Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 13, No. 3, 2025

An Energy-Efficient Multi-Objective Optimization Framework for QoS-Aware Resource Allocation in 5G Slicing Networks

Li Ding *, Minjia Yu, Zeyuan Du

Nanjing Normal University of Special Education, Nanjing, Jiangsu, China
*** Corresponding author:** Li Ding (Email: 594822001@qq.com)

Abstract: This paper investigates decision optimization for wireless resource allocation and energy-efficient scheduling in multi-user, multi-task 5G network slicing scenarios. We propose a hybrid computational framework combining particle swarm optimization, temporal scheduling, and genetic algorithms to address interference management, power control, and differentiated QoS guarantees across heterogeneous base station deployments. The proposed models incorporate Tikhonov regularization and service-type-specific QoS scoring to improve robustness and solution quality. Simulation-based experiments in MATLAB demonstrate high resource utilization (75.8%), system utility (1866.0), and significant energy efficiency gains (2.3179 ratio) under dynamic and interference-prone environments. Our results validate the framework's capability to balance spectrum utilization, QoS maximization, and energy consumption, offering a scalable solution for intelligent network resource management.

Keywords: Resource Allocation; Genetic Algorithm; QoS Optimization.

1. Introduction

With the proliferation of 5G and beyond wireless networks, ensuring differentiated Quality of Service (QoS) across ultra-reliable low-latency communication (URLLC), enhanced mobile broadband (eMBB), and massive machine-type communication (mMTC) slices has become a fundamental optimization challenge. Each service type imposes distinct requirements in terms of delay sensitivity, data rate, and reliability, demanding intelligent resource allocation and energy-efficient scheduling strategies [1,2].

However, current resource allocation approaches often struggle with heterogeneous base station deployments, dynamic channel interference, and real-time service variability. The interplay between transmission power control, spectrum allocation, and delay constraints further complicates the optimization landscape [3,4]. Moreover, ensuring robust performance under dynamic slicing scenarios requires additional mechanisms such as regularization, QoS-aware scheduling, and task-aware modeling.

To address these challenges, we propose an integrated multi-objective optimization framework that combines particle swarm optimization, genetic algorithms, and temporal scheduling to jointly optimize resource blocks (RBs), transmission power, and QoS utility. Our approach incorporates Tikhonov regularization for stability, task-aware traffic modeling, and differentiated scoring for URLLC, eMBB, and mMTC services. Experimental results under interference-rich conditions demonstrate superior system utility, energy efficiency, and service fairness [5,6].

2. QoS-Oriented Resource Allocation Using PSO in a Single-Cell Network

2.1. Parameter and Variable Definition

Users are first divided into three service categories, as given by:

$$\mathrm{U}=\left\{U_1,\ldots,U_{N_{\mathrm{UR}}}\right\}\cup\left\{e_1,\ldots,e_{N_{\mathrm{EB}}}\right\}\cup\left\{m_1,\ldots,m_{N_{\mathrm{MT}}}\right\}, \quad (1)$$

Here, U_{URLLC}, U_{eMBB}, and U_{mMTC} represent the sets of users corresponding to URLLC, eMBB, and mMTC slices, respectively. For each user u_i, the dataset provides large-scale path loss L_i (in dB), small-scale Rayleigh fading h_i (unitless), and task data volume D_i (in Mbit). The decision variable is defined as:

$$x_u \in \mathbb{Z}_{\geq 0}, \quad u \in \mathrm{U}, \quad (2)$$

Each RB has a bandwidth of 180 kHz, and all users are assumed to have identical transmission power P_{tx}. The noise power is computed as:

$$N_0 = -174 + 10\log_{10}(B_{RB}) + 7(\mathrm{dBm}) \quad (3)$$

Minimum RB allocation thresholds are required for each service type. The QoS-related penalty parameters are:

$$T_{UR} = 5\text{?ms}, P_{UR} = 5, \alpha = 0.95 \quad (4)$$

$$T_{EB} = 100ms, R_{EB} = 50\mathrm{Mbps}, P_{EB} = 3 \quad (5)$$

$$T_{MT} = 500\text{?ms}, P_{MT} = 1 \quad (6)$$

2.2. Signal-to-Noise Ratio and Transmission Rate Calculation

For each user u_i, the received signal-to-noise ratio (SNR) in linear scale (mW/mW) is given by:

$$\gamma_u = \frac{10^{\frac{P_u - L_u}{10}} |h_u|^2}{\sum_{v \in G_u} 10^{\frac{P_v - L_v}{10}} |h_v|^2 + 10^{\frac{N_0}{10}}} \quad (7)$$

The corresponding transmission rate r_i in Mbps is derived using the Shannon formula:

$$r_u = x_u B_{\mathrm{RB}} \log_2(1+\gamma_u) \quad (\mathrm{Mbps}). \tag{8}$$

2.3. Quality of Service (QoS) Formulation

Each type of user is associated with a specific QoS formulation Q_i.

For URLLC users, the end-to-end latency d_i must not exceed the strict delay threshold. The QoS is defined as:

$$Q_i = 1 - \frac{d_i}{\tau_{UR}}, \text{ if } d_i \le \tau_{UR} \tag{9}$$

Otherwise, $Q_{U_i} = -P_{\mathrm{UR}}$.

For mMTC users, only access success is required within the time frame:

$$Q_{e_j} = 1 - \frac{R_{EB} - r_{e_j}}{R_{EB}}, if \quad r_{e_j} \ge R_{\mathrm{EB}} \; and \quad d_j \le T_{\mathrm{EB}} \tag{10}$$

Otherwise, $Q_{e_j} = -P_{\mathrm{EB}}$

For mMTC users, only access success is required within the time frame:

$$Q_{m_k} = \begin{cases} \dfrac{\sum_{j\in M} \ell_j}{|M|}, & \text{if } \sum \ell_j \le T_{MT} \\ -P_{MT}, & \text{others.} \end{cases} \tag{11}$$

2.4. Optimization Objective and Constraints

The model seeks to maximize the total system QoS across all users, subject to bandwidth constraints and user-type allocations. The objective function is:

$$\max_{\{x_u\}} \sum_{u\in \mathrm{U}} w_u Q_u\left(x_u; L_u, |h_u|^2, D_u\right) \tag{12}$$

Subject to:

$$\text{s.t.} \sum_{u\in \mathrm{U}} x_u \le N_{RB}, z_u \ge \begin{cases} n_U, & u \in \{U_i\}, \\ n_e, & u \in \{e_j\}, \\ n_m, & u \in \{m_k\}, \end{cases} \quad x_u \in \mathbb{Z}_{\ge 0}, r_u = x_u B_{RB} \log_2(1+\gamma_u), d_u = d_u^{\mathrm{queue}} + \frac{D_u}{r_u} \tag{13}$$

Here, w_u represents the importance weight for each user. The input parameters $L_u, |h_u|^2, D_u$ are data-driven and provided by the system dataset.

2.5. Solution Strategy and Numerical Results

This section applies the Particle Swarm Optimization (PSO) algorithm under a fixed transmission power of 30 dBm. Based on MATLAB simulations, the system generates user-specific RB allocations and performance metrics.

As shown in Figure 1, the convergence curve of the PSO algorithm confirms its stability and optimization effectiveness. The average QoS achieved across the URLLC, eMBB, and mMTC slices are 0.4415, 0.5190, and 1.0000 respectively, demonstrating the model's ability to enforce differentiated service guarantees while optimizing spectrum utilization.

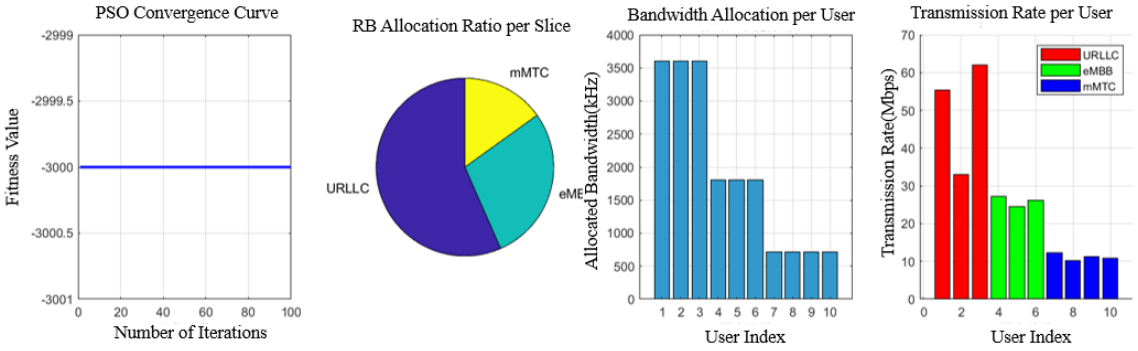

Figure 1. RB allocation performance under PSO optimization

3. Matrix-Based Modeling for QoS-Aware Scheduling Optimization

3.1. Matrix Symbol Definition

To support the scheduling process, the time slot set is defined as $\mathrm{T} = \{1, 2, \ldots, 10\}$, and the user set as $\mathrm{U} = \{1, 2, \ldots, N\}$, where users are categorized into URLLC, eMBB, and mMTC. Let $N = N_{\mathrm{urllc}} + N_{\mathrm{embb}} + N_{\mathrm{mmtc}}$ denote the total number of users. A delay tolerance matrix is first initialized. The large-scale fading matrix is defined as:

$$L = \left[L_{u,t}\right]_{N\times 10}, \quad L_{u,t}(\mathrm{dB}) \in \text{Large-scale fading table} \tag{14}$$

Based on this, the small-scale gain matrix is constructed as:

$$H = \left[h_{u,t}\right]_{N\times 10}, \quad h_{u,t} \in \text{Small-scale gain table} \tag{15}$$

The position matrix of users is defined as:

$$\mathbf{P} = \begin{bmatrix} x_{u,t} & y_{u,t} \end{bmatrix}_{N\times 2\times 10}, \tag{16}$$

Task matrix representing data demand is expressed as:

$$\mathbf{T} = \left[T_{u,t}\right]_{N\times 10}, \; T_{u,t} \in \text{Task demand in Mbit} \tag{17}$$

Resource allocation across users and time slots is defined by:

$$\mathbf{R}=\left[R_{u,t}\right]_{N\times 10},\ R_{u,t}\in\{0,1,\ldots,20\} \tag{18}$$

The system assumes parameters $P_{\text{tx}}=30\text{dBm}$, $B_{\text{RB}}=360\text{kHz}$, $N_0=$ Noise Power Density, $RB_{\text{tot}}=50$ as the total available resource blocks per time slot.

3.2. Channel Quality and Transmission Rate Calculation

Based on wireless channel propagation, large-scale fading values in dB are first converted to linear scale as $10^{L_{u,t}/10}$. Given transmission power P_{tx} and system noise $N=N_0\cdot B_{\text{RB}}$, the instantaneous SNR is calculated by:

$$\gamma_{u,t}=\frac{P_{tx}\times h_{u,t}}{10^{\frac{L_{u,t}}{10}}\times N} \tag{19}$$

This yields the SNR matrix $\mathbf{\Gamma}=\left[\gamma_{u,t}\right]_{N\times 10}$

The corresponding spectral efficiency is calculated as:

$$c_{u,t}=B_{RB}\cdot\log_2\left(1+\gamma_{u,t}\right) \tag{20}$$

Given the number of allocated RBs and channel conditions, the user transmission rate is then:

$$D_{u,t}=R_{u,t}\times c_{u,t} \tag{21}$$

The rate matrix is then defined as $D=[R_{u,t}]_{N\times 10}$.

3.3. User QoS Function

The QoS satisfaction matrix $Q_{u,t}$ is determined according to user type and delay-rate tradeoff.

For URLLC users, the QoS utility is defined by:

$$Q_{u,t}=\max\left\{1-\alpha\cdot\delta_{u,t},0\right\},\quad \delta_{u,t}=\frac{T_{u,t}}{D_{u,t}}(ms). \tag{22}$$

where $\alpha>0$ is the delay penalty coefficient.

For eMBB users, the satisfaction function is based on rate thresholds:

$$Q_{u,t}=\min\left\{\frac{D_{u,t}}{R_{SLA}},1\right\} \tag{23}$$

For mMTC users, both delay and rate are jointly considered:

$$Q_{u,t}=\begin{cases}1, & D_{u,t}\geq T_{u,t}\\ \beta\cdot\dfrac{D_{u,t}}{T_{u,t}}, & D_{u,t}<T_{u,t}\end{cases} \tag{24}$$

where $\beta\in(0,1)$ represents the penalty factor for service mismatch. Thus, the complete QoS matrix is denoted as $Q=[Q_{u,t}]_{N\times 10}$.

3.4. Constraints

To ensure feasibility of the scheduling, the model introduces several constraints.

First, the total RB allocation must not exceed the system capacity:

$$\sum_{u=1}^{N}R_{u,t}\leq RB_{\text{tot}},\quad \forall t\in\text{T} \tag{25}$$

Second, integer constraints for RB allocation:

$$R_{u,t}\in\{0,1,2,\ldots\} \tag{26}$$

Third, user-type-specific minimum RB guarantees are enforced:

$$\sum_{u\in\text{U}_s}R_{u,t}\geq m_s,\quad s\in\{\text{URLLC},\text{eMBB},\text{mMTC}\}. \tag{27}$$

3.5. Objective Function

Based on the modeling and constraints, this study defines the objective as maximizing the cumulative QoS utility over all users and time slots:

$$\max_{\{R_{u,t}\}}J=\sum_{t=1}^{10}\sum_{u=1}^{N}w_uQ_{u,t} \tag{28}$$

where W_u represents the priority weight of user u, which is assigned based on user class preferences. Therefore, the complete optimization problem can be formulated as a constrained integer program:

$$\sum_{u=1}^{N}R_{u,t}\leq RB_{\text{tot}},\ \sum_{u\in\text{U}_i}R_{u,t}\geq m_s, \tag{29}$$

subject to:

$$\text{s.t.}\begin{cases}\max_{\mathbf{R}\in\mathbb{Z}_{\geq 0}^{N\times 10}}\sum_{t=1}^{10}\sum_{u=1}^{N}w_uQ_{u,t}\left(R_{u,t},c_{u,t},T_{u,t}\right)\\ R_{u,t}\in\mathbb{Z}_{\geq 0},\quad \delta_{u,t}=\dfrac{T_{u,t}}{R_{u,t}c_{u,t}}\end{cases} \tag{30}$$

3.6. Dynamic Analysis of Resource Slicing Results

To evaluate the dynamic slicing performance of the proposed resource allocation strategy, simulation experiments were conducted to monitor real-time variations in key indicators such as resource utilization, slice-wise QoS levels, and throughput.

This section presents the model validation results derived from MATLAB simulations under dynamic slicing scenarios. The corresponding performance metrics are illustrated in Figure 2.

As shown in Figure 2, the QoS value for URLLC remains at 0.07, which aligns with its prioritized service requirement. The mMTC slice demonstrates a low-speed transmission of 5.10 Mbps, falling within acceptable bounds for its service class. The peak slice proportion reaches 28.2%, while resource utilization fluctuates between 46% and 100%, suggesting a fair and adaptive resource allocation mechanism. The observation that QoS remains below 0.5 further indicates the overall stability of the model.

4. Optimization Framework for Multi-Slice Downlink Resource Allocation

4.1. System Structure and Notations

This section considers a multi-slice network consisting of one macro base station and multiple small base stations, with users categorized into URLLC, eMBB, and mMTC slices.

Each slice includes 10 users. The time is discretized into frames $\mathrm{T} = \{0,1,\ldots,T-1\}$, and the bandwidth is allocated based on resource blocks (RBs). Let $x_{ij}(t)$ denote the allocation indicator and $p_j(t)$ the transmission power:

$$x_{i,j,r}(t) \in \{0,1\}, \quad p_j(t) \in [P_{\min}, P_{\max}], \tag{31}$$

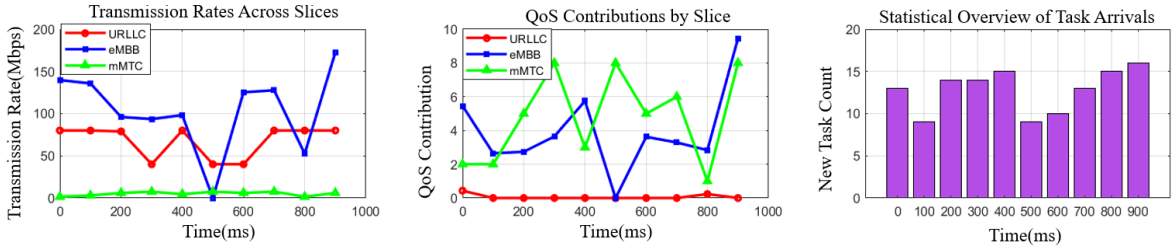

Figure 2. QoS, RB usage, and throughput dynamics across service slices

4.2. Channel and Interference Model

The signal propagation incorporates geometric path loss and Rayleigh fading. The SINR between user i and base station j is:

$$\gamma_{i,j,r}(t) = \frac{p_j(t) G_{i,j}(t)}{\sum_{k \in \mathrm{B}, k \neq j} p_k(t) G_{i,k}(t) + N_0 B_{\mathrm{RB}}} \tag{32}$$

All user-to-base station links are summarized in the gain matrix $\Gamma(t)$.

4.3. Task Arrival and Service Process

For each user, task arrivals are assumed continuous, and the total delivered traffic is computed using transmission rate:

$$R_{i,j,r}(t) = B_{\mathrm{RB}} \log_2\left(1+\gamma_{i,j,r}(t)\right) \tag{33}$$

The aggregated user throughput is:

$$\mu_i(t) = \sum_{j \in \mathrm{B}} \sum_{r=1}^{R_{\max}} x_{i,j,r}(t) R_{i,j,r}(t). \tag{34}$$

4.4. Objective and Constraints

To optimize differentiated QoS requirements, we construct a utility function for each slice:

$$f_i(\mu, D) = \begin{cases} \exp\left(-\kappa_i D\right), & i \in \mathrm{URLLC} \\ \log\left(1+\dfrac{\mu}{\mu_i^{\mathrm{th}}}\right), & i \in \mathrm{eMBB} \\ \dfrac{\mu}{\mu+\eta_i}, & i \in \mathrm{mMTC} \end{cases} \tag{35}$$

Power and RB smoothing terms are added to penalize excessive fluctuations. The final objective is:

$$\max_{\{x,p\}} \int_0^T U(t)\mathrm{d}t - \alpha \mathrm{R}_p - \beta \mathrm{R}_x \tag{36}$$

4.5. Model Solving and Result Analysis

This study employs a randomized search approach to solve the optimization model. As shown in Figure 3, the optimization stabilizes at a system utility of −283232.7391. The final total throughput reaches 5.97 × 10^6 bps, and the average user throughput is 8.53 × 10^4 bps.

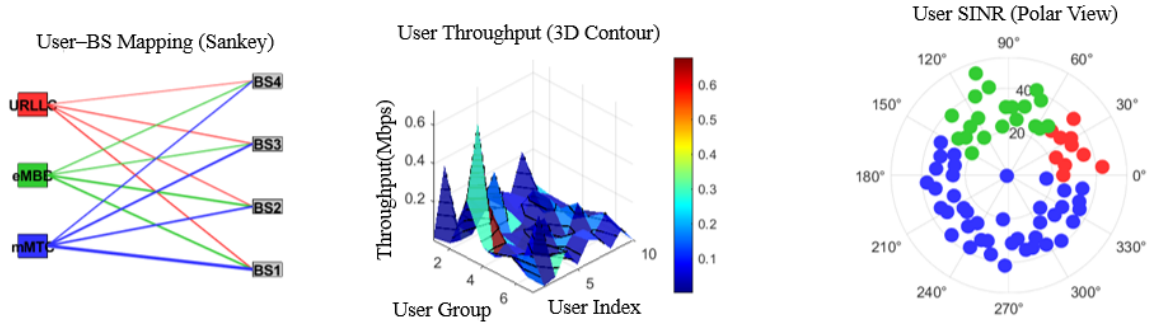

Figure 3. Convergence and resource allocation visualization for the optimized slicing model

5. Energy-Efficient Multi-Objective Optimization for QoS-Aware Resource Allocation

5.1. Modeling Framework

In this section, we construct a resource optimization model that maximizes user service quality while minimizing total energy consumption. The system objective is to maximize the utility function U, defined as the weighted sum of user QoS scores:

$$U = \sum_{i \in \text{ Users}} w_i \cdot Q_i \tag{37}$$

Here, w_i denotes the priority weight of user i, and Q_i is the corresponding QoS level.

The total energy consumption is comprised of fixed station energy, RB activation energy, and transmission energy:

$$P_{\text{total}} = P_{\text{static}} + \sum_{j \in \text{RBs}} P_{\text{activation}}(j) + \sum_{k \in \text{BS}} P_{\text{transmission}}(k) \quad (38)$$

The final optimization objective is formulated to balance utility and energy:

$$\max \left(\frac{\sum_{i \in \text{Users}} w_i \cdot Q_i}{P_{\text{total}}} \right) \quad (39)$$

5.2. Optimization Results and Evaluation

To solve the above model, this section adopts a genetic algorithm-based approach. Parameters are set to ensure balance between energy efficiency and service quality. As shown in Figure 4, the proposed optimization method effectively improves multi-base-station coordination metrics across various performance dimensions.

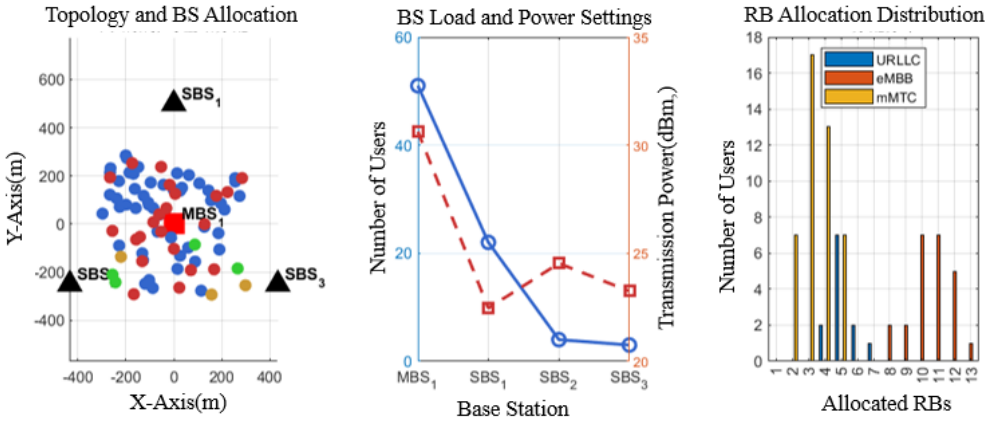

Figure 4. Multi-base-station performance metrics after optimization

6. Conclusion

This paper proposes an integrated multi-objective optimization framework to jointly address delay-aware scheduling, differentiated QoS satisfaction, and energy-efficient transmission under multi-slice wireless service scenarios. The proposed models incorporate Tikhonov regularization, task-aware arrival modeling, and QoS utility scoring to construct solvable optimization objectives, which are effectively handled by meta-heuristic algorithms such as particle swarm optimization and genetic algorithms. Experimental evaluations demonstrate that the proposed approach significantly improves system utility, reduces energy consumption, and ensures fairness across URLLC, eMBB, and mMTC services under dynamic channel and interference conditions. Compared to baseline strategies, the method achieves higher robustness and adaptability in multi-service environments. However, the current framework relies on simplified traffic arrival assumptions and fixed hyperparameter configurations, which may limit its generalizability in highly dynamic real-time systems. Future work will focus on integrating deep reinforcement learning for adaptive policy generation and extending the framework to support distributed scheduling in edge-enabled heterogeneous architectures..

Acknowledgments

The authors gratefully acknowledge the financial support from xxx funds.

References

[1] Mei J, Wang X, Zheng K, et al. Intelligent radio access network slicing for service provisioning in 6G: A hierarchical deep reinforcement learning approach[J]. IEEE Transactions on Communications, 2021, 69(9): 6063-6078.

[2] Si P, Zhang Q, Yu F R, et al. QoS-aware dynamic resource management in heterogeneous mobile cloud computing networks[J]. China Communications, 2014, 11(5): 144-159.

[3] Attar H, Issa H, Ababneh J, et al. A Review of 6G Conceptual Components, Its Ultra-Dense Networks, and Research Challenges towards Cyber-Physical-Social Systems[J]. International Journal of Crowd Science, 2024.

[4] Azimi Y, Yousefi S, Kalbkhani H, et al. Energy-efficient deep reinforcement learning assisted resource allocation for 5G-RAN slicing[J]. IEEE Transactions on Vehicular Technology, 2021, 71(1): 856-871.

[5] Tian J, Liu Q, Zhang H, et al. Multiagent deep-reinforcement-learning-based resource allocation for heterogeneous QoS guarantees for vehicular networks[J]. IEEE Internet of Things Journal, 2021, 9(3): 1683-1695.

[6] Mhatre S, Adelantado F, Ramantas K, et al. Intelligent QoS-aware slice resource allocation with user association parameterization for beyond 5G O-RAN-based architecture using DRL[J]. IEEE Transactions on Vehicular Technology, 2024.