Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Joint 3D Trajectory Planning and Task Offloading in Multi-UAV MEC Systems: A Stackelberg Game and Decomposed MARL Approach

Wei Wang
Putian University
Yuanmo Lin
linyuanmo@ptu.edu.cn

Minjiang University
Yuanyuan Han
Putian University
Zhiyong Fan
Putian University
Shien Chen
Putian University

Research Article

Keywords: UAV-MEC, 3D Trajectory Planning, Stackelberg Game, Multi-Objective Reinforcement Learning, Task Offloading

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11625340/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Joint 3D Trajectory Planning and Task Offloading in Multi-UAV MEC Systems: A Stackelberg Game and Decomposed MARL Approach.

Wei Wang[1], Yuanmo Lin[2,3*], Yuanyuan Han[3], Zhiyong Fan[4], Shien Chen[3]

[1]School of Intelligent Manufacturing, Putian University, Putian, Fujian, China.
[2]Fujian Provincial Key Laboratory of Information Processing and Intelligent Control, Minjiang University, Fuzhou, Fujian, China.
[3]School of Artificial Intelligence, Putian University, Putian, Fujian, China.
[4]Information Construction and Management Center, Putian University, Putian, Fujian, China.

*Corresponding author(s). E-mail(s): linyuanmo@ptu.edu.cn;

Abstract

To address the tight coupling between UAV 3D trajectory planning, user association, and task offloading, alongside the multi-objective trade-offs among latency, energy consumption, and throughput in UAV-assisted mobile edge computing (UAV-MEC) systems, this paper proposes a hierarchical two-level optimization framework based on a Stackelberg leader-follower game. Specifically, the 3D trajectory and offloading strategies of the UAV swarm are modeled as the leader layer, aiming to identify Pareto-optimal solutions within a complex non-convex space. To mitigate the gradient interference prevalent in multi-objective reinforcement learning, we develop a Decomposition-based Independent Learning Strategy (DIL-MADDPG). By decomposing the original problem into scalarized sub-tasks and deploying specialized expert agents, gradient conflicts are resolved at the algorithmic level. Simultaneously, the dynamic association of ground users is formulated as the follower layer, where a matching game is utilized to achieve a locally optimal response to UAV positions, ensuring stable and real-time association. A cross-layer implicit feedback mechanism is further integrated to facilitate closed-loop coordination and co-evolution between the leader's maneuvering and the follower's association. Simulation results demonstrate that the

proposed framework significantly outperforms NSGA-II and traditional heuristics by achieving an IGD of 0.159 for convergence and an HV of 6.455 for diversity.

Keywords: UAV-MEC, 3D Trajectory Planning, Stackelberg Game, Multi-Objective Reinforcement Learning, Task Offloading

1 Introduction

With the rapid proliferation of Internet of Things devices and the surging demand for latency-sensitive services, Mobile Edge Computing (MEC) has emerged as a pivotal paradigm for providing proximity-based computing and low-latency processing [1]. Despite its advantages, traditional terrestrial MEC infrastructure is often constrained by limited coverage flexibility and high deployment costs. Furthermore, it suffers from poor adaptability in highly dynamic environments, particularly in challenging terrains such as mountainous regions, disaster zones, or during temporary large-scale events [2].

To circumvent these limitations, UAV-MEC has been introduced, leveraging aerial platforms as mobile nodes to provide on-demand deployment and dynamic coverage [3]. By optimizing their 3D trajectories, UAVs can provide ubiquitous computation offloading and communication services to ground users . However, the system architecture exhibits intricate cross-layer coupling among UAV mobility planning, user association, and offloading decisions [4]. Specifically, 3D trajectories dictate link quality and service latency, while offloading strategies govern the UAV's payload distribution and energy depletion. This tight interdependency significantly escalates the complexity of joint optimization.

Moreover, UAV-MEC systems are inherently characterized by multi-objective trade-offs. There exists a fundamental tension between minimizing latency, reducing energy consumption, and maximizing mission throughput [5–7]. For instance, aggressive latency reduction often necessitates higher flight speeds or increased transmission power, thereby accelerating propulsion and communication energy consumption. Conversely, prioritizing data collection may overload the system and compromise service stability. In dynamic scenarios, the synergy of multi-objective conflicts and variable couplings results in optimization problems that are high-dimensional, non-linear, and non-convex [8–10].

Existing literature, such as the works by Wang *et al.* [11] and Jiang *et al.* [12], has modeled UAV-MEC resource scheduling using mixed-integer linear programming or non-linear programming . These studies employ optimization techniques like successive convex approximation and block coordinate descent , which provide theoretical guarantees for convergence and optimality. However, these methods often rely on idealized static channel assumptions. Furthermore, in large-scale and highly dynamic 3D environments, their computational complexity scales exponentially with the system size, making it challenging to satisfy real-time decision-making requirements.

To enhance adaptability, several studies have explored reinforcement learning using scalarized reward functions, where multiple objectives, such as latency and energy consumption, are combined into a single scalar via weighted summation [13]. Although easy to implement with fast inference, this approach is essentially a single-point search. As noted by Felten *et al.* [14], such schemes are heavily dependent on predefined weights; a single training session yields only one trade-off solution, failing to characterize the entire Pareto front or adapt to dynamically shifting preferences. While Multi-Objective Reinforcement Learning (MORL) aims to approximate the full Pareto front by training conditional policy networks, Zhang *et al.* [15–18] identified that severe gradient interference occurs during ensemble training when objectives conflict. This often results in training oscillations and hinders the approximation of complex, non-convex fronts.

In light of these challenges, this paper proposes a hierarchical two-layer optimization framework for joint trajectory and payload optimization in UAV-MEC systems. Specifically, UAV trajectory control and offloading decisions are formulated as the leader layer of a Stackelberg game, while the user association problem constitutes the follower layer, solved via a matching mechanism. To resolve multi-objective conflicts, we introduce a DIL mechanism. By decomposing the original problem into independent sub-tasks and assigning specialized expert agents to each, our approach eliminates gradient interference at the structural level and significantly improves the approximation of intricate Pareto fronts.

The primary contributions of this paper are summarized as follows:

1. We develop a two-tier optimization model based on the Stackelberg game. This framework effectively characterizes the intricate interplay between high-level UAV trajectory planning and low-level user association, ensuring a stable and coordinated system response.
2. We propose the DIL framework, which introduces a decomposition strategy into multi-objective reinforcement learning. By isolating conflicting gradients at the algorithmic level, the proposed method significantly improves the convergence and diversity of the learned Pareto solutions.
3. Comprehensive simulations are conducted to evaluate the proposed framework. The results demonstrate that our approach consistently outperforms NSGA-II and conventional heuristic baselines across key metrics, including IGD, HV, and multi-objective balancing capability.

2 Related Work

The joint optimization of trajectory planning and task offloading in UAV-MEC has emerged as a pivotal research frontier in recent years. This section provides a comprehensive review of existing literature across four key dimensions: UAV communication optimization, the evolution of offloading paradigms, traditional optimization methodologies, and DRL approaches.

Owing to their superior mobility and agile deployment capabilities, UAVs have been extensively integrated into wireless communication networks as aerial base stations or relays. Initial studies predominantly focused on enhancing ground coverage

or maximizing system throughput by optimizing UAVs' 3D positioning and flight altitudes [19]. For instance, several works proposed fly-hover-communicate protocols, which achieve enhanced downlink multicast performance through the joint optimization of altitude and antenna beamwidth [20, 21]. However, these early investigations were largely restricted to pure communication scenarios, often overlooking the dynamic distribution and scheduling of computational resources within a 3D spatial-temporal context [22].

MEC effectively augments the processing capabilities of resource-constrained smart terminals by shifting computational tasks to the network edge [23]. Current task offloading paradigms are generally classified into binary offloading and partial offloading [24]. While conventional terrestrial MEC infrastructures significantly reduce latency and energy consumption, their fixed-location base stations struggle to provide flexible coverage in unpredictable or infrastructure-limited environments, such as mountainous regions, disaster zones, or tactical battlefields. To address these gaps, the concept of Flying MEC (F-MEC) has been introduced, which integrates computing servers directly onto UAV platforms to enable on-demand, mobile adaptation of computing services [25].

In multi-UAV collaborative F-MEC systems, the intricate coupling between user association, resource allocation, and trajectory control presents significant optimization challenges [26, 27]. Conventional research typically formulates these interactions as mixed-integer non-linear programming problems, which are solved via iterative techniques such as block coordinate descent , successive convex approximation , or branch-and-bound algorithms [28]. While these methods offer mathematical rigor and convergence guarantees, they often suffer from prohibitive computational overhead and slow convergence when deployed in high-dimensional, dynamic environments. Consequently, their applicability is often limited in scenarios requiring real-time decision-making and rapid response to environmental shifts.

To circumvent the limitations of traditional optimization and enhance adaptability to environmental stochasticity [29], DRL has been increasingly integrated into UAV trajectory control. For instance, Deep Q-Networks (DQN) and Deep Deterministic Policy Gradient have been successfully applied to navigate discrete and continuous action spaces, respectively [30]. Furthermore, the introduction of advanced mechanisms such as target networks, experience replay, and Prioritized Experience Replay has significantly improved the sample efficiency and training stability of these models [31]. However, most existing DRL-based solutions focus on single-objective optimization or simple weighted-sum scalarization, which struggle to handle the complex trade-offs inherent in multi-objective UAV-MEC systems.

Despite extensive research, UAV-MEC systems remain hindered by dynamic constraints and the friction between competing objectives. Current approaches often fall short in three key areas: balancing latency against energy consumption, accurately capturing non-convex Pareto fronts, and modeling the system's hierarchical structure across layers. Bridging these gaps requires a unified optimization framework—one capable of navigating hierarchical interactions while simultaneously neutralizing multi-objective gradient interference.

Therefore, there is a compelling need for a unified optimization framework that can simultaneously characterize the hierarchical interaction between system variables and mitigate multi-objective gradient interference.

3 System Model and Problem Reformulation

3.1 System Overview

This paper considers a multi-tier UAV-assisted edge computing system, as illustrated in Fig 1. The network architecture is organized into three functional layers:

1. The Ground Layer consists of A set $\mathcal{K} = \{1, 2, \ldots, K\}$ of mobile ground devices (MDs), which generate heterogeneous and latency-sensitive computing tasks.
2. The Aerial Edge Layer comprisesA set $\mathcal{N} = \{1, 2, \ldots, N\}$ of rotary-wing UAVs serving as mobile edge servers. These UAVs provide localized computation offloading services by dynamically adjusting their positions in 3D space to optimize service coverage and processing efficiency.
3. The Infrastructure Layer includes a terrestrial Base Station (BS) situated at a fixed coordinate. The BS is equipped with substantial computational resources and functions as a cloud access node to process tasks that exceed the localized capacities of the UAV swarm.

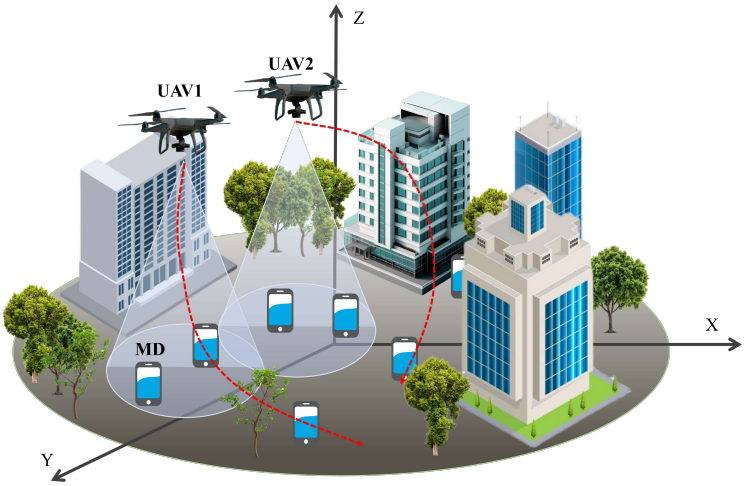


Fig. 1 The UAV-MEC system architecture.

The entire system is deployed within a 3D Cartesian coordinate system. The horizontal service area is defined by a rectangular region in the $X \times Y$ plane, while the UAVs operate within a regulated altitude range $[H_{\min}, H_{\max}]$. Due to the intrinsic characteristics of wireless signal propagation, the UAVs create dynamic coverage footprints on the ground. Specifically, an increase in flight altitude broadens the coverage range at the expense of potentially increased path loss and diminished signal strength.

3.2 Sub-model Definitions

This section breaks the system down into four submodels—drone movement, communication channels, mission queuing and computation, and energy consumption—and provides quantitative definitions for each. See Table 1 for detailed parameters.

3.2.1 UAV Mobility Model

We consider a 3D Cartesian coordinate system where the spatial coordinates of UAV $n \in \mathcal{N}$ at time slot t are denoted by the vector $\mathbf{q}_n[t] = [x_n[t], y_n[t], z_n[t]]^T \in \mathbb{R}^3$. The maneuverability of each UAV is governed by its mechanical constraints, with position updates decoupled into horizontal displacement and vertical motion (ascent/descent). Following the discrete-time kinematic principles [32], the trajectory evolution is formulated as:

$$x_n[t+1] = x_n[t] + d_n[t]\cos(2\pi\theta_n[t]) \tag{1}$$

$$y_n[t+1] = y_n[t] + d_n[t]\sin(2\pi\theta_n[t]) \tag{2}$$

$$z_n[t+1] = z_n[t] + \Delta h_n[t] \tag{3}$$

The decision variables governing the UAV's maneuver comprise the normalized horizontal azimuth $\theta_n[t] \in [0, 1]$, which represents the heading angle mapped to the physical range $[0, 2\pi]$, the horizontal displacement $d_n[t] \in [0, d_{\max}]$ constrained by the maximum traversable distance $d_{\max} = V_{xy} \cdot \Delta t$ per time slot, and the vertical displacement $\Delta h_n[t] \in [-h_{\text{step}}, h_{\text{step}}]$ dictated by the maximum climb/descent rate while ensuring the altitude remains within the operational bounds $z_n[t] \in [H_{\min}, H_{\max}]$.

3.2.2 Air-to-Ground Channel Model

The wireless link between a UAV and a ground node is primarily governed by the probability of Line-of-Sight propagation, which is highly sensitive to the UAV's elevation angle. Let $r_{n,k}[t]$ and $z_n[t]$ denote the horizontal distance and the vertical altitude between UAV n and target k at time slot t, respectively. The elevation angle $\psi_{n,k}[t]$ is defined as $\psi_{n,k}[t] = \arctan(z_n[t]/r_{n,k}[t])$. To characterize the average path loss $L_{n,k}[t]$ (in dB) while accounting for the urban environment's morphology, we adopt an elevation-dependent log-distance model [33]:

$$L_{n,k}[t] = A_0 + 10\gamma \log_{10}(d_{n,k}[t]) - B(\psi_{n,k}[t] - \psi_0)e^{-C\psi_{n,k}[t]} \tag{4}$$

where $d_{n,k}[t] = \sqrt{r_{n,k}^2[t] + z_n^2[t]}$ represents the 3D Euclidean distance. The parameters $\{A_0, B, C, \gamma, \psi_0\}$ are environment-specific constants typically calibrated via empirical measurements in urban or suburban scenarios. Consequently, the achievable transmission rate $R_{n,k}[t]$ for the link between UAV n and target k is formulated based on the Shannon-Hartley theorem:

$$R_{n,k}[t] = B_w \log_2\left(1 + \frac{P_{tx} \cdot 10^{-L_{n,k}[t]/10}}{\sigma^2}\right) \tag{5}$$

where B_w denotes the channel bandwidth, P_{tx} is the transmission power, and σ^2 represents the additive white Gaussian noise (AWGN) power. This model implicitly captures the trade-off between coverage expansion and signal attenuation as the UAV's spatial coordinates evolve.

3.2.3 Propulsion and Computing Energy Model

The total energy consumption of UAV n in slot t, denoted as $E_n^{\text{total}}[t]$, comprises propulsion, computation, and communication components:

$$E_n^{\text{total}}[t] = (P_n^{fly} + P_n^{com} + P_n^{exe})\Delta t \tag{6}$$

To determine the propulsion power (P_n^{fly}), we adopt the standard rotary-wing power model.[33]:

$$P_n^{fly} = P_0\left(1 + \frac{3v_n^2}{v_{tip}^2}\right) + P_i\left(\sqrt{1 + \frac{v_n^4}{4v_0^4}} - \frac{v_n^2}{2v_0^2}\right)^{1/2} + \frac{1}{2}d_0\rho sAv_n^3 \tag{7}$$

where v_n is the flight speed. Parameters $\{P_0, P_i, v_{tip}, v_0\}$ are constants related to the UAV's aerodynamic profile.

Regarding the energy consumption profile, the system accounts for both computing and communication power. Specifically, the execution power P_n^{exe} is dictated by the CMOS logic model, defined as:

$$P_n^{\text{exe}} = \kappa f_n^3 \tag{8}$$

where κ represents the effective capacitance. Parallel to this, the communication power P_n^{com} encompasses the energy demands of the underlying circuitry as well as the RF transmission overhead tied to the relay rate $R_n^{\text{relay}}[t]$.

Table 1 Hyperparameter Configuration for DRL Agent

Parameter Definition	Symbol	Value
Learning Rate	α	10^{-4}
Discount Factor	γ	0.99
Target Update Rate	τ	0.005
Replay Buffer Size	$\lvert\mathcal{D}\rvert$	10^5
Mini-batch Size	B	256
Exploration Noise	σ	0.1
Maximum Episodes	$E_{\max}$	5000
Steps per Episode	$T_{\max}$	200

3.3 Problem Formulation

We model the joint 3D trajectory planning, task offloading, and user association as a Constrained Multi-Objective Optimization Problem . Given the decision set $\chi[t] =$

$\{\mathbf{Q}[t], \rho[t], \mathcal{A}[t]\}$, we define three conflicting objectives to balance service quality and sustainability:

Minimizes the service bottleneck between relaying and local computation:

$$J_1 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{N} \max\left(D_n^{relay}, D_n^{local}\right) \tag{9}$$

Reduces cumulative propulsion, communication, and computation energy:

$$J_2 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{N} E_n^{total} \tag{10}$$

Maximizes total mission data ingestion:

$$J_3 = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{N} N_n^{in} \tag{11}$$

The optimization problem $\mathcal{P}_1$ aims to identify the optimal policy π^* to minimize the objective vector $\mathbf{F}(\chi) = [J_1, J_2, -J_3]$ subject to:

$$\begin{aligned}
\mathcal{P}_1 : \min_{\chi} \quad & \mathbf{F}(\chi) = [J_1(\chi), J_2(\chi), -J_3(\chi)] && \text{(12a)}\\
\text{s.t.} \quad & C1 : 0 \le d_n[t] \le d_{\max}, |\Delta h_n[t]| \le h_{step}, && \text{(12b)}\\
& C2 : \mathbf{q}_n[t] \in \mathcal{S}_{safe}, \forall n, t, && \text{(12c)}\\
& C3 : 0 \le \rho_n[t] \le 1, \forall n, t, && \text{(12d)}\\
& C4 : \mathcal{A}[t] \in \mathcal{A}_{feasible}(\mathbf{Q}[t]), \forall t, && \text{(12e)}\\
& C5 : \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[B_n[t]\right] < \infty, \forall n. && \text{(12f)}
\end{aligned}$$

Specifically, constraints $C1$ and $C2$ enforce UAV kinematic limits and spatial safety, $C3$ defines the continuous offloading space, $C4$ ensures binary association validity within the communication footprint, and $C5$ guarantees long-term queue stability via mean rate stability; furthermore, since aggressive throughput maximization typically compromises energy and latency, the intrinsic conflicts among objectives J_1, J_2, and J_3 render the formulated problem $\mathcal{P}_1$ a high-dimensional, non-convex NP-hard challenge.

4 Decomposition-Based Multi-Objective Joint Optimization Algorithm

4.1 Bi-level Hierarchical Interaction Framework

To address the inherent hierarchical dependencies and the intricate coupling between trajectory control, task offloading, and user association in UAV-MEC systems, we

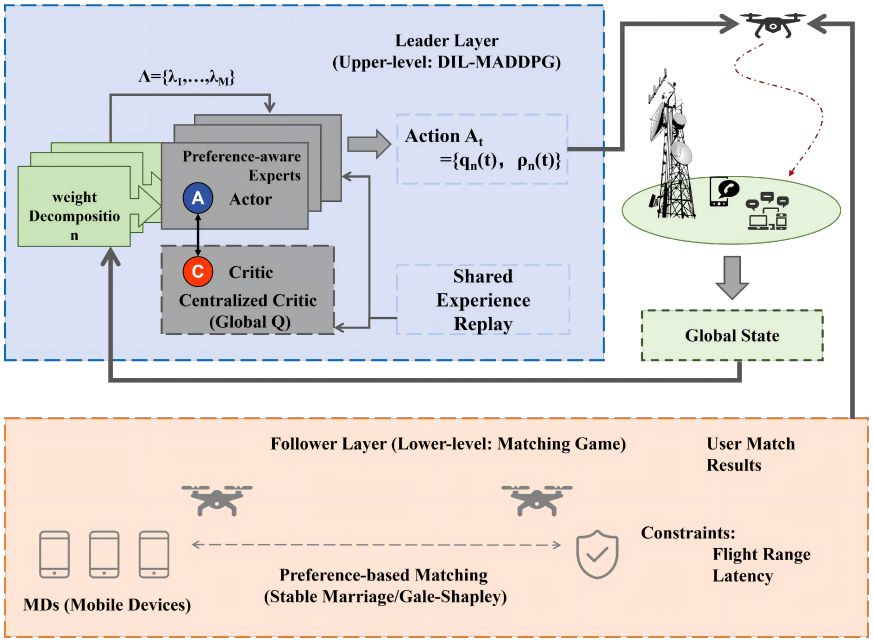


Fig. 2 The proposed hierarchical optimization framework based on Stackelberg game and DIL-MADDPG.

formulate the joint optimization problem as a Stackelberg leader-follower game, as illustrated in Fig. 2.

4.2 Leader Layer: Joint Decision-Making on 3D Trajectories and Task Offloading

4.2.1 Decomposition-Based Independent Learning Strategy

At the leader level of the hierarchical game, the UAV swarm collaboratively optimizes the long-term multi-objective system performance. In each time slot t, each UAV $n \in \mathcal{N}$ generates a continuous action vector $\mathbf{a}_n[t]$ based on its policy π_n and local observation $s_n[t]$. The action vector $\mathbf{a}_n[t]$ encompasses four key decision variables:

$$\mathbf{a}_n[t] = [\theta_n[t], \tilde{d}_n[t], \rho_n[t], \Delta\tilde{h}_n[t]] \tag{13}$$

where:$\theta_n[t] \in [0,1]$ is the normalized horizontal heading angle, mapped to the physical range $[0, 2\pi]$.$\tilde{d}_n[t] \in [0,1]$ represents the horizontal displacement ratio, where the actual movement is $d_n[t] = \tilde{d}_n[t] \cdot d_{\max}$. $\rho_n[t] \in [0,1]$ is the task offloading ratio, determining the proportion of traffic relayed to the BS.$\Delta h_n[t] \in [0,1]$ signifies the vertical altitude adjustment, which is scaled to the physical range $[-h_{\text{step}}, h_{\text{step}}]$.

To address the multi-objective nature of the problem, we employ a decomposition-based strategy. For a specific preference or sub-task defined by the weight vector $\lambda_i = [\lambda_{1,i}, \lambda_{2,i}, \lambda_{3,i}]$, where $\sum \lambda = 1$, the instantaneous utility function U_{leader} for the

Algorithm 1 Decomposition-based Independent Learning (DIL-MADDPG) Framework

Require: Weight vectors $\Lambda = \{\lambda_1, \ldots, \lambda_M\}$, Max episodes E_{max}, Max steps T_{max}, System parameters.
Ensure: Set of optimized policies $\Pi^* = \{\pi^*_{\lambda_1}, \ldots, \pi^*_{\lambda_M}\}$.

for each preference weight vector $\lambda \in \Lambda$ **do**
  Initialize N Actor networks μ_n and Critic networks Q_n with random weights θ^μ, θ^Q
  Initialize Target networks μ'_n, Q'_n and Replay Buffer $\mathcal{D}$
  for episode $e = 1$ to E_{max} **do**
    Initialize environment and observe initial state $S_1 = \{s^1_1, \ldots, s^1_N\}$
    for step $t = 1$ to T_{max} **do**
      // Leader Level: UAV Decision
      For each UAV n, select action $a^t_n = \mu_n(s^t_n|\theta^\mu_n) + \mathcal{N}_t$
      Update UAV positions $P(t+1)$ based on action set $A^t = \{a^t_1, \ldots, a^t_N\}$
      // Follower Level: User Association
      Obtain optimal user association matrix $\mathcal{A}^*(t)$ using **Algorithm 2**
      // Environment Feedback
      Calculate System Objectives: $D_{sys}, E_{sys}, N_{sys}$
      Compute scalarized reward: $r_t = \lambda \cdot [D_{sys}, E_{sys}, N_{sys}]^T$
      Observe next state S_{t+1} and done flag
      Store transition (S_t, A^t, r_t, S_{t+1}) into $\mathcal{D}$
      // MADDPG Model Update
      if buffer $\mathcal{D}$ is sufficient **then**
        Sample random minibatch of transitions from $\mathcal{D}$
        Update Critics by minimizing Loss $L(\theta^Q)$
        Update Actors by maximizing Policy Gradient $\nabla_{\theta^\mu} J$
        Soft update target networks: $\theta' \leftarrow \tau\theta + (1-\tau)\theta'$
      end if
      $S_1 \leftarrow S_{t+1}$
    end for
  end for
  Save trained policy π^*_λ into Π^*
end for
return Π^*

leader layer is formulated as:

$$U_{leader}(\mathbf{a}[t], \mathcal{A}^*[t]) = \lambda_{1,i}\hat{J}_1(\mathbf{a}[t], \mathcal{A}^*[t]) + \lambda_{2,i}\hat{J}_2(\mathbf{a}[t], \mathcal{A}^*[t]) + \lambda_{3,i}\hat{J}_3(\mathbf{a}[t], \mathcal{A}^*[t]) \quad (14)$$

Here, $\mathbf{a}[t] = \{\mathbf{a}_n[t]\}_{n\in\mathcal{N}}$ denotes the joint action set, and $\mathcal{A}^*[t]$ represents the optimal user association matrix fed back from the follower layer. The terms $\hat{J}_1, \hat{J}_2, \hat{J}_3$ denote the normalized objective functions for latency, energy consumption, and throughput, respectively. This decomposition-based utility guides the UAV swarm toward a Pareto-optimal policy π^* that respects the kinematic constraints $C_1 - C_3$. The detailed

implementation of this multi-objective independent learning process is summarized in Algorithm 1.

4.2.2 Centralized Training and Distributed Execution (CTDE) Mechanism

To mitigate the non-stationarity and policy interdependency inherent in multi-agent environments, the proposed framework adopts a CTDE architecture. This mechanism allows agents to leverage global information during the training phase while maintaining decentralized decision-making during execution.

For Centralized Critic Optimization, each agent $n \in \mathcal{N}$ is equipped with a centralized critic network Q_n optimized using the joint information of the global state $\mathbf{S} = \{s_1, \ldots, s_N\}$ and the joint action $\mathbf{A} = \{a_1, \ldots, a_N\}$, wherein the temporal difference target y_j for the j-th sample in a minibatch of size B is defined as:

$$y_j = r_{n,j} + \gamma(1 - d_j)Q'_n(\mathbf{S}_{j+1}, \mathbf{A}'_{j+1}|\theta_n^{Q'}) \tag{15}$$

where $\mathbf{A}'_{j+1} = \{\mu'_m(s_{m,j+1})|m \in \mathcal{N}\}$ denotes the joint action set estimated by the target actor networks for the subsequent state, and γ is the discount factor. The critic network is updated by minimizing the mean squared error (MSE) loss:

$$L(\theta_n^Q) = \frac{1}{B}\sum_{j=1}^{B}\left(y_j - Q_n(\mathbf{S}_j, \mathbf{A}_j|\theta_n^Q)\right)^2 \tag{16}$$

For Distributed Actor Optimization, the actor network μ_n generates actions based solely on local observations s_n during the execution phase to facilitate decentralized control, with its policy parameters updated according to the deterministic policy gradient theorem to maximize the centralized critic's evaluation of the current policy:

$$\nabla_{\theta_n^\mu} J \approx \frac{1}{B}\sum_{j=1}^{B} \nabla_{a_n} Q_n(\mathbf{S}_j, \mathbf{A}|\theta_n^Q)\big|_{a_n=\mu_n(s_{n,j})} \cdot \nabla_{\theta_n^\mu}\mu_n(s_{n,j}|\theta_n^\mu) \tag{17}$$

where $\mathbf{A} = \{a_1, \ldots, a_n, \ldots, a_N\}$ incorporates the actions of other agents as predicted by their respective actors.

Regarding stability and exploration, we implement Target Networks with a soft update mechanism to ensure convergence and training stability:

$$\theta_n^{Q'} \leftarrow \tau\theta_n^Q + (1-\tau)\theta_n^{Q'}, \quad \theta_n^{\mu'} \leftarrow \tau\theta_n^\mu + (1-\tau)\theta_n^{\mu'} \tag{18}$$

where $\tau \ll 1$ is the smoothing coefficient. Furthermore, to enhance exploration within the continuous action space, time-decaying Gaussian noise $\mathcal{N}(0, \sigma^2)$ is added to the actor's output during the training phase.

4.2.3 Architecture of Expert Agents

For each decomposed sub-problem i, a set of dedicated MADDPG-based expert agents is deployed. Each UAV $n \in \mathcal{N}$ is equipped with two primary networks: a policy network (Actor) $\mu(\cdot|\theta_n^\mu)$ and a value network (Critic) $Q(\cdot|\theta_n^Q)$.

Regarding state space design, to mitigate the non-stationarity inherent in multi-agent environments and enhance spatial-temporal awareness, the state $\mathbf{s}_n[t]$ encapsulates self-information, mission environment perception, and teammate coordination features:

$$\mathbf{s}_n[t] = \left[\mathbf{q}_n, B_n, \{\Delta\mathbf{q}_{n,k}\}_{k=1}^K, \mathbf{v}_{global}, \{\Delta\mathbf{q}_{n,m}\}_{m\neq n}\right] \tag{19}$$

Specifically, the self-state comprises the UAV's normalized 3D coordinates $\mathbf{q}_n$ and current computational payload B_n; the mission perception includes the relative coordinates $\{\Delta\mathbf{q}_{n,k}\}$ to ground devices alongside the geometric centroid $\mathbf{v}_{global}$ of the K most critical task points to help the UAV identify task-dense hotspots and optimize long-term coverage; and teammate awareness utilizes the relative positions $\{\Delta\mathbf{q}_{n,m}\}$ of neighboring UAVs to facilitate collaborative obstacle avoidance and prevent redundant coverage overlap.

Regarding the action space design, the Actor network outputs a 4D continuous action vector $\mathbf{a}_n[t]$, which is first squashed into the $[-1, 1]$ interval via a tanh activation function to ensure numerical stability during training and subsequently mapped to actual physical ranges through a linear transformation:

$$\mathbf{a}_n[t] = [\theta_n, \tilde{d}_n, \rho_n, \Delta\tilde{h}_n] \tag{20}$$

These components correspond to the horizontal heading angle, the horizontal displacement ratio, the task offloading ratio, and the vertical altitude adjustment, respectively. This mapping ensures that all generated maneuvers strictly adhere to the kinematic constraints defined in Section 3.2.1.

4.3 Follower Layer: Responsive User Association

4.3.1 User Association Mechanism Based on Matching Games

To address the short-term association problem at the follower layer, we use a matching game approach. Assuming co-channel interference within the UAV-MEC system is mitigated through an orthogonal frequency division scheme, the matching game has no externalities; a ground device's preference is thus independent of the association decisions of other devices. This property ensures the Deferred Acceptance algorithm consistently converges to a stable matching, which serves as a refinement of the Nash equilibrium in bilateral markets.

In this hierarchical game, the UAVs' spatial coordinates $\mathbf{q}[t+1]$ are directly dictated by the leader's action $\mathbf{a}[t]$. Upon perceiving the updated network topology, the MDs select their optimal service nodes. Let the mapping operator $\mathcal{M}(\cdot)$ represent the matching outcome. The optimal association matrix $\mathcal{A}^*[t]$ is formulated as a function of the UAVs' maneuver:

$$\mathcal{A}^*[t] = \mathcal{M}(\mathbf{q}[t+1]) = \mathcal{M}(f_{motion}(\mathbf{q}[t], \mathbf{a}[t])) \tag{21}$$

Algorithm 2 Preference-based User Association Matching Game

Require: Current UAV positions $\mathbf{Q}(t)$, User set $\mathcal{K}$, UAV set $\mathcal{N}$.
Ensure: User Association Matrix $\mathcal{A}^*(t)$.

Initialize unassigned users set $\mathcal{K}_{un} \leftarrow \mathcal{K}$
Initialize current matching $\mathcal{M} \leftarrow \emptyset$
Construct preference lists for Users $\succ_k$ and UAVs $\succ_n$ based on channel quality and load
while $\mathcal{K}_{un} \neq \emptyset$ **do**
    for each user $k \in \mathcal{K}_{un}$ **do**
        User k proposes to its most preferred UAV n in $\succ_k$ that hasn't rejected it yet
    end for
    for each UAV $n \in \mathcal{N}$ **do**
        Let $\mathcal{R}_n$ be the set of new proposals received by UAV n
        Let $\mathcal{C}_n = \mathcal{M}_n \cup \mathcal{R}_n$ be the candidate set
        Sort $\mathcal{C}_n$ according to UAV preference $\succ_n$
        if $|\mathcal{C}_n| > C_n$ (Capacity quota of UAV n) **then**
            Keep the top-C_n users in $\mathcal{C}_n$ to update $\mathcal{M}_n$
            Reject remaining users and add them back to $\mathcal{K}_{un}$
        else
            Accept all users in $\mathcal{C}_n$ to update $\mathcal{M}_n$
        end if
    end for
end while
Construct binary matrix $\mathcal{A}^*(t)$ based on stable matching $\mathcal{M}$
return $\mathcal{A}^*(t)$

To maximize preference satisfaction under resource constraints that limit each UAV to a single high-bandwidth link per time slot to ensure QoS, the interaction is modeled as a bilateral one-to-one matching game detailed in Algorithm 2, wherein a ground device k establishes an association $a_{n,k} = 1$ with UAV n strictly when the device resides within the UAV's effective communication footprint and ranks it highest in its current feasible preference set.

4.3.2 Modeling User Preferences

To quantify the desirability of available service nodes, each ground device $k \in \mathcal{K}$ establishes a preference list based on a comprehensive evaluation of communication quality and potential service latency. We define a composite cost function $\mathcal{C}_{k,n}[t]$ to evaluate the overhead associated with user k selecting UAV n as its service provider:

$$\mathcal{C}_{k,n}[t] = \omega_1 B_n[t] + \omega_2 d_{n,BS}[t] + \omega_3 d_{n,k}[t] \tag{22}$$

where $B_n[t]$ represents the current task backlog or queue length of UAV n. A higher $B_n[t]$ implies a more congested buffer and longer waiting times for new missions. $d_{n,BS}[t]$ denotes the Euclidean distance between UAV n and the base station. This term serves as a surrogate for the relay transmission rate, as larger distances typically lead to lower backhaul throughput. $d_{n,k}[t]$ represents the 3D distance between user k and UAV n, which directly dictates the path loss and the attainable uplink transmission rate for the access link.

The coefficients $\omega = [\omega_1, \omega_2, \omega_3]$ are normalization weights that balance the trade-off between local congestion, backhaul efficiency, and access link quality. Based on this cost function, the preference relationship $\succ_k$ for user k over the set of UAVs is defined as:

$$n \succ_k n' \Leftrightarrow \mathcal{C}_{k,n}[t] < \mathcal{C}_{k,n'}[t] \tag{23}$$

This preference structure ensures that users prioritize UAVs that offer a superior balance of proximity and processing capacity, thereby facilitating an efficient matching that aligns with the system's global optimization goals.

4.4 Follower Layer: Responsive User Association

Table 2 compares the computational overhead of the evaluated algorithms. DIL-MADDPG achieves superior real-time efficiency by shifting the primary training burden, $\mathcal{O}(MNI_{train})$, to the offline phase. Consequently, the online execution complexity is reduced to only $\mathcal{O}(NL)$ for neural network inference and $\mathcal{O}(KN)$ for user matching (N: UAVs, M: expert agents, I_{train}: iterations, N_p: network parameter scale, K: devices).

Table 2 Algorithm Complexity Comparison

Algorithm	**Training Complexity**	**Decision-making Complexity**	**Real-time Assessment**
DIL-MADDPG (Ours)	$\mathcal{O}(M \cdot N \cdot I_{train})$	$\mathcal{O}(N \cdot N_p + K \cdot N)$	High
NSGA-II	N/A	$\mathcal{O}(G \cdot P^2)$	Low
Greedy	N/A	$\mathcal{O}(K \cdot N)$	Moderate
APF	N/A	$\mathcal{O}(N)$	High

Furthermore, since the inference complexity is effectively decoupled from the environmental scale (i.e., the number of mission points), the proposed method exhibits robust computational performance even in large-scale user scenarios. Here, N denotes the number of UAVs, K represents the count of ground terminal devices, N_p is the parameter scale of the deep neural network, M signifies the number of decomposed expert agents, P is the population size of NSGA-II, G is the number of evolutionary generations, and I_{train} represents the total reinforcement learning training iterations.

5 Simulation Results and Analysis

5.1 Experimental Configuration and Baselines

We evaluate the DIL-MADDPG framework in the UAV-MEC scenario defined in Section 3.1. The simulation is conducted in a $1000 \times 1000 \times 100$ m^3 3D airspace with $N = 3$ UAVs and $K = 100$ ground devices randomly distributed. The BS is located at $[0, 0, H_{BS}]$. All physical and network parameters, including hardware constraints and channel models, are summarized in Table 3.

Table 3 Simulation Parameters

Parameter Definition	Symbol	Value
Area Size	$L \times W$	400×400 m^2
Number of UAVs	N	3
Number of MDs	K	100
Flight Altitude	H	$[30, 100]$ m
Maximum Velocity	$V_{\max}$	15 m/s
Channel Bandwidth	W	10 MHz
Noise Power	σ^2	10^{-9} W
Task Data Size	$s_{k,j}$	5 MB
Task Computation Cycles	$c_{k,j}$	10^9 cycles
CPU Frequency	f_n	1 GHz

To demonstrate the effectiveness of the DIL-MADDPG approach, we compare it against the following four baseline strategies:

1. **NSGA-II:** A classical multi-objective optimization algorithm that identifies Pareto-optimal solutions through evolutionary search. We employ Simulated Binary Crossover and Polynomial Mutation for the offline global optimization of UAV trajectories and offloading decisions.
2. **Greedy Strategy:** A heuristic approach where each UAV identifies the nearest unvisited task point based on its current coordinates. The UAV navigates toward the target at maximum velocity while maintaining a constant altitude until the mission is completed or boundary constraints are met.
3. **Artificial Potential Field (APF):** A force-based navigation method where unvisited mission points exert attractive forces on the UAV, while the mission boundaries and neighboring UAVs generate repulsive forces. The UAV moves in the direction of the resultant force, with its velocity magnitude proportional to the force's intensity.
4. **2D-MADDPG :** To validate the benefits of 3D spatial optimization, this baseline constrains the UAVs to a fixed operational altitude. The UAVs perform path planning exclusively in the 2D horizontal plane, utilizing a network architecture and reward function identical to the proposed 3D framework.

5.2 Performance Evaluation

To validate the DIL-MADDPG framework within the complex UAV-MEC landscape, we perform a multi-dimensional quantitative analysis. This evaluation specifically examines its effectiveness in mitigating gradient interference and its convergence toward the Pareto front.

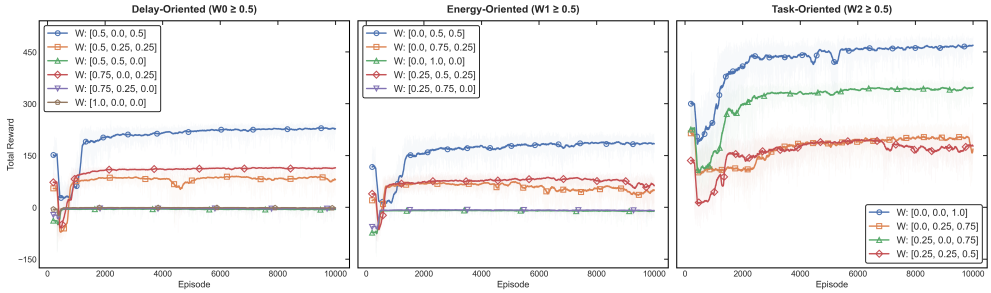


Fig. 3 The UAV-MEC system architecture.

Fig 3 illustrates the convergence profiles of DIL-MADDPG across 16 heterogeneous weight vectors λ_i. Despite intrinsic conflicts between energy saving and throughput maximization, all expert agents achieve stable convergence within 4,000 episodes. This confirms that the DIL mechanism structurally eliminates gradient interference through task decomposition, ensuring sub-strategy robustness.

The results reflect a dynamic reward shift aligned with preference evolution: in throughput-prioritized scenarios, rewards surge above 400 as agents adopt a coverage-expansion strategy. By increasing flight altitude to broaden the communication footprint, the UAVs maximize mission ingestion at the cost of marginal propulsion increases, demonstrating the framework's ability to exploit 3D spatial degrees of freedom.

To quantitatively evaluate the quality of the solution set, HV and IGD are adopted as performance metrics, with the results illustrated in Fig 4.

In terms of convergence, the proposed method achieves an IGD of 0.159, significantly outperforming NSGA-II with 0.294, the Greedy algorithm with 0.776, and APF with 0.823 to yield non-dominated solutions closer to the global optimum, whereas the slightly lower IGD of 0.154 from the 2D-Baseline is merely a result of its restricted, less complex two-dimensional search space.

In terms of diversity, where a higher HV value indicates a broader coverage of the solution set, the proposed method achieves an HV of 6.455 to surpass NSGA-II with 6.300 and the Greedy algorithm with 5.152, demonstrating its ability to generate a more uniformly distributed and comprehensive set of non-dominated solutions while ensuring high convergence accuracy.

DIL-MADDPG exhibits high policy flexibility by accurately capturing Pareto front shifts to adapt to latency-priority scenarios with a delay of 9.84 s and task-priority scenarios completing 92.19 tasks. Regarding energy efficiency, DIL-MADDPG consumes

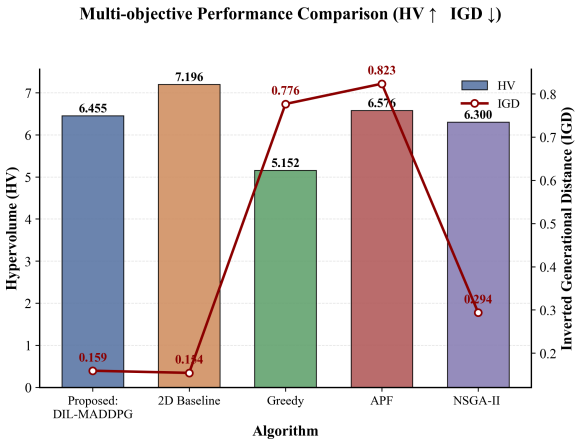


Fig. 4 Analysis of Diversity Metrics for Different Algorithms

Table 4 Performance comparison of various algorithms under different preference weight vectors

Weight \ Algorithm	Proposed (MADDPG)	2D-MADDPG	Greedy	APF
$[\mathbf{1.0}, \mathbf{0.0}, \mathbf{0.0}]$	9.84 / 56k / 45.5	5.71 / 16.0k / 25.1	82.9 / 476k / 63.4	27.2 / 481k / 20.5
$[\mathbf{0.0}, \mathbf{1.0}, \mathbf{0.0}]$	26.8 / 19.6k / 39.8	15.0 / 17.4k / 27.8	79.5 / 477k / 61.4	28.2 / 481k / 21.0
$[\mathbf{0.0}, \mathbf{0.0}, \mathbf{1.0}]$	101.2 / 440k / 92.2	159.9 / 432k / 84.7	81.4 / 476k / 62.2	28.0 / 481k / 20.7
$[\mathbf{0.25}, \mathbf{0.25}, \mathbf{0.5}]$	96.8 / 206k / 84.6	102.2 / 294k / 80.7	82.5 / 475k / 63.2	27.6 / 481k / 20.6

Note: Each cell contains three metrics in the format: *Total Delay (s) / Total Energy (kJ) / Collected Tasks*. Weights are defined as $\lambda = [\lambda_{delay}, \lambda_{energy}, \lambda_{task}]$.

only 19.6×10^3 J, representing just 4.1% of the energy used by the Greedy algorithm, which it outperforms by optimizing long-term hovering and relay paths; furthermore, in terms of Pareto superiority under equilibrium preferences, DIL-MADDPG dominates the 2D-MADDPG that completes 84.55 tasks with an energy consumption of 206.6×10^3 J, proving that 3D trajectory optimization significantly expands the system's service boundary.

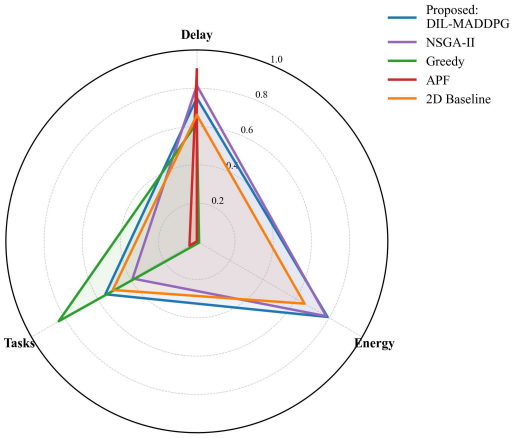


Fig. 5 Analysis of Radar Charts Showing the Overall Performance of Different Algorithms

Fig 5 compares the algorithms across three normalized objectives. The proposed method demonstrates the best comprehensive performance with the largest coverage area, indicating superior multi-objective coordination. In contrast, the Greedy algorithm prioritizes short-term performance at the cost of energy efficiency. Both NSGA-II and APF are outperformed by the proposed method, exhibiting either a performance gap or an insufficient balance across all metrics.

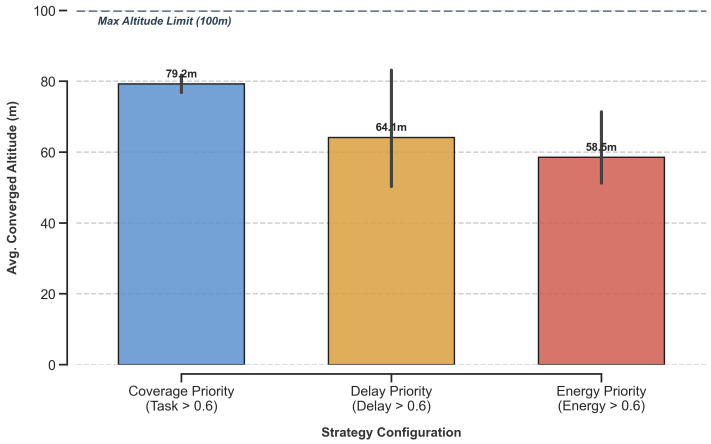


Fig. 6 Convergence Altitude of UAVs Under Different Preference Weights

Converged altitudes demonstrate the physical interpretability of DIL-MADDPG, with agents occupying distinct strata, such as a high altitude of $\bar{H} = 79.2$ m for coverage, a moderate altitude of $\bar{H} = 64.1$ m for LoS-path-loss balance, and a low

altitude of $\bar{H} = 58.5$ m for energy conservation. Such "physical intelligence" confirms the framework internalizes environmental dynamics rather than performing mere black-box reward optimization, ensuring context-aware, optimal maneuvering.

6 Conclusions

This paper develops a hierarchical collaborative framework for 3D UAV-MEC systems, integrating Stackelberg game theory with DIL-MADDPG to resolve high-dimensional variable coupling and multi-objective conflicts. By decoupling the joint optimization into a low-complexity matching game and scalarized sub-tasks, the framework structurally eliminates the gradient interference bottleneck in multi-objective reinforcement learning, ensuring a high-fidelity approximation of the Pareto-optimal front. Simulation results substantiate that the proposed method significantly outperforms NSGA-II and conventional baselines in both convergence accuracy (IGD) and solution set diversity (HV). Furthermore, the learned flight strategies show strong alignment with underlying physical dynamics, confirming the framework's interpretability. This study provides a rigorous theoretical foundation for resource scheduling in the Low-altitude Economy and offers a scalable reference for 6G-enabled aerial networks. Future research will focus on interference coordination for ultra-large-scale UAV swarms and adaptive environmental generalization via meta-learning.

Declarations

Availability of data and materials

All data supporting the findings of this study are included in this article.

Competing interests

The authors declare that they have no competing interests.

Funding

This research is supported by the Putian Science and Technology Bureau (Grant No. 2024GZ3001PTXY19), the Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University) (Grant No. MJUKF-IPIC202406), and the Putian University High-Level Talent Start-up Fund (Grant No. XN2026036).

Authors' contributions

Conceptualization, W.W. and Y.L.; Formal analysis, W.W., Y.L. and Y.H.; Funding acquisition, Y.L., W.W., Y.H. and C.S.; Investigation, W.W., Y.L., Y.H. and C.S.; Methodology, Y.L., W.W., Y.H. and C.S.; Project administration, Y.L., W.W., Y.H. and C.S.; Resources, W.W., Y.H. and C.S.; Software, W.W., Y.L. and C.S.; Supervision, Y.L., Y.H. and Z.F.; Validation, W.W. and Z.F.; Visualization, W.W. and Y.H.; Writing—original draft, W.W. and Y.L.; Writing—review & editing, W.W., Y.L.,

Z.F., Y.H. and C.S. All authors have read and agreed to the published version of the manuscript.

References

[1] Zeng, Y., Chen, S., Cui, Y., Yang, J., Fu, Y.: Joint resource allocation and trajectory optimization in uav-enabled wirelessly powered mec for large area **10**, 15705–15722 (2023)

[2] Lin, N., Tang, H., Zhao, L., Wan, S., Hawbani, A., Guizani, M.: A pddqnlp algorithm for energy efficient computation offloading in uav-assisted mec **22**, 8876–8890 (2023)

[3] Chen, J., Cao, X., Yang, P., Xiao, M., Ren, S., Zhao, Z., Wu, D.O.: Deep reinforcement learning based resource allocation in multi-uav-aided mec networks **71**, 296–309 (2023)

[4] Li, B., Yang, R., Liu, L., Wang, J., Zhang, N., Dong, M.: Robust computation offloading and trajectory optimization for multi-uav-assisted mec: A multi-agent drl approach **11**, 4775–4786 (2024)

[5] Song, F., Deng, M., Xing, H., Liu, Y., Ye, F., Xiao, Z.: Energy-efficient trajectory optimization with wireless charging in uav-assisted mec based on multi-objective reinforcement learning **23**, 10867–10884 (2024)

[6] Gao, Z., Fu, J., Jing, Z., Dai, Y., Yang, L.: Moipc-maac: Communication-assisted multi-objective marl for trajectory planning and task offloading in multi-uav assisted mec **11**, 18483–18502 (2024)

[7] Wu, P., Xiao, F., Sha, C., Huang, H.: Service-oriented segmented trajectory design for low-altitude uav-assisted mec networks **25**, 2027–2044 (2026)

[8] Pan, H., Liu, Y., Sun, G., Wang, P.: Resource scheduling for uavs-aided d2d networks: A multi-objective optimization approach **23**, 4691–4708 (2024)

[9] Chen, Y., Wang, B., Yin, H., Yang, S.: Multi-objective optimization for multi-uav-assisted mec systems **PP**, 1–1 (2026)

[10] He, Y., Gan, Y., Cui, H., Guizani, M.: Fairness-based 3-d multi-uav trajectory optimization in multi-uav-assisted mec system. **10**, 11383–11395 (2023)

[11] Wang, C., Zhai, D., Zhang, R., Cai, L., Liu, L., Dong, M.: Joint association, trajectory, offloading, and resource optimization in air and ground cooperative mec systems **73**, 13076–13089 (2024)

[12] Jiang, L., Li, Y.: Trajectory, computation offloading, and transmission optimization strategy based on age of information in uav-assisted mec system **PP**, 1–16

(2025)

[13] Qi, H., Wu, M., Zhao, M.: Trajectory optimization for multi-uav-enabled wireless-powered mec with graph-based madrl (2025)

[14] Felten, F., Talbi, E.-G., Danoy, G.: Multi-objective reinforcement learning based on decomposition: A taxonomy and framework. **79**, 679–723 (2024)

[15] Ju, T., Li, L., Liu, S., Zhang, Y.: A multi-uav assisted task offloading and path optimization for mobile edge computing via multi-agent deep reinforcement learning. Journal of Network and Computer Applications **229**, 103919 (2024)

[16] Zhang, L., Tian, C., Liu, T., Li, X., Mumtaz, S., Khan, W.U.: Multi-objective decomposition evolutionary drl for uav-assisted mec in internet of vehicles. IEEE Transactions on Vehicular Technology (2025)

[17] Li, T., Meng, Y., Tang, L.: A decomposition optimization-based multiobjective reinforcement learning algorithm for obtaining nonconvex pareto fronts. IEEE Transactions on Neural Networks and Learning Systems **37**(1), 219–232 (2026)

[18] Zheng, L., He, J., Zhu, X., Zhang, Y., Shen, Y., Taleb, T.: Rsma-enabled multi-uav secure communication via marl with multi-task attention drnn. IEEE Transactions on Information Forensics and Security **20**, 12624–12639 (2025)

[19] Yu, X., Zhang, X., Rui, Y., Dang, X., Jia, G., Guizani, M.: Joint resource allocation and 3d-position optimization for uav-assisted mec network with noma **12**, 1440–1456 (2025)

[20] Wang, M., Li, R., Jing, F., Gao, M.: Multi-uav assisted air–ground collaborative mec system: Drl-based joint task offloading and resource allocation and 3d uav trajectory optimization **8** (2024)

[21] Zheng, Y., Li, A., Wen, Y., Wang, G.: A uav trajectory optimization and task offloading strategy based on hybrid metaheuristic algorithm in mobile edge computing **17** (2025)

[22] Zeng, X., Yan, G., Tan, H., Du, J., Deng, X., Wu, H.: Uav trajectory planning for computational offloading and resource allocation optimizing in uav-assisted mec with unavailable base station. IEEE Internet of Things Journal, 1–1 (2026)

[23] Li, J., He, Q., Wang, X., Hawbani, A., Yu, K., Bi, Y., Zhao, L.: Uav-assisted microservice mobile edge computing architecture: Addressing post-disaster emergency medical rescue **74**, 2635–2648 (2025)

[24] Jia, M., Zhang, L., Wu, J., Guo, Q., Zhang, G., Gu, X.: Deep multiagent reinforcement learning for task offloading and resource allocation in satellite edge computing. IEEE Internet of Things Journal **12**(4), 3832–3845 (2025)

[25] Al-Bakhrani, A.A., Li, M., Obaidat, M.S., Amran, G.A.: Moalf-uav-mec: Adaptive multi-objective optimization for uav-assisted mobile edge computing in dynamic iot environments. IEEE Internet of Things Journal (2025)

[26] Chen, Y., Yang, Y., Wu, Y., Huang, J., Zhao, L.: Joint trajectory optimization and resource allocation in uav-mec systems: A lyapunov-assisted drl approach. IEEE Transactions on Services Computing (2025)

[27] Zakaryia, S.A., Meaad, M., Nabil, T., Hussein, M.K.: Task offloading and resource allocation for multi-uav asset edge computing with multi-agent deep reinforcement learning. Computing **107**(5), 126 (2025)

[28] Liu, Q., Zhou, X., Yang, H., Zhang, J., Chen, J.: Deep reinforcement learning-based cross-layer energy efficiency optimization in uav-mec systems. Mobile Networks and Applications, 1–14 (2026)

[29] Shukla, P., Shukla, S., Kumar Singh, A.: Trajectory-prediction techniques for unmanned aerial vehicles (uavs): A comprehensive survey. IEEE Communications Surveys & Tutorials **27**(3), 1867–1910 (2025)

[30] Wang, C., Liu, K., Yuan, Y., Peng, S., Li, G.: Joint trajectory and offloading optimization in uav-assisted mec via federated multi-agent reinforcement learning and potential fields. Computer Networks **272**, 111681 (2025)

[31] Wu, Q., Liu, Q., He, Y., Wu, Z.: Uav-enabled energy-efficient aerial computing: A federated deep reinforcement learning approach. IEEE Transactions on Reliability **74**(3), 3559–3572 (2025)

[32] Song, F., Xing, H., Wang, X., Dai, P., Xiao, Z., Zhao, B.: Evolutionary multi-objective reinforcement learning based trajectory control and task offloading in uav-assisted mobile edge computing **22**, 7387–7405 (2023)

[33] Yu, Tang, J., Huang, J., Zhang, X., So, D.K.C., Wong, K.-K.: Multi-objective optimization for uav-assisted wireless powered iot networks based on extended ddpg algorithm. IEEE Transactions on Communications **69**(9), 6361–6374 (2021)