputation mechanism. This modification aims to foster a fairer mining process and enhances overall system security.

The remainder of the article is structured into three main sections: Section 2 provides a review of related research, Section 3 outlines the proposed model and algorithms, Section 4 presents the results of the study, and Section 5 concludes the paper.

2. Related Work

A survey of the literature is included in this section. For our analysis, we looked at a wide range of research articles, some of which are discussed below. G. Haveri and B. Deewan (2023) conducted a comparative study of Ethereum's Proof of Work (PoW) and Proof of Authority (PoA) within OpenStack SWIFT, analyzing provenance data such as RecordID, UserName, DateTime, FileName, and Action. Their research, performed in the OpenStack SWIFT cloud environment, explored how scalability varies with network size and hardware specifications. Using Ethereum's blockchain technology, they evaluated transaction throughput and efficiency under both consensus mechanisms and concluded that PoA offers superior scalability over PoW in decentralized cloud environments [2].In another study, researchers investigated IoT token and input provenance within Ethereum Rinkeby and Ropsten networks, capturing data like tokenID, inputProvenanceID, Context (datapoint, time, location), and index. Testing scenarios simulated real-world conditions, considering factors critical for applications such as vaccine supply chains, including latency, throughput, network load, and gas prices. Ethereum's blockchain infrastructure provided secure and transparent transactions, ensuring data integrity and traceability in dynamic IoT environments [3].

Almamun et al. proposed SciChain, a blockchain framework tailored for scientific applications deployed on High-Performance Computing (HPC) systems. Their study focused on managing provenance data associated with scientific experiments, including metadata on processing time, speedup, Cumulative Distribution Function (CDF), and overall performance metrics. Testing was conducted in local blockchain environments suitable for HPC, demonstrating improved scalability and performance compared to traditional ledger systems. SciChain utilized POST and POET consensus mechanisms to validate transactions and ensure data integrity across distributed scientific workflows [4].Celik et al. explored surveillance data provenance in collaborative workflows using Ethereum's Kovan test network and public blockchain (Kovan). Their research focused on monitoring and verifying data records within collaborative environments, particularly in industries like Building Information Modeling (BIM). Ethereum's Kovan network provided a scalable cloud environment for blockchain-based applications, enhancing transparency and accountability in data management processes. The research leveraged blockchain technology to improve trust and reliability in surveillance data provenance across distributed networks [5].

This study investigated blockchain's role in validating file operations within ownCloud, focusing on metadata such as RecordID, DateTime, Username, Filename, AffectedUser, Action, TxHash, and BlockHash. While specific blockchain platforms were not explicitly mentioned, the research highlighted decentralized validation mechanisms to ensure data integrity without reliance on a specific validation network. Testing aimed to enhance the security and reliability of file operations in cloud-based environments, showcasing blockchain's potential to improve data management practices [6].Metadata of file operations on Ethereum within ownCloud was examined, comparing the scalability of Proof of Stake (PoS) versus Proof of Work (PoW). The study focused on scalability challenges influenced by network load, size, and blockchain difficulty. Ethereum's blockchain technology validated and secured metadata associated with file operations, demonstrating improved transaction efficiency and data provenance in cloud storage solutions. The research underscored blockchain's capability to enhance reliability and scalability in managing metadata across diverse cloud platforms [7].

This study explored metadata of file operations on Ethereum within AWS S3 cloud environments, evaluating scalability issues affected by network load, size, and blockchain complexity. Ethereum's blockchain technology was leveraged to validate and secure metadata associated with file operations, addressing challenges in transaction efficiency and data provenance in cloud-based storage solutions. The research highlighted blockchain's potential to improve scalability and reliability in managing metadata across diverse cloud computing infrastructures [8].

These studies collectively contribute to advancing blockchain technology's application in enhancing scalability, security, and data provenance across various computing environments and domains. Each paper employs specific blockchain solutions tailored to unique cloud environments and data management challenges, demonstrating the versatility and potential of decentralized ledger technologies in modern computing infrastructures.

In the work by Hu et al. [9], the authors focus on enhancing Delegated Proof of Stake (DPoS) by introducing reputational factors such as trusted state value (TS), number of votes (V), rank (S), reward (R), and transaction fee reward (F). It addresses scalability and fairness issues by dynamically adjusting node selection criteria.In another study [10], the authors examine the integration of reputation mechanisms with Proof of Work (PoW), using statistical measures like arithmetic mean ($\bar{X}$) and standard deviation (σ) of miner reputations to assess reliability. This study emphasizes the role of reputation metrics in improving consensus efficiency and reliability based on historical performance data.

The paper by various authors [11] implements Practical Byzantine Fault Tolerance (PBFT) and cumulative reputation (CuR) to measure node contributions over time, promoting fairness in consensus group selection. It enhances reliability and trustworthiness in blockchain networks through reputation-driven consensus mechanisms.Another study [12] introduces a comprehensive reputation evaluation framework (Retu), integrating historical reputation (RHu), consensus reputation (RCtu), and transaction reputation (RTtu) with weighted factors (α, β, γ). This framework applies across various blockchain protocols (PoS, PoW, PoA), integrating Byzantine Fault Tolerance to ensure continuous node engagement and fairness in consensus activities.

The study by various authors [13] develops a framework for accumulating and evaluating historical reputation records (PoR) in blockchain operations, focusing on metadata and transactional data integrity to enhance transparency and integrity.Lastly, in the paper by various authors [14], the authors address security concerns in Bitcoin mining pools with reputation-based mechanisms to mitigate dishonest mining behavior, defining miner lifetime (Lj) and attack probability (Q) to safeguard network integrity in Proof-of-Work consensus models. Together, these studies highlight the evolving strategies and methodologies aimed at improving reliability, fairness, and scalability of blockchain technologies within cloud computing contexts.

3. Methodology

Within this section, we discuss the precise methodology adopted in the proposed model. This include the system model and the algorithms in the implementation.

3.1. Mathematical Formulation

Dataset Description

The Table 1 gives the attributes of dataset along with their meaning.

Logistic Regression Model

Initially, we collected historical data during the mining process, considering the attributes mentioned in the above table. Using this collected data, we trained a model using logistic regression to predict a node's capability to act as a miner. The prediction model leverages these attributes to make its predictions. Train the logistic regression model using the historical dataset D. The model predicts the probability $P(\text{Miner}_i)$ that a node i is capable of being a miner based on the attributes in D.

$$P(\text{Miner}_i) = \frac{1}{1 + e^{-(\beta_0 + \sum_{k=1}^{15} \beta_k x_{ik})}} \tag{1}$$

Table 1. Input Parameters and Historical Data

Attribute	**Explanation**
Timestamp, T	Time of event occurrence (e.g., block sealing).
Miner Address, A	Address of the responsible miner.
Miner Stakes, S	Amount staked by the miner in Proof of Stake.
Block Number, B	Identifier of the blockchain block.
Gas Limit, GL	Maximum gas usage allowed for block transactions.
Gas Used, GU	Gas are consumed by block transactions.
Transaction Time, TT	Time taken for a transaction to be processed and included.
Seal Hash, SH	Cryptographic hash of the block's seal information.
Difficulty, D	Computational complexity of mining the block.
Parent Hash, PH	Cryptographic hash of the previous block's header.
Root, R	Root hash summarizing all transactions in the block.
No. of Malicious Acts, M	Count of detected malicious behaviors.
Number of Blocks Sealed, N	Count of blocks sealed by the miner.
Rewards, RW	Amount earned by the miner for participation.
Malicious Behavior, MB	Impact of detected malicious acts on reputation or rewards.

where x_{ik} represents the k-th attribute value for node i, and $\beta_0, \beta_1, \ldots, \beta_{15}$ are coefficients from the model.

Gini Coefficient Calculation

The Gini coefficient ranges from 0 to 1, where a value closer to 0 indicates a fairer network. Nodes are selected as miners based on their reputation scores, with a lower RS indicating a higher probability of being chosen as miner. Conversely, nodes with higher RS are less likely to be selected and are subjected to penalties to discourage unfair behavior. Calculate the Gini coefficient G to measure stake distribution fairness among nodes:

$$G = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} |(S_i + RW_i - \mathrm{MB}_i) - (S_j + RW_j - \mathrm{MB}_j)|}{2n \sum_{i=1}^{n} (S_i + RW_i - \mathrm{MB}_i)} \tag{2}$$

where S_i is stake, RW_i is rewards, MB_i is malicious behavior impact, and n is node count.

Reputation Score Calculation

Once the prediction model is in place, it is used to calculate the probability of a node being selected as a miner. Additionally, another parameter, the Gini coefficient, is calculated for all nodes to assess how evenly stakes and rewards are distributed among the nodes, ensuring fairness. The predicted value is then combined with the Gini coefficient, which is calculated based on factors such as stakes, rewards, and the number of malicious activities by the node, to compute the reputation score (RS) of a node. Calculate Reputation Score (RS) for each node i:

$$\mathrm{RS}_i = \frac{(1 - P(\mathrm{Miner}_i)) + G}{2} \tag{3}$$

Selection Criteria and Penalties

Based on RS:

- If $\mathrm{RS}_i > 0.70$: $S_i = S_i - (S_i \times 0.75)$
- If $0.70 \geq \mathrm{RS}_i > 0.45$: $S_i = S_i - (S_i \times 0.50)$
- If $0.45 \geq \mathrm{RS}_i > 0.25$: $S_i = S_i - (S_i \times 0.25)$

- If $RS_i \leq 0.25$: No penalty.

The thresholds for applying penalties based on the Reputation Score (RS) were chosen by systematically varying the values and observing their impact on network performance and security. Through a series of controlled experiments and simulations, different threshold values were tested to identify the optimal balance between maintaining high network performance and ensuring robust security. By analyzing the results, the thresholds that resulted in the best overall network behavior were selected. This approach ensured that the chosen values effectively minimized malicious activities while promoting fairness and efficiency across the network

By integrating reputation into the PoS algorithm, rPOS aims to promote fairness and discourage malicious behavior, thereby creating a more balanced and secure blockchain network.

3.2. System model

Fig. 1, shows the system architecture of the proposed model.

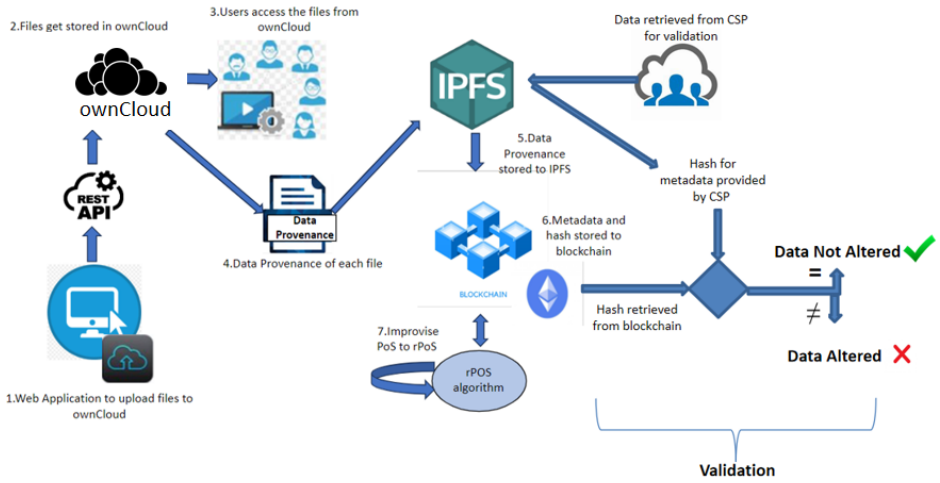


Fig. 1. System model

The proposed system consists of the following steps:

1. User Registration: When the user visits the system for the first time, the user needs to register themselves.
2. File Upload via Web Application: Users utilize a web application to upload files to the ownCloud storage system. This interaction is facilitated through a REST API.
3. File Storage in ownCloud: The uploaded files are securely stored within the ownCloud platform, which supports file synchronization and sharing.
4. File Access by Users: Authorized users can retrieve and download these files from ownCloud as needed.
5. Generation of Data Provenance: For each file in ownCloud, provenance data is created. This data captures the file's history, including its origin, modifications, and access records.
6. Storage of Provenance Data in IPFS: The generated data provenance information is stored in the InterPlanetary File System (IPFS), a decentralized file storage system that ensures data availability and verifiability.

7. Storage of Metadata and Hash on Blockchain: The metadata of the provenance data along with its hash (a cryptographic representation) are stored on a blockchain. This guarantees the integrity and immutability of the provenance records.

8. Enhanced PoS to rPoS Algorithm: An improved Proof of Stake (PoS) algorithm, known as revised PoS (rPoS), is utilized for validation processes within the blockchain, ensuring secure and efficient consensus.

Validation Process

1. Retrieval of Data from CSP for Validation: When validation is necessary, data is fetched from the Cloud Service Provider (CSP) to verify its integrity.

2. Provision of Metadata Hash by CSP: The CSP supplies the hash of the metadata, which is then compared with the hash stored on the blockchain.

3. Verification of Data Integrity: The hash obtained from the blockchain is compared to the hash provided by the CSP. If the hashes match, it confirms that the data remains unaltered (Data Not Altered). If the hashes do not match, it indicates data tampering (Data Altered).

The proposed system starts with user registration. Users upload files via a web application that interacts with ownCloud through a REST API, storing files securely. Authorized users can access and download these files. Detailed data provenance, including the file's origin, modifications, and access history, is generated and stored in IPFS. The metadata and hash of this provenance data are recorded on a blockchain for integrity and immutability. The system uses a revised PoS (rPoS) algorithm for efficient and secure consensus. To validate data integrity, information is retrieved from the CSP and compared with the blockchain-stored hash. Matching hashes confirm the data is unaltered; mismatches indicate potential tampering. This approach leverages ownCloud, IPFS, and blockchain for secure data storage, retrieval, and verification.

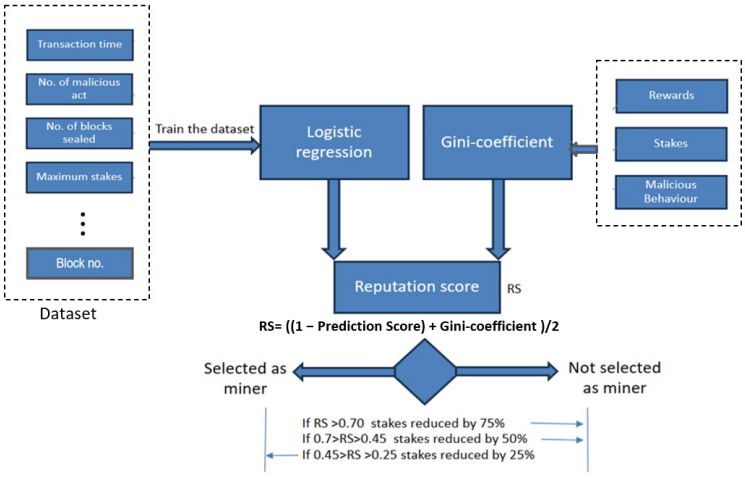


Fig. 2. rPOS consensus design

The provided figure 2 illustrates a methodology for assigning reputation scores (RS) to miners within a blockchain network, determining their eligibility and stake reduction. The process begins with compiling a dataset comprising transaction time, the number of malicious acts, the number of blocks sealed, maximum stakes, and other relevant attributes. This dataset is utilized to train a logistic regression model, which generates a prediction score. Concurrently, a Gini coefficient is calculated to assess the inequality in the distribution of rewards, stakes, and malicious behavior among miners.

Once the RS is determined, it influences the selection process of miners. Miners with higher RS values are preferred, with specific thresholds dictating the extent of stake reduction. If a miner's RS is greater than 0.70, their stakes are reduced by 75%. For RS values between 0.45 and 0.70, stakes are reduced by 50%, and for RS values below 0.45 but above 0.25, stakes are reduced by 25%. This system aims to incentivize honest behavior and equitable reward distribution, ensuring the integrity and security of the blockchain network by penalizing malicious actions proportionately.

3.3. Algorithm for provenance and consensus

Algorithm 1 outlines the process for provenance data collection and validation, ensuring the integrity and traceability of data through systematic verification. Algorithm 2 details the implementation of the consensus mechanism, enabling agreement among distributed nodes to maintain data consistency and reliability in a decentralized system. These algorithms collectively enhance the security and transparency of the system.

Algorithm 1 Provenance Data Collection and Validation

procedure CollectProvenanceData
  Step 1: Initialize GUI for user interactions (similar to Dropbox).
  Users register and log in to the application.
  while user is logged in **do**
    User performs file operations (upload, download, delete, share).
    Capture provenance data:

```
{
"username":  "user1",
"filename":  "document.txt",
"timestamp":  "2024-06-11T10:00:00Z",
"action":  "upload"
}
```

    Convert provenance data to JSON format and upload JSON file to IPFS network.
    Retrieve unique hash from IPFS.
    Add IPFS hash to Ethereum private blockchain.
  end while
end procedure
procedure ValidateProvenanceData
  Retrieve original provenance data from blockchain.
  Retrieve tampered provenance data from CSP and Compare original and tampered data.
  if original data ≠ tampered data **then**
    Identify data discrepancies and Confirm data tampering.
    Identify the original culprit.
  else
    Data is not tampered.
  end if
end procedure

Algorithm 2 Reputation-based Proof of Stake (rPOS) Algorithm

Require: Timestamp as T, Miner Address as A, Miner Stakes as S, Block Number as B, Gas Limit as GL, Gas Used as GU, Transaction Time as TT, Seal Hash as SH, Difficulty as D, Parent Hash as PH, Root as R, Malicious Number as M, Number of Blocks Sealed as N

1: Train the prediction model with the historical dataset
2: Predict the score for the node based on the values of T, A, S, B, GL, GU, TT, SH, D, PH, R, M, N
3: Calculate the Gini coefficient value based on S, rewards, and M
4: Calculate the Reputation Score (RS) as:

$$\text{RS} = \frac{(1 - \text{Prediction Score}) + \text{Gini Coefficient}}{2}$$

RS > 0.70
5: $S = S - (S \times 0.75)$
6: The node is not allowed to be selected as the miner $0.70 \geq \text{RS} > 0.45$
7: $S = S - (S \times 0.50)$
8: The node is not allowed to be selected as the miner $0.45 \geq \text{RS} > 0.25$
9: $S = S - (S \times 0.25)$
10: The node is allowed to be selected as the miner **return**

4. Results and Discussion

4.1. Experimental Setup

In this section, we discuss the configuration and system used to carry out experiments. The i7 processor with 32GB RAM machine with Ubuntu 22.04.4 LTS is used . We used Geth 1.10.17 for Ethereum blockchain. Solidity is used for smart contract programming to write access control rules. Table 2 shows the detailed configurations of the systems used. Used IPFS 0.28.0 for intermediate storage and ownCloud 10.9.1 as cloud environment.

Table 2. Experimental Setup

Components	Software/Language	Version
OS	Ubuntu	22.04.4 LTS
Processor	Intel	i7-9300H
Cloud	OwnCloud	10.9.1
Blockchain Client	Geth	1.10.17
Intermediate Storage	IPFS	0.28.0

4.2. Consensus Algorithm results analysis

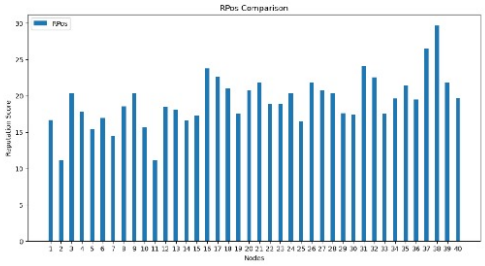
The implementation of data provenance using blockchain technology offers significant insights into file integrity and traceability within a cloud environment. In our scenario, a user, "abcd@gmail.com," deletes an important file, "new.sh." The Cloud Service Provider (CSP) then alters the username to "XYZ@gmail.com" and provides this information to the file owner. By comparing the original untampered data from the blockchain with the tampered data from the CSP, discrepancies are revealed, indicating data alteration.

Our results demonstrate that blockchain effectively maintains an immutable, transparent record of all file operations, including uploads, deletions, and modifications, each associated with specific timestamps and user identities. This secure and tamper-proof recording allows for the identification of the original user responsible for modifications, enhancing transparency and accountability. The ability to detect and

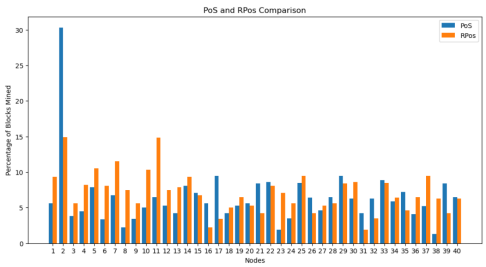
prove tampering underscores blockchain's effectiveness in maintaining data integrity in cloud environments, promoting trust among users.

For testing the performance of our designed RPoS (Reputation-based Proof of Stake) algorithm, we set up a blockchain network with varying numbers of nodes, such as 10, 20, and 40. We obtained results for the block mining activity of nodes, average block creation time, and average transaction time. These results were then compared with the PoS (Proof of Stake) algorithm to ensure better performance in terms of security and fairness.

Figure 3(a) illustrates the reputation scores of all nodes in the blockchain network. The reputation scores for each node are calculated based on the RPOS algorithm we designed. These reputation scores vary from one node to another based on their performance in the network. The node with the lowest reputation score will be chosen as the miner. In this scenario, Node 2 has the lowest reputation score, so it is chosen as the miner. Figure 3(b) compares the percentage of blocks mined by nodes using the PoS and rPoS algorithms. The data indicates that blocks are mined almost uniformly under the rPoS algorithm, as represented by the relatively even height of the orange bars. Nodes with lower reputation scores have a higher percentage of blocks mined, evidenced by the taller orange bars for rPoS. Conversely, nodes with higher reputation scores are mined less frequently, as shown by the shorter orange bars compared to the blue bars for PoS. This demonstrates the effectiveness of the rPoS algorithm in promoting fairness by reducing the mining advantage of nodes with high reputation scores.



((a)) Percentage of Blocks Mined

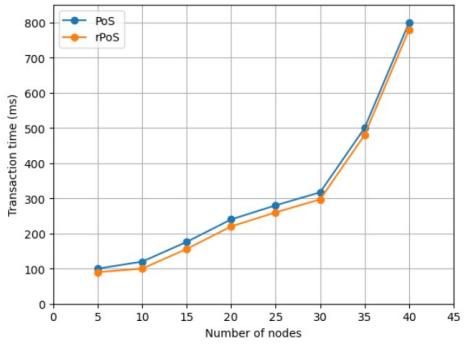


((b)) Average Block Creation Time

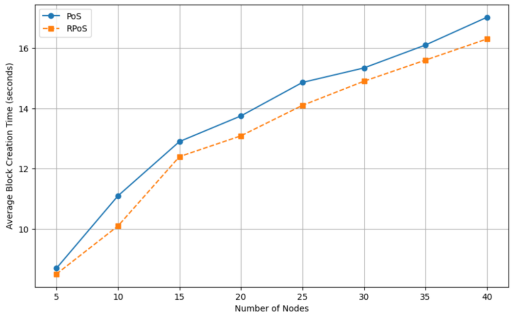
Fig. 3. Results

4(a) compares the average block creation time in seconds as the number of nodes increases, for both the traditional PoS (Proof of Stake) and the proposed rPoS (Reputation-based Proof of Stake) algorithms. The data shows that the average block creation time generally increases with the number of nodes. However, the rPoS algorithm consistently achieves lower average block creation times compared to the PoS algorithm, as indicated by the orange dashed line. This suggests that rPoS not only enhances fairness in block mining but also improves efficiency in terms of block creation time. Nodes with the best reputations are chosen to create blocks, ensuring that the most capable and efficient nodes handle the task, thereby reducing block creation time. This demonstrates that the rPoS mechanism is more effective in maintaining a faster and more reliable blockchain network. Figure 4(b) depicts the relationship between the average transaction time and the number of nodes for two systems: PoS (Proof of Stake) and rPoS (revised Proof of Stake). As the number of nodes increases, both systems experience a rise in transaction time. This trend is more pronounced beyond 30 nodes, where transaction times increase sharply. The rPoS system demonstrates a consistently better performance, with slightly lower transaction times compared to PoS. The underlying reason for rPoS's superior performance can be attributed to its reputation system. In rPoS, nodes with a history of malicious behavior are less likely to participate in the network. This filtering mechanism reduces the chances of network disruptions and slowdowns, ensuring smoother operation. By excluding poorly performing nodes, the reputation system ensures that only high-quality nodes handle transactions. This selective participation enhances the network's efficiency, leading to reduced processing times. Consequently, rPoS maintains lower transaction times, especially as the network scales up, highlighting the effectiveness of its reputation-based

approach in maintaining network performance and reliability.



((a)) Average Transaction Time



((b)) Reputation Score vs. Number of Nodes

Fig. 4. Results

5. Conclusions

The reputational Proof of Stake (rPoS) system outperforms the traditional Proof of Stake (PoS) system in terms of scalability and efficiency. The rPoS system maintains higher and more consistent reputation scores, ensuring a more reliable network by filtering out poorly performing nodes. It achieves a lower average block creation time, which increases more gradually as the number of nodes rises, demonstrating superior efficiency. For instance, with 40 nodes, rPoS achieves an average block creation time of around 15 seconds compared to PoS's 17 seconds. These advantages result in rPoS maintaining lower transaction and block creation times, making it a more scalable and effective solution for larger, more complex blockchain networks. In summary, rPoS proves to be more scalable and effective than PoS. Its reputation system filters out low-performing nodes, resulting in lower transaction and block creation times, making rPoS a better choice for larger, more complex blockchain networks.

References

[1] C. Gong, J. Liu, Q. Zhang, H. Chen, Z. Gong, The characteristics of cloud computing, in: 2010 39th International Conference on Parallel Processing Workshops, 2010, pp. 275–279. doi:10.1109/ICPPW.2010.45.

[2] N. G, P. Haveri, R. B, Y. Deewan, A framework for data provenance assurance in cloud environment using ethereum blockchain: Data provenance assurance in cloud using blockchain, EAI Endorsed Transactions on Scalable Information Systems 11 (2). doi:10.4108/eetsis.3536.
URL https://publications.eai.eu/index.php/sis/article/view/3536

[3] M. Sigwart, M. Borkowski, M. Peise, S. Schulte, S. Tai, Blockchain-based data provenance for the internet of things, in: Proceedings of the 9th International Conference on the Internet of Things, IoT '19, Association for Computing Machinery, New York, NY, USA, 2019. doi:10.1145/3365871.3365886.
URL https://doi.org/10.1145/3365871.3365886

[4] A. Al-Mamun, D. Zhao, Scichain: Trustworthy scientific data provenance (2020). arXiv:2002.00141.

[5] Y. Celik, I. Petri, M. Barati, Blockchain supported bim data provenance for construction projects, Computers in Industry 144 (2023) 103768. doi:https://doi.org/10.1016/j.compind.2022.103768.
URL https://www.sciencedirect.com/science/article/pii/S0166361522001646

[6] X. Liang, S. Shetty, D. Tosh, C. Kamhoua, K. Kwiat, L. Njilla, Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability, in: 2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID), 2017, pp. 468–477. doi:10.1109/CCGRID.2017.8.

[7] P. Abhishek, Y. Akash, D. G. Narayan, A scalable data provenance mechanism for cloud environment using ethereum blockchain, in: 2021 IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER), 2021, pp. 1–6. doi:10.1109/DISCOVER52564.2021.9663425.

[8] A. Patil, A. Jha, M. M. Mulla, D. Narayan, S. Kengond, Data provenance assurance for cloud storage using blockchain, in: 2020 International Conference on Advances in Computing, Communication Materials (ICACCM), 2020, pp. 443–448. doi:10.1109/ICACCM50413.2020.9213032.

[9] Q. Hu, B. Yan, Y. Han, J. Yu, An improved delegated proof of stake consensus algorithm, Procedia Computer Science 187 (2021) 341–346, 2020 International Conference on Identification, Information and Knowledge in the Internet of Things, IIKI2020. doi:https://doi.org/10.1016/j.procs.2021.04.109.
URL https://www.sciencedirect.com/science/article/pii/S1877050921009133

[10] C. Tang, L. Wu, G. Wen, Z. Zheng, Incentivizing honest mining in blockchain networks: A reputation approach, IEEE Transactions on Circuits and Systems II: Express Briefs 67 (1) (2020) 117–121. doi:10.1109/TCSII.2019.2901746.

[11] T. Zhang, Z. Huang, Fpor: Fair proof-of-reputation consensus for blockchain, ICT Express 9 (1) (2023) 45–50. doi:https://doi.org/10.1016/j.icte.2022.11.007.
URL https://www.sciencedirect.com/science/article/pii/S2405959522001655

[12] X. Qiu, Z. Qin, W. Wan, J. Zhang, J. Guo, S. Zhang, J. Xia, A dynamic reputation-based consensus mechanism for blockchain., Computers, Materials & Continua 73 (2).

[13] H. Chai, S. Leng, K. Zhang, S. Mao, Proof-of-reputation based-consortium blockchain for trust resource sharing in internet of vehicles, IEEE Access 7 (2019) 175744–175757.

[14] C. Tang, L. Wu, G. Wen, Z. Zheng, Incentivizing honest mining in blockchain networks: a reputation approach, IEEE Transactions on Circuits and Systems II: Express Briefs 67 (1) (2019) 117–121.