Journal of Data Processing

DLINE JOURNALS

Print ISSN: 2278 – 6481
Online ISSN: 2278 – 649X

JDP 2025: 15 (4)
https://doi.org/10.6025/jdp/2025/15/4/151-158

A Rough Set Based Parallel Verification Model for Accounting Mining Systems

Wei Zeng
School of Physical Education
Hubei University of Technology
Wuhan, Hubei, 430068. China
zengwei800104@126.com

ABSTRACT

The paper proposes a cloud based data integrity verification algorithm for accounting informationization systems, leveraging rough set based data mining techniques. With the integration of cloud computing into financial management, data security and integrity have become critical concerns due to risks like data breaches and accidental loss as illustrated by real world incidents involving Huawei and Tencent Cloud. Traditional verification methods, such as HMAC or RSA based Provable Data Possession (PDP), suffer from high computational and communication overheads, especially under dynamic data operations.

To address these limitations, the author introduces a parallel verification algorithm that supports multi user batch validation, significantly reducing communication costs and verification time. By employing rough set theory, the system performs efficient feature selection and attribute reduction, enhancing computational efficiency without compromising data integrity. The verification process uses BLS signatures with homomorphic properties, allowing multiple files to be verified at the communication cost of a single file. Experimental results show that the proposed algorithm outperforms conventional single point verification methods, especially as user scale and data volume increase. The system also offers tiered audit levels that balance detection accuracy, verification frequency, and cost. Implemented in a Java/SQL Server environment, the approach demonstrates improved scalability and suitability for accounting cloud environments, offering a practical solution for secure, efficient financial data management in the era of digital transformation.

Keywords: Accounting Informationization, Cloud Computing, Data Integrity Verification, Rough Set Theory, Data Mining, BLS Signature, Parallel Verification Algorithm

Received: 18 July 2025, Revised 3 September 2025, Accepted 19 September 2025

1. Introduction

With the rise of computerized accounting and the sharing economy, financial sharing systems developed through basic digital analysis methods have gained widespread adoption. These systems not only assist small, medium, and large businesses in enhancing their financial management practices but also significantly improve their operational capabilities, reducing operational costs. Recently, leveraging state of the art digital transformation technologies, some multinational corporations have adopted a holistic, sustainable approach to financial sharing management, yielding positive outcomes. In China, the swift pace of economic growth has prompted numerous firms to seek innovative and advantageous financial management strategies to bolster their overall performance [1]. As cloud computing becomes increasingly prevalent, a multitude of companies and organizations have begun utilizing their server infrastructure for data and document storage. Compared with conventional internal storage methods, this modern service significantly improves efficiency and reduces equipment maintenance costs. Presently, the convergence of accounting information systems with cloud computing has emerged as a noteworthy focus. Sudersan employs an advanced cloud computing platform, integrated with big data analysis, to develop a novel and efficient accounting information management experience [2]. Barba-Sánchez V thoroughly examines the obstacles faced by small and medium enterprises in financial information management, integrates real world scenarios, and offers practical, actionable solutions to help them achieve superior information management [3]. However, with the merging of information systems and cloud computing, certain potential risks have surfaced. For instance, malicious actors could compromise cloud security, and personal data breaches may occur easily during information system operations. As a result, cloud service providers must remain vigilant and implement effective measures to safeguard information security. To enhance cloud server management, it is essential to conduct regular audits and validations of the stored data to ensure its accuracy and dependability. In this context, rough set based data mining techniques, centered on cloud computing, can be utilized to effectively verify data integrity and address the security challenges associated with financial sharing models. Our experiments have demonstrated that the algorithm we proposed is notably more effective than conventional single point testing approaches.

2. Related Work

Cloud computing technology has become a crucial element in the current computing landscape, facilitating information exchange across various nations and regions, thereby enabling rapid transmission and secure, efficient information management. Cloud storage of accounting information is a vital aspect of this, enabling real time transmission and effective information management, thereby fulfilling the requirements of enterprise information development. Despite the clear benefits of cloud storage, its extensive capacity also introduces specific security vulnerabilities that cannot be overlooked. Recent incidents involving the improper handling of accounting data within cloud storage services have shown that these services are not entirely trustworthy. For example, in September 2017, Huawei technicians inadvertently erased 800,000 user records from Guangxi Mobile during an expansion. In August 2018, a malfunction occurred in Tencent Cloud's client, Beijing QINGBO CNC Technology Co., Ltd.'s "Frontier CNC" platform, resulting in damage to its file system data, leading to economic losses estimated at around 11 million yuan [4].

The concept of "Vague" was first introduced in linguistics in 1904, referring to entities that cannot be entirely classified into a particular category and cannot be accurately defined. Consequently, for these intricate ideas, it is essential to investigate them thoroughly to provide better explanations, descriptions, and interpretations,

thereby enriching our linguistic comprehension. However, the membership function of this theory is indeed shaped by individual biases and preferences [5]. Researchers have refined the idea of boundary regions by segmenting indistinguishable units into separate ranges, known as the upper and lower approximation sets [6]. Its expression through precise mathematical models lends it considerable reliability [6]. The cloud data integrity verification scheme known as PDP was initially utilized in grid computing and P2P networks. Given the advantages of cloud storage services, most users now prefer storing their data in the cloud. However, this shift has also introduced a range of security challenges. Initially, researchers contemplated employing the HMAC hash function to ensure the integrity of remote data. Before uploading, users calculate the MAC value of their data and store it [7]. When verification is required, the data is retrieved from the remote node, its MAC is computed, and the computed MAC is compared with the previously stored MAC to verify the data's integrity. This verification method requires downloading the entire data file during verification, consuming considerable computational resources and communication bandwidth, making it unfeasible. Later, scholars explored utilizing the homomorphic characteristics of RSA signatures to develop a PDP scheme that accommodates unlimited data checks. However, this approach requires calculating the entire file, resulting in higher computational costs. Researchers first proposed segmenting the data file into blocks and computing signatures for the divided files to alleviate computational overhead. Despite having a large number of data blocks, verification still demands significant computational resources [8]. They then introduced a technique for extracting a fixed number of data blocks per verification and leveraged the homomorphic properties of RSA signatures to consolidate the evidence into a single value, greatly reducing the scheme's communication overhead while ensuring a high probability of successful file detection. Subsequently, they proposed the MR-PDP scheme, which allows for multi replica data verification but does not accommodate fully dynamic data operations. With the growing variety of storage requirements, researchers recognized that users expect dynamic operations on cloud data, such as additions, deletions, and modifications. To address this need, scholars adapted the conventional PDP scheme by implementing dynamic data structures to organise data blocks, creating a scheme that facilitates fully dynamic block level operations. They incorporated the Merkle hash tree to verify the accuracy of data blocks' locations and employed the BLS signature mechanism to confirm the correctness of the content within data blocks. This scheme includes specific optimizations to minimize computational and communication overheads and to lessen the user's burden during verification. They introduced authoritative third party auditors to initiate verification challenges on behalf of users, but this brings the risk of data leakage. To protect user privacy, Wang and others introduced random masking to generate the returned evidence, effectively hiding the data and preventing curious third parties from obtaining it through calculations.

3. Rough Set based Data Mining Technology and Its Application in Accounting Informationization System

3.1 Cloud Computing Theory and Related Technologies for Accounting Informationization

With the ongoing evolution of computer science, cloud computing has emerged as a trending subject, widely acknowledged across diverse sectors, and a necessary progression for future computing networks. An effective cloud computing platform [6] must possess a well designed storage framework, encompassing efficient storage management, robust security measures, adequate computational power, and the ability to offer adaptable storage solutions for users' long-term storage and oversight. By merging accounting information technology with cloud computing, we can enhance financial sharing, significantly boost operational efficiency, and reduce hardware costs. Utilizing cloud computing technology alongside SOA architecture enables us to provide more accessible accounting and financial information services for businesses, thus lowering information technology

expenses. Building on prior research accomplishments, we propose a novel cloud based accounting information system, comprising five distinct components: process oversight, SAP, document management, procurement oversight, and contract oversight.

3.2 Data Mining Based on Rough Set Theory

Feature selection plays a crucial role in data mining, focusing on identifying effective features from a multitude of models and model collections while simplifying the model set to attain optimal model performance. Rough set theory also addresses this challenge within the realm of feature selection, with the goal of minimizing the complexity of the model set through careful selection. The strength of rough set theory lies in its ability to effectively maintain and enhance the relationship between the condition set and the decision set, thereby improving their performance in specific areas. An information entity with n features can be segmented into numerous parts, and the disparities between these parts will increase, leading to variations among the n features. In light of this scenario, even with multiple features, it remains impossible to eradicate the differences among them entirely. Most feature selection algorithms aim to identify the smallest reduction, meaning retaining as many attributes as feasible. However, this endeavor is quite challenging, leading to the frequent use of heuristic algorithms to achieve reduction as swiftly as possible under constrained resources for optimal outcomes. By calculating the simplification and relative simplification of attributes, several effective strategies can be employed: 1) Altering the values of the Discernibility Matrix can substantially decrease computational complexity, facilitating precise reduction of the information system. 2) By assessing attribute importance, we can extract valuable information and progressively eliminate the least representative attributes to accomplish attribute simplification. 3) Merging genetic algorithms with other technological tools can lead to effective reductions. 4) Dynamic reduction serves as a means of transforming intricate information systems into manageable subsystems, enabling us to complete tasks more efficiently and predict outcomes with greater accuracy.

3.3 The Application of Cloud Data Parallel Verification Algorithm in Accounting Information ization

By using rough digital analysis methods, we can integrate a company's financial data into the Internet across geographical and temporal boundaries, thereby improving the company's efficiency. However, the success of this method depends on how we use the latest accounting information system to ensure the company's cash

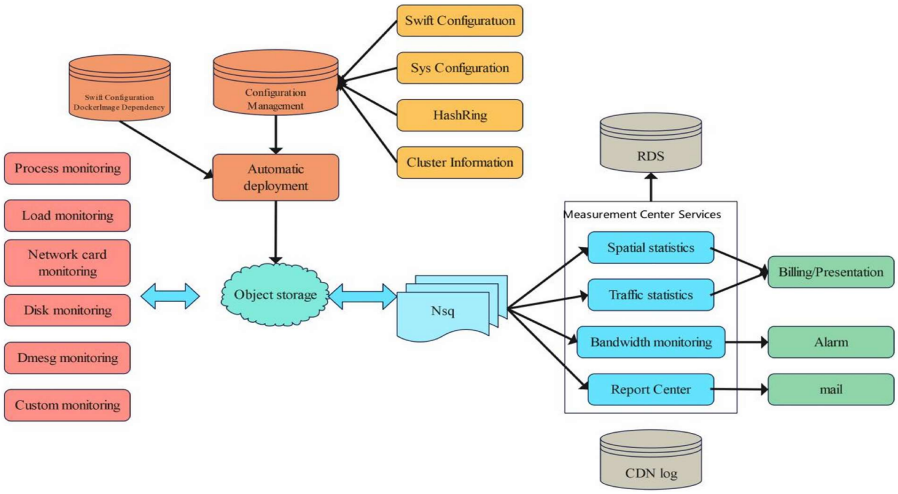


Figure 1. Network topology of Cloud storage system

flow. With the continuous improvement of technology, financial sharing has emerged from the original three changes: First, its organizational structure has undergone major adjustments; second, its application has shifted from previous single financial management to diversified financial services; and finally, its security has also been greatly improved, ensuring the reliability of financial information. Therefore, we have established a complete cloud based data storage system to provide information exchange among various nodes, as shown in Figure 1.

By adopting new technologies, we have developed a new validation algorithm that can simultaneously validate the data of a single user or multiple users. This new algorithm greatly simplifies the verification traffic, and also greatly shortens the communication time between the verification organization, cloud storage and the verifier. The detailed calculation process is as follows.

1) To ensure security, we need to create a public and private password for each user, and then place them in a reliable third party inspection mechanism for inspection. In addition, we also need to check the signature of each password block ($F=(m_1, m_2,..., m_n)$ to ensure their security H (i) represents a hash function, while x and u are a set of random numbers.

$$s_i = \left(H(i) \cdot u^{m_i}\right)^x \qquad (1)$$

2) Send a verification request. Each file block generates a verification request sequence, expressed as follows:

$$CHAL = \left\{\left(i, v_i\right)\right\} \qquad (2)$$

3) Generate verification information. To achieve parallel verification of multiple users, it is first necessary to group and integrate all user profiles:

$$m_k = \sum_{i=1}^{n} v_i m_{k,i} + m_1 \qquad (3)$$

In the equation, u represents the random number generated by the cloud storage server for each user during each verification process.

4) Verify the results of group merging to ensure the correctness of cloud storage.

4. Experimental Design and Result Analysis

4.1 Experimental Design and Operating Methods

By utilizing the Intel Core i5-8400 processor, 4GB of RAM, and a Seagate 500GB hard drive, we can effectively assess the efficiency of rough set based data mining techniques alongside traditional single user data verification methods. We implemented Eclipse as our development environment, Java as our coding language, and SQL Server 2010 as our database management system. Presently, the emphasis in data integrity verification has shifted from conventional block level data placement to more precise methods for detecting data corruption and adaptable data analysis techniques, aimed at achieving swift, cost effective identification of data. Among these methods, traversal detection has emerged as the most prevalent technique for detecting data integrity, enabling more accurate pinpointing of damaged data areas and facilitating precise data identification. Binary search is considered a more complex detection method that segments data blocks into several independent

regions for faster identification and comparison, thereby enhancing the accuracy and speed of the verification process. However, due to varying corruption rates, the effectiveness and speed of this method can often be compromised, leading to increased computational demands.

4.2 Experimental Results and Analysis

We assessed the performance of two distinct algorithms: the feature based clustering algorithm and the optimized clustering algorithm. Both algorithms were derived from classical methods, and their performance was evaluated based on the outcomes they generated. As illustrated in Figure 2, the performance comparison indicates that, with the advancement of cloud computing, the time required for each file update using the single user verification algorithm remains relatively constant, whereas under the multi user parallel verification algorithm, this time is considerably reduced. By leveraging cloud computing's sharing capabilities, parallel detection strategies can significantly enhance the system's computational efficiency compared to traditional single-point detection methods.

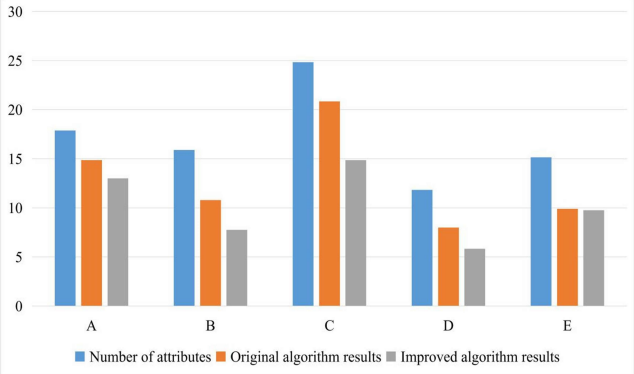
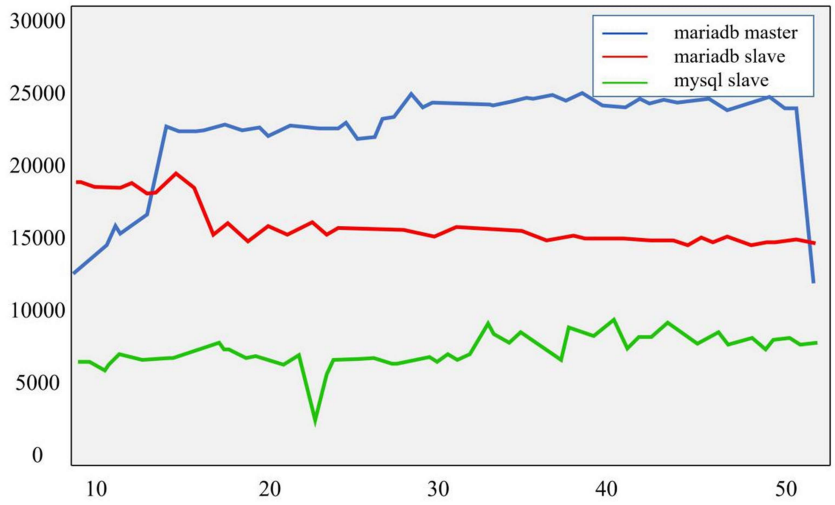


Figure 2. Performance Comparison of Data Integrity Verification Algorithms

After processing the fifth dataset using both algorithms, the overall reduction outcomes are nearly identical, and within the fifth dataset, there are merely 60 markers and 15 conditional attributes; yet, it still achieves its maximum utility, showcasing the algorithm's impressive performance. Following the analysis of the third dataset with the two algorithms, completely contrasting results can be observed, indicating that even with minimal processing, high efficiency can still be attained once the processing volume reaches a certain threshold, thereby emphasizing the strengths of both algorithms in handling processing volume. This system allows users to select the audit level independently; the higher the audit level, the more rapid and precise the detection of data block damage becomes. In this framework, the audit level is categorized into three tiers based on the probability of detection and the audit cycle.. Assuming the user data error rate is 1%, to achieve the requirements of audit level 1 within 24 hours, the user data needs to be verified once, with 230 data blocks verified each time; to achieve the requirements of audit level 2, it needs to be verified twice, with 299 data blocks verified each time; to achieve the requirements of audit level 3, it needs to be verified four times, with 459 data blocks verified each time. The verification time required for different file sizes and different audit levels is shown in Figure 3, and it can be seen that the higher the audit level, the more verification time is required, and accordingly, the higher the cost incurred by the user.

From the chart, it is evident that batch verification consistently requires less computing time than traversal verification. In terms of communication overhead, thanks to the system's implementation of the BLS signature



homomorphism, validating multiple files incurs the exact communication cost as validating a single file. If a user needs to validate 50 files, the communication overhead for traversal verification is 50 times greater than that of batch verification.

5. Conclusion

In this study, we examined a novel approach to identifying data integrity issues. We developed a framework utilizing cloud computing to facilitate this method. Additionally, we investigated a new technique for handling large scale data to enhance support for complex business processes. We also aimed to implement a more effective strategy for detecting large scale data to assist business operations further. Through this investigation, we discovered that, compared with the conventional single point verification approach, the rough set based data mining verification algorithm not only effectively ensures data completeness but also significantly reduces the cost of each update, thereby enhancing verification efficiency.

Acknowledgements

The study was supported by the Application Research of Big Data Technology in Financial Budget Performance Evaluation, Soft Science Research Project in 2022 in Henan Province (Accounting Special Project).

References

[1] Huang, J., Tao, B., Zeng, F. (2022). Point cloud registration algorithm based on ICP algorithm and 3D-NDT algorithm. *International Journal of Wireless and Mobile Computing,* 22 (2), 125-130.

[2] Sudersan, S., Abhijith, V. S., Thangavel, M., et al. (2021). A Review on Data Integrity Verification Schemes Based on TPA and Blockchain for Secure Cloud Data Storage. Security, Privacy and Data Analytics: *Select Proceedings of ISPDA*, 2022, 117-129.

[3] Barba-Sánchez, V., Martínez-Ruiz, M. P., Jiménez-Zarco, A I. (2007). Drivers, benefits and challenges of ICT adoption by small and medium sized enterprises (SMEs): a literature review. *Problems and Perspectives in Management,* 5(1), 103-114.

[4] Huang, C. (2020). Data parallel clustering algorithm based on mutual information mining of joint condition. IOP Conference Series: *Materials Science and Engineering, 914(1), 01-03.*

[5] Huang, Y., Cheng, Z., Zhou, Q., et al. (2020). Data Mining Algorithm for Cloud Network Information Based on Artificial Intelligence Decision Mechanism. *IEEE Access,* (8), 53394-53407.

[6] Li, X., Wang, Y., Liu, G. (2020). Structured Medical Pathology Data Hiding Information Association Mining Algorithm Based on Optimized Convolutional Neural Network. *IEEE Access,* (8), 1443-1452.

[7] Raheem, S., Shehabi, S. A., Nassief, A. M. (2022). MIGR: A Categorical Data Clustering Algorithm Based on Information Gain in Rough Set Theory. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems,* 30(05), 757-771.

[8] Chen, M. (2021). Accounting Data Encryption Processing Based on Data Encryption Standard Algorithm. *Complexity,* (5), 1-12.

[9] Xingyu, G., Ke, C., Pengtao, J, et al. (2021). K modes Algorithm Based on Rough Set and Information Entropy. *Journal of Physics: Conference Series,* 754 (1), 01-04.

[10] Dai, N. (2023). Application and functional simulation of data mining technology in Hadoop cloud platform based on improved algorithm. *Soft Computing,* 27 (12), 8381-8389.