



DOI:10.12404/j.issn.1671-1815.2408552

引用格式: 翟若雪, 林鹏, 程方, 等. MEC 辅助的无人机系统中基于多智能体强化学习的分布式任务卸载策略[J]. 科学技术与工程, 2025, 25(20): 8543-8551.

Zhai Ruoxue, Lin Peng, Cheng Fang, et al. Distributed task offloading for MEC-assisted UAVs using multi-agent reinforcement learning[J]. Science Technology and Engineering, 2025, 25(20): 8543-8551.

电子技术、通信技术

MEC 辅助的无人机系统中基于多智能体强化学习的分布式任务卸载策略

翟若雪, 林鹏*, 程方, 纪阳, 张治中

(南京信息工程大学电子与信息工程学院, 南京 210044)

摘要 无人机(unmanned aerial vehicle, UAV)系统具有部署灵活、视距传播等优势,成为辅助移动通信应对高密度数据处理、应急通信等业务的重要工具。然而,复杂环境下 UAV 的计算处理能力及续航问题是其长期面临的技术瓶颈。移动边缘计算(mobile edge computing, MEC)技术的发展为解决 UAV 计算与能耗问题提供了有效方法。针对 MEC 辅助的 UAV 系统,提出一种基于多智能体强化学习算法的分布式任务卸载策略。将 UAV 的任务卸载和资源分配过程建模为多个 MEC 节点马尔可夫博弈过程(Markov game process, MGP)。为了解决 MGP 问题,提出一种面向多智能体协作的分布式强化学习算法。该算法能够使智能体基于局部观测信息,通过在线协作学习方式寻找最优策略。在对比实验中,评估了所提方案的收敛性和系统性能。结果显示该方案在收敛速度、能耗和卸载率方面均优于对比方案。

关键词 无人机(UAV); 能量消耗; 移动边缘计算; 任务卸载; 强化学习

中图分类号 TN929.5;

文献标志码 A

Distributed Task Offloading for MEC-assisted UAVs Using Multi-agent Reinforcement Learning

ZHAI Ruo-xue, LIN Peng*, CHENG Fang, JI Yang, ZHANG Zhi-zhong

(School of Electronics and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China)

[Abstract] The unmanned aerial vehicle (UAV) system, with its advantages of flexible deployment and line-of-sight propagation, has become an essential tool for assisting mobile communications in handling high-density data processing and emergency communications. However, the computational processing capabilities and endurance issues of UAVs under complex environments remain significant technological bottlenecks. The development of mobile edge computing (MEC) technology provides an effective solution to address UAVs' computational and energy consumption challenges. A distributed task offloading strategy based on a multi-agent reinforcement learning algorithm was proposed for MEC-assisted UAV systems. The task offloading and resource allocation process of UAVs was modelled as a Markov game process (MGP) involving multiple MEC nodes. To solve the MGP problem, a distributed reinforcement learning algorithm for multi-agent collaboration was proposed. The algorithm enabled agents to find the optimal strategies through online collaborative learning based on local observation information. In comparative experiments, the convergence and system performance of the proposed scheme were evaluated. The results show that the proposed scheme outperforms the comparison schemes in terms of convergence speed, energy consumption, and unloading rate.

[Keywords] unmanned aerial vehicle(UAV); energy consumption; mobile edge computing; task offloading; reinforcement learning

无人机(unmanned aerial vehicle, UAV)凭借其灵活性和易于部署等特点,近年来在交通监控、目

标识别、应急通信与导航等领域广泛应用^[1-3]。无人机的服务性能取决于其计算能力和服务响应时

收稿日期: 2024-11-16; 修订日期: 2025-04-28

基金项目: 国家自然科学基金(6220011089)

第一作者: 翟若雪(1996—),女,汉族,安徽亳州人,硕士研究生。研究方向:无人机边缘计算和智能优化算法。E-mail:1986573175@qq.com。

*通信作者: 林鹏(1991—),男,汉族,山东临沂人,博士,副教授。研究方向:边缘计算与资源管理技术、数字/网络孪生技术、强化学习算法、边缘智能网络。E-mail:linpeng@nuist.edu.cn。

投稿网址:www.stae.com.cn

间。轻量化计算及有限的电池容量仍然是无人机系统面临的瓶颈。无人机的电池容量低且计算能力有限,长期执行计算密集型任务会缩短续航时间,进而影响无人机运行性能。移动边缘计算(mobile edge computing, MEC)技术^[4-5]将无人机计算密集型任务卸载到具备更强计算能力的边缘节点上,从而缩短计算时延并降低无人机能耗,有效解决上述问题。

随着无人机和 MEC 技术的融合发展^[6-8], MEC 辅助的无人机计算服务系统研究成为热点,相关领域得到广泛关注。文献[9]使用最优传输理论和经典粒子群优化算法,通过优 UAV 部署和任务卸载决策改善用户的任务服务延迟。文献[10]设计一种预先调度的 UAV 辅助车辆边缘计算系统,并利用 DQN 算法实现 UAV 部署决策,从而确定 UAV 最佳悬停位置。文献[11]建立一个 UAV-MEC 系统,通过综合考虑任务卸载量、UAV 功率和飞行轨迹,提出一种交替迭代优化算法,以确保用户体验的基础上最小化系统能耗。文献[12]针对移动边缘卸载场景,提出缓存辅助的动态任务卸载与多维资源分配联合优化策略,以时延和能耗为优化目标,设计一种深度强化学习的求解算法。文献[13]构建一种基于异构 UAV 的协同计算网络新框架,通过联合优化用户关联、CPU 循环频率和发射功率等方式,最大化网络中计算任务的服务量。文献[14]研究了 MEC 系统中计算服务缓存问题,其利用 UAV 提供的计算和缓存资源进行协作,最小化所有终端设备间的任务完成时延。文献[15]针对物联网设备在偏远地区面临的时延优化问题,提出了一种多无人辅助的任务卸载模型,并制定了混合整数非线性规划问题来优化通信和计算资源分配,以减少执行延迟。文献[16]提出了一种基于 Dinkelbach 方法和块坐标下降技术的双环结构迭代算法,旨在优化重构智能表面辅助无人机支持的 MEC 系统中的比特分配、发射功率、相移和飞行轨迹,从而实现能源效率的最大化。

上述研究为 MEC 辅助的 UAV 系统提供了丰富的解决方案,一定程度上解决了相关问题。然而,在面对多 MEC 节点服务的多无人机系统时,相关研究仍面临一些挑战。一是现有的研究主要通过基于完备的网络环境信息优化用户终端和无人机之间的任务卸载决策、无人机部署以及系统总能耗等,而忽视了无人机与 MEC 节点的分布式特性,从而引发了优化成本问题;二是传统的集中式优化或中心式强化学习算法依赖于网络全局参数的收集,忽略了节点之间协作和协调的必要性。为解决上

述问题,提出一种具有负载感知能力的自适应 UAV 任务卸载与资源分配方案。在该方案中,通过多个 MEC 节点协同优化卸载决策、传输功率分配和计算资源分配策略,在满足任务服务时延要求的条件下,降低 UAV 能耗。将任务卸载与资源分配问题建模 MEC 节点之间的马尔可夫博弈过程(Markov game process, MGP)。为解决分布式多 MEC 节点的马尔科夫博弈问题,设计一种多智能体强化学习(multi-agent DDPG, MADDPG)算法,通过中心化训练,使 MEC 节点能够利用局部观测并采用自适应协作机制实现线学习并做出最佳策略。在仿真实验过程中,评估 MADDPG 算法收敛性,并将任务卸载与资源分配方案与现有方案在能耗、任务卸载率和计算资源占用率等方面进行比较分析,仿真实验结果表明所提方案具有明显优势。

1 系统建模和问题表述

1.1 系统模型

如图 1 所示,构建一个系统模型,涵盖 N 架无人机 $\forall n \in \mathcal{N} = \{1, 2, \dots, N\}$ 与 M 个 MEC 节点 $\forall m \in \mathcal{M} = \{1, 2, \dots, M\}$ 。该系统的核心目标是为分布在不同位置的无人机提供任务卸载计算与通信服务。每个 MEC 节点均被视作一个小型边缘服务器,具备为无人机提供计算任务卸载服务的能力。邻近的 MEC 通过有线链路连接实现信息交互。MEC 节点与 UAV 间则借助无线通信进行数据传输。无人机实时处理相关计算任务,且对响应时间具有较高的敏感度。为提升无人机任务处理的效率与精确度,无人机可以将计算任务卸载到 MEC 节点上进行处理。MEC 节点执行计算任务,并将结果回传至无人机,从而有效地降低无人机自身能耗、延长续航及服务时间。

在考虑的环境中,MEC 节点能够根据无人机所执行的任务规模及其计算能力,灵活且动态地分配

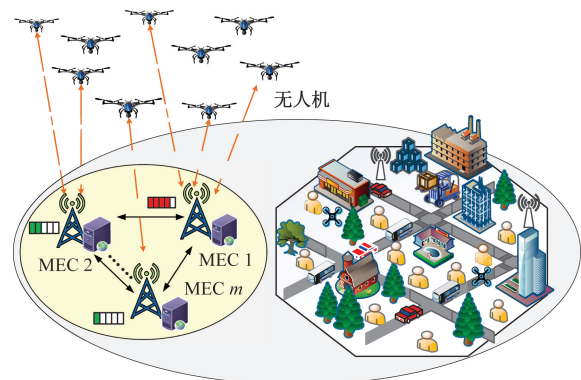


图 1 MEC 辅助的无人机系统
Fig. 1 MEC-supported UAV system

相应的计算资源。在业务高峰时段,若某一 MEC 节点的计算任务负载过重,将增加无人机计算任务的处理时延,同时也会导致其他 MEC 节点资源闲置,对无人机用户体验产生严重的负面影响。为避免上述问题,MEC 节点可通过多节点间的协同运作,综合考量网络状态、节点剩余资源状态以及任务时延需求,确定最优策略,以进行无人机任务卸载和资源分配,从而满足多无人机用户场景对于高计算能力和通信需求。

1.2 网络模型

考虑无人机持续性的产生计算任务。为更好地描述 MEC 辅助的无人机系统,将系统时间划分为离散时隙 $t \in \{1, 2, \dots, T\}$ 。鉴于所采集数据内容的差异性,每个时隙内所有无人机生成的计算任务集合可以表示为 $A = \{A_1, A_2, \dots, A_n, \dots, A_N\}$, 其中, $A_n = \{D_n, C_n, T_n^{\max}\}$ 为 UAV n 生成的计算任务; D_n 为执行任务所需输入数据的字节数, bits; C_n 为完成任务所需的计算资源, CPU cycles; T_n^{\max} 为任务最大时延限制。

在 t 时隙,每架无人机针对其产生的任务,向附近的 MEC 节点发送任务请求。随后,MEC 节点根据自身的资源状况和无人机的请求数量决定是否为该无人机提供计算和通信服务。定义 $x_{mn}^t = \{0, 1\}$ 代表 MEC 节点 m 在 t 时隙对 UAV n 的任务卸载决策,若 $x_{mn}^t = 1$, 则表明任务由无人机 n 卸载至 MEC 节点 m 执行,若 $x_{mn}^t = 0$, 则任务在无人机本地执行。

假定每个无人机任务仅能卸载到一个 MEC 节点进行处理,可得

$$\sum_{m=1}^M x_{mn}^t = 1, \quad \forall n, t \quad (1)$$

当多个无人机卸载到同一 MEC 节点时,由于其共享同一通信信道,彼此间会产生通信干扰。因此,当 UAV n 在 MEC 节点 m 上执行任务卸载时,根据香农公式,在 t 时隙,UAV n 的上行传输速率可表示为

$$r_{mn}^t = \frac{B_{\text{up}}}{\sum_{n \in N} x_{mn}^t} \log_2 \left(1 + \frac{h_{mn}^t p_{mn}^t}{\eta_{mn}^t + \sigma^2} \right) \quad (2)$$

式(2)中: η_{mn}^t 为 UAV n 在 t 时隙受到来自其他无人机的通信干扰, $\eta_{mn}^t = \sum_{n' \neq n} x_{mn'}^t p_{mn'}^t h_{mn'}^t$, 其中, $x_{mn'}^t$ 为 m 在 t 时隙对无人机 n' 的卸载决策; $p_{mn'}^t$ 为无人机 n' 的发射功率; $h_{mn'}^t$ 为无人机 n' 与 m 之间的信道增益; B_{up} 为上行链路可用频谱带宽; p_{mn}^t 为无人机 n 的发射功率; σ^2 为接收位置的加性高斯白噪声,且服从复数正态分布 $\sigma \sim \text{CN}(0, 1)$; h_{mn}^t 为 UAV n 与 MEC

节点 m 之间的信道增益。

在 t 时隙,定义下行传输速率为

$$r_{mn}^{\text{dw}} = \frac{B_{\text{dw}}}{\sum_{n \in N} x_{mn}^t} \log_2 (1 + \psi_{mn}^{\text{dw}}) \quad (3)$$

式(3)中: B_{dw} 为下行带宽; ψ_{mn}^{dw} 为无人机下行链路信噪比。

1.3 计算模型

在 MEC 辅助的无人机应用场景中,对于轻量级计算任务,各无人机可选择在本地执行;而对于产生的计算密集型任务,则可通过提交任务请求的方式将其卸载至 MEC 节点进行远程计算。考虑任务处理过程中的时延及能量消耗两方面因素,具体细节如下。

1.3.1 服务时延模型

对于轻量级计算任务,无人机 UAV n 会选择利用其自身配备的本地计算单元进行处理。因此,任务数据无需通过无线传输,从而避免传输时延。

任务 A_n 在 UAV n 的计算时延可表示为

$$T_n^{\text{loc}} = \frac{C_n}{f_n} \quad (4)$$

式(4)中: f_n 为 UAV n 的计算能力, CPU cycles/s。

当 UAV n 产生计算任务并将其卸载至 MEC 节点 m 时,服务延迟由两部分构成。一部分为无线链路的传输时延,具体涵盖任务输入数据的上传时延以及任务输出结果的下载时延。

上传时延可表示为

$$T_{mn}^{\text{up}} = \frac{D_n}{r_{mn}^t} \quad (5)$$

当 MEC 节点 m 处理完任务后,下载任务输出结果的时延可表示为

$$T_{mn}^{\text{dw}} = \frac{\beta D_n}{r_{mn}^{\text{dw}}} \quad (6)$$

式(6)中: β 为任务处理后的数据大小与输入数据大小的比率。

另一部分是 UAV n 在 MEC 节点 m 处进行任务计算时产生的计算时延,可表示为

$$T_{mn}^{\text{com}} = \frac{C_n}{F_{mn}^t} \quad (7)$$

式(7)中: F_{mn}^t 为在 t 时隙 MEC 节点 m 分配给无人机 n 的计算资源, CPU cycles/s。

1.3.2 能量消耗模型

在无人机系统中,能量消耗是影响任务执行效率和系统持续运行时间的重要因素。

任务 A_n 在 UAV n 本地执行的能耗可表示为

$$E_n^{\text{loc}} = k f_n^2 C_n \quad (8)$$

式(8)中: k 为能量转换效率,该效率取决于芯片

架构。

当任务被卸载到 MEC 节点 m 执行, UAV n 计算能耗将得以消除。然而,任务卸载过程中仍会产生额外的能量消耗,具体包括将任务上传至 MEC 节点 m 的传输能耗,可表示为

$$E_{mn}^{\text{up}} = \frac{p_{mn}^t D_n}{r_{mn}^t} \quad (9)$$

以及下载计算结果所产生的能耗,其计算公式为

$$E_{mn}^{\text{dw}} = \frac{\beta p_{mn}^t D_n}{r_{mn}^{\text{dw}}} \quad (10)$$

2 联合资源与任务卸载

2.1 部分可观察马尔可夫博弈

在考量无人机与地面 MEC 节点间的动态交互及其分布式特性的基础上,采用多智能体强化学习算法,旨在确保移动边缘计算系统高效处理无人机任务卸载及资源分配,并促进 MEC 节点间的协作。在多智能体算法中,每个 MEC 节点都被视为一个智能体,能根据当前系统状态和历史经验,自主决策以优化关键性能指标。具体而言,在满足无人机任务时延要求的前提下,MEC 节点需要最小化 UAV 系统能耗。为实现这一目标,智能体需要在本地观测信息的基础上,通过与周围环境的交互和其他 MEC 节点的协作,学习到最优的任务卸载决策、上行传输功率策略和资源分配策略。为进一步提高 MEC 辅助的无人机系统性能,采用基于分布式多智能体协作的 MADDPG 卸载算法。MADDPG 算法将任务卸载和资源分配过程建模为 MEC 节点之间的马尔可夫博弈过程(Markov game process, MGP),该 MGP 涉及 4 个关键要素:观测、动作、状态和奖励。

(1) 观测。MEC 辅助的无人机系统是一个动态系统。为优化资源分配并提高系统性能,将连续信道增益 h_{mn}^t 划分为 K 个等间隔的离散级别,即 $L_v: L_v^{v-1} \leq h_{mn}^t \leq L_v^v$, 其中, $v = 1, 2, \dots, K$; L_v 为划分出的第 v 个离散级别,满足 $L_1 = 0$ 和 $L_K = \infty$; L_v^v 为第 v 个离散级别 L_v 的边界值。在任务卸载过程中, MEC 节点 m 根据其获取的信息自主决策。MEC 节点 m 的观测值可表示为 $o_m(t) = \{h_{m1}^t, h_{m2}^t, \dots, h_{mN}^t\}$, 其中 h_{mn}^t 的取值范围为 $\{L_1, L_2, \dots, L_K\}$ 。

(2) 动作。智能体的动作包括任务卸载决策 x_{mn}^t 、上行传输功率策略 p_{mn}^t 和资源分配策略 F_{mn}^t 。根据 MGP,在 t 时隙 MEC 节点 m 的系统动作可表示为

$$a_m(t) = \{x_{mn}^t, p_{mn}^t, F_{mn}^t\} \quad (11)$$

(3) 状态。在 MEC 辅助的无人机系统中,系统

状态是所有移动边缘计算服务器局部观测的集合,可表示为

$$s(t) = \{o_1(t), o_2(t), \dots, o_M(t)\} \quad (12)$$

2.2 MGP 奖励模型

系统奖励是评价无人机在任务卸载和资源分配决策过程中能量消耗的重要指标,总能耗越低系统奖励越高。无人机产生的计算任务具有不同的时延要求,但每个卸载任务都应在时延限制内完成。任务服务时延限制可表示为

$$\max \{T_n^{\text{loc}}, T_{mn}^{\text{up}} + T_{mn}^{\text{com}} + T_{mn}^{\text{dw}}\} < T_n^{\text{max}}, \quad \forall n \in N \quad (13)$$

在 t 时隙,MEC 节点 m 根据局部观测 $o_m(t)$ 采取动作 $a_m(t)$ 的系统奖励可表示为

$$r_m(t) = \frac{1}{\sum_{n=1}^N E_n^t(\mathbf{x}, \mathbf{p}, \mathbf{F})} \quad (14)$$

式(14)中: $\mathbf{x} = \{x_{mn}^t\}$ 为任务卸载决策; $\mathbf{p} = \{p_{mn}^t\}$ 为上行传输功率策略, $\mathbf{F} = \{F_{mn}^t\}$ 为计算资源分配策略; $E_n^t(\mathbf{x}, \mathbf{p}, \mathbf{F})$ 为 UAV n 处理计算任务 A_n 的能耗成本,可表示为

$$E_n^t(\mathbf{x}, \mathbf{p}, \mathbf{F}) = \sum_{m=1}^M [(1 - x_{mn}^t) E_n^{\text{loc}} + x_{mn}^t (E_{mn}^{\text{up}} + E_{mn}^{\text{dw}})] \quad (15)$$

为促进 MEC 节点间的信息共享和协作,最大程度地避免对抗行为,确保每个智能体都获得相同的当前奖励,即 $r(t) = r_m(t)$ 。由于 MEC 辅助无人机系统的动态性和复杂性,为了让智能体做出最优决策,MEC 节点智能体的目标是最大化预期累积奖励

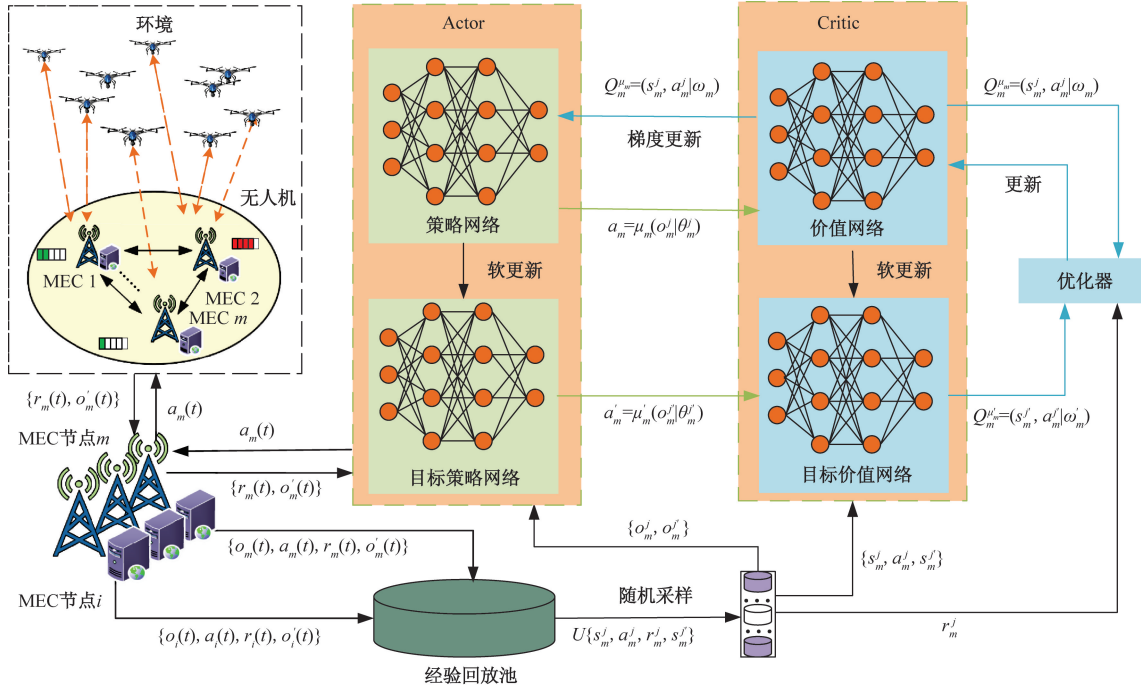
$R_m(t) = E \left[\sum_{t=1}^T \gamma^t r(t) \right]$, 其中 $\gamma \in (0, 1)$ 为折扣因子,用于平衡当前奖励和未来奖励的重要性。智能体能够随着策略优化不断改善行为,以最大化累积奖励,从而提高任务卸载效率和可靠性。

2.3 MADDPG 框架

MADDPG 是一种结合 DDPG 与多智能体的策略梯度算法,通过中心化训练与去中心化执行的方式,实现在复杂动态环境下的多智能体策略优化。针对无人机与 MEC 节点的分布式特性, MADDPG 算法相较于 DDPG 算法,允许智能体之间交换观测信息,从而增强 MEC 节点间的协作能力。因此,系统能够自适应地调整策略,实现整个无人机系统能耗最小化。首先阐述针对单一 MEC 智能体的 DDPG 算法,随后详细介绍适用于多 MEC 节点的 MADDPG 算法。

2.3.1 单 MEC 节点 m 的 DDPG 算法

如图 2 右侧所示,DDPG 框架中的 MEC 节点智



$o_i(t)$ 、 $a_i(t)$ 、 $r_i(t)$ 和 $o'_i(t)$ 分别为在 t 时隙智能体 i 的观测状态、动作、奖励和下一个观测状态; s_m^j 为第 j 个全局信息样本中所有节点智能体的状态信息; a_m^j 为在第 j 个样本中 MEC 节点 m 在观察状态 o_m^j 下执行的动作; r_m^j 为 MEC 节点 m 在第 j 个样本中, 执行 a_m^j 后获得的即时奖励; s_m^j 为在第 j 个样本中, 所有 MEC 节点智能体执行动作后转移到下一个状态; $\mu'_m(o_m^j | \theta'_m)$ 为 Actor 执行的目标策略网络; $Q_m^{\mu'_m}(s_m^j, a_m^j | \omega'_m)$ 为 Critic 执行的目标价值网络; ω_m 为价值网络参数; ω'_m 为第 m 个智能体的目标价值网络参数; $U\{s_m^j, a_m^j, r_m^j, o_m^j\}$ 为随机抽取 U 个随机样本 $\{s_m^j, a_m^j, r_m^j, o_m^j\}$

图 2 基于 MADDPG 的分布式卸载决策算法结构

Fig. 2 Structure diagram of distributed offloading decision algorithm based on MADDPG

能体 m 是由两个独立模块 Actor 网络和 Critic 网络组成。Actor 网络基于当前环境观测 $o_m(t)$, 通过具有网络参数 θ_m 的策略网络 $\mu_m(o_m(t) | \theta_m)$ 输出一个确定性动作 $a_m(t)$ 。而 Critic 网络则将环境的局部观测(即状态)和动作作为输入, 利用动作价值函数 $Q_m(s_m, a_m)$ 评估动作的优劣, 从而为 Actor 网络提供反馈, 协助其选择最优策略 π_m^* 。

动作价值函数为

$$Q_m(s_m, a_m) = E_{\mu_m} \left[\sum_{t=1}^T \gamma^t r(t) \mid s_m = o_m(t), a_m = a_m(t) \right] \quad (16)$$

根据贝尔曼方程, 动作价值函数的递归关系可表示为

$$Q_m(s_m, a_m) = E_{\mu_m} [r(s_m, a_m) + \gamma Q_m(s'_m, a'_m)] \quad (17)$$

式(17)中: s'_m 为智能体 m 在执行动作 $a_m(t)$ 后, 环境转移到的下一个状态 $o'_m(t)$, 即 $s'_m = o'_m(t)$; a'_m 为 t 时隙在下一个状态 s'_m 下, 由策略网络输出的动作 $a'_m(t)$, 即 $a'_m = a'_m(t)$ 。

为保持算法稳定性并减少过拟合风险, Actor 网

络与 Critic 网络都具有一个与自身结构相同的目标网络, 即带参数 θ'_m 的目标策略网络和带参数 ω'_m 的目标价值网络。同时, DDPG 算法引入了经验回放机制, 允许 MEC 节点智能体 m 在与环境交互时将收集到的经验 $\{o_m(t), a_m(t), r_m(t), o'_m(t)\}$ 存储至经验回放池。通过上述方式, 智能体能够学习过往经验, 进而提升决策的稳定性。

作为一种确定性策略梯度算法, DDPG 算法的核心思想在于通过 Actor 网络生成最优决策策略, 并通过 Critic 网络评估当前策略的价值, 从而调整策略以最大化累计期望奖励 $R_m(t)$ 。为寻找最优策略, 该算法通过对策略网络 $\mu_m(o_m(t) | \theta_m)$ 中的参数 θ_m 以及价值网络 $Q_m(s_m, a_m | \omega_m)$ 中的参数 ω_m 展开迭代训练, 从而实现持续优化。具体而言, 从经验回放池中随机抽取一批样本, 并将其输入到智能体 m 中。Actor 和 Critic 网络在训练过程中基于输入的样本数据实时更新 θ_m 和 ω_m 参数。以第 j 样本数据 $\{s_m^j, a_m^j, r_m^j, s'_m^j \mid s'_m^j = o'_m(t), s'_m^j = o'_m(t)\}$ 为例, MEC 节点智能体 m 中的 Critic 网络先计算 TD 误差, 然后更新价值网络参数 ω_m 。Critic 价值网络的均方误差损失函数可表示为

$$L(\omega_m) = \frac{1}{U} \sum_{j=1}^U [Q_m^{\mu_m}(s_m^j, a_m^j | \omega_m) - y_m^j]^2 \quad (18)$$

式(18)中: $y_m^j = r_m^j + \gamma Q_m^{\mu_m}(s_m^j, a_m^j | \omega_m)$ 为智能体 m 的预测动作价值; U 为从经验回放池中抽取的样本量。预测动作价值与当前动作价值 $Q_m^{\mu_m}(s_m^j, a_m^j | \omega_m)$ 之间的差值为 TD 误差。然后 Critic 价值网络参数更新公式可表示为

$$\omega_m = \omega_m - \delta \nabla_{\omega_m} L(\omega_m) \quad (19)$$

式(19)中: δ 为 Critic 价值网络的学习率。

DDPG 算法通过最大化动作价值,对 Actor 网络参数 θ_m 进行梯度更新,从而优化智能体的策略。

具体来说,DDPG 算法利用策略梯度方法,通过估计每个动作在给定状态下的期望回报,来指导策略的改进。在确定性策略的框架下,智能体 m 的策略梯度函数可表示为

$$\begin{aligned} \nabla_{\theta_m} J_m(\theta_m) &\approx E_{\mu'} [\nabla_{\theta_m} Q_m^{\mu_m}(s_m^j, a_m^j | \omega_m)] \\ &= E [\nabla_{\theta_m} \mu_m[o_m(t)] \nabla Q_m^{\mu_m}(s_m^j, a_m^j | \omega_m)] \end{aligned} \quad (20)$$

然后,采用梯度下降法更新策略网络参数 θ_m , 其表达式为

$$\theta_m = \theta_m - \alpha \nabla [-J_m(\theta_m)] \quad (21)$$

式(21)中: α 为 Actor 策略网络中的学习率。随着 θ_m 和 ω_m 的实时更新,目标网络参数 θ'_m 和 ω'_m 可以进行软更新,可表示为

$$\theta'_m = (1 - \tau_\theta) \theta'_m + \tau_\theta \theta_m \quad (22)$$

$$\omega'_m = (1 - \tau_\omega) \omega'_m + \tau_\omega \omega_m \quad (23)$$

式中: τ_θ 和 τ_ω 为目标网络的超参数,通过超参数控制软更新进程,缓慢更新目标网络参数,以增加训练的稳定性。

2.3.2 多 MEC 节点的 MADDPG 算法

如图 2 所示, MADDPG 是由 M 个 MEC 节点组成的多智能体 DDPG 协作框架,其核心优势在于中心化训练能够采用自适应协作机制实现多 MEC 节点间的信息共享,而在执行过程的去中心化则是每个 MEC 节点利用自身的局部信息来提升智能体的决策效率。例如,在执行过程中,MEC 节点 m 在 t 时隙根据局部观测 $o_m(t)$, 通过 DDPG 算法来选择动作 $a_m(t) = \mu_m[o_m(t)]$ 。随后,MEC 节点 m 的状态 $o_m(t)$ 转移到下一个状态 $o'_m(t)$, 并根据式(14)计算系统奖励 $r_m(t)$ 。为了最大化所有智能体的奖励值,在中心化训练阶段充分利用 MEC 节点的全局信息。具体而言,通过节点间的无线传输,MEC 节点 m 在收集到其他 MEC 的局部信息后,将观测到的全部信息 $\{o_i(t), a_i(t), r_i(t), o'_i(t), \forall i \in M\}$ 存储在经验回放池中。然后,从经验回放池中随机抽取 U 个全局信息样本 $\{s_m^j, a_m^j, r_m^j, s'_m^j | j \in U\}$ 进行训

练,以更新 Actor 和 Critic 网络参数。 $s_m^j = \{o_1^j, o_2^j, \dots, o_m^j, \dots, o_M^j | o_m^j = o_m^j(t)\}$ 和 $s'_m^j = \{o_1^j, o_2^j, \dots, o_m^j, \dots, o_M^j | o_m^j = o_m^j(t)\}$ 为全部 MEC 节点的智能体信息状态。由于抽取的样本数据为离线数据,因此整个训练过程也是离线过程。这种离散训练方法有效解决了在马尔可夫博弈过程中多智能体决策所面临的动态环境挑战。基于 MADDPG 算法的决策过程在算法 1 中列出。

算法 1 基于 MADDPG 的系统总能耗最小化

1. 初始化:网络参数 θ_m 、 ω_m 、 θ'_m 和 ω'_m ;
2. 设置最大迭代次数 E 和经验回放池大小;
3. for $e = 1 \rightarrow E$ do
4. 获取所有代理的初始观察值,并设置初始奖励值 $r' = 0$;
5. for $t = 1 \rightarrow T$ do
6. 对于智能体 m , 根据当前策略 $\mu_m[o_m(t) | \theta_m]$ 选择动作 $a_m(t)$;
7. 执行动作 $a_m(t)$, 然后得到奖励 $r_m(t)$ 和新观测值 $o'_m(t)$;
8. MEC 节点间无线传输,将所有智能体全局信息储存 $\{s_m, a_m, r_m, s'_m\}$ 在经验回放池;
9. 从经验回放池中随机选择一小批量样本 $\{s_m^j, a_m^j, r_m^j, s'_m^j | j \in U\}$;
10. 中心化训练阶段,根据式(18)计算损失函数 $L(\omega_m)$, Critic 价值网络的参数更新公式为

$$\omega_m = \omega_m - \delta \nabla_{\omega_m} L(\omega_m);$$
11. 根据式(20)计算 Actor 网络的策略梯度 $\nabla_{\theta_m} J_m(\theta_m)$, 然后更新参数 θ_m , 可表示为

$$\theta_m = \theta_m - \alpha \nabla [-J_m(\theta_m)];$$
12. 根据式(22)和式(23)更新目标 Actor 网络 θ'_m 和目标 Critic 网络 ω'_m 。

3 性能评估

3.1 仿真设计

本实验软件为 Python 3.7 和 Minicond 3。实验的参数设置如下:在 MEC 辅助的无人机系统中,无人机数量取值范围为 5 ~ 20, 其中 3 个 MEC 节点随机分布在特定区域范围内。无线信道增益表示为 $h_{mn}^t = \frac{\beta_0}{d_{mn}^{\varphi(t)}}$ 其中, β_0 为参考距离等于 1 的信道增益; $d_{mn}^{\varphi(t)}$ 为在 t 时隙无人机 n 与其关联的 MEC 节点 m 间的欧式距离; $\varphi(t)$ 为路径损失函数,即 $\varphi(t) = 2 + 0.5 \sin\left(\frac{2\pi}{T}\right)$ 。上行链路带宽 B_{up} 和下行链路带宽 B_{dw} 分别为 200 MHz 和 50 MHz; 每个任务的计算需求 C_n 为 300 M CPU cycles; 并且系统噪声功率 σ^2 为 -100 dBm。本地计算时,任务 D_n 的数据量设置为 55 000 bits; 无人机 n 的计算能力为 $f_n = 100$ M CPU cycle/s, $\forall n \in N$; 每个

无人机任务完成时间上限为 $T_n^{\max} = 3$ s; MEC 节点的最大处理能力为 $F_m^{\max} = 20$ GHz。对于 MADDPG 算法,折扣因子 γ 设置为 0.95; Actor 和 Critic 网络学习率分别为 10^{-2} 和 10^{-3} 。

为评估 MADDPG 方案,与 DDPG 和策略梯度等对比算法进行比较,并配置以下方案进行比较。

(1) 基于近距离计算的优化卸载 (optimized offloading in proximity-based computing, OOPBC)。在此方案中,无人机将计算任务分配给附近 MEC 节点处理。然而,卸载决策主要取决于邻近度,无法全面评估 MEC 节点是否具有足够计算能力。

(2) 具有离散决策的随机资源分配 (stochastic resource allocation with discrete decision-making, SRADD)。在该方案中,MEC 节点采用自主卸载策略将计算资源随机分配至无人机。

(3) 无卸载策略的机载优化 (onboard optimization without offloading strategy, OOWOS)。无人机在其机载设备上独立执行和处理任务,无需任务卸载。因此,OOWOS 在比较方案中需建立最大能耗基准。

3.2 仿真结果

所提 MADDPG 算法与对比算法性能比较如图 3 所示,可以看出,3 种算法奖励值均随着迭代次数增加而增加,并最终收敛。MADDPG 只需约 40 次迭代即可快速收敛达到最优状态,而 DDPG 和策略梯度算法分别需要 140 次和 210 次迭代。DDPG 和策略梯度算法虽然取得了不错的效果,但在收敛速度和奖励方面仍然存在不足。这主要是因为 DDPG 和策略梯度算法中 MEC 节点仅使用本地信息进行动作选择。与 DDPG 相比,策略梯度算法可避免估计值函数计算,直接优化策略,表现出更高的样本效率。相比较而言,MADDPG 通过中心训练使所有 MEC 节点能获得其余 MEC 节点策略,促进 MEC 节点间的紧密合作和全局信息共享,从而提高算法整体性能,并实现更快的收敛速度和更高的奖励(即更低的无人机能耗)。例如,在大规模灾后紧急通信场景中,MADDPG 算法通过优化全局共享策略加速收敛,有效处理无人机网络中的通信任务,能够在最短的时间内为灾区提供稳定的通信服务,并快速恢复灾后网络运行。MADDPG 算法在动态且复杂的环境中展现显著优势,能够确保紧急响应的及时性并延长无人机的运行时间。

图 4 为不同无人机数量对系统总能耗的影响。所提方案在各种无人机数量配置下能耗均最低。与 SRADD 方案相比,本文方案优化卸载决策、传输功率策略和计算资源分配策略,同时实现多个 MEC

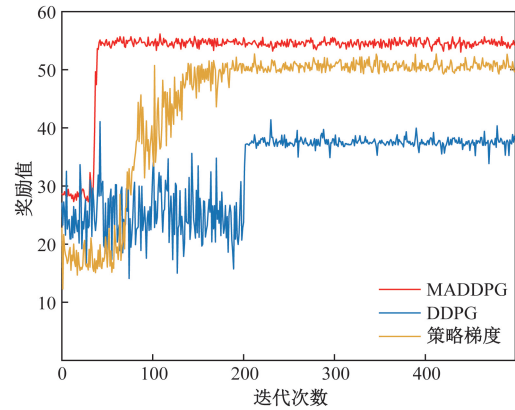


图 3 不同算法下智能体的平均奖励

Fig. 3 Average reward of agents under different algorithms

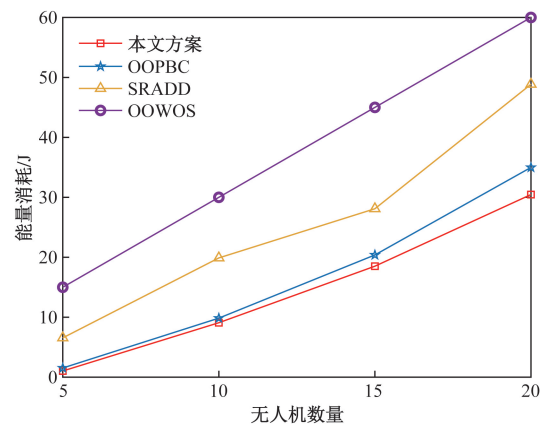


图 4 不同无人机数量下的系统能量消耗

Fig. 4 Energy consumption at different numbers of UAVs

节点间的协作,有效降低能耗。在 OOPBC 方案中,邻近卸载导致过多任务卸载至 MEC 节点,造成计算资源的冗余使用,进而引起较高的能耗。与 OOWOS 方案相比,基于自适应协作机制的边缘计算方案能显著提高无人机处理任务的整体性能。这表明本文提出的方案在多无人机系统中具有重要的实用价值。

如图 5 所示,当 MEC 节点分配给无人机最大计算资源增加时,各方案能量消耗呈现不同趋势。具体而言,由于 OOWOS 方案完全依赖本地计算,其能量消耗处于最高水平。其余方案随 MEC 节点计算资源增加无人机总能耗逐渐降低并最终趋于稳定。这是因为在初期随着 MEC 节点分配给 UAV 的计算能力增强,MEC 服务器能够响应更多的 UAV 卸载请求,从而降低无人机的能耗。最终能耗稳定是因为所有无人机的任务都被卸载到了 MEC 节点上,此时能耗主要来源于数据传输。相比之下,本文方案通过利用 MEC 节点间的信息交互和协作,为多无人机用户提供最优的卸载决策和资源分配策略,从而实现了最低的能量消耗。因此,在系统能耗方面,本文方案表现

最为出色,可有效减轻无人机计算负担,适用于大规模无人机机群的卸载和任务调度。

图6为无人机数量对无人机任务卸载性能的影响。显而易见, OOWOS方案的卸载率保持不变,这是因为该方案完全依赖于无人机机载设备处理本地计算任务。除 OOWOS方案外,其他方案的卸载率随着无人机数量增加而下降。在初期阶段,由于无人机数量不够, MEC节点拥有充足的资源来满足无人机任务的卸载需求,所提出的方案甚至实现了100%卸载率。然而,随着无人机数量的增多导致任务量增加,因此所需的计算和通信资源持续增长。然而, MEC节点计算能力有限导致 UAV任务卸载率下降。与其他对比方案相比,本文方案展现出了更优的卸载性能。例如,在无人机物流场景中,多个无人机负责监控交通状况并采集数据。通过本文方案系统能够根据每个无人机的计算能力动态调整任务卸载,确保无人机能够有效地卸载计算任务,从而在复杂的交通场景下实时监控和数据分析。

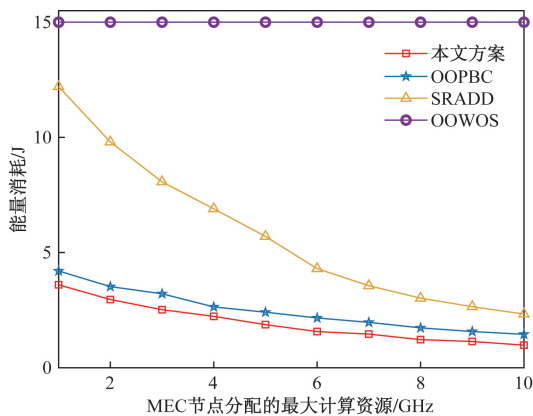


图5 系统能耗与 MEC 分配最大资源的关系

Fig. 5 System energy consumption versus maximum resources allocated by MEC nodes

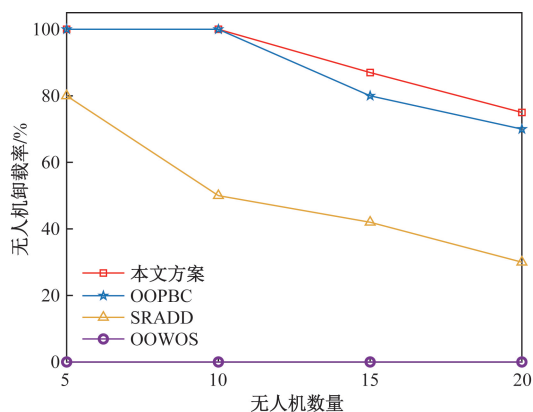


图6 不同无人机数量下 UAV 的卸载率

Fig. 6 UAV offloading rate for different numbers of UAVs

图7为 MEC节点数量对分配给无人机的计算资源的影响。可以明显看出,在 OOWOS方案中, MEC服务器分配给 UAV的计算资源为零。这是因为该方案并未涉及任务卸载功能。与另外两个对比方案相比,所提方案实现了 MEC节点向无人机卸载任务时的最大计算资源分配。这得益于本方案能够高效利用和协调各 MEC节点的计算资源,从而使无人机能够灵活地卸载任务。通过这种方式, MEC节点能够为任务卸载提供更丰富的计算资源,从而提升系统性能。这进一步表明所提出的方案具有良好的实用性。

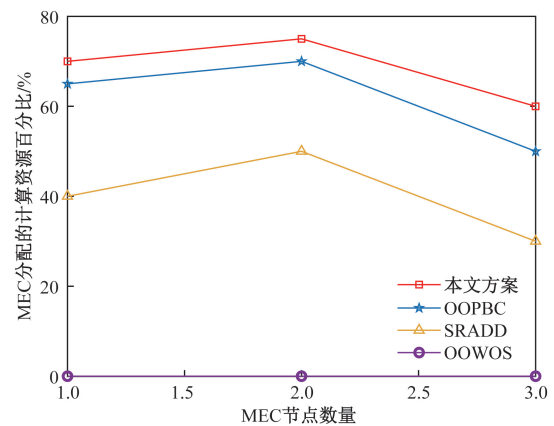


图7 MEC 分配的资源百分比与 MEC 节点数的关系
Fig. 7 Percentage of resources allocated by MEC versus the number of MEC nodes

4 结论

研究在 MEC 辅助的无人机系统中计算密集型任务卸载以及 MEC 节点间的协作。提出一种负载感知任务卸载和资源分配方案以最小化无人机能耗。在确保无人机任务满足时延要求和 MEC 节点资源得到合理利用的前提下,将任务卸载与资源分配问题建模为多个 MEC 节点间的马尔可夫博弈过程,并设计一种分布式 MDDPG 算法,该算法使每个 MEC 节点都能获取其他 MEC 节点策略信息,实现在线数据协同。在智慧交通系统中,无人机需要实时采集并分析大量的交通数据。通过本文方案,多个 MEC 节点协同工作,动态调整计算资源和任务卸载策略,确保无人机能够高效完成数据处理任务,从而提高了整个交通监控系统的处理能力和响应速度。在仿真实验中,评估 MDDPG 算法与 DDPG 和策略梯度算法的收敛性能,从系统能耗、任务卸载率和分配资源方面分析本文方案的整体性能。仿真结果表明,与其他对比方案相比,本文方案具有显著优势,进而说明该方案的有效性和可行性。随着无人机、MEC 和智能通信网络不断发展,

本文方案能优化现有无人机系统, 还能对未来复杂、动态环境中的无人机网络发展提供新思路。特别是在未来通信网络、无人机物流和智慧交通等领域, 本文方案能有效解决计算任务卸载和能量优化问题, 具有广泛的现实应用场景。

参 考 文 献

- [1] Tarik T, Nassim S, Zinelaabidine N, et al. VR-based immersive service management in B5G mobile systems: a UAV command and control use case[J]. *IEEE Internet of Things Journal*, 2023, 10(6): 5349-5363.
- [2] Lin P, Liu Y, Zhang Z Z, et al. Cost-aware task offloading and migration for wireless virtual reality using interactive A3C approach[J]. *IEEE Transactions on Vehicular Technology*, 2024, 73(7): 10850-10855.
- [3] Shen H, Lin D F, Yang X W, et al. Vision-based multiobject tracking through UAV swarm[J]. *IEEE Geoscience and Remote Sensing Letters*, 2023, 20: 1-5.
- [4] 张晓龙, 吴巍, 周彬. 基于移动边缘计算的任务卸载策略研究[J]. *科学技术与工程*, 2022, 22(11): 4434-4439.
Zhang Xiaolong, Wu Wei, Zhou Bin. Task unloading strategy based on mobile edge computing[J]. *Science Technology and Engineering*, 2022, 22(11): 4434-4439.
- [5] Ning Z L, Zhang K Y, Wang X J, et al. Intelligent edge computing in internet of vehicles: a joint computation offloading and caching solution[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22(4): 2212-2225.
- [6] 张广驰, 何梓楠, 崔苗. 基于深度强化学习的无人机辅助移动边缘计算系统能耗优化[J]. *电子与信息学报*, 2023, 45(5): 1635-1643.
Zhang Guangchi, He Zinan, Cui Miao. Energy consumption optimization of unmanned aerial vehicle assisted mobile edge computing systems based on deep reinforcement learning[J]. *Journal of Electronics & Information Technology*, 2023, 45(5): 1635-1643.
- [7] Ibrahim S, Chadi A, Dariush E, et al. Optimizing information freshness for MEC-enabled cooperative autonomous driving[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(8): 13127-13140.
- [8] 岳文静, 李可, 陈志. 无人机辅助车联网边缘计算卸载与资源分配方案[J/OL]. (2024-10-29) [2024-11-05]. <http://kns.cnki.net/kcms/detail/10.1491.TP.20241028.1739.002.html>.
- Yue Wenjing, Li Ke, Chen Zhi. UAV-assisted vehicular network edge computing offloading and resource allocation scheme[J/OL]. (2024-10-29) [2024-11-05]. <http://kns.cnki.net/kcms/detail/10.1491.TP.20241028.1739.002.html>.
- [9] Han Z H, Zhou T, Xu T H, et al. Joint user association and deployment optimization for delay-minimized UAV-aided MEC networks[J]. *IEEE Wireless Communications Letters*, 2023, 12(10): 1791-1795.
- [10] Wu Z W, Yang Z L, Yang C, et al. Joint deployment and trajectory optimization in UAV-assisted vehicular edge computing networks[J]. *Communications and Networks*, 2022, 24(1): 47-58.
- [11] 薛建彬, 武清清, 张海军. 面向用户的无人机辅助 MEC 网络性能优化[J]. 2024, 46(6): 166-173.
Xue Jianbin, Wu Qingqing, Zhang Haijun. User-oriented UAV-aided MEC network performance optimization[J]. *Journal of the National University of Defense Science and Technology*, 2024, 46(6): 166-173.
- [12] 赵婵婵, 郭晓敏, 海晓伟, 等. 缓存辅助移动边缘计算的任务卸载与资源分配联合优化策略[J]. *科学技术与工程*, 2023, 23(9): 3812-3819.
Zhao Chanchan, Guo Xiaomin, Hai Xiaowei, et al. Joint optimization strategy of task offloading and resource allocation for cache-assisted mobile edge computing[J]. *Science Technology and Engineering*, 2023, 23(9): 3812-3819.
- [13] Hu H, Chen Z, Zhou F H, et al. Joint resource and trajectory optimization for heterogeneous-UAVs enabled aerial-ground cooperative computing networks[J]. *IEEE Transactions on Vehicular Technology*, 2023, 72(7): 8812-8826.
- [14] Zheng G Y, Xu C, Wen M W, et al. Service caching based aerial cooperative computing and resource allocation in multi-UAV enabled MEC systems[J]. *IEEE Transactions on Vehicular Technology*, 2022, 71(10): 10934-10947.
- [15] Qu L, Wang J M, Chadi A. Resource scheduling and delay optimization of IoT devices in UAV-assisted multi-access edge computing[J]. *IEEE Internet of Things Journal*, 2025, 12(11): 16998-17011.
- [16] Qin X T, Song Z Y, Hou T W, et al. Joint optimization of resource allocation, phase shift, and UAV trajectory for energy-efficient RIS-assisted UAV-enabled MEC systems[J]. *IEEE Transactions on Green Communications and Networking*, 2023, 7(4): 1778-1792.