Journal of Electronics and Information Science (2025)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2025.100120
ISSN 2371-9524 Vol. 10 Num. 1

Intelligent Data Collection Strategy for Distributed Storage Systems Based on Security Optimization Guidelines

Xiaoyu Deng

School of Engineering and Applied Science, University of Pennsylvania, Systems Engineering Philadelphia, PA 19104, USA
xiaoyud@alumni.upenn.edu

Keywords: Security optimization guidelines; Distributed storage system; Intelligent data; Collection strategy

Abstract: In the era of big data, distributed storage systems (DSS) are gradually replacing traditional storage systems and becoming the mainstream solution in the field of data storage due to their excellent storage performance and relatively low construction costs. In the widespread application of DSS, how to ensure the security, integrity, and efficiency of data has become a key issue that urgently needs to be addressed. Therefore, implementing intelligent data collection strategies under the guidance of security optimization is particularly important. This article proposes an innovative DSS intelligent data collection strategy that innovatively combines Ant Colony Optimization Algorithm (ACO) with Blockchain Technology (BT). ACO, with its powerful search capability and adaptability, can intelligently select the optimal path during data collection, thereby improving the efficiency of data collection. BT, on the other hand, provides strong guarantees for data security and integrity with its decentralized and tamper proof features. The experimental results show that this DSS intelligent data collection strategy combining ACO and BT not only significantly improves the efficiency of data collection, but also effectively ensures the security and integrity of data.

1. Introduction

With the rapid development of technology, especially the continuous innovation of mobile communication and IoT technology, we are in an era of data explosion [1]. The rapid increase in the number of mobile devices has enabled more and more devices to be interconnected, resulting in unprecedented massive data traffic. These data not only cover people's daily lives, but also delve into various fields of social development, such as economy, education, healthcare, environment, et al [2]. The widespread deployment and application of monitoring systems provide us with comprehensive insights into social life and ecological development, assisting various fields in continuously improving the accuracy of data analysis, rational allocation of resources, and real-time monitoring capabilities of the environment [3]. In the field of data storage, traditional centralized storage systems are inadequate when faced with such a large amount of data. Centralized storage systems typically rely on a single or limited storage device, and when the amount of data exceeds their processing capacity, performance bottlenecks can occur, even leading to system crashes [4].

In addition, centralized storage systems have poor performance in scalability, and once hardware devices reach their storage or computation limit, complex upgrade or replacement operations are required, which not only increases maintenance costs but may also lead to service interruptions. In contrast, DSS has gradually become a new favorite in the field of data storage due to its unique advantages [5]. DSS achieves redundant backup and load balancing of data by storing it on multiple independent physical nodes, greatly improving the reliability and performance of the system [6]. When the existing storage hardware devices cannot meet the storage requirements, DSS can expand cluster capacity by adding or upgrading underlying hardware facilities [7]. This flexible expansion method enables the system to easily cope with rapid growth in data volume without the need for complex system reconstruction [8]. DSS adopts standardized interfaces and protocols, allowing different types of hardware facilities to be perfectly compatible in the system. This not only reduces the procurement and maintenance costs of hardware equipment, but also improves the flexibility and scalability of the system.

DSS adopts a multi-tenant isolation design, optimizing the mechanism for handling concurrent requests from users. By distributing read and write requests to multiple physical nodes, the system can achieve higher read and write efficiency, thereby meeting the requirements of high concurrency and low latency applications. For suppliers, the underlying hardware infrastructure of DSS can be built using a large number of inexpensive storage devices, which greatly reduces the system's construction cost. However, these inexpensive hardware storage devices also bring certain security risks, such as data node failure, data loss, and other issues. To address these challenges and ensure the security, integrity, and efficiency of data in DSS, this paper proposes an innovative DSS intelligent data collection strategy. This strategy cleverly combines ACO with BT, aiming to achieve intelligent path selection and decentralized, tamper proof storage of data during the data collection process.

2. ACO and BT

2.1. ACO

ACO, the core concept of intelligent technology that simulates ant foraging behavior originates from biological evolution and collective intelligence [9]. In the field of multi-sensor data fusion, this algorithm not only breaks through the limitations of traditional data processing methods, but also significantly improves the accuracy and efficiency of data fusion [10]. In this application, sensor nodes are regarded as nodes in the ant colony search path, and the correlation and importance between data are reflected through the concentration of pheromones. At the beginning, each ant can explore all potential paths and choose the next step based on the concentration of pheromones on the path. The higher the concentration of pheromones, the greater the likelihood of the path being selected. This mechanism converts the correlation strength of data into path preference, making pheromone concentration the key to evaluating data importance. As the search deepens, ants leave pheromones on their paths, which gradually dissipate over time, simulating the natural updating of pheromones. This feature enables the algorithm to adaptively adjust the search direction, focusing on the path that is most likely to contain useful information, i.e. the optimal data fusion solution.

The Elite Ant Colony Algorithm (EACA) is an innovative upgrade of traditional ACO, with the adoption of a more optimized strategy to screen and optimize ant colony composition. This algorithm identifies and maintains the top "elite ants" in the search, who, with enhanced pheromone release power, lead the ant colony to quickly approach the optimal solution in subsequent iterations, thereby significantly improving efficiency and accuracy in processing complex multidimensional sensor data. This algorithm simulates the collaborative and competitive behavior among ants, and can autonomously restructure multidimensional data such as medical equipment in a regular manner without the need for preset rules, demonstrating excellent adaptability and flexibility. Relying on the parallel search capability of ant colonies, the algorithm can process massive sensor data in parallel, quickly extract and integrate their features, and significantly reduce the data processing cycle. Meanwhile, by utilizing a positive feedback mechanism, the algorithm continuously corrects fusion errors to ensure the accuracy and correctness of the final results.

2.2. BT

Since its inception, BT has developed rapidly and is gradually becoming a key force in promoting national new technology innovation and industrial upgrading. Its notable features - decentralization and tamper resistance - provide a solid foundation for data security and integrity, leading a new transformation in data management and trust systems. This technology is based on a unique distributed "ledger" architecture, using a chain data structure to record data. Each block contains transactions or information within a specific time period, and is connected to the previous block through cryptographic techniques to form a continuous and immutable data chain. This design ensures the timing and integrity of data, and utilizes consensus mechanisms to ensure consensus and fairness among all nodes when adding data, achieving transparency and tamper resistance of data.

In addition, another advantage of BT is its data redundancy storage mechanism. The data is stored in multiple replicas to ensure consistency between replicas, effectively prevent single point of failure (SPOF) and data loss, and achieve high reliability and persistent storage of data. This distributed storage method means that even if some nodes are attacked or fail, the data security of the entire blockchain network is not affected, because the integrity and authenticity of the data are jointly guaranteed by the entire network. At the same time, BT also widely applies cryptographic technology to provide security guarantees for data transmission and storage. Whether it is data encryption storage or transaction authentication, cryptography is the core of blockchain security. With the help of complex encryption algorithms and key management, BT ensures that data can only be accessed and modified by authorized users, effectively preventing data leakage and illegal tampering.

3. Algorithms and Experiments

3.1. Algorithm Design

In the context of integrating multi-sensor data, firstly we define a probability t_{ij}^{k} to represent the possibility that ant individual k selecting node j as per EACA. On this basis, the specific expression of ant individual inspiration value μ_{ij} is refined as follows:

$$\mu_{ij} = \frac{1}{w(i,j)} \tag{1}$$

In this formula, $w(i,j)$ represents the elements of the sparse weight matrix. In the operation mechanism of EACA, the neighboring nodes of ant individuals have a certain degree of influence on the pheromone volatilization behavior of elite ant individuals. When t represents the volatility of pheromones, the specific mathematical expression of the pheromone volatility function τ_{ij} is presented as follows:

$$\tau_{ij} = (1-\rho)\,\tau_{ij} \tag{2}$$

In this formula, ρ represents the pheromone attenuation coefficient.

In the application scenario of the blockchain model, we regard all mobile edge computing (MEC) servers as the nodes of the blockchain, which shoulder the responsibility of providing storage and management services for the DSS edge network. By ingeniously integrating smart contracts into the blockchain system, we can achieve self-organizing completion of data content caching and credibility evaluation. This not only deepens the mutual trust between edge servers, but also significantly improves the operational efficiency of edge networks. During each time slot period, the DSS server will send a bill containing billing content to the MEC server. Subsequently, each MEC server e will broadcast the collected transaction information to the entire network. The blockchain system will randomly designate a master node, which is responsible for integrating transaction information into new data blocks based on the preset packaging timeout $T_i(t)$ and the maximum block size $S(t)$ limit. After completing the data packaging, the master node will further perform strict verification on the signature and message authentication code (MAC). The calculation cycle of this series of operations is clearly defined as $B_{p1}(t)$.

$$B_{p1}(t) = \frac{d(t)}{\delta(t)}(\alpha + \beta) \tag{3}$$

In the formula, $d(t)$ represents the total transaction size in the block, and $\delta(t)$ represents the average transaction size.

3.2. Experimental Result

In order to fully verify the effectiveness and innovation of the proposed strategy, we carefully planned and executed a series of comparative tests, comparing it in detail with the traditional Particle Swarm Optimization (PSO) Algorithm strategy. Figure 1 presents a visual comparison of the two in terms of data collection time. As clearly shown in the figure, our strategy exhibits excellent reduction in data collection time compared to traditional PSO. Specifically, our strategy significantly reduces the time required, which strongly demonstrates its efficient performance in dealing with complex data collection tasks. The root of this efficiency lies in our clever integration of EACA and BT. EACA, with its excellent search capability and self-organizing properties, can quickly capture key information in data, shorten the search process, and thus improve data collection speed. The addition of BT further consolidates the security and reliability of the data, ensuring the transparency and credibility of the collection process. The synergistic effect of the two not only maximizes their respective advantages, but also stimulates additional synergistic effects, resulting in significant optimization of our strategy in terms of data collection time.

Figure 2 visually presents the outstanding performance of the proposed strategy in terms of data collection integrity compared to traditional PSO. It is evident that the data integrity obtained by this strategy is significantly better than traditional methods, effectively reducing the risk of information loss or damage during the collection process and ensuring the comprehensiveness and accuracy of the data. This outstanding achievement is attributed to the innovative integration of EACA and BT in the strategy proposed in this article. EACA, with its excellent search capabilities and high adaptability, delves deep into data sources and accurately captures key information, not only improving collection efficiency but also ensuring data integrity from the source. The addition of BT has built an unbreakable security line for data collection. Its decentralized and tamper proof features ensure the security and reliability of data transmission and storage. Even in the face of tampering or leakage threats, it can be quickly identified and corrected through the verification mechanism of blockchain, effectively maintaining the authenticity and integrity of data.

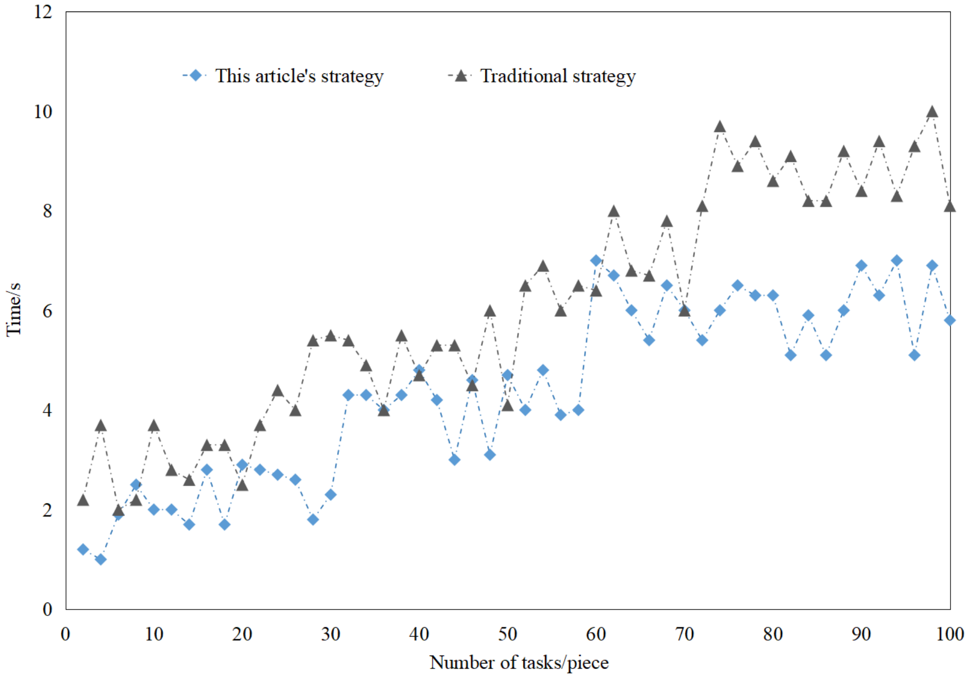

Figure 1 Time comparison

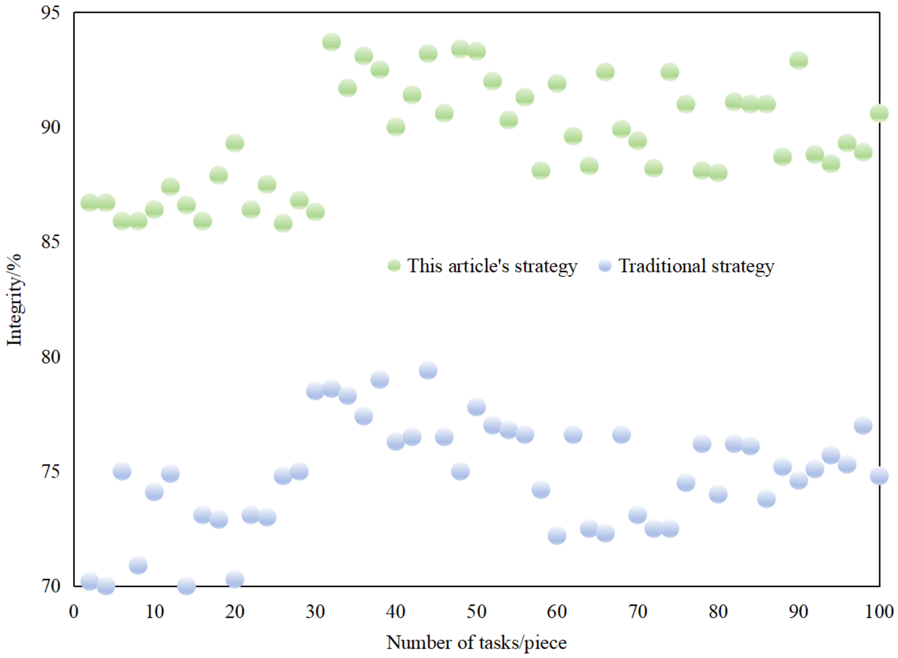

Figure 2 Comparison of data integrity

4. Conclusions

This article proposes a groundbreaking DSS intelligent data collection strategy that ingeniously integrates EACA and BT, aiming to address the efficiency and security issues currently faced in the field of data collection. EACA, with its excellent search capability and high adaptability, can intelligently identify and select the optimal path in complex data collection environments, greatly improving the efficiency of data collection. At the same time, BT's decentralized and tamper proof features have built an unbreakable security line for data transmission and storage processes, ensuring the integrity and authenticity of data. The experimental data fully verified the effectiveness of this strategy, and the results showed that compared with traditional methods, the DSS intelligent data collection strategy combining EACA and BT achieved significant improvement in data collection efficiency. At the same time, the security and integrity of the data were effectively guaranteed. This achievement not only provides new ideas and methods for data collection in the DSS field, but also provides useful references for data processing in other fields. However, no strategy is perfect, and this strategy also has certain limitations. For example, there is still room for further optimization in terms of algorithm complexity, resource consumption, and other aspects. In addition, with the continuous development of technology, how to better adapt to new application scenarios and needs is also a direction that needs to be deeply researched and explored in the future. Nevertheless, the innovation and practicality of the strategy presented in this article are still worthy of recognition, injecting new vitality into the development of the field of data collection.

References

[1] Li T. Optimization of Clinical Trial Strategies for Anti-HER2 Drugs Based on Bayesian Optimization and Deep Learning[C]//Proceedings of the 2025 5th International Conference on Bioinformatics and Intelligent Computing. 2025: 163-168.

[2] Huang S, Liang Y, Shen F, et al. Research on Federated Learning's Contribution to Trustworthy and Responsible Artificial Intelligence[C]//Proceedings of the 2024 3rd International Symposium on Robotics, Artificial Intelligence and Information Engineering. 2024: 125-129.

[3] Huang S, Diao S, Wan Y, et al. Research on multi-agency collaboration medical images analysis and classification system based on federated learning[C]//Proceedings of the 2024 International Conference on Biomedicine and Intelligent Technology. 2024: 40-44.

[4] Gao K, Yoo Y, Schecter A. Open-Source AI Community as' Trading Zone': The Role of Open-Source Models in the Diffusion of Artificial Intelligence Innovation Completed Research Paper[C]//Forty-Fifth International Conference on Information Systems, Bangkok, Thailand. 2024.

[5] Zhou Yefen. Construction of Intelligent Management Platform for Chemical Industry Parks Based on Internet of Things and Artificial Intelligence [J]. Technology and Market, 2024, 31 (02): 41-44+49.

[6] Xiao Ping, Liu Jingxin, Wang Yan, et al. A Video Multimodal Data Adaptive Acquisition Strategy [J]. Small Microcomputer System, 2023, 44 (02): 383-391.

[7] Chen Kelong, Zhong Jiansheng, Shen Yajun. Research on Optimization Method of Medical Equipment Operation Data Collection Based on Wireless Sensor Network [J]. Medical and Health Equipment, 2023, 44 (09): 83-87.

[8] Nie Yongquan, Li Jianshe, He Yubin, et al. Optimization strategy for unmanned aerial vehicle assisted intelligent grid fault terminal data acquisition [J]. Computer Application Research, 2023, 40 (12): 3723-3727.

[9] Guo Yong'an, Zhou Yi, Wang Quan, et al. Optimization Strategy for Multi Edge Server Collaborative Video Stream Caching Assisted by Blockchain [J]. Data Collection and Processing, 2023, 38 (06): 1353-1368.

[10] Qin Nan, Zheng Jingli, Wu Chi, et al. Architecture Design and Key Strategy Research of Intelligent Recommendation System for University Information [J]. Modern Educational Technology, 2023, 33 (12): 100-110.