Research on Key Technologies of Edge Computing

Jiarong Chen[a]
College of Arts and Sciences, Beijing Normal University, No. 18 Jinfeng Road, Zhuhai, China

Keywords: Edge Computing, Task Scheduling and Resource Management, Information System Security, Cloud Computing, Internet of Things.

Abstract: Edge computing, as a key technology for IoT applications, has become an important driver of digital transformation due to its ability to provide low-latency and high-bandwidth data processing at the edge of the network. With the explosive growth of IoT devices, the traditional cloud computing model is facing many challenges, especially in low-latency, high-bandwidth, real-time data processing scenarios. Edge computing can significantly reduce data transmission time, reduce network congestion, and improve quality of service by pushing computing resources and data processing capabilities to the edge of the network. This paper reviews the key technologies of edge computing, focusing on the research progress in task scheduling and resource management, as well as information system security. By analyzing the current mainstream methods and models, this paper summarizes the advantages and limitations of edge computing in task scheduling, resource management, and security evaluation, etc. Finally, it looks forward to the future research direction and proposes potential ways to optimize the performance of the edge computing system, including multi-source data fusion and intelligent scheduling mechanism.

1 INTRODUCTION

The rise of edge computing is driven by the explosive growth of IoT devices and the resulting demand for data processing. Traditional cloud computing models face many challenges when dealing with low latency, high bandwidth, and massive data processing scenarios in real time (Alrowaily & Lu, 2018). For example, vehicle-to-vehicle (V2V) or vehicle-to-infrastructure (V2I) communication in intelligent transportation systems requires low latency at the millisecond level (Wang, 2024) to ensure collision warnings and traffic flow optimization, while real-time control and monitoring aspects in industrial automation require extremely high time-to-moment data processing, where a slight delay can trigger production failures (Garg, Singh, Kaur, et al., 2019). Based on these demands, edge computing has emerged as a key enabler to drive digital transformation by pushing computing resources and data processing capabilities to the edge of the network to significantly reduce the time and distance of data transmission, alleviate network congestion, reduce latency, and improve service quality and user experience (Wu, 2024).

In edge computing networks, the optimization of task scheduling and resource management directly affects the overall performance of the system and the efficiency of resource utilization. Since edge nodes have limited computational resources, the key to ensuring efficient processing is how to rationally distribute tasks based on the characteristics of the tasks, such as computational requirements, delay constraints, and priorities. In terms of resource management, the heterogeneous resources involved in edge computing (e.g., compute, storage, and network bandwidth) need to be dynamically allocated and co-optimized to avoid idle or overloaded resources and ensure that the tasks can be successfully executed under resource constraints (Feng, 2023). This is critical to improving system stability and reliability, e.g., in smart home scenarios, where multiple tasks generated by smart devices require precise scheduling and resource allocation to ensure the stability of the home network and the smoothness of user interactions (Mijuskovic, Chiumento, Bemthuis, et al., 2021).

[a] https://orcid.org/0009-0002-1342-1173

Chen, J.
Research on Key Technologies of Edge Computing.
DOI: 10.5220/0013677500004670

In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)*, pages 26-32
ISBN: 978-989-758-765-8

At the same time, the security of edge computing information systems has become a challenge that cannot be ignored. Due to its complex architecture, limited resources, and dynamic and changing application scenarios, the security posture of edge computing is severe and complex. Therefore, it is of great importance to construct an accurate security evaluation model. Traditional security evaluation methods often have limitations in edge computing scenarios, for example, the single-assignment method is easily disturbed by subjective factors, the first-order gray clustering method has fuzzy degree determination, and the fuzzy evaluation model is computationally expensive and poorly adapted (Guo, 2024). Therefore, designing a security evaluation model with high accuracy, high efficiency and dynamic adaptability can more accurately assess the security status of the system, timely warn potential risks, reasonably allocate security resources and optimize protection strategies. This will provide solid theoretical and technical support for the application of edge computing in key areas such as finance, healthcare, energy, etc., ensure the security of data assets and privacy, and promote the in-depth application of edge computing technology in various industries (Chen, 2020).

The application of edge computing in IoT significantly improves the speed and efficiency of data processing, but it also brings new challenges in task scheduling, resource management, and information security. The next section addresses these key areas. Chapter 2 details task scheduling and resource management strategies for edge computing, including prediction-based resource allocation, priority-aware task scheduling, and bursty traffic response schemes, and focuses on information system security, introducing several efficient security evaluation models and their application scenarios. Chapter 3 discusses the commonly used evaluation criteria and datasets in experiments. Chapter 4 summarizes the current research results, and the last two chapters point out future directions and potential research avenues. Through these chapters, this thesis aims to lead the reader to a comprehensive understanding of the current state of edge computing technology and its future development.

2 OVERVIEW OF MAINSTREAM APPROACHES IN RECENT YEARS

2.1 Task Scheduling and Resource Management Aspects of Edge Computing

・Edge-cloud collaborative task offloading scheme based on resource utilization prediction: machine learning models (e.g., LSTM, GRU) are used to predict the future resource utilization and combined with Deep Deterministic Policy Gradient Algorithm (DDPG) to decide the task offloading strategy. When the edge nodes are resource-constrained, the tasks are intelligently offloaded to the cloud or other edge nodes to balance the load, improve resource utilization, and reduce task latency.

・Multilateral cloud collaborative task scheduling strategy based on task priority awareness: through the multi-intelligent body deep Q network (MADQN) framework, each intelligent body is responsible for the edge node scheduling, collaborating through information sharing and learning to optimize scheduling strategy according to the reward mechanism to ensure that high-priority tasks are prioritized and resources are reasonably allocated to reduce latency and improve the task completion rate, which is suitable for multi-user multi-task resource competition scenarios.

・Joint arithmetic slicing strategy based on business intent constraints: maps business intent to network state demand, generates arithmetic slicing strategy by means of Proximal Policy Optimization (PPO) algorithm, allocates resources to different business demands, realizes dynamic and flexible resource allocation, reduces high load blocking rate and intent violation rate, and improves resource utilization efficiency and business adaptability.

・Network survivability assurance solution based on burst traffic prediction: Uses a machine learning model to predict burst traffic, sets protection thresholds, and schedules traffic using a particle swarm optimization algorithm based on policy gradient (PSO-PG). As network traffic fluctuates, it plans ahead and precisely allocates resources according to the predicted feature variables to ensure low-latency processing of tasks, control the blocking rate, maintain network stability and survivability, and withstand unexpected traffic surges (Sun, 2024).

2.2 Information System Security Evaluation Model Aspects of Edge Computing

• High-precision security evaluation grading model based on combinatorial assignment gray clustering (C-SG model): through security policy decomposition and security index mapping, construct a pool of security evaluation indexes, propose subject-objective combinatorial assignment method, improve the first-order gray clustering method, and construct a second-order gray clustering security evaluation method, so as to improve the accuracy of security evaluation grading.

• Efficient safety evaluation model based on Spearman's PCA distinction combination screening (PSDC-CVF model): improve the PCA indicator screening method, propose the PCA-S indicator screening method, and combine the indicator distinction screening method to reduce resource consumption, improve the indicator screening efficiency, and achieve a better balance between efficiency and accuracy.

• Dynamic safety evaluation model based on time-varying motorized screening of indicator validity within gray classes (GFCIV-CGTOPSIS model): separating subjective scoring type indicators and objective collection type indicators, improving gray rough set screening method, forming gray F-statistical clustering screening method of indicator validity within gray F-statistical clustering, ensuring high efficiency of the dynamic evaluation model, and using gray TOPSIS method to dynamically evaluate the safety state of the system (Guo, 2024). Evaluation of system security status using gray TOPSIS method (Guo, 2024).

3 FREQUENTLY USED DATA SETS AND EVALUATION CRITERIA

3.1 Edge Computing Task Scheduling and Resource Management

3.1.1 Datasets

Mostly simulation data sets are used to construct an edge computing network simulation environment with a three-layer architecture (terminal layer, edge layer, cloud layer). The terminal layer covers all kinds of mobile and sensor devices, randomly generating data volume and task request time according to the task characteristics of the device; the edge layer consists of performance heterogeneous servers; the cloud layer is a powerful and stable computing node. The task request includes key characteristics such as computation demand, latency constraints, task size, etc. Meanwhile, historical time-series data such as edge node CPU utilization, task queue length, network communication latency, bandwidth utilization, etc. are recorded for model training and testing, which comprehensively simulates the dynamic change of task load and complex scenarios of network state (Yu, 2024).

3.1.2 Evaluation Criteria

• Average task latency: reflects the task processing speed, which directly affects user experience and system real-time responsiveness.

• Offload failure rate: reflects the reliability of the task offloading strategy, which must be strictly controlled at a low level to ensure smooth task execution.

• Resource utilization rate: evaluates the degree of resource utilization of edge nodes, and efficient utilization can avoid wasting idle resources.

• Task completion rate: emphasizes the comprehensive ability of the system to handle tasks, and a high completion rate ensures that task requirements are effectively met.

3.2 Security Evaluation of Edge Computing Information Systems

3.2.1 Datasets

The data set is obtained from the fusion of actual system monitoring data and simulated security event scenario data, collecting data on system architecture, device configuration, network connection, user access behavior, application operation logs, etc.; simulating various types of network attacks, data leakage risk scenarios, and system failure status data, comprehensively covering edge information system security elements and potential risk dimensions.

3.2.2 Evaluation Criteria

• Security evaluation accuracy: accurately measure the accuracy of system security status and risk level determination, and quantify the accuracy of the evaluation model in identifying and classifying security events by using confusion matrix, accuracy rate, recall rate, F1 value and other indicators.

• Model efficiency: Focusing on resource-constrained scenarios, researchers evaluate the degree of computational resource consumption by model computation time and memory usage, and shorten the evaluation cycle and reduce the dependence on hardware resources with efficient models.

• Dynamic evaluation efficiency: Given the spatial and temporal dynamic changes of the system, the model is tested for its ability to dynamically track and predict the security status, such as calculating the deviation of dynamic security indicators and predicting the accuracy of events, to ensure that the model dynamically adapts to the evolution of the system and warns of the risk trend in real time (Deng, Liu, Wang, et al., 2020).

4 ANALYSIS OF CURRENT RESULTS

4.1 Results of Edge Computing Task Scheduling and Resource Management Strategy

• Edge-cloud cooperative task offloading scheme based on resource occupancy prediction: in a simulation environment with 20 edge nodes and 2 cloud servers covering 10 square kilometers of network topology, the performance of processing 500-2000 task requests per round is excellent. The average task latency is significantly reduced compared with edge-only or cloud-only processing, the offload failure rate is controlled within 5.1%, and the resource utilization rate of edge nodes is increased to about 70%, effectively relieving the resource bottleneck, proving its significant advantages in optimizing task latency, ensuring offload reliability, and improving resource efficiency.

• Multi-cloud cooperative scheduling strategy based on task priority awareness: Tested in static and dynamic load scenarios with remarkable results. The average processing latency of tasks under dynamic load is stabilized at 28.4 milliseconds, the completion rate of tasks exceeds 98%, and the resource utilization rate of severe resource-guzzling scenarios still exceeds 60%. This strategy accurately detects task priority, efficiently schedules resources, and significantly improves system service quality and resource utilization performance.

• Common arithmetic slicing strategy based on business intent constraints: Excellent results in single-service (e.g., live video streaming) and mixed-services (e.g., IoT and VR in parallel) scenarios. Task blocking rate is as low as 0.08% under heavy load, and intent violation rate is within 8.43%. Dynamic arithmetic slicing improves resource utilization, reduces network congestion, effectively improves service processing efficiency, and guarantees service quality.

• Network survivability guarantee solution based on burst traffic prediction: Under the simulation of low-burst, high-burst and multi-node failure scenarios, the traffic prediction accuracy reaches 94.2%, the average task completion delay is only 137 milliseconds, and the blocking rate of high-burst traffic is continuously controlled within 9.2%. The solution significantly enhances the network's ability to withstand sudden impacts and guarantees the timeliness of task processing and network stability (Sun, 2024).

4.2 Edge Computing Information System Security Evaluation Model Results

• A high-accuracy security evaluation grading model based on combined-assignment gray clustering (C-SG model): a multi-level system with 8 first-level and 42 second-level indicators is constructed to improve the global and scalability. Experiments show that the difference coefficient of assignment degree is increased by more than 40%, and the first and last consistency rate is more than 85%, which significantly improves the accuracy of safety evaluation and grading, and accurately determines the safety level of the system.

• Fuzzy comprehensive and efficient safety evaluation model based on indicator selection (PSDC-CVF model): 42 core safety indicators are optimized and selected to build the evaluation system, resource consumption is reduced by 67%, and the offset of fuzzy evaluation results is only 1.8%. While ensuring the accuracy of fuzzy evaluation, it also greatly improves the efficiency and efficiently handles the security evaluation of edge information systems.

• Dynamic security evaluation model based on indicator motorized screening (GFCIV-CGTOPSIS model): The test of telematics edge system shows that compared with the traditional gray rough set screening method, the reasonableness of indicator weight distribution is improved by more than 30%, the deviation of evaluation results is reduced by about 40%, and the time consumption of the same resource is reduced by 25%. Dynamic adaptation system changes accurate screening indicator real-time evaluation, greatly improve the dynamic safety evaluation accuracy and timeliness (Guo, Lu, Tian, et al, 2023).

5 CHALLENGES AND OUTLOOK

5.1 Edge Computing Task Scheduling and Resource Management

Edge computing task scheduling and resource management strategies face several key issues, the first of which is the limitation of resource prediction accuracy. Due to the interference of complex factors in reality, it is often difficult to accurately predict resource utilization by relying only on historical data and traditional machine learning models, which affects the scientific nature of task offload scheduling decisions and system performance optimization. To solve this problem, feature dimensions can be expanded by fusing data from multiple sources (e.g., weather, holidays, geographic location, etc.), and advanced architectures (e.g., Transformer, BERT, etc.) can be used to capture spatiotemporal dependencies and build hybrid models to improve the accuracy of resource prediction in complex scenarios (Cen, Hu, Cai, et al., 2022).

Second, edge computing faces a dilemma in large-scale task scheduling, especially when using the MADQN framework, the computational cost of handling massive tasks skyrockets, the efficiency drops sharply, and it is difficult to meet real-time requirements, which in turn limits the system's task processing scale and efficiency improvement. In this regard, a hierarchical partition scheduling architecture can be introduced to stream processing based on task characteristics and resource hierarchy, and at the same time combined with distributed training to accelerate the learning convergence of MADQN intelligences, thus improving the efficiency of large-scale task processing and system scalability.

In addition, there is a lag problem in the dynamic adaptation of arithmetic slices, and the PPO algorithm is slow to adapt to sudden changes in business demand, which affects the delay of business processing and fluctuations in service quality. To solve this problem, it is proposed to construct an adaptive arithmetic slicing mechanism, design an intelligent module that monitors changes in business demand in real time, and adaptively adjust the arithmetic allocation strategy combined with reinforcement learning to dynamically optimize the slicing to ensure smooth business and improve resource efficiency.

Predicting burst traffic is also a current problem. Existing models are difficult to accurately capture extreme burst traffic, and relying on predictive models when burst traffic occurs leads to scheduling delays that affect task timeliness and exacerbate network congestion. Therefore, it is critical to strengthen the synergy between traffic prediction and contingency scheduling. Real-time monitoring and predictive models can be integrated to activate an early warning mechanism when unexpected traffic thresholds are encountered, combined with reinforcement learning for online optimal scheduling to ensure uninterrupted task processing and maintain smooth network performance.

Finally, current models are less generalizable and the evaluation system is too one-sided, with many models based on specific simulations and lacking validation on multi-industry and multi-regional data, resulting in weak generalizability. The evaluation system also often ignores energy consumption and costs, which affects the long-term sustainable operation and market competitiveness of the system. Therefore, it is necessary to expand the validation scope of the model, collect multi-industry and multi-region data for training and optimization, and improve its generalizability. At the same time, a multidimensional evaluation system should be established to introduce energy consumption (e.g., computing and communication energy consumption) and cost (e.g., equipment and operation cost) indicators, and the dynamic environment should be simulated to evaluate the performance of the system in the whole life cycle, so as to provide comprehensive and accurate decision support for the optimization and operation of the system (Luo, Hu, Li, et al., 2021).

5.2 Security Evaluation Model of Edge Computing Information System

In the security evaluation model of edge computing information system, one of the first problems is the difficulty of dynamic development of the index system. With the rapid development of edge computing technology and the continuous improvement of security standards, the existing security evaluation index system often fails to reflect the new risks and protection needs in a timely manner. For example, with the emergence of new IoT protocols, the lack of security indicators leads to a lag in the timeliness of the model, which affects the accurate metrics of the system's security status, thus weakening the effectiveness of security decision-making. To this end, a dynamic update mechanism for adaptive indicators can be constructed to track the latest developments in technical standards and security research in real time, and iteratively update the indicators according to emerging risks and technologies to ensure that the model can adapt to the latest security needs. In addition, machine learning methods are used to select key indicators that have a significant impact on system security, and to ensure that the indicator system has good scalability and can be continuously enriched and improved as technology evolves.

Another problem is the data collection dilemma and quality bottleneck. The complex architecture and limited resources of edge computing systems make it difficult to collect safety data, and expert evaluation is highly subjective, and there are omissions and errors in objective collection, resulting in uneven data quality and limited scale, which affects the accuracy and generalization of the evaluation model. For example, in industrial control networks, it is impossible to comprehensively collect key security parameters, which leads to the model misjudging risks and affecting the allocation of security resources. To solve this problem, a multifaceted fusion data collection strategy can be innovated to combine active detection and passive listening means to comprehensively collect safety data. In addition, blockchain technology is used to ensure the trustworthiness and non-tamperability of the data, improve the security and reliability of the data, and expand the scope of data collection by introducing a third-party data source to improve the quality of the data and the generalization ability of the model.

Finally, the security evaluation model of edge computing information system also faces the problem of complexity of effectiveness validation. Existing models lack unified validation standards and have a single metric, which makes it difficult to comprehensively measure the performance advantages and disadvantages of the model, and hinders the iterative improvement and application promotion of the technology. To this end, a comprehensive and standardized validation and evaluation system can be established, and common validation guidelines can be formulated to ensure that different models can be effectively compared and integrated with each other. At the same time, a multidimensional evaluation index system can be established, covering multiple dimensions such as performance, safety, economic cost, etc., to comprehensively evaluate the advantages and disadvantages of the models. On this basis, a public testing platform is established to provide a standardized testing environment and data set, which facilitates the validation and optimization of the models by researchers and enterprises, and promotes the rapid development and application promotion of the technology (Ma & Li, 2018).

6 CONCLUSIONS

This paper provides a comprehensive overview of the key areas of edge computing technology, especially the latest research progress in task scheduling and resource management, and information system security. First, in terms of task scheduling and resource management, this paper analyzes the challenges facing edge computing, such as resource occupancy prediction, task priority awareness, and arithmetic slicing, and proposes solutions based on machine learning, deep learning, and multi-intelligence body collaboration. These solutions effectively improve resource utilization, reduce task latency, and enhance system reliability. Second, in the field of information system security, with the rapid development of edge computing, security issues are becoming more and more prominent. Aiming at the dynamic security risks in edge computing environment, this paper explores new security evaluation models such as combined-empowerment gray clustering, PCA indicator screening, etc., which have achieved remarkable results in improving the accuracy and efficiency of security evaluation. However, there are still some challenges in the existing research, such as insufficient resource prediction accuracy and difficulty in dynamically evolving system security evaluation metrics. In the future, edge computing will face problems such as large-scale task scheduling, traffic burst prediction, and system generalization capability, which require the fusion of multi-source data, the improvement of model adaptability, and the exploration of efficient security policies. In conclusion, the research of edge computing technology is still in continuous development, and in the future, it will better serve the Internet of Things, Industrial Internet and other fields through intelligent and dynamic optimization solutions to promote the process of digital transformation.

REFERENCES

Alrowaily, M., & Lu, Z. 2018. Secure edge computing in IoT systems: review and case studies. *2018 IEEE/ACM symposium on edge computing* (pp. 440-444). IEEE.

Cen, B., Hu, C., Cai, Z., Wu, Z., Zhang, Y., Liu, J., & Su, Z. 2022. A configuration method of computing resources for microservice-based edge computing apparatus in smart distribution transformer area. *International Journal of Electrical Power & Energy Systems*, 138, 107935.

Chen, X. 2020. A security integration model for private data of intelligent mobile communication based on edge computing. *Computer Communications*, 162, 204-211.

Deng, X., Liu, J., Wang, L., & Zhao, Z. 2020. A trust evaluation system based on reputation data in mobile edge computing network. *Peer-to-Peer Networking and Applications*, 13, 1744-1755.

Feng, C. 2023. Task Scheduling and Resource Management in Edge Computing Networks (Ph.D. dissertation, Xi'an University of Electronic Science and Technology). Ph.D.

Garg, S., Singh, A., Kaur, K., Aujla, G. S., Batra, S., Kumar, N., & Obaidat, M. S. 2019. Edge computing-based security framework for big data analytics in VANETs. *ieee network*, 33(2), 72-81.

Guo, Z. 2024. Research on Information System Security Evaluation Model for Edge Computing (Ph.D. dissertation, Beijing University of Posts and Telecommunications). Ph.D.

Guo, Z., Lu, Y., Tian, H., Zuo, J., & Lu, H. 2023. A security evaluation model for multi-source heterogeneous systems based on IOT and edge computing. *Cluster Computing*, 1-15.

Luo, Q., Hu, S., Li, C., Li, G., & Shi, W. 2021. Resource scheduling in edge computing: a survey. *IEEE Communications Surveys & Tutorials*, 23(4), 2131-2165.

Ma, X., & Li, X. 2018. Trust evaluation model in edge computing based on integrated trust. *2018 International Conference on Algorithms, Computing and Artificial Intelligence* (pp. 1-6).

Mijuskovic, A., Chiumento, A., Bemthuis, R., Aldea, A., & Havinga, P. 2021. Resource management techniques for cloud/fog and edge computing: an evaluation framework and classification. *sensors*, 21(5), 1832.

Sun, Z. 2024. Research on Task Scheduling and Resource Management Strategies in Edge Computing Networks (Ph.D. dissertation, Beijing University of Posts and Telecommunications). Ph.D.

Wang, G. 2024. Research on key technology of multi-access edge computing based computing service for Telematics (Ph.D. dissertation, Beijing University of Posts and Telecommunications, Beijing, China). Ph.D.

Wu, Z. 2024. Research on key technology of blockchain-based offloading for edge distributed computing (Ph.D. dissertation, Beijing University of Posts and Telecommunications). Ph.D.

Yu, X. 2024. Research on task scheduling and resource management for heterogeneous edge computing with cloud-side-end collaboration (Ph.D. dissertation, Beijing University of Posts and Telecommunications). Ph.D.