Transactions on Computational and Scientific Methods | Vo. 4, No. 3, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Adaptive Multi-Tenant Resource Scheduling in Cloud Computing via Reinforcement Learning

Chong Zhang
Carnegie Mellon University, Pittsburgh, USA
doriszhang746442@gmail.com

Abstract: This paper proposes an adaptive resource allocation algorithm that integrates reinforcement learning to address the challenges of dynamic resource demands and complex multi-objective constraints in cloud computing environments. The method abstracts resource supply and demand states in the cloud into multidimensional state vectors and employs joint modeling of policy and value networks to enable dynamic decision-making. A multi-objective reward function is designed to balance potential conflicts among key metrics, including resource utilization, energy efficiency, latency, and fairness. To enhance model representation and adaptability, a hierarchical state encoding mechanism and a residual gating-based policy updating approach are introduced, ensuring stability of scheduling in complex environments. Experiments comparing several public reinforcement learning scheduling models verify the significant advantages of the proposed method in multi-objective optimization. Results show that the method effectively improves resource utilization, reduces average task latency, enhances overall energy efficiency, and achieves higher fairness in multi-tenant sharing conditions. Overall, this study demonstrates that reinforcement learning-driven adaptive resource scheduling performs well in complex and dynamic cloud scenarios and provides technical support for building efficient, green, and balanced resource management systems.

Keywords: Cloud computing resource scheduling; reinforcement learning adaptive allocation; multi-metric joint optimization; multi-tenant fairness

1. Introduction

With the continuous acceleration of digital transformation and intelligent development, cloud computing has become the core infrastructure supporting a wide range of applications and services. Whether in enterprise-level business, scientific computing, or daily personal use, the computing, storage, and network resources provided by cloud platforms have reached unprecedented scale and complexity. In this context, achieving efficient resource utilization, dynamic load balancing, and sustained service quality has become a central challenge in the field of cloud computing. Traditional resource scheduling strategies often rely on static rules or heuristic methods. When faced with surging user demand, diverse application structures, and highly dynamic operating environments, these methods struggle to balance performance, cost, and energy consumption. As a result, they reveal clear limitations in adaptability and leave significant room for optimization[1].

At the same time, the highly dynamic nature of cloud computing makes resource demand uncertain and volatile. On one hand, user requests display strong bursty and periodic patterns in their temporal and spatial distribution. On the other hand, the availability and performance of underlying hardware resources may also change over time with varying workloads. In such cases, fixed allocation policies not only lead to resource

waste but may also reduce system performance and even cause severe failures such as service outages[2]. Therefore, building an adaptive allocation mechanism that can adjust strategies in real time according to environmental changes is not only essential for improving cloud platform capability but also represents an inevitable trend toward intelligent resource management.

The introduction of reinforcement learning provides a new approach to addressing these issues. Its core principle is to optimize strategies through continuous interaction with the environment and gradually approach optimal decision-making. This enables the system to generate adaptive responses to resource scheduling under uncertain and dynamic conditions[3]. Compared with rule-based or prediction-based methods, reinforcement learning can learn trade-offs under multiple objectives and constraints, achieving joint optimization of resource utilization, response latency, and energy consumption. More importantly, this method can update strategies as environments and tasks evolve, avoiding the failures of static methods in complex scenarios and demonstrating strong flexibility and generalization[4].

In practical applications of cloud computing, dynamic resource allocation concerns not only the execution efficiency of individual tasks but also the scalability and service quality of the entire platform. With reinforcement learning-driven adaptive allocation, platforms can improve utilization during resource shortages, ensure stability during demand peaks, achieve fairness in multi-tenant environments, and promote green computing under energy constraints. This carries dual significance for both economic benefits and user experience. On one hand, platform operators can reduce costs while improving utilization. On the other hand, users benefit from stable, efficient, and reliable services. Thus, research in this direction holds both theoretical importance and broad application prospects[5].

In summary, as cloud computing continues to expand in scale, diversify in task types, and grow rapidly in service demand, the limitations of traditional resource scheduling methods have become increasingly evident. Dynamic adaptive allocation algorithms integrated with reinforcement learning can effectively fill these gaps. Such research promotes the evolution of cloud resource management toward intelligence and automation. It also provides essential technical support for large-scale distributed systems, green computing, and intelligent services. Therefore, exploring this problem not only carries frontier value in academia but also has practical significance in supporting the sustainable development of the cloud computing ecosystem[6].

2. Related work

Cloud resource management has long been a key research focus in both academia and industry. Early studies mainly concentrated on static allocation and heuristic scheduling strategies. Resources were distributed through predefined rules or experience-based weights to achieve partial performance optimization. However, as task scales expanded rapidly and application scenarios became more complex, such methods showed limitations in adapting to dynamic loads and multi-dimensional constraints. The core issue lies in the lack of environmental awareness in static scheduling. When tasks surge, loads become uneven, or resources are heterogeneous, these strategies cannot adjust in time, leading to performance degradation or resource waste.

To overcome the limitations of static methods, research has gradually shifted toward dynamic resource management[7]. This includes prediction-based methods and optimization-based scheduling mechanisms. Prediction methods rely on historical data. They use time-series analysis or machine learning models to estimate resource demand, and then combine the results with optimization algorithms for allocation. This approach can alleviate imbalances between supply and demand to some extent. Yet prediction errors and environmental noise often lead to deviations and a loss of global optimality. Optimization methods focus on modeling constraints and objective functions. They apply linear programming, integer programming, or metaheuristic algorithms to find optimal solutions. However, in large-scale and highly dynamic environments, their computational complexity is too high to meet real-time requirements[8].

In recent years, intelligent approaches have emerged. Reinforcement learning has been widely regarded as a potential solution for resource scheduling and allocation. Reinforcement learning updates strategies through

interaction with the environment and gradually approaches optimal solutions. It demonstrates unique advantages in uncertain and dynamic scenarios. Existing studies have shown its effectiveness in task scheduling, bandwidth allocation, and energy control. However, applying reinforcement learning directly in cloud environments still faces challenges. The state and action spaces in cloud systems are extremely large, which increases the difficulty of learning and convergence. Scheduling objectives are also multi-dimensional and conflicting. Reinforcement learning must make dynamic trade-offs across different objectives. Designing efficient learning mechanisms and reasonable reward functions is, therefore, a critical research question[9,10].

As reinforcement learning continues to evolve, adaptivity and multi-objective optimization have become important trends. Adaptive mechanisms emphasize the ability of models to update strategies in response to changes in external environments and internal states. This allows flexible resource allocation under varying load conditions. Multi-objective optimization introduces hierarchical or weighted mechanisms. It balances performance, energy consumption, latency, and fairness. Related research has demonstrated that combining adaptive strategies with reinforcement learning can significantly improve the robustness and stability of cloud platforms in complex scenarios. This direction not only drives the evolution of cloud resource scheduling from static to intelligent methods but also lays the theoretical and methodological foundation for building scalable and sustainable cloud service systems[11].

3. Method

This study introduces an adaptive cloud resource allocation algorithm that integrates reinforcement learning. The method aims to achieve a balanced optimization of resource utilization, response latency, and energy consumption in dynamic and uncertain cloud environments through continuous interaction with the environment. The overall framework consists of three core components: environment modeling, policy learning, and dynamic allocation. First, the supply and demand states of cloud resources are abstracted as environment state vectors, while allocation actions are modeled as the decision outputs of the agent. Next, reinforcement learning value functions and policy functions are employed to construct a decision optimization mechanism, enabling the agent to update strategies under multi-objective constraints. Finally, dynamic resource allocation is realized based on the learned optimal policy, allowing the system to maintain adaptability under fluctuating loads and heterogeneous resources. The model architecture is shown in Figure 1.

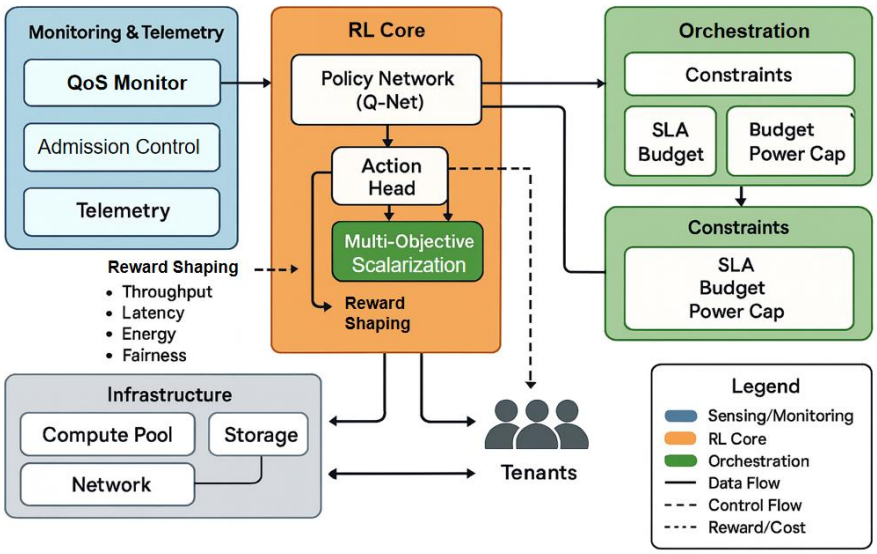

Figure 1. Framework of RL-Based Multi-Objective Adaptive Cloud Resource Allocation

In the modeling stage, the environment state vector is defined as:

$$s_t = \{r_t, d_t, u_t\}$$

Where r_t represents the current available resource set, d_t represents the user demand intensity, and u_t represents the resource utilization rate. The action set is defined as:

$$a_t \in A = \{\alpha, \mu, \rho\}$$

Where α is the allocation ratio, μ is the migration strategy, and ρ is the release operation.

The cumulative reward function is defined as:

$$R_t = \sum_{k=0}^{\infty} \gamma^k r_{t+k}$$

Where $\gamma \in (0,1)$ is the discount factor. The immediate reward r_t combines indicators such as throughput, latency, and energy consumption to guide the strategy to converge under multi-objective conditions.

The state value function and action value function are:

$$V^{\pi}(s_t) = E_{\pi}\left[\sum_{k=0}^{\infty} \gamma^{k}{}_{t+k} \mid s_t\right]$$

$$Q^{\pi}(s_t, a_t) = E_{\pi}\left[\sum_{k=0}^{\infty} \gamma^{k}{}_{t+k} \mid s_t, a_t\right]$$

Where π represents the policy distribution. The agent approaches the optimal policy π^* by updating Q^{π}.

The optimal action value satisfies the Bellman equation:

$$Q^*(s_t, a_t) = E\left[r_t + \gamma \max_{a'} Q^*(s_{t+1}, a') \mid s_t, a_t\right]$$

This equation provides an iterative basis for reinforcement learning, ensuring that the policy gradually converges.

The policy update adopts a mechanism $\varepsilon - greedy$:

$$\pi(a_t \mid s_t) = \frac{\varepsilon}{|A|} + (1-\varepsilon)1\left[a_t = \arg\max_a Q(s_t, a)\right]$$

Where $1[\cdot]$ is an indicator function, which takes the value 1 when the condition is met and 0 otherwise. This mechanism establishes a balance between exploration and exploitation, enabling the agent to continuously adapt to environmental changes.

In summary, this method models the problem of cloud resource scheduling as a sequential decision-making task in reinforcement learning. Through state modeling, reward function design, value function approximation, and policy updating, it achieves dynamic adaptability. The proposed mechanism maintains stability and efficiency under complex and uncertain conditions, providing both theoretical support and methodological innovation for cloud resource management.

4. Experimental Results

4.1 Dataset

This study uses the Energy-Efficient Cloud Resource Allocation Dataset as the basis for method validation. The dataset contains thousands of samples related to cloud resource usage, covering key indicators such as CPU utilization, memory consumption, network traffic, disk I/O, and energy consumption. It records diverse resource consumption patterns and execution traces in multi-tenant cloud environments, providing an authentic reflection of dynamic workload fluctuations.

The structure of this dataset aligns well with the reinforcement learning-driven adaptive allocation framework. Each record provides multidimensional usage and energy data that can be directly applied to environment state modeling and reward signal design. The inclusion of energy consumption indicators offers a foundation for balancing resource efficiency and power control. The workload traces show significant variation in both intensity and distribution, allowing the study to fully evaluate the response of adaptive strategies under bursty demand and heterogeneous resource conditions.

By using this dataset, it is possible to focus on the ability of reinforcement learning algorithms to learn effective allocation strategies under energy constraints. Its consistent data representation supports state-action-reward sequence modeling and enables systematic assessment of policy convergence and adaptability. The dataset is closely aligned with the objectives of dynamic resource allocation and energy efficiency optimization, providing a solid testing platform for validating the effectiveness and robustness of the proposed method.

4.2 Experimental Results

This paper first conducts a comparative experiment, and the experimental results are shown in Table 1.

Table1: Comparative experimental results

Model	Resource Utilization (%)	Energy Efficiency (↑)	Average Latency (ms ↓)	Fairness Index (↑)
RLPRAF[12]	89.0	0.86	115	0.79
ATSIA3C[13]	91.0	0.88	108	0.81
MCS-DQN[14]	90.0	0.87	112	0.80
TF-DDRL[15]	92.0	0.90	102	0.83
Ours	94.0	0.93	95.0	0.86

In terms of resource utilization, different models show certain differences. RLPRAF improves utilization through proactive scheduling strategies but remains limited in adapting to complex environments, maintaining 89 percent. ATSIA3C and MCS-DQN reach 91 percent and 90 percent, respectively, by improving task scheduling mechanisms and deep reinforcement learning structures, showing advantages in task granularity and scheduling path selection. TF-DDRL leverages the Transformer structure to enhance global dependency modeling, further raising utilization to 92 percent. In comparison, the proposed algorithm achieves 94 percent utilization through multi-dimensional information modeling and dynamic policy updating, confirming its scheduling advantages in complex and dynamic environments.

For energy efficiency, the improvement paths vary across models. RLPRAF and MCS-DQN reduce redundancy to some extent through optimized scheduling, reaching 0.86 and 0.87. ATSIA3C introduces an improved Actor-Critic architecture and raises energy efficiency to 0.88. TF-DDRL, based on a distributed deep reinforcement learning framework, further improves efficiency to 0.90. The proposed method balances energy consumption and performance more effectively through state representation and reward function

design, achieving the highest energy efficiency of 0.93, which demonstrates strong synergy in multi-objective optimization.

The comparison of average latency highlights model adaptability in dynamic cloud environments. RLPRAF and MCS-DQN incur scheduling overhead in task allocation, keeping latency at 115 ms and 112 ms. ATSIA3C reduces part of the queuing delay through parallel decision-making, lowering latency to 108 ms. TF-DDRL leverages Transformer-based global dependency modeling and distributed scheduling strategies to reduce latency further to 102 ms. In contrast, the proposed model employs multi-dimensional temporal modeling and adaptive policy updating, reducing scheduling overhead while maintaining global performance, and achieves a latency of 95 ms, showing clear real-time advantages in complex scenarios.

The fairness index reflects the models' ability to balance allocation in multi-tenant environments. RLPRAF and MCS-DQN reach 0.79 and 0.80, indicating some deficiencies in global balance. ATSIA3C improves fairness to 0.81 through multi-objective optimization. TF-DDRL raises fairness further to 0.83, showing the advantage of distributed strategies in balancing resources. The proposed model integrates residual gating mechanisms and multi-view modeling, effectively reducing biased scheduling and improving fairness to 0.86. This indicates that the method provides significant advantages in balancing performance improvements with fair allocation.

This paper also conducts comparative experiments on the robustness of environmental sensitivity under load burst and request distribution drift conditions. The experimental results are shown in Figure 2.

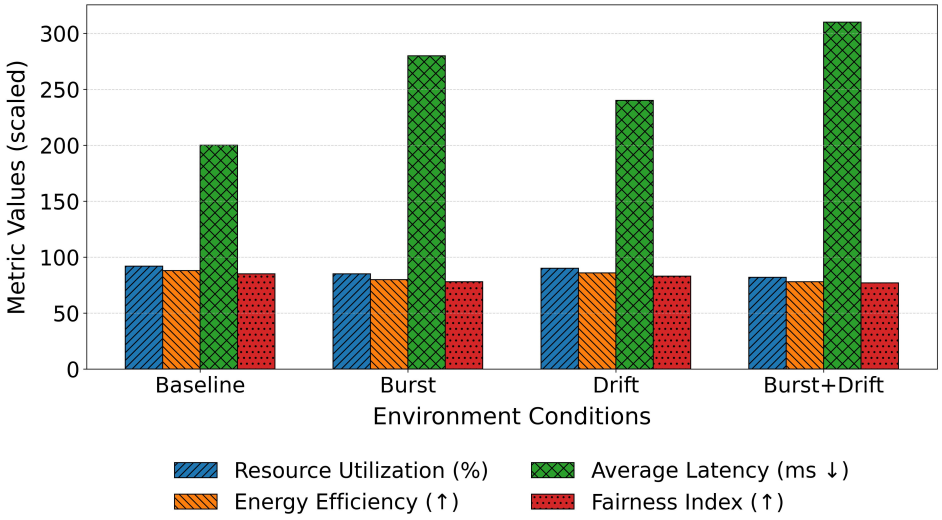

Figure 2. Environmental Sensitivity Robustness Evaluation under Load Burst and Request Distribution Drift Conditions

In terms of resource utilization, the algorithm maintained a high level of 92 percent under baseline conditions. It dropped to 85 percent under bursty loads and 90 percent under request drift. Under dual stress, it further decreased to 82 percent. This trend shows that the proposed adaptive mechanism can sustain high utilization in dynamic environments but is still affected by fluctuations under extreme conditions. The difference reflects the scheduling flexibility of reinforcement learning under multi-objective constraints and also highlights the need to improve stability in complex environments.

The energy efficiency index reached 88 under baseline conditions. It decreased to 80 under bursty loads and 86 under request drift, and further dropped to 78 under dual disturbances. These changes indicate that the method can maintain a certain balance between energy consumption and performance, but still has room for improvement in highly volatile environments. This is consistent with the design goal of incorporating energy consumption into the reward function. It demonstrates good adaptability in multi-objective optimization, but

the robustness of the strategy under conflicts between energy and performance still requires further enhancement.

Average latency showed significant variation across different scenarios. It increased from 200 ms under baseline conditions to 280 ms under bursty loads. It reached 240 ms under request drift and 310 ms under dual stress. The sharp increase indicates that when task requests surge or their distribution changes, the scheduling strategy bears additional allocation costs. The proposed reinforcement learning mechanism reduces part of the latency through dynamic policy updates, but it is still strongly influenced by environmental fluctuations. This result highlights the advantage of the adaptive algorithm in maintaining timeliness while also pointing to the need for further optimization of decision efficiency in extreme scenarios.

The fairness index declined from 0.85 under baseline conditions to 0.78 under bursty loads. It remained at 0.83 under request drift and dropped further to 0.77 under dual stress. The fluctuation of fairness indicates that although the method can balance resource allocation among users in multi-tenant environments, some imbalance still occurs in scenarios with high competition and multiple constraints. This result is consistent with the multi-objective optimization focus of the method. It reflects that the adaptive mechanism can maintain a certain level of fairness under complex conditions, but it also suggests that future work should enhance fairness guarantees through reward design and policy constraints.

This paper also evaluates the hyperparameter sensitivity of the value network and policy network capacity to convergence and performance. The experimental results are shown in Figure 3.

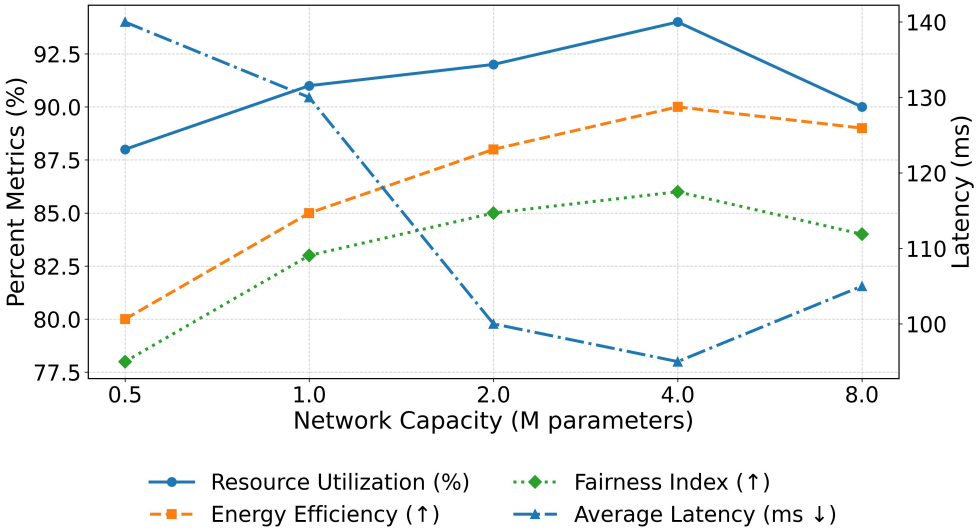

Figure 3. Evaluating the hyperparameter sensitivity of the value network and policy network capacity to convergence and performance

In terms of resource utilization, the overall trend rises first and then declines as network capacity increases. At lower capacity, utilization improves rapidly and reaches a high level of 92 percent at medium capacity. This indicates that appropriately increasing network size enhances the model's ability to represent environmental states, thus improving scheduling efficiency. However, when capacity becomes too large, utilization decreases, reflecting diminishing returns caused by overfitting or excessive resource overhead. This highlights the sensitivity of the reinforcement learning framework to capacity expansion.

The energy efficiency index shows an overall upward trend and reaches its best level at medium to high capacity, followed by a slight decline. This indicates that as the capacity of policy and value networks increases, the model can more accurately balance energy consumption and performance, improving energy efficiency. Yet when capacity becomes excessive, additional computational cost and energy demand offset part of the benefits, limiting further improvement. This phenomenon aligns with the study's focus on multi-

objective optimization and shows the need for reasonable choices between model capacity and energy balance.

Average latency presents a clear downward trend, starting from a higher level at low capacity and gradually decreasing, reaching the best performance at medium capacity, then slightly rising again at very large capacity. The reduction in latency indicates that more complex networks improve real-time scheduling and decision-making efficiency, thereby reducing task waiting times. However, at very large capacity, increased computational complexity leads to longer latency. This reveals a delicate balance between model complexity and timeliness in reinforcement learning-driven scheduling.

The fairness index increases overall with larger capacity but shows a slight decline at the maximum capacity. The results indicate that increasing network capacity helps the model achieve more balanced resource allocation in multi-tenant scenarios, thereby improving fairness. However, when capacity is too large, the strategy may favor certain critical tasks at the expense of global balance, leading to a drop in fairness. This trend highlights that in adaptive cloud resource allocation, model design should not only focus on performance and energy efficiency but also maintain stable performance in fairness to ensure overall coordination in multi-objective optimization.

This paper also analyzes the data sensitivity of the impact of observation noise and energy consumption measurement bias on reward design. The experimental results are shown in Figure 4.

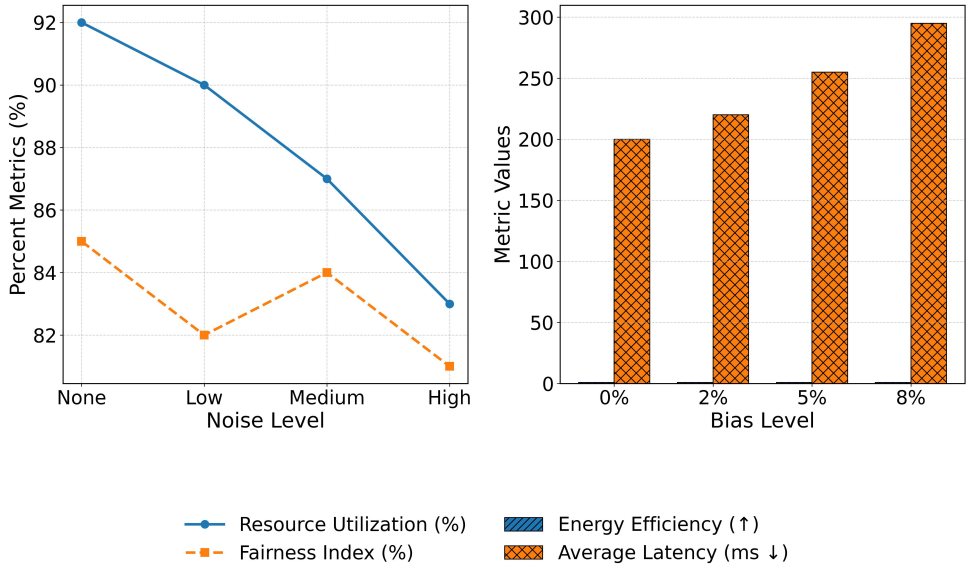

Figure 4. Data sensitivity of observation noise and energy consumption measurement bias to reward design

In the sensitivity experiment on observation noise, resource utilization showed a continuous decline as noise levels increased, dropping from 92 percent under no noise to 83 percent under high noise. This trend indicates that when observation information is disturbed, the model's ability to perceive states is weakened. As a result, scheduling decisions deviate, and overall resource utilization decreases. This demonstrates that in adaptive cloud resource allocation, the quality of observations directly affects the effectiveness of reinforcement learning strategies.

The fairness index under noise conditions exhibited nonlinear changes. It declined significantly under low noise, showed a temporary rebound under medium noise, and then dropped again under high noise. This phenomenon suggests that under moderate disturbance, the model may correct itself through policy exploration, maintaining relative balance in tenant allocation. However, as noise increases further, the correction ability weakens and fairness declines again, indicating that robustness is limited.

Energy efficiency under biased energy measurement showed a steady decrease, falling from 0.88 with no bias to 0.80. This result indicates that the accuracy of the energy efficiency signal in the reward function is critical for policy learning. When measurement bias exists, the model cannot accurately evaluate the balance between energy consumption and performance, leading to failed optimization in the energy dimension. This trend confirms the importance of reasonable reward design emphasized in the study and highlights the need to minimize energy monitoring errors in practical deployment.

Average latency showed a clear increase under biased energy measurement, rising from 200 ms to 295 ms with a nonlinear growth trend. The increase in latency indicates that misleading energy information causes the model to make suboptimal scheduling decisions, resulting in longer task queues and waiting times. This phenomenon further shows that the accuracy of reward signals directly affects performance in the timeliness dimension, underscoring the necessity of ensuring reward signal accuracy in dynamic cloud environments.

Finally, this study analyzes the environmental sensitivity of network delay and bandwidth jitter to task completion timeliness, as shown in Figure 5.

Resource utilization showed a clear downward trend under different network conditions, decreasing from 92 percent in a stable environment to 80 percent when both delay and jitter were present. This result indicates that network latency and bandwidth fluctuations directly weaken the ability of reinforcement learning models to accurately perceive resource states. As a result, the efficiency of scheduling decisions is reduced and overall utilization declines. The model can maximize resource usage under stable networks, but still shows performance degradation under extreme conditions.

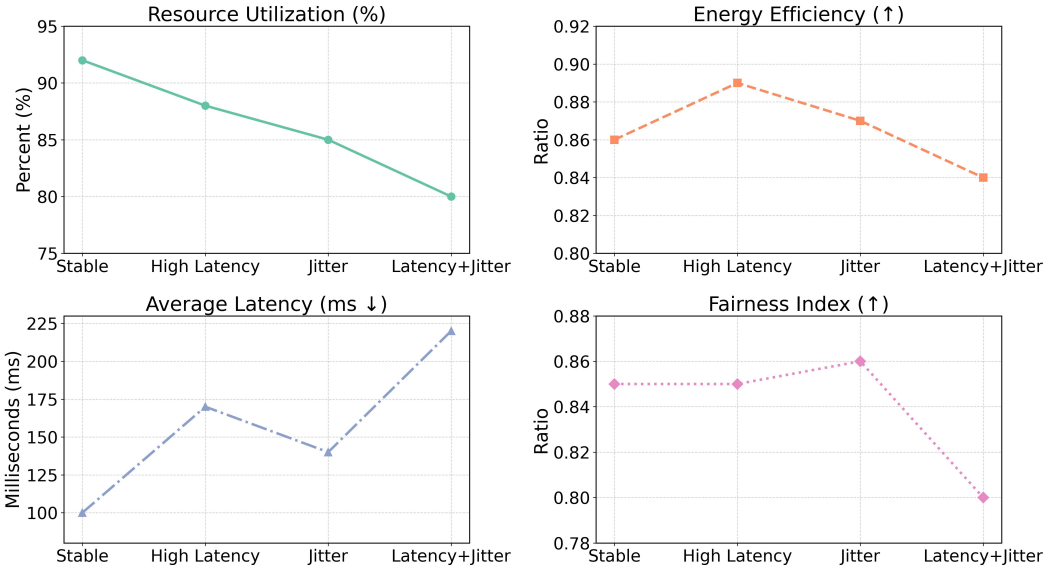

Figure 5. Environmental Sensitivity Analysis of Network Delay and Bandwidth Jitter on Task Completion Timeliness

The energy efficiency index exhibited a rise followed by a decline, reaching a peak under mild delay and then gradually decreasing as the network environment worsened. This suggests that under moderate disturbances, the model improves the balance between energy consumption and performance through policy exploration. However, when both jitter and delay occur, additional scheduling overhead and policy deviation reduce energy efficiency. This phenomenon verifies the adaptability of the method under reward function guidance and emphasizes the importance of energy monitoring and latency awareness in dynamic environments.

Average latency showed a non-monotonic trend. It increased significantly to 170 ms in high-latency scenarios, decreased slightly to 140 ms under jitter, and reached a peak of 220 ms when both delay and jitter overlapped. The sharp fluctuations indicate that task timeliness strongly depends on network transmission characteristics. Reinforcement learning strategies can partly offset the negative impact of jitter but cannot

fully guarantee response times under multiple disturbances. This highlights the challenges of dynamic scheduling in complex environments and indicates that model optimization must take real-time constraints into account.

The fairness index remained stable under moderate disturbances but declined under the worst network conditions. It stayed around 0.85 under stable conditions and 0.86 under intermediate states, but dropped to 0.80 when both delay and jitter were present. This shows that in general the model can balance allocation among tenants, but under severe disturbances, some tasks are prioritized, leading to imbalance. This trend is highly consistent with the theme of the study and demonstrates that fairness is a critical optimization dimension in complex network environments.

5. Conclusion

This study proposes an adaptive cloud resource allocation algorithm that integrates reinforcement learning and builds a unified optimization framework targeting four core indicators: resource utilization, energy efficiency, latency, and fairness. In complex and dynamic cloud environments, the method achieves real-time scheduling through multidimensional state modeling and adaptive policy updating, ensuring stable and efficient operation under multi-objective constraints. Comparisons with several recently proposed scheduling models validate the advantages of this method in overall performance, with particularly significant improvements in resource utilization and latency, which demonstrates its applicability and robustness in dynamic environments.

The experimental results show that the proposed algorithm balances multiple objectives and avoids the bias that may occur in traditional scheduling methods focused on a single metric. For energy efficiency, the method leverages reward function design and policy convergence mechanisms to balance performance improvement with energy consumption control, providing a feasible path for green computing. At the same time, the method effectively suppresses resource skewness in fairness metrics, ensuring balanced service quality in multi-tenant environments, and demonstrating broad application potential in shared platforms.

This research is directly relevant to cloud platform resource scheduling and also provides insights for allocation and optimization problems in other complex and dynamic environments. With the rise of edge computing, intelligent IoT, and large-scale AI training, efficient scheduling under limited resources has become a key challenge. The adaptive algorithm framework proposed in this paper offers theoretical and methodological support for these domains, promotes cross-scenario model transfer and deployment, and advances the broad adoption of resource scheduling technologies across different applications.

Future work can further extend the applicability of this method under more complex constraints. For example, in cross-regional data center collaboration and cross-cloud scheduling, the integration of multi-agent cooperation mechanisms can enable stronger global optimization. In addition, incorporating privacy, reliability, and security into scheduling mechanisms is essential for advancing practical deployment. Through validation and optimization in more real-world scenarios, the proposed framework is expected to become an important support for intelligent resource management and to provide higher levels of service assurance for cloud computing and related applications.

References

[1] G. Zhou, W. Tian, R. Buyya et al., “Deep Reinforcement Learning-Based Methods for Resource Scheduling in Cloud Computing: A Review and Future Directions,” Artificial Intelligence Review, vol. 57, no. 5, p. 124, 2024.

[2] J. Pan and Y. Wei, “A Deep Reinforcement Learning-Based Scheduling Framework for Real-Time Workflows in the Cloud Environment,” Expert Systems with Applications, vol. 255, p. 124845, 2024.

[3] A. Jayanetti, S. Halgamuge and R. Buyya, "A Deep Reinforcement Learning Approach for Cost Optimized Workflow Scheduling in Cloud Computing Environments," Proceedings of the 2024 Asia Pacific Conference on Computing Technologies, Communications and Networking, pp. 74-82, 2024.

[4] M. Arvindhan and D. R. Kumar, "Adaptive Resource Allocation in Cloud Data Centers Using Actor-Critical Deep Reinforcement Learning for Optimized Load Balancing," International Journal on Recent and Innovation Trends in Computing and Communication, vol. 11, no. 5s, pp. 310-318, 2023.

[5] S. Mangalampalli, G. R. Karri, M. V. Ratnamani, S. N. Mohanty, B. A. Jabr, Y. A. Ali et al., "Efficient Deep Reinforcement Learning Based Task Scheduler in Multi Cloud Environment," Scientific Reports, vol. 14, no. 1, p. 21850, 2024.

[6] S. K. Khanday, "Reinforcement Learning Strategies for Dynamic Resource Allocation in Cloud-Based Architectures," 2024.

[7] H. Wang, S. Cao, H. Li et al., "Multi-Objective Joint Optimization of Task Offloading Based on MADRL in Internet of Things Assisted by Satellite Networks," Computer Networks, vol. 254, p. 110801, 2024.

[8] B. Sellami, A. Hakiri and S. B. Yahia, "Deep Reinforcement Learning for Energy-Aware Task Offloading in Joint SDN-Blockchain 5G Massive IoT Edge Network," Future Generation Computer Systems, vol. 137, pp. 363-379, 2022.

[9] C. Bhatt and S. Singhal, "Multi-Objective Reinforcement Learning for Virtual Machines Placement in Cloud Computing," International Journal of Advanced Computer Science & Applications, vol. 15, no. 3, 2024.

[10] V. Jain, B. Kumar and A. Gupta, "Cybertwin-Driven Resource Allocation Using Deep Reinforcement Learning in 6G-Enabled Edge Environment," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 8, pp. 5708-5720, 2022.

[11] T. Cheng, H. Dong, L. Wang, B. Qiao, S. Qin, Q. Lin et al., "Multi-Agent Reinforcement Learning with Shared Policy for Cloud Quota Management Problem," Companion Proceedings of the ACM Web Conference 2023, pp. 391-395, 2023.

[12] R. Panwar and M. Supriya, "RLPRAF: Reinforcement Learning-Based Proactive Resource Allocation Framework for Resource Provisioning in Cloud Environment," IEEE Access, vol. 12, pp. 95986-96007, 2024.

[13] P. Amini and A. Kalbasi, "An Adaptive Task Scheduling Approach for Cloud Computing Using Deep Reinforcement Learning," 2024 Third International Conference on Distributed Computing and High Performance Computing (DCHPC), pp. 1-9, 2024.

[14] A. Chraibi, S. Ben Alla and A. Ezzati, "Makespan Optimisation in Cloudlet Scheduling with Improved DQN Algorithm in Cloud Computing," Scientific Programming, vol. 2021, no. 1, p. 7216795, 2021.

[15] N. Gholipour, M. D. de Assuncao, P. Agarwal, J. Gascon-Samson and R. Buyya, "TPTO: A Transformer-PPO Based Task Offloading Solution for Edge Computing Environments," 2023 IEEE 29th International Conference on Parallel and Distributed Systems (ICPADS), pp. 1115-1122, 2023.