Journal of Information Technology Review

DLINE JOURNALS

Print ISSN: 0976-3511
Online ISSN: 0976-2922

JITR 2025: 16 (4)
https://doi.org/10.6025/jitr/2025/16/4/138-145

Enhancing Cloud Accounting Security: A Rough Set-Based Parallel Algorithm for Efficient Data Integrity Verification

Bing Xu[1],Yu Song [2]
[1]Network Security Department, Henan Police College
Zhengzhou, Henan, 450046. China
[2]School of Computer and Artificial Intelligence, Zhengzhou University
Zhengzhou, Henan, 450001. China
xyx@hnp.edu.cn

ABSTRACT

This paper proposes a rough set-based data mining algorithm to enhance data integrity verification within cloud-based accounting information systems. It addresses security risks inherent in cloud storage, such as data breaches and corruption, which traditional verification methods handle inefficiently due to high computational and bandwidth costs. The authors introduce a novel parallel verification algorithm that can simultaneously check data for single or multiple users, significantly reducing communication time and computational overhead compared to single point testing. Experimental results demonstrate that this approach, leveraging cloud computing's parallel processing capabilities, greatly improves computational efficiency and reduces update costs while ensuring data completeness. The system also allows users to select different audit levels, balancing verification speed, precision, and cost. The research aims to provide a more secure and efficient framework for financial data management in the era of accounting informationization.

Keywords: Cloud Accounting Security, Data Integrity Verification, Rough Set Theory, Parallel Verification Algorithm, Computational Efficiency, Cloud Storage, Accounting Informationization, Data Mining, Third-Party Auditing (TPA), Audit Level Customization

Received: 3 May 2025, Revised 30 July 2025, Accepted 10 August 2025

1. Introduction

With the advent of digital accounting and the sharing economy, financial sharing systems that have been developed using fundamental digital analysis techniques have seen widespread implementation. This method not only aids a multitude of small, medium, and large enterprises in improving their financial management strategies but also significantly boosts their management capabilities, resulting in decreased operational costs. Recently, fueled by advanced digital transformation technologies, numerous global multinational corporations have adopted a

comprehensive and sustainable model of financial sharing management, leading to favorable results. In China, the rapid economic growth has led many companies to pursue innovative and beneficial financial management approaches to enhance their overall performance [1]. As cloud computing becomes more widespread, many businesses and organizations have started to leverage their server infrastructure for data and document storage. In comparison to traditional internal storage, this contemporary service significantly improves work efficiency and reduces equipment upkeep expenses. Currently, the integration of accounting information systems with cloud computing has become a point of significant interest. Sudersan S utilizes an advanced cloud computing platform, combined with big data analytics, to provide a fresh and effective accounting information experience [2]. Barba-Sánchez *V* conducts an in depth examination of the challenges faced by small and medium sized businesses in managing financial information, incorporating real world examples and suggesting practical, attainable solutions to help them achieve better information management [3]. However, the fusion of information systems and cloud computing has introduced certain potential risks. For example, malicious entities could exploit cloud security vulnerabilities, and personal data breaches may easily occur during the operation of information systems. Therefore, to protect information security, cloud service providers must be vigilant and implement robust measures. To optimize the management of cloud servers, it is crucial to conduct regular audits and validations of the stored data to ensure its accuracy and dependability. In this context, we can employ rough set-based data mining techniques, utilizing cloud computing as the backbone of information management, to facilitate efficient assessments of data integrity and effectively address the security risks associated with financial sharing models. Through a series of evaluations, we found that our proposed algorithm exhibits significantly higher efficiency compared to traditional single point testing methods.

2. Related Work

Cloud computing technology has become a crucial component in today's computing environment, enabling the flow of information across various countries and regions, thus allowing for swift transmission and secure, effective information management. The storage of accounting information in the cloud is a significant aspect of this, facilitating real time transmission and dependable information management, thereby meeting the needs of enterprise information growth. Despite the numerous evident advantages of cloud storage, its extensive capacity also poses certain security threats that cannot be ignored. Recent events involving issues in accounting information cloud storage have underscored that cloud storage services are not completely reliable. In September 2017, Huawei technicians accidentally deleted 800,000 user records from Guangxi Mobile during an expansion.. In August 2018, a failure occurred in Tencent Cloud's client, the "Frontier *CNC*" platform of Beijing *QINGBO CNC* Technology Co., *Ltd*., resulting in damage to its file system data, leading to economic losses estimated at around 11 million yuan [4].

The term "Vague" was first introduced to the domain of linguistics in 1904, referring to entities that cannot be completely categorized and lack clear definitions. Consequently, for these complex ideas, it is vital to explore them in depth to elucidate, describe, and interpret them, thereby enhancing our grasp of language. However, the membership function of this theory is indeed influenced by personal biases and preferences [5]. Researchers have refined the concept of the boundary region, classifying ambiguous units into different ranges, referred to as upper and lower approximation sets [6]. Since it can be expressed using precise mathematical frameworks, it possesses considerable reliability [6]. The cloud data integrity verification PDP scheme for accounting information was initially employed in grid computing and *P2P* networks. Due to the benefits of cloud storage, most users now opt to keep their data in the cloud. Nevertheless, this transition has also introduced various security challenges.

Initially, researchers considered using the HMAC hash function to safeguard the integrity of remote data. Before uploading any data, users calculate the MAC value and store it [7]. When verification is necessary, the data is fetched from the remote node, its MAC is recalculated, and this new value is compared with the previously stored MAC value to evaluate data integrity. This verification process requires retrieving the complete data file, which consumes substantial computational resources and communication bandwidth, making it impractical. Subsequently, scholars examined the potential of utilizing the homomorphic properties of RSA signatures to create the PDP scheme, allowing for unlimited data checks. However, this method involves transforming the entire file into a large number, resulting in increased computational costs. They suggested initially dividing the data file into blocks and calculating signatures for the segmented files to reduce computational demands. Yet, when the number of data blocks becomes excessively large, verification still requires significant computational resources [8]. They then introduced a method for extracting a specific number of data blocks for verification during each check. They employed the homomorphic characteristics of RSA signatures to condense the evidence into a smaller value, markedly decreasing the communication overhead of the scheme while ensuring a high probability of successful file detection. Later, they proposed the MR-PDP scheme, which supports multi replica data verification but does not fully accommodate dynamic data operations. With the growing diversity in storage needs, researchers acknowledged that users require dynamic operations on cloud data, such as adding, removing, and modifying data. To meet this demand, scholars adapted the traditional PDP scheme by integrating dynamic data structures to organize data blocks, creating a solution that enables block level fully dynamic operations. They implemented the Merkle hash tree to verify the accuracy of data blocks' locations and used the BLS signature mechanism to guarantee the correctness of the content within data blocks. This scheme has certain optimizations in computational and communication overheads, and to reduce the burden on users 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 technology to construct the returned evidence, effectively hiding the data information and avoiding the risk of curious third parties obtaining user data 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 progress in computer science, cloud computing has emerged as a widely discussed subject, gaining recognition across various sectors and establishing itself as an unavoidable trend for the future of computer networks. An effective cloud computing platform must feature an advanced storage structure, which includes proficient storage administration, robust security protocols, adequate processing capabilities, and the provision of adaptable storage options for users' long term data storage and management needs. By merging accounting digitization with cloud computing, we can enhance financial sharing, significantly boost operational efficiency, and reduce hardware costs. Leveraging cloud computing technology alongside SOA architecture, we can offer more accessible accounting and financial digitization services to businesses, thus cutting down on information technology expenses. Building on prior research accomplishments, we propose a novel cloud-based platform for accounting digitization, comprising five distinct components: process oversight, SAP, document management, procurement oversight, and contract management.

3.2 Data Mining Based on Rough Set Theory

Feature selection plays a crucial role in data mining technology, focusing on identifying valuable features from

an extensive range of models and model collections while efficiently decreasing the complexity of the model set to optimize performance. Rough set theory also addresses this challenge within the realm of feature selection, aiming to lessen the complexity of the model set through careful selection. The strength of rough set theory lies in its ability to effectively sustain and enhance the correlation between the condition set and the decision set, leading to superior performance in certain areas. An information entity with n features can be partitioned into multiple segments, with the disparities among these features increasing, resulting in variations among the n features. In light of this situation, despite having numerous features, it remains impossible to erase the distinctions between them. Most feature selection algorithms strive to find the most minor reduction, meaning they aim to retain as many attributes as possible. However, this endeavor is quite challenging, which is why heuristic algorithms are frequently employed to achieve reduction swiftly under resource constraints to realize the desired outcome. By calculating the simplification and relative simplification of attributes, various effective strategies can be implemented: 1) Modifying the values of the Discernibility Matrix can significantly lower computational complexity, thereby attaining precise reduction of the information system. 2) By assessing the significance of attributes, we can extract valuable insights and progressively eliminate the least representative attributes to achieve attribute simplification. 3) The integration of genetic algorithms with other software technologies can lead to effective reduction. 4) Dynamic reduction serves as a means to abstract intricate information systems into manageable subsystems, facilitating faster task completion and more accurate result forecasting.

3.3 The Application of Cloud Data Parallel Verification Algorithm in Accounting Informationization

Utilizing basic digital analysis techniques, we can merge a company's financial information into the Internet across different geographical locations and timeframes, thus enhancing the company's operational efficiency. Nevertheless, the effectiveness of this approach hinges on our ability to leverage the most recent accounting information systems to ensure the company's cash flow. With the ongoing advancements in technology, financial sharing has evolved from its initial three transformations: First, its organizational framework has experienced significant modifications; second, its focus has transitioned from previously isolated financial management to a variety of financial services; and lastly, its security has seen considerable enhancements, guaranteeing the dependability of financial data. Consequently, we have developed a comprehensive cloud based data storage system to facilitate information exchange among various nodes, as illustrated in Figure 1.

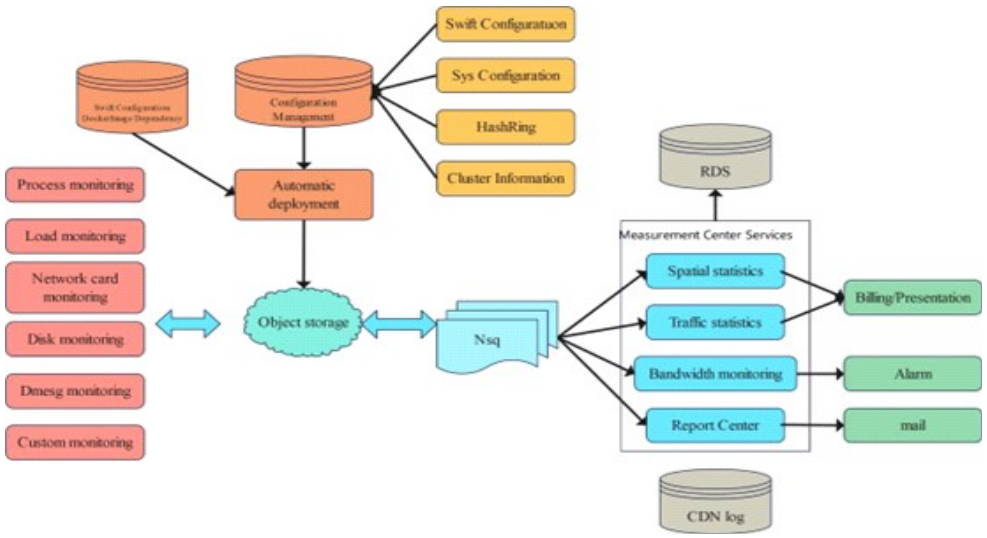


Figure 1. Network topology of Cloud storage system

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*=(*mi, m2,..., mn*) to ensure their security

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

H (*i*) represents a hash function, while x and u are a set of random numbers.

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

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

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

$$\mu_k = \sum_{i=1}^{n} v_i m_{k,i} + \mu_r \quad (3)$$

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

With the use of an Intel Core i5-8400 processor, 4GB of RAM, and a Seagate 500GB hard drive, we can effectively assess the performance of rough set based data mining techniques alongside traditional single user data verification methods. We employed Eclipse as the development environment, Java as the programming language, and SQL Server 2010 as our database management system. Presently, the emphasis on data integrity verification has shifted from conventional data block level positioning to more precise data corruption detection methods and versatile data analysis strategies, aiming to achieve swift and cost effective identification of data. Among these methods, traversal detection has emerged as the most prevalent data integrity identification technique, adept at accurately pinpointing damaged data areas, thus facilitating precise data identification. Binary search is recognized as a more complex detection technique, which segments data blocks into several independent regions for quicker identification and comparison, thereby enhancing the accuracy and speed of the verification process. However, the presence of corruption rates often influences the accuracy and speed of this technique, 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 previous classical methods, and their effectiveness was gauged based on the outcomes of these algorithms. As indicated in the performance comparison in Figure 2,

with the advancement of cloud computing environments, the time required for each file update using the single-user verification algorithm remains relatively stable. In contrast, the time is considerably diminished under the multi user parallel verification algorithm. By harnessing cloud computing's sharing capabilities, the implementation of parallel detection strategies can greatly enhance the system's computational efficiency in contrast to traditional single point detection methods.

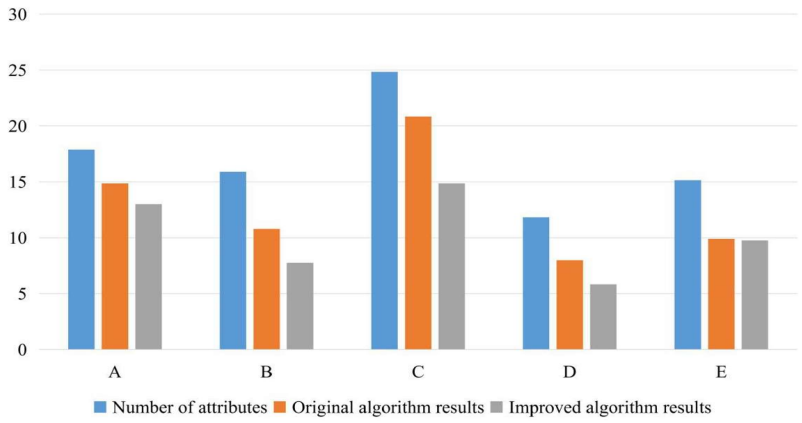
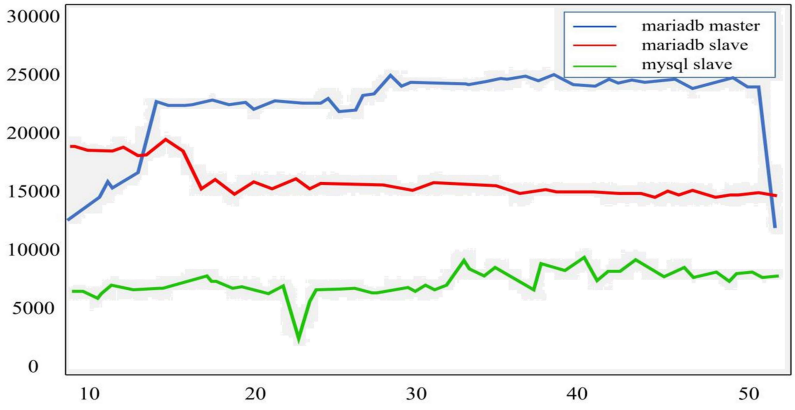


Figure 2. Performance Comparison of Data Integrity Verification Algorithms

After analyzing the fifth data set using two different algorithms, the result shows a remarkably similar reduction effect. This dataset comprises only 60 markers and 15 conditional attributes, yet it still reaches its maximum utility, highlighting the exceptional performance of the algorithm. On the other hand, an evaluation of the third data group with both algorithms yielded opposite results, indicating that even with limited processing, high efficiency can still be achieved once the workload surpasses a certain threshold, emphasizing the advantages of both algorithms in terms of processing capability. This system empowers users to choose their desired audit level independently, where opting for a higher audit level results in faster and more accurate detection of data block damage. Within this structure, audit levels are divided into three categories based on the likelihood of discovery and the frequency of audits. Assuming a user data error rate of 1%, to fulfill the requirements of audit level 1 within 24 hours, user data must undergo validation once, checking 230 data blocks each time; for audit level 2, it requires two validations, each time verifying 299 data blocks; and for audit level 3, four validations are necessary, with 459 data blocks checked at each instance. The verification time across different file sizes and audit levels is depicted in Figure 3, illustrating that a rise in audit level is associated with an increase in verification time, leading to greater costs for the user.



The graph indicates that batch verification consistently demands less computational time than traversal verification. Concerning communication overhead, thanks to the system's use of BLS signature homomorphism, the communication overhead for verifying multiple files is equivalent to that of ascertaining a single file. If a user needs to verify 50 files, the communication overhead for traversal verification is 50 times greater than that for batch verification.

5. Conclusion

This paper explored an innovative approach to data integrity detection. A model was developed utilizing cloud computing to support this approach. We also examined a new method for managing large scale data to improve support for complex business processes. Additionally, we aimed to establish a more efficient way of detecting extensive data to enhance business operations. Through this research, we found that, in contrast to traditional single-point verification methods, the rough set-based data mining verification algorithm not only ensures data completeness but also significantly decreases the costs involved in each update, thereby greatly improving 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. (2022). A review on data integrity verification schemes based on TPA and blockchain for secure cloud data storage. In *Security, Privacy and Data Analytics: Select Proceedings of ISPDA 2021* (pp. 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*(5), 757-771.

[8] Chen, M. (2021). Accounting data encryption processing based on data encryption standard algorithm. *Complexity, 2021*(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, 1754*(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.