Information System Reliability and Fault Root Location Based on PageRank Iterative Algorithm

Yushan Zhao[1*], Na Xiao[1], De Meng[1]

1 Platform Operation and Security Center, Information and Communication Branch of State Grid Jibei Electric Power Co., Ltd., Beijing 100053, China

Email Information:

Yushan Zhao: zzzyyy96@126.com

Na Xiao: xnpaper2023@163.com

De Meng: mengde8801@163.com

***Corresponding author, Email: zzzyyy96@126.com**

Abstract: To improve the information system's reliability and stability, this paper combines the PageRank iterative algorithm and Genetic Algorithm (GA) to optimize the information system. Firstly, this paper analyzes the structure of the risk assessment model of the information security system. Secondly, the application of the PageRank iterative algorithm and GA in information systems is expounded. Finally, intelligent computing and GA are combined to optimize the risk assessment system, and horizontal comparative experiments prove the rationality of this study. The experimental results show that the accuracy of the optimized model in fault root cause location is much higher than that of the traditional model. The accuracy of locating network risk and equipment risk reaches 97.2% and 96.9%, which verifies the reliability of the optimized model in this paper. Then, through the simulation experiment, the 10-13 cycle of the assessment cycle is attacked, and it shows that the situation value of the information system optimized in this paper has improved during this cycle, with the highest value reaching 24.7. At the same time, the response time of the time-consuming operating system has also decreased to an average of 3.76s, which shows that the performance of the information system optimized in this paper has been dramatically improved. The optimization model in this paper has specific reference significance for the research of information system reliability and fault root cause location.

Keywords: PageRank iterative algorithm; Information system; Fault root cause location; Genetic algorithm; Intelligent computing

1. Introduction

The reliability of information systems and the fault root cause location have always been essential research directions in computer science [1]. Information systems reliability enhancement necessitates a comprehensive strategy that includes handling complexity, integrating technologies, resolving technical debt, cybersecurity concerns, data quality, performance optimization, and regulatory compliance. With the rapid development of the Internet and the popularization of large-scale network applications, the reliability of information systems has become particularly critical. Users have increasingly high expectations for the availability and stability of information systems while facing an increasing number of failures and security threats [2]. Using risk-based decision-making, incident response integration,

and data fusion, algorithms employ behavioral analysis, biometric identification, surveillance, and geofencing to improve physical security measures, minimize risks, and expedite the response to security incidents.

Regarding information systems, organizations use tactics like risk assessment, strong security measures, incident response planning, backup, disaster recovery, redundancy, constant monitoring, staff training, frequent updates, patch management, and cooperation to combat security threats. These steps support system resilience, failure impact minimization, identification of possible failures, prioritization, and continuance of vital business processes. In the information system, failures will lead to the system being unavailable. It also may cause serious consequences, such as data loss and user privacy leakage [3]. Therefore, timely and accurate location of the root cause of failure is essential to maintain the reliability of the information system. Information systems root cause analysis is challenging because of linked systems, heterogeneous technologies, dynamic nature, volume of data, failures, legacy systems, technical debt, cross-functional cooperation, and time and resource limitations. However, fault root cause location is often challenging due to the complexity and sheer size of information systems. Enhancing fault root cause location efforts requires cooperation and knowledge exchange. They provide a culture of continuous development, knowledge, validation, brainstorming, cross-functional ideas, and shared learning. To improve system resilience and dependability, this enables teams to investigate system failures, pinpoint underlying causes, and dismantle silos.

As a classic link analysis algorithm, the PageRank iterative algorithm has been widely used in information retrieval and network analysis. One search engine tool that ranks websites according to the number and quality of connections is the PageRank algorithm. It employs a random surfer model, links-based scoring, and a dampening factor to avoid infinite loops and spider traps. This algorithm has transformed web search and information organization. The principle is based on the link relationship between nodes in the network diagram, and the weight of nodes is iteratively calculated to evaluate the importance of nodes [4]. In recent years, researchers have begun to apply the PageRank algorithm to the reliability of information systems and the fault root cause location. Component identification, graph creation, relationship definition, weight assignment, normalization, algorithm application, outcome analysis, and interpretation are all steps in the process. Accuracy and relevancy must be improved over time via iteration and refinement. They have achieved some encouraging results. Based on the PageRank iterative algorithm, this paper proposes a new method for locating information systems and finding root causes of faults. The PageRank algorithm is used to analyze the correlation between various components in the information system, and the probability of the fault root cause is determined by calculating the node's importance. The method gives each node an equal weight and recalculates values depending on the scores of arriving neighbors. At high inbound link counts, a damping factor avoids buildup by iterating numerous times until convergence. Node relevance is represented by the final PageRank scores, where higher scores denote a more significant impact. Meanwhile, other related technologies and methods will be combined to improve the accuracy and efficiency of fault root cause location.

The research in this paper is of great significance for improving the reliability of information

systems and reducing the impact of faults on systems. By pinpointing the root cause of failures, system administrators can take action more quickly to fix failures, reducing system downtime and data loss. In addition, the research results can provide valuable reference and guidance for designing and maintaining information systems.

In summary, this paper first discusses the Information Security System (ISS) risk assessment model, including structure, PageRank iterative algorithm, and Genetic Algorithm (GA) in information systems. Second, intelligent computing and GAs are collected to optimize the risk assessment system. Finally, through horizontal comparison experiments, the rationality of the research in this paper is proved.

This study optimizes information systems using a combination of the Genetic method (GA) and the PageRank iterative method. It uses GA to increase stability and reliability while analyzing the structure of risk assessment models. With 97.2% and 96.9% accuracy for network and equipment risks, respectively, the optimized model outperforms the other models in failure root cause location. The optimized model is essential for information system reliability and problem root cause localization study as simulation trials demonstrate better situation value and response time. It suggests a PageRank iterative approach to enhance fault root location and information system dependability. It helps with defect root cause analysis, resource prioritization, and essential component identification. Performance optimization, risk reduction, and proactive maintenance are all aided by this strategy.

2. Literature review

In the study of information system failures, Pandey et al. (2020) analyzed the impact of information network faults on the three stages of fault location, isolation, and recovery in detail based on the general process of fault self-healing of distribution physical systems. The time of each stage of the physical system fault self-healing process was corrected when the information system backbone, access network, and interface layer failed. An information system effectiveness assessment model was established. Based on the improved least-pass method, the reliability of the Communications Processing Distribution System (CPDS), considering the influence of the information transmission process, was evaluated. A network design known as the Communications Processing Distribution System (CPDS) manages protocol conversion, distributes communication loads, assures fault tolerance, facilitates scalability, boosts security, and offers efficient administration and monitoring. It uses technical ideas such as middleware, routing algorithms, protocol adherence, and message queuing. The proposed method was verified by example, and the influence of information system access network structure and failure rate of information system components on the reliability of CPDS was analyzed [5]. Wiedenhöft et al. (2020) designed a fault diagnosis model for command information system equipment based on the Self Organizing Map (SOM) network algorithm and simulated the sample data.

An unsupervised neural network model for visualizing and organizing high-dimensional data is called the Self-Organizing Map (SOM). All of its neurons are arranged in a grid and trained to recognize input patterns by repeatedly changing their weights. Important aspects include unsupervised learning, clustering, dimensionality reduction, and topology preservation. They verified the failure mode recognition ability using the sample set of different characteristic parameters. They used the test samples to be diagnosed in the power supply system to analyze the trained SOM network model. The original

SOM network model was improved to solve the problems of low diagnosis rate and insufficient classification information due to an unsatisfactory clustering effect. The SOM network optimized by particle swarm was connected in series to form a composite neural network, and the improved network model could be verified by case simulation to improve the fault diagnosis rate and meet the equipment fault diagnosis requirements [6]. Ranjha and Kaddoum (2020) proposed a black-box testing method based on code coverage, introducing white-box measurement methods into black-box testing techniques. It was guided by black box testing techniques and measured by code coverage to gradually achieve the completeness of test coverage through iteration of test case design execution and code coverage analysis. This approach was combined with the design strategy of the test case to visualize the test process and results. It ensured the communication subsystem and the acquisition subsystem quality of the protection and recorder in the most core part of the power grid fault information system. It improved the stability and reliability of the system. A set of test strategies suitable for the power grid fault information system was proposed through research on the test strategy at each stage of the software life cycle and the analysis of the characteristics of the power grid fault information system. The black-box testing strategy based on code coverage effectively solved the problem of test result measurement [7].

Therefore, the root cause of failure can be effectively located using the link relationship between nodes and the importance index of nodes, and the reliability and stability of the information system can be improved. However, traditional research still has some challenges, such as the complexity and scale of information systems and the optimization and improvement of algorithms, which require further study and exploration.

3. Information system reliability and fault root cause location based on PageRank algorithm

3.1 Structural design of security information system risk assessment model

The ISS needs to realize the two main functions of risk assessment and network security situation awareness, complete the security guarantee of the information system, and include auxiliary functional modules, such as the information collection module and system management module [8, 9]. The information-gathering module gathers data from several sources, whereas the system management module manages the system's functionality, emphasizing optimization, configuration, and monitoring. While one module collects data and the other ensures system functionality, both guarantee efficient and effective working. The functional and non-functional requirements of the ISS are shown in Figure 1. A vulnerability and risk assessment, pattern recognition, threat intelligence integration, vulnerability and risk assessment, security incident monitoring, behavioral analytics, report generation, and alert generation are all functions of an information security system (ISS). Taking a proactive stance reduces security risks, protecting resources, information, and operations from online attacks.

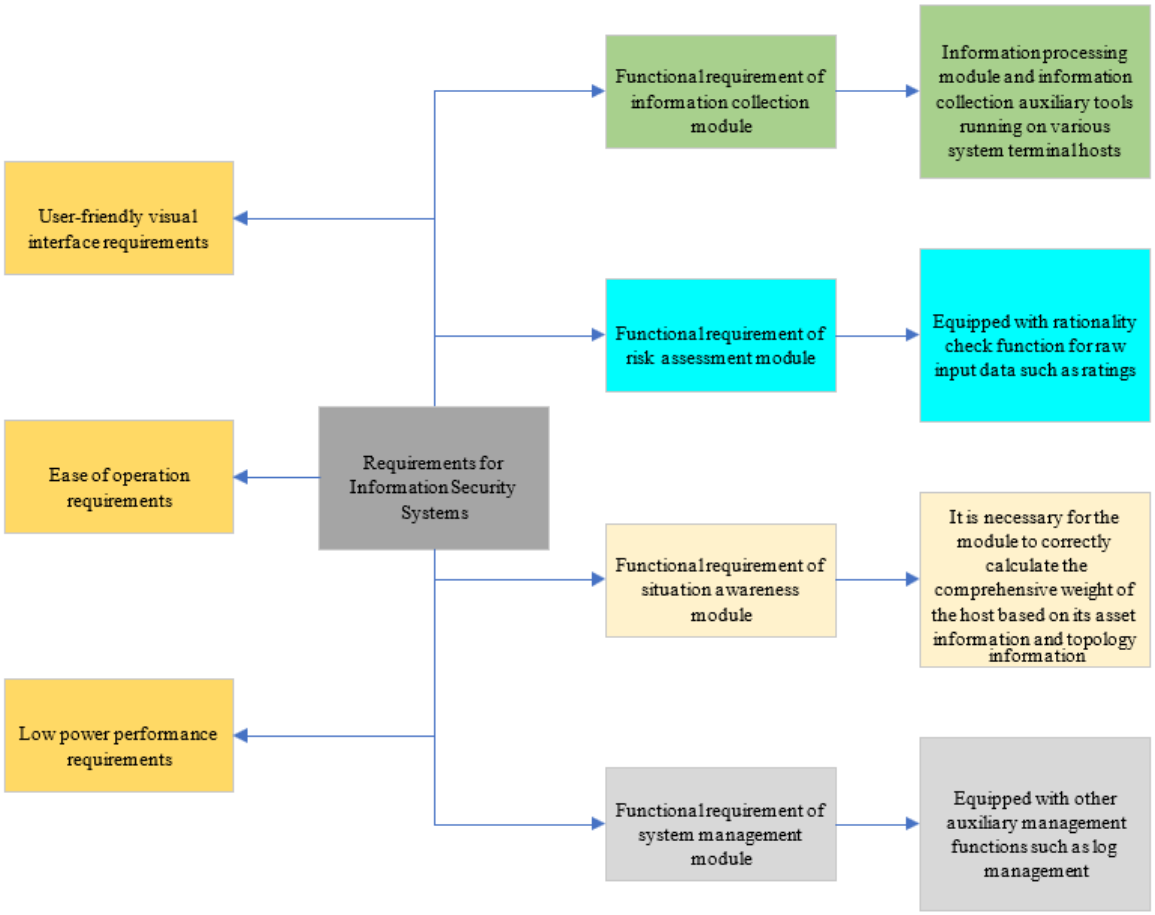

Figure 1: Functional requirement and non-functional requirements of ISS

There are three main stages in the information collection module process: the information collection judgment stage, the information collection stage, and the information reporting and storage stage [10-12]. The real-time data of intrusion detection systems is collected automatically all the time, so there is no active collection or other methods [13]. The flow of the information collection module is shown in Figure 2.

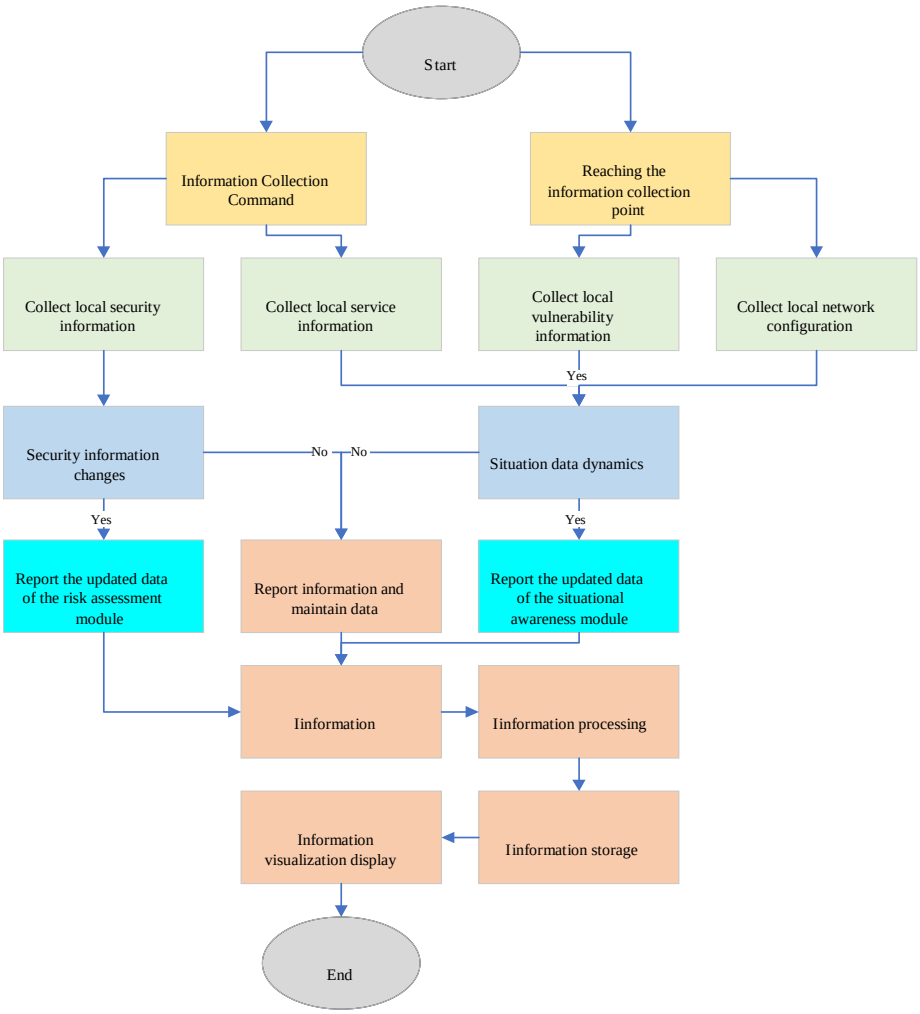

Figure 2: The process of information collection module

The information system risk assessment module is one of the key links to ensure the security of the information system, which interacts with the expert group participating in the assessment to complete the risk assessment function of the information system [14-17]. The information system risk assessment module works with expert groups to increase the precision and potency of risk assessments. Experts offer domain-specific expertise, assist with data gathering and analysis, choose risk indicators, determine weights, and estimate risk values. This module is mainly divided into four stages: assessment preparation stage, risk indicator scoring stage, weight calculation stage, and risk value calculation stage [18]. The company builds a team with knowledge of information security and risk management, determines the scope of the assessment, and establishes criteria and procedures before the evaluation. By doing this, compliance with legal and industry norms is guaranteed. The assessment team finds and rates risk indicators, including threats, software vulnerabilities, access restrictions, and problems with regulatory compliance, according to how likely they are to occur and how they could affect the company's goals. Both qualitative and quantitative analysis may be used in the scoring process. Using input from stakeholders, subject matter experts, and organizational leadership, weighting is a risk management method that ranks risk indicators according to importance and ensures that risk management priorities are appropriately reflected. Risk value calculation involves scoring and

weighting risk indicators and then calculating the overall risk value for each identified risk. This quantitative measure helps prioritize mitigation, allocate resources effectively, and inform decision-making at tactical and strategic levels within an organization. In the assessment preparation stage, evaluators must collect and organize information, collect various equipment, networks, data, and other relevant information in the information system, and organize and classify them to provide basic data for subsequent assessment [19]. In the risk indicator scoring stage, the evaluator needs to select the corresponding risk indicator for scoring based on the collected data [20, 21]. Making informed decisions, maximizing resource allocation, improving validity and accuracy, and enabling customized risk management strategies depend on selecting pertinent risk indicators throughout the risk assessment stage. Assuring an effective and efficient procedure guarantees that the evaluation is in line with the organization's strategic objectives and priorities. These risk indicators can include the stability of the device, the reliability of the network, and the confidentiality of the data, and the evaluator needs to score these indicators according to the actual situation to calculate the risk value later [22]. It is advised to have clear evaluation criteria, evaluator training, standardized scales, multiple evaluators for peer review, documentation, regular review processes, quality assurance checks, and continuous feedback to preserve uniformity and dependability in scoring subjective factors such as device stability and connectivity. In the weight calculation stage, the evaluator needs to calculate the weight of different risk indicators to determine their risk contribution to the entire information system [23]. These weight calculations can be performed by expert scoring methods and analytic hierarchy methods [24]. Selected experts with pertinent information about an information system are tasked with finding risk indicators as part of the expert scoring technique. Scoring standards are established for every indication, and each expert assigns a score independently. Disputes are settled through consensus-building meetings. Compiling the weighted average or median of the scores and ensuring their sum equals one determines the final weight. The Analytical Hierarchy Process (AHP) is a systematic approach to risk assessment, goal definition, criteria definition, alternative evaluation, significance determination, consistency, and correct weight representation of the organization's risk management objectives, all while maintaining validity and dependability. In the risk value calculation stage, the evaluator needs to calculate the risk value of each risk indicator and the overall risk value of the entire information system based on the scoring and weight calculation results. To foster consensus-building and well-informed decision-making in risk management, evaluators maintain accuracy and reliability in risk value calculation by using aggregation models, conducting sensitivity analyses, adopting a consistent scoring methodology, offering training, putting peer review into practice, guaranteeing data quality, and upholding transparency. These risk values can be calculated by certain mathematical equations to subsequently formulate corresponding risk control plans [25, 26]. The flow of the risk assessment module is shown in Figure 3.

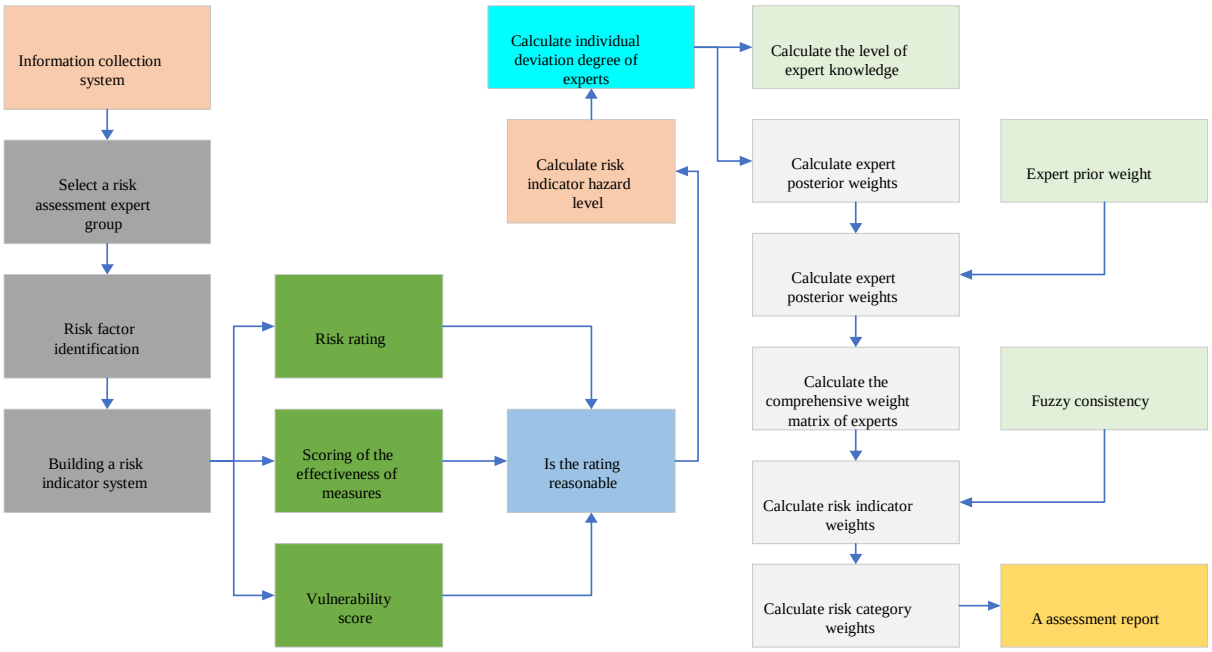

Figure 3: Process of risk assessment model

The situation awareness module mainly completes the relevant functions of the situation awareness framework [27, 28]. The process consists of three stages: the data acquisition stage, the host weight calculation stage, and the network security situation calculation stage [29]. Its module flow is displayed in Figure 4.

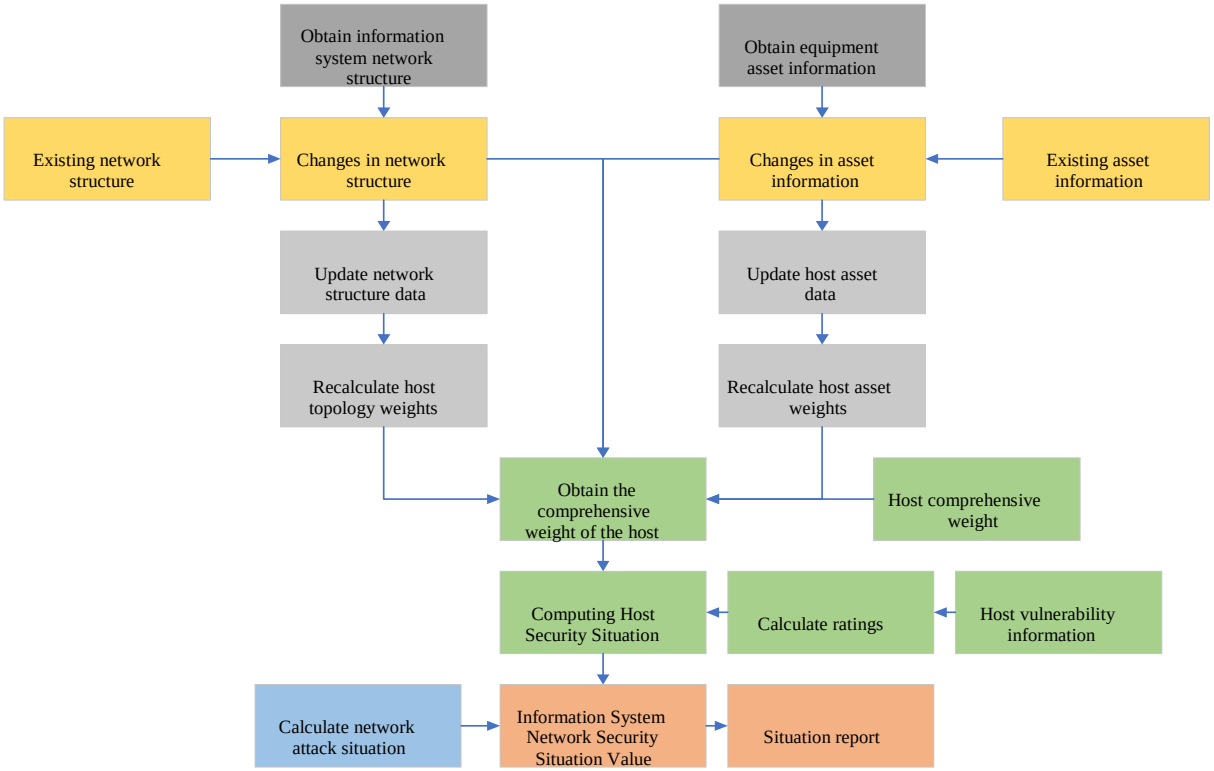

Figure 4: The process of situation awareness module

The main functions of the information acquisition stage are initialization and information gathering functions [30]. A security risk assessment's information acquisition stage includes setting up the system,

defining the scope, deploying asset discovery tools, automating scanning tools, gathering and analyzing log data, interviewing and surveying stakeholders, reviewing current policies and documentation, and compiling the information for analysis and assessment. The information gathering aids running on each host are first started. Then, the Security Risk Assessment (SRA) system is begun [31, 32]. According to the collected system information, the comprehensive weight of the host is calculated, and the security situation of the host is calculated [33]. According to the alarm data of the intrusion detection system, the network attack situation is calculated. Finally, the information system's overall network security situation assessment value is fused. An information system's fused network security situation assessment value offers a thorough risk assessment, improves situational awareness, supports well-informed decision-making, speeds up incident detection, and guarantees ongoing insight into the changing threat landscape. It also proves that internal security guidelines and legal requirements are being followed. The risk assessment phase assists experts in completing the risk assessment function of the information system [34]. According to the risk index system and scoring established by the assessment expert group, the risk degree of the risk index is first calculated. Next, all experts' combined weights and risk indicators' weights are calculated. The selection of risk indicators for risk assessment considers several factors, including interdependencies, historical data, expert judgment, influence on business operations, chance of occurrence, severity of repercussions, and availability of data. Then, the risk value of the information system is calculated based on the risk degree and weight information [35]. The overall flow of the SRA system for information systems is given in Figure 5.

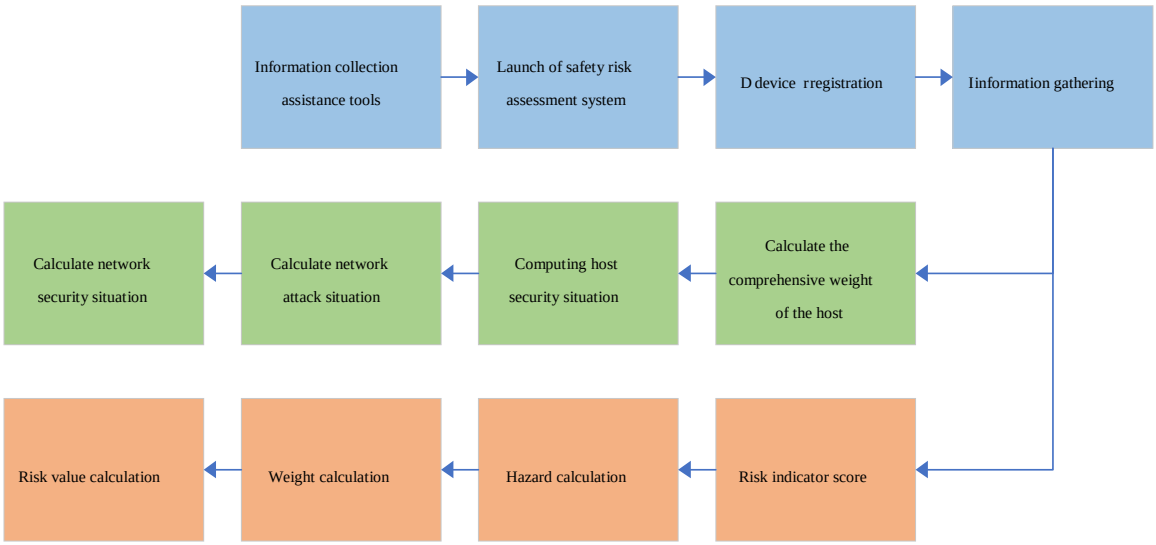

Figure 5: The overall process of the SRA system for information systems

3.2 Application of PageRank iterative algorithm and GA in information system

In the context of the increasing Internet scale, some traditional security protection methods, such as firewall technology, intrusion detection systems, virtual private networks, and other technologies, resist network attacks to a certain extent and protect the security of information systems. However, these technologies do not guarantee absolute security [36, 37]. Besides, with the increasing size of networks, these technologies cannot visually quantify the system's security status. For example, continuously

operating intrusion detection systems tend to generate massive amounts of alarm data, but a large amount of this data is often irrelevant alarms [38]. As a result, managers still need a comprehensive understanding of the security threat situation of the system in the face of a large amount of data, resulting in the inability to manage the system better or take defensive measures. Therefore, it is necessary to increase awareness of information system security situations [39].

The PageRank algorithm is a well-known web page sorting algorithm used in search engine-directed networks based on quantitative and quality assumptions. If a webpage is linked to more web pages, the webpage becomes more critical. When several other websites link to a webpage, it becomes more important. This is because it becomes more noticeable, draws more traffic, gains credibility from search engine algorithms, gets traffic from referrals, has better exposure in search results, and functions as social validation, which raises its reputation even further. The more critical a webpage is, the more weight it will pass to other web pages through directed edges [40, 41]. The calculated host topology weights are as follows.

$$w_k(i) = \alpha \sum_{j \in indeg_i} \frac{w_{k-1}(j)}{outdeg(j)} + \frac{1-\alpha}{N} \quad (1)$$

In Eq. (1), w is the topology weight, i is the host, and α is the recommended value. $indeg_i$ is the list of running services, and $outdeg$ is the collection of outbound hosts. N is the total number of hosts, j is the number of out-of-chains, and k is the parameter item. Then, the weights are normalized.

$$w(i) = \frac{1}{2}(w_a(i) + w_k(i)) \quad (2)$$

In Eq. (2), $w(i)$ is the overall weight, and w_a is the initial weight of the host. The PageRank algorithm is a computational process that helps hosts improve their cybersecurity situation awareness [42]. A GA is a search algorithm based on natural selection and genetic mechanisms to solve complex problems. In GAs, the solution to the problem is represented as chromosomes, and different chromosomes are evaluated using a fitness function [43]. The fitness function measures the performance of chromosomes for selection, crossing, and mutation operations to generate new solutions [44]. The advantage of GAs is that they can find the optimal solution in the large-scale search space without prior knowledge of the structure of the search space [45, 46]. In addition, GAs can effectively deal with multimodal problems, which contain multiple optimal solutions. Genetic Algorithms (GAs) are solid and adaptable for real-world optimization because of their population-based approach, variety, resistance to noise, balance of exploration and exploitation, and capacity to solve multimodal issues. GAs have been widely used in various fields, such as engineering design, economics, and computer science [47, 48]. The use of GAs to find potential root causes of failures in information systems is expected to improve the system's reliability and efficiency [49].

3.3 Fault location optimization of information system

This paper designs a new information system to meet the functions of security risk identification, system risk assessment, and network security situation awareness of the information system and effectively protect the information asset security of the information system. The specific system architecture is given in Figure 6.

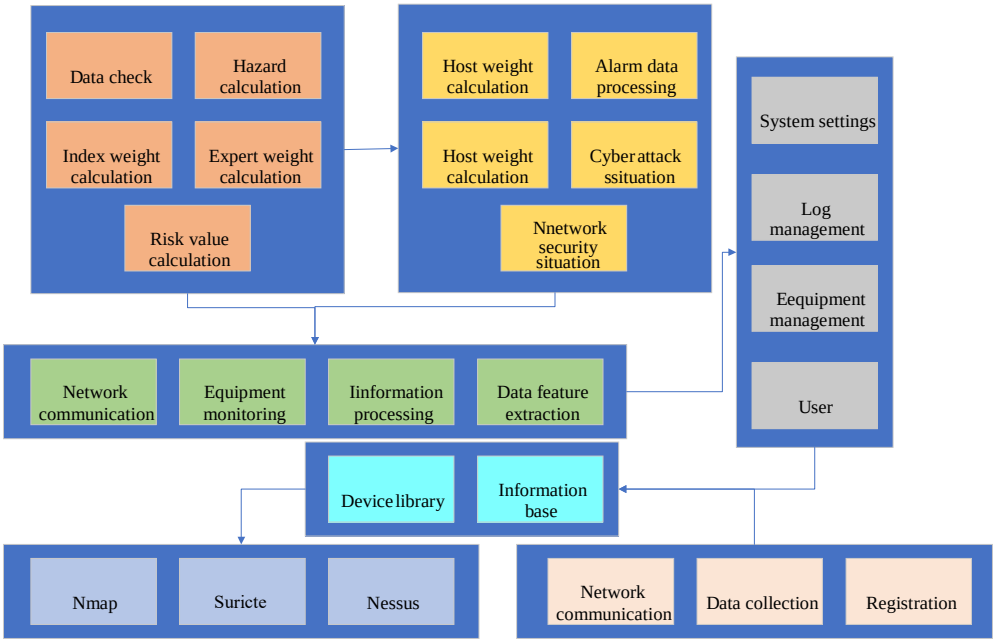

Figure 6: Overall architecture of SRA system for information systems

The risk assessment module interacts with the user the most and completes the risk assessment function of the information system. In implementing this module, the steps are divided into five steps: system information, evaluators, risk indicator system, risk index scoring, and risk assessment. The flow is presented in Figure 7.

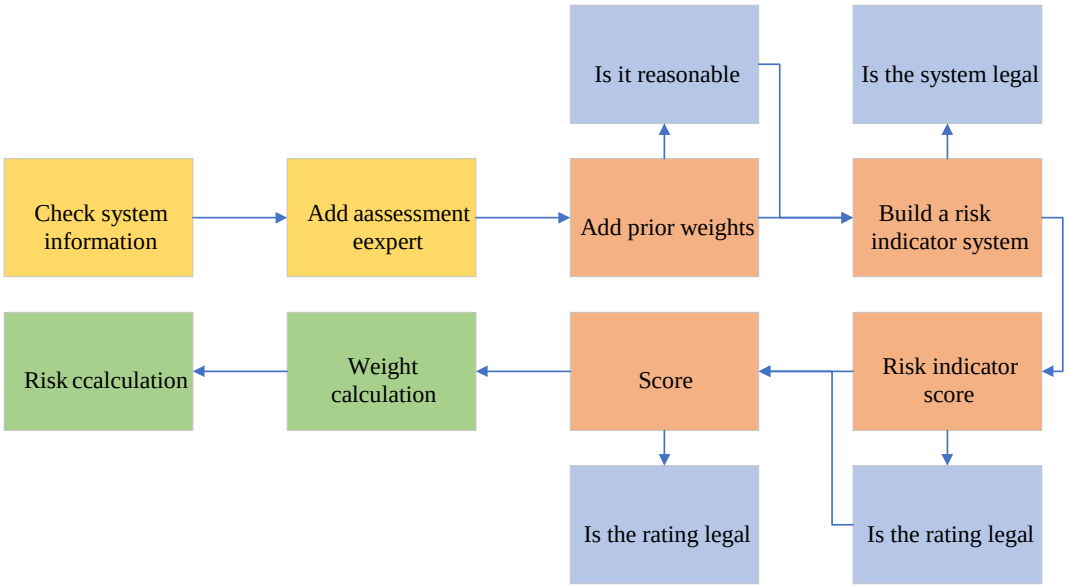

Figure 7: Risk assessment operation process

The risk degree of the risk indicator, the expert's knowledge level, and the expert's individual deviation are first calculated by entering the expert group and the risk index that already has scoring information. Next, the expert posterior weight is calculated based on the above results. Finally, the comprehensive weight of the expert for each risk indicator is calculated according to the prior weight vector [50].

The dataset selected for the experiment is the National Vulnerability Database (NVD) dataset. It is

a public, global vulnerability database maintained by the National Institute of Standards and Technology. The database contains detailed information on vulnerabilities and security issues worldwide, including vulnerability descriptions, risk assessments, the scope of impact, and solutions. The National Vulnerability Database (NVD) tracks vulnerability timeline stamps well because of its regular updates, precise data, uniform layout, and metadata. Its value is increased by integrating external datasets, facilitating well-informed vulnerability management decision-making. NVD datasets can be used to study and analyze information security issues, including vulnerability detection, risk assessment, and security policy formulation. Sources of NVD datasets include various vulnerability bulletins, security vendor reports, and analysis by security agencies. Each vulnerability has a unique identifier and is classified and rated according to different classification criteria. Additionally, NVD provides a timeline of vulnerabilities, which can track the discovery, disclosure, repair, and other processes of vulnerabilities. The environment and parameter settings of the experiment are shown in Table 1.

Table 1: Experimental environment

Facilities	Model
Central Processing Unit	2.5G
Operating system	Windows7
Web	Apache-tomecat6
Memory	12G
Damping coefficient	0.85
Maximum number of iterations	100
Population size	50
Crossover probability	0.1

4. Information system reliability and fault root cause location analysis based on PageRank iterative algorithm

4.1 Analysis of the accuracy of the fault root cause location of the information system

The experiment uses traditional models for comparison. Management, physical, equipment, network, data, and personnel risk are indicators. The experimental results are plotted in Figure 8.

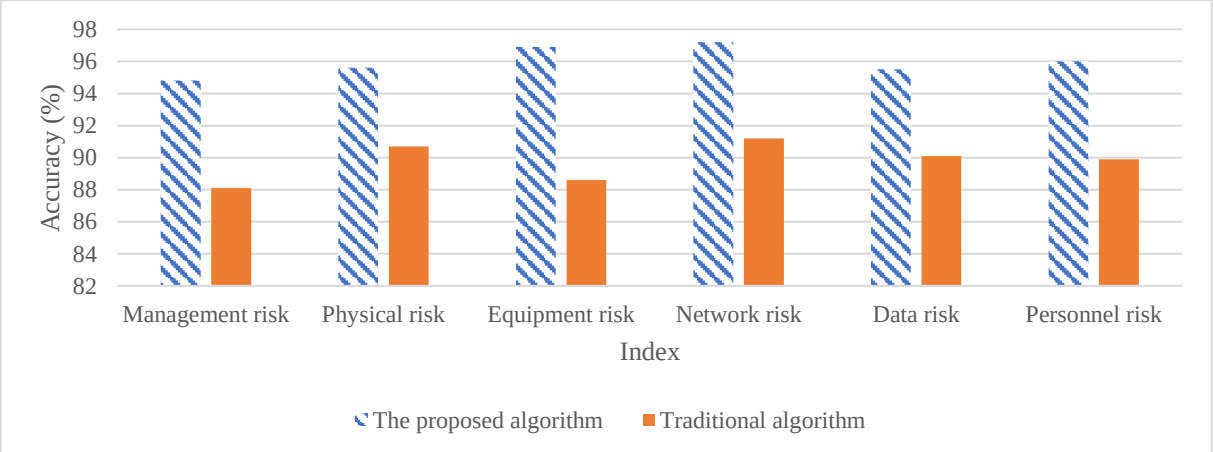

Figure 8: Result of fault root cause location accuracy

As shown in Figure 8, in terms of managing risk, physical risk, equipment risk, network risk, personnel risk, and data risk, the proposed algorithm performs better than the traditional algorithm. Specifically, regarding managing risk, the proposed algorithm scores 94.8%, higher than the conventional algorithm of 88.1%. This means that the optimized algorithm can more effectively identify and respond to potential risk factors in managing risk. In terms of physical risk, the proposed algorithm scores 95.6%, which is higher than the traditional algorithm of 90.7%, indicating that the optimized algorithm can reduce the damage caused by potential risks to the system by identifying physical risk factors more accurately and taking corresponding measures. Scalability, flexibility, accuracy, automation, real-time monitoring, interaction with current systems, customization, visualization, predictive analytics, and regulatory compliance are among the fundamental competencies of the optimized risk management algorithm.

Regarding equipment risk, the proposed algorithm has a score of 96.9%, higher than the traditional algorithm's 88.6%, indicating that the new algorithm can better evaluate and manage equipment-related risks. The reliability and stability of the system can be improved by detecting equipment failures and weaknesses in a timely and taking appropriate maintenance and protection measures. Through identifying warning indicators, ease of maintenance, downtime reduction, performance optimization, and support for data-driven decision-making, proactive equipment failure detection enhances system stability, safety, and dependability while guaranteeing continuous operations. Regarding network risk, the proposed algorithm has a score of 97.2%, higher than the traditional algorithm's 91.2%. Regarding data risk, the proposed algorithm has a score of 95.5%, which is higher than the conventional algorithm's 90.1%.

Regarding personnel risk, the proposed algorithm scores 96%, higher than the traditional algorithm of 89.9%. This means the optimized algorithm can more accurately assess and manage personnel risks. The risk caused by human factors can be reduced by strengthening employee training, establishing a sound authority management system, and monitoring internal threats. Employee training and internal threat monitoring are the two most important tactics for lowering cybersecurity risks. While internal monitoring identifies insider threats and triggers incident response, strengthening the organization's security posture, training raises awareness, compliance, and behavior.

4.2 Analysis of the results of the simulation experiment of fault root cause location of the information system

The risk assessment module mainly tests whether the module can correctly implement the risk assessment process of the risk assessment model using the same model scoring data for the same information system. The prototype system is continuously run in a simple network environment deployed in the lab. The system calculates the cybersecurity situation assessment for 25 assessment cycles, and attacks are carried out in assessment cycles 10-13. The experimental results are shown in Figure 9. It is essential to comprehend the cause of the optimized algorithm's faster information-gathering time to assess performance gains, spot design errors, spot optimization possibilities, and successfully direct further optimization efforts.

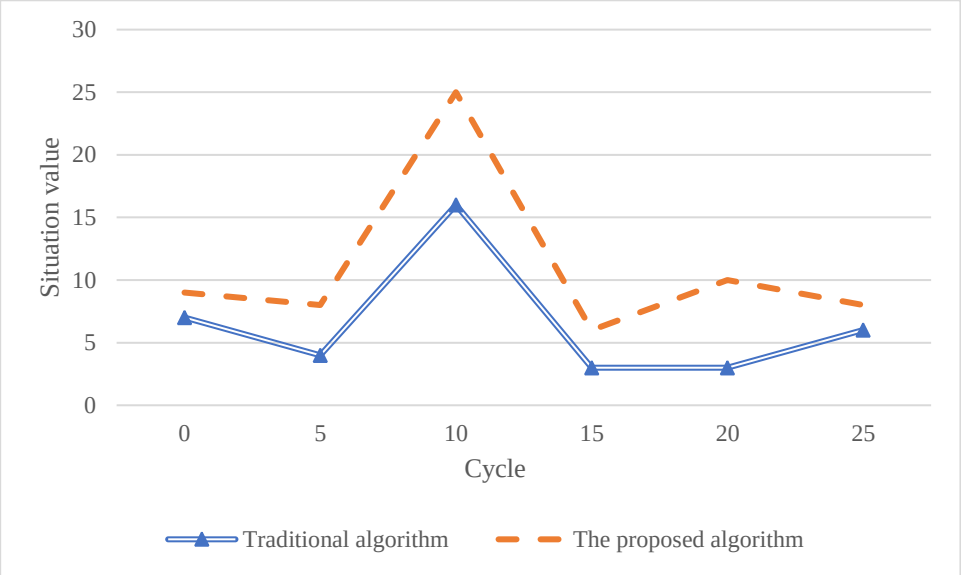

Figure 9: Cycle situation value results

As shown in Figure 9, the optimized system shows good performance in the cybersecurity situation awareness of the information system. Real-time monitoring, threat intelligence integration, behavioral analytics, complete visibility, automatic reaction, scalability, user-friendly interface, continuous monitoring, and regulatory compliance are all essential components of an optimal cybersecurity system. During cycles 10 to 13, the system is attacked, which causes the system's situation value to rise. Increased system situation value during cycles 10–13 results in heightened cybersecurity awareness, re-evaluation, enhanced stakeholder collaboration, heightened security alertness, intensified response, and investments in monitoring and detection capabilities to lessen attacks and fortify defenses. The optimized model shows the highest situation value in this process, reaching 24.7. It shows strong cybersecurity situation awareness ability. This means that the optimized system can effectively identify and respond to network attacks and adjust the security strategy and defense measures promptly to protect the information system from potential threats. These results further verify the effectiveness and feasibility of the optimized model in improving cybersecurity awareness of information systems. Information systems must be protected from possible attacks with solid cybersecurity. It seeks to safeguard essential assets and sensitive data's availability, confidentiality, and integrity. Guarding against unauthorized changes, tampering, or corruption guarantees data and systems integrity. The response time results for the time-consuming operating system are revealed in Figure 10.

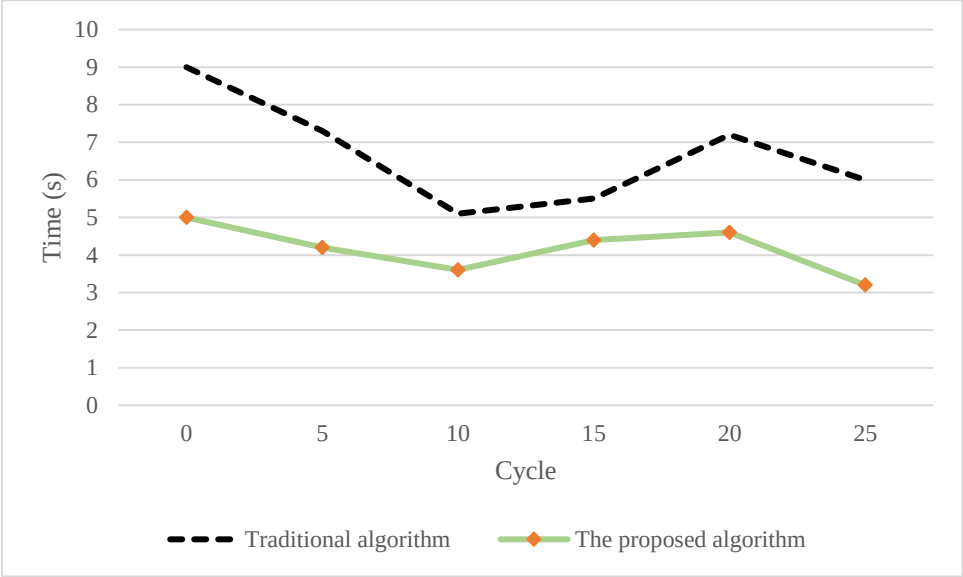

Figure 10: Time-consuming operating system response time results

In Figure 10, the information collection time of the optimized algorithm is much lower than that of the traditional model, and the command of collecting all device information in the active information system is carried out in each cycle, which takes a long time. Numerous devices, complexity, network latency, resource consumption, overhead associated with authentication, data processing needs, and error handling can all make gathering device information time-consuming. Identifying and fixing these problems may make the process more efficient, which will also shorten execution durations. The information collection time of the optimized algorithm is an average of 3.76 seconds, which is within the acceptable range. The operation is infrequent and only initiated when needed to meet the system's performance requirements.

5. Discussion

Through the comparison of the accuracy of fault root cause location, it can be found that the fault root cause location method based on the PageRank iterative algorithm has high practical application value in information system risk assessment, and the overall positioning accuracy is higher than 94%. When assessing the risk of an information system, fault root cause location techniques are essential since they provide light on the variables that lead to security events. They promote industry standards and improve stakeholder confidence, decision-making, risk management, regulatory compliance, and incident response effectiveness. It shows that the optimized algorithm can better help enterprises and organizations evaluate the security of their information systems and take corresponding measures to protect their information systems. An essential tool for locating and reducing cybersecurity risks is an open-source intrusion detection system (IDS). They carry out tasks like payload inspection, statistical analysis, packet capture, analysis, anomaly identification, protocol analysis, pattern matching, and decoding encrypted traffic. In the simulation experiment, the situation value of the optimized model is up to 24.7. The deployed open-source intrusion detection system detects the attack packets. The system's network attack situation is also improved after processing, resulting in the overall cybersecurity situation of the system. In the analysis of time-consuming responses, the practicability of the optimized

algorithm is fully verified. The efficiency of information collection can be effectively improved without affecting the system's performance using this algorithm. This is important for businesses and organizations that need access to as much information as possible in the shortest possible time to protect their information systems better. In addition, the optimized algorithm can be used in other fields, such as search engine optimization and social network analysis, and has a wide range of application prospects.

6. Conclusion

With the development of science and technology, information systems' reliability and fault root cause location have been widely concerned. This paper combines the PageRank iterative algorithm and GA to optimize the information system. This paper studies the structure of the risk assessment model of ISS and discusses the application of the PageRank iterative algorithm and GA in information systems. Subsequently, the risk assessment system is optimized by ensemble intelligent computing and GA, and horizontal comparison experiments verify the rationality of the proposed study. Experimental results show that the optimized model shows higher accuracy than the traditional model in fault root cause location. In addition, when locating network and device risks, the accuracy rate reaches 97.2% and 96.9%, respectively. Then, after the attack is carried out in cycles 10-13 of the system assessment cycle, the situation value of the information system optimized here is significantly improved. Additionally, the operating system's response time has been reduced to an average of 3.76s. This proves that the performance of the information system optimized here has been dramatically improved. Many things could be improved. On the one hand, the risk assessment model designed here does not consider the retention of historical assessment data in the assessment process and cannot thoroughly combine and apply these data in the ongoing risk assessment process. Follow-up research will explore how to effectively use expert historical assessment data, such as weight information, to reduce the amount of computation in later assessments and further improve the accuracy of assessment results. On the other hand, in calculating the topology weight of the host by the PageRank algorithm, only the connection relationship between the hosts is currently considered, and there is no further demonstration to explore the influence of the size of the data transmission between these connections on the importance of the host. This can result in a slight discrepancy between the calculated host topology weights and the actual situation. Subsequent studies will refine these two issues in the situation awareness framework.

Declarations

Funding

Any of the authors received no funds or grants.

Conflict of interest

There is no conflict of interest among the authors.

Data Availability

All data generated or analyzed during this study are included in the manuscript.

Code Availability

Not applicable.

Author’s contributions

All Authors contributed to the design and methodology of this study, the assessment of the outcomes, and the writing of the manuscript.

References

[1] Jiang D. The construction of a smart city information system based on the Internet of Things and cloud computing. Computer Communications, 2020, 150(11): 158-166.

[2] Firouzi F, Farahani B, Marinšek A. The convergence and interplay of edge, fog, and cloud in the AI-driven Internet of Things (IoT). Information Systems, 2022, 107(17): 101840.

[3] Fazlollahtabar H, Kazemitash N. Green supplier selection based on the information system performance evaluation using the integrated Best-Worst Method. Facta Universitatis, Series: Mechanical Engineering, 2021, 19(3): 345-360.

[4] Niknejad N, Ismail W, Ghani I, et al. Understanding Service-Oriented Architecture (SOA): A systematic literature review and directions for further investigation. Information Systems, 2020, 91(21): 101491.

[5] Pandey S, Srivastava A K, Amidan B G. A real-time event detection, classification and localization using synchrophasor data. IEEE Transactions on Power Systems, 2020, 35(6): 4421-4431.

[6] Wiedenhöft G C, Luciano E M, Pereira G V. Information technology governance institutionalization and the behavior of individuals in the context of public organizations. Information Systems Frontiers, 2020, 22(6): 1487-1504.

[7] Ranjha A, Kaddoum G. URLLC facilitated by mobile UAV relay and RIS: A joint design of passive beamforming, blocklength, and UAV positioning. IEEE Internet of Things Journal, 2020, 8(6): 4618-4627.

[8] González-García J, Gómez-Espinosa A, Cuan-Urquizo E, et al. Autonomous underwater vehicles: Localization, navigation, and communication for collaborative missions. Applied Sciences, 2020, 10(4): 1256.

[9] Gassar A A A, Cha S H. Review of geographic information systems-based rooftop solar photovoltaic potential estimation approaches at urban scales. Applied Energy, 2021, 291(6): 116817.

[10] Dong L, Sun D, Han G, et al. Velocity-free localization of autonomous driverless vehicles in underground intelligent mines. IEEE Transactions on Vehicular Technology, 2020, 69(9): 9292-9303.

[11] Cherif H, Benakcha A, Laib I, et al. Early detection and localization of stator inter-turn faults based on discrete wavelet energy ratio and neural networks in induction motor. Energy, 2020, 212(39): 118684.

[12] Mustafa S Z, Kar A K, Janssen M. Understanding the impact of digital service failure on users: Integrating Tan's failure and DeLone and McLean's success model. International Journal of Information Management, 2020, 53(20): 102119.

[13] Vom Brocke J, Winter R, Hevner A, et al. Special issue editorial–accumulation and evolution of design knowledge in design science research: a journey through time and space. Journal of

the Association for Information Systems, 2020, 21(3): 9.

[14] Subedi S, Pyun J Y. A survey of smartphone-based indoor positioning system using RF-based wireless technologies. Sensors, 2020, 20(24): 7230.

[15] Fayyad J, Jaradat M A, Gruyer D, et al. Deep learning sensor fusion for autonomous vehicle perception and localization: A review. Sensors, 2020, 20(15): 4220.

[16] Herraiz Á H, Marugán A P, Márquez F P G. Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure. Renewable Energy, 2020, 153(41): 334-348.

[17] Zhu J W, Gu C Y, Ding S X, et al. A new observer-based cooperative fault-tolerant tracking control method with application to networked multiaxis motion control system. IEEE Transactions on Industrial Electronics, 2020, 68(8): 7422-7432.

[18] Pappas I O, Woodside A G. Fuzzy-set Qualitative Comparative Analysis (fsQCA): Guidelines for research practice in Information Systems and marketing. International Journal of Information Management, 2021, 58(11): 102310.

[19] Daradkeh Y I, Tvoroshenko I, Gorokhovatskyi V, et al. Development of effective methods for structural image recognition using the principles of data granulation and apparatus of fuzzy logic. IEEE Access, 2021, 9(1): 13417-13428.

[20] Ham Y, Kim J. Participatory sensing and digital twin city: Updating virtual city models for enhanced risk-informed decision-making. Journal of Management in Engineering, 2020, 36(3): 04020005.

[21] Liu S, Guo C, Al-Turjman F, et al. Reliability of response region: a novel mechanism in visual tracking by edge computing for IIoT environments. Mechanical systems and signal processing, 2020, 138(12): 106537.

[22] Kim B, Park J, Suh J. Transparency and accountability in AI decision support: Explaining and visualizing convolutional neural networks for text information. Decision Support Systems, 2020, 134(12): 113302.

[23] Chen M, Liu Q, Huang S, et al. Environmental cost control system of manufacturing enterprises using artificial intelligence based on value chain of circular economy. Enterprise Information Systems, 2022, 16(8): 1856422.

[24] AlShorman O, Alkahatni F, Masadeh M, et al. Sounds and acoustic emission-based early fault diagnosis of induction motor: A review study. Advances in Mechanical Engineering, 2021, 13(2): 1687814021996915.

[25] Zhang Z, Yang Z, Lu S, et al. Strain localization and failure at twin-boundary complexions in nickel-based superalloys. Nature Communications, 2020, 11(1): 4890.

[26] Wang M, Lin Y, Tian Q, et al. Transfer learning promotes 6G wireless communications: Recent advances and future challenges. IEEE Transactions on Reliability, 2021, 70(2): 790-807.

[27] Ali L, Alnajjar F, Jassmi H A, et al. Performance evaluation of deep CNN-based crack detection and localization techniques for concrete structures. Sensors, 2021, 21(5): 1688.

[28] Aloisio A, Di Battista L, Alaggio R, et al. Sensitivity analysis of subspace-based damage indicators under changes in ambient excitation covariance, severity, and location of damage. Engineering Structures, 2020, 208(22): 110235.

[29] Qiu S, Zhao H, Jiang N, et al. Sensor network-oriented human motion capture via wearable intelligent system. International Journal of Intelligent Systems, 2022, 37(2): 1646-1673.

[30] Tavitiyaman P, Qu H, Tsang W L, et al. The influence of smart tourism applications on perceived destination image and behavioral intention: The moderating role of information search behavior. Journal of Hospitality and Tourism Management, 2021, 46(19): 476-487.

[31] Faccia A, Petratos P. Blockchain, enterprise resource planning (ERP) and accounting information systems (AIS): Research on e-procurement and system integration. Applied Sciences, 2021, 11(15): 6792.

[32] Roy A M, Bhaduri J, Kumar T, et al. WilDect-YOLO: An efficient and robust computer vision-based accurate object localization model for automated endangered wildlife detection. Ecological Informatics, 2023, 75(18): 101919.

[33] Ibrahim M S, Dong W, Yang Q. Machine learning driven smart electric power systems: Current trends and new perspectives. Applied Energy, 2020, 272(23): 115237.

[34] Li Y, Zhuang Y, Hu X, et al. Toward location-enabled IoT (LE-IoT): IoT positioning techniques, error sources, and error mitigation. IEEE Internet of Things Journal, 2020, 8(6): 4035-4062.

[35] Alzoubi H M, Aziz R. Does emotional intelligence contribute to quality of strategic decisions? The mediating role of open innovation. Journal of Open Innovation: Technology, Market, and Complexity, 2021, 7(2): 130.

[36] Anagnostopoulou E, Urbančič J, Bothos E, et al. From mobility patterns to behavioural change: leveraging travel behaviour and personality profiles to nudge for sustainable transportation. Journal of Intelligent Information Systems, 2020, 54(7): 157-178.

[37] Han S, Xi Z. Dynamic scene semantics SLAM based on semantic segmentation. IEEE Access, 2020, 8(1): 43563-43570.

[38] Wang J, Xu C, Zhang J, et al. Big data analytics for intelligent manufacturing systems: A review. Journal of Manufacturing Systems, 2022, 62(4): 738-752.

[39] Chen Y, Xu D, Chen N, et al. FRL-MFPG: Propagation-aware fault root cause location for microservice intelligent operation and maintenance. Information and Software Technology, 2023, 153(7): 107083.

[40] Wang S, Zhao Q, Han Y, et al. Root cause diagnosis for process faults based on multisensor time-series causality discovery. Journal of Process Control, 2023, 122(27): 27-40.

[41] Song P, Zhao C, Huang B. MPGE and RootRank: A sufficient root cause characterization and quantification framework for industrial process faults. Neural Networks, 2023, 161(2): 397-417.

[42] Wang S, Zhao Q, Han Y, et al. Root cause diagnosis for complex industrial process faults via spatiotemporal coalescent-based time series prediction and optimized Granger causality.

Chemometrics and Intelligent Laboratory Systems, 2023, 233(31): 104728.

[43] Oliveira E, Miguéis V L, Borges J L. Automatic root cause analysis in manufacturing: an overview & conceptualization. Journal of Intelligent Manufacturing, 2023, 34(5): 2061-2078.

[44] Hyun S, Song J, Jee E, et al. Timed pattern-based analysis of collaboration failures in system-of-systems. Journal of Systems and Software, 2023, 6(3): 111613.

[45] Ortega S A, Martin-Delgado M A. Generalized quantum PageRank algorithm with arbitrary phase rotations. Physical Review Research, 2023, 5(1): 013061.

[46] Zhu D, Wang H, Wang R, et al. Identification of key nodes in a power grid based on modified PageRank algorithm. Energies, 2022, 15(3): 797.

[47] Hua Z, Fei L, Jing X. An improved risk prioritization method for propulsion system based on heterogeneous information and PageRank algorithm. Expert Systems with Applications, 2023, 212(42): 118798.

[48] Tanglay O, Young I M, Dadario N B, et al. Eigenvector PageRank difference as a measure to reveal topological characteristics of the brain connectome for neurosurgery. Journal of Neuro-Oncology, 2022, 157(1): 49-61.

[49] Hajarathaiah K, Enduri M K, Anamalamudi S, et al. Computing influential nodes using the nearest neighborhood trust value and PageRank in complex networks. Entropy, 2022, 24(5): 704.

[50] Bautista E, Latapy M. A local updating algorithm for personalized PageRank via Chebyshev polynomials. Social Network Analysis and Mining, 2022, 12(1): 31.