

无人机集群联合拓扑控制的智能路由规划方法

颜志, 易正伦, 欧阳博, 王耀南

(湖南大学电气与信息工程学院, 湖南 长沙 410082)

摘要: 针对现有无人机集群路由协议拓扑自适应能力弱, 易产生包重传、能量空洞和高时延, 严重恶化了数据路由性能的问题, 针对无人机集群中集群拓扑与路由的耦合特性, 提出了一种联合拓扑控制的智能路由规划 (IRPJTC) 方法。该方法由基于虚拟力的自适应拓扑控制 (VFATC) 和基于近端策略优化的地理路由规划 (PPO-GRP) 组成。其中, VFATC 使各无人机根据邻居运动状态信息自适应调整与邻居的距离, 保证集群中链路的稳定连接; 进一步, PPO-GRP 引入 VFATC 中的链路稳定性指标, 并结合端到端时延与能耗指标, 设计多目标奖励函数, 采用深度强化学习中的近端策略优化算法训练路由策略。仿真实验结果表明, IRPJTC 相比于现有路由方法, 能在保证分组传输成功率的同时, 使端到端时延降低 12.11%, 无人机集群能耗降低 4.56%, 且具备更强的能耗均衡能力。

关键词: 无人机集群; 路由协议; 拓扑控制; 近端策略优化; 深度强化学习

中图分类号: TN915.04

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2024032

Intelligent route planning method with jointing topology control of UAV swarm

YAN Zhi, YI Zhenglun, OUYANG Bo, WANG Yaonan

College of Electrical and Information Engineering, Hunan University, Changsha 410082, China

Abstract: Existing routing protocols without awareness of the topology causes excessive retransmissions, energy holes, and long delay, data routing performance was seriously deteriorated. Considering the relation of topology and routing, an intelligent route planning with jointing topology control (IRPJTC) method was proposed. IRPJTC consisted of two part, the virtual force-based adaptive topology control (VFATC), and the PPO-based geographic routing protocol (PPO-GRP). Based on neighbor's mobility information, the distance between UAVs was adaptively adjusted by VFATC to provide stable links between UAVs. Combined with link stability metric in VFATC, end-to-end delay and energy consumption, a multi-objective reward function was designed by PPO-GRP to train optimal routing strategy. According to the performance study, the proposed IRPJTC reduces existing routing protocols by 12.11% of end-to-end delay, and 4.56% of energy consumption, and has a better energy balance ability.

Keywords: UAV swarm, routing protocol, topology control, proximal policy optimization, deep reinforcement learning

收稿日期: 2023-08-21; 修回日期: 2023-11-07

基金项目: 国家自然科学基金资助项目 (No.62293511); 湖南省科技重大专项基金资助项目 (No.2021GK1010); 网络与交换技术国家重点实验室 (北京邮电大学) 开放课题基金资助项目 (No.SKLNST-2021-2-03)

Foundation Items: The National Natural Science Foundation of China (No.62293511), Special Funding Support for the Construction of Innovative Provinces in Hunan Province (No.2021GK1010), The State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, China (No.SKLNST-2021-2-03)

0 引言

随着无人机 (UAV, unmanned aerial vehicle) 技术的飞速发展, 无人机集群已被广泛应用于区域监视^[1]、野火监测^[2]、空中基站^[3]、灾后应急通信^[4]以及移动边缘计算^[5]等场景。

无人机集群区域监视场景中, 集群节点的高移动性导致机间通信链路频繁中断, 如何动态调控机间距离, 使机间通信链路维持稳定连接, 是亟待解决的一大挑战。与此同时, 拓扑高频变化将加重数据路由中的包重传、能量空洞和高时延等问题, 如何综合考虑各项通信指标, 针对性设计路由协议, 以保证稳定高效的数据路由, 是亟待解决的另一大挑战。针对这两大挑战, 国内外学者在无人机集群拓扑控制^[6]和路由协议设计两方面开展了大量工作。

在无人机集群拓扑控制方面, 文献[7]利用粒子群优化算法, 在信干噪比 (SINR, signal to interference plus noise ratio) 与安全距离约束下实现拓扑构建, 并采用梯度上升法进行动态拓扑调整, 以增强任务与网络性能。文献[8]面向应急救援场景下的用户覆盖场景, 提出一种基于虚拟弹簧的拓扑控制方法, 在保证空中链路连接质量的同时, 最大化对地面用户的覆盖。文献[9]提出了一种基于虚拟力的分布式拓扑控制算法, 以有效控制无人机集群拓扑, 增强连接性与扩大覆盖范围。

在无人机集群路由协议方面, 现有方法可分为三类。第一类为静态路由协议, 其基于静态路由表进行路由决策, 由于缺乏路由表更新策略, 仅适用于固定拓扑的无人机编队, 难以应用于高动态的分布式无人机集群; 第二类为先验式路由协议, 其在网络拓扑改变时动态更新路由表, 但在大规模无人机集群场景中, 维护路由表将产生极大的路由开销; 第三类为基于地理位置的路由协议, 其不需要为整个拓扑建立并维护路由表, 仅需根据邻居位置信息进行路由决策, 这避免了大量额外开销。因此, 基于地理位置的路由协议在拓扑频繁变化的无人机集群场景中有良好的应用前景。

在基于地理位置的路由协议方面, 文献[10]提出了经典的贪婪周边无状态路由 (GPSR, greedy perimeter stateless routing) 协议, 其通过贪婪策略选取距离目的节点最近的下一跳节点进行路由转发。进一步, 研究者将强化学习算法引入基于地理位置的路由协议中。文献[11]提出了一种基于 Q 学

习的地理路由协议, 通过结合链路和位置误差辅助进行路由选择。但是, 其只考虑了时延指标, 缺乏对链路稳定性与能耗的考虑, 在拓扑动态变化的大规模无人机集群场景中具有局限性。文献[12]提出了一种基于 Q 学习的多目标优化路由协议, 在路由过程中考虑了能耗因素, 并针对路由空洞问题改进了奖惩机制, 有效缓解了路由空洞导致的数据包重传问题。然而上述方法均基于 Q 学习方法, 随着无人机集群规模扩大, Q 表的更新变得烦琐且低效, 其路由性能难以保证, 故研究者开始将深度强化学习算法与传统路由协议结合。文献[13]基于深度 Q 网络 (DQN, deep Q-network) 提出一种可信地理位置路由协议, 考虑了对异常节点的感知, 以保障网络性能。文献[14]基于 DQN 提出了一种去中心化智能路由策略, 在奖励函数设置中综合考虑了节点运动信息、负载程度和链路质量, 在拓扑动态变化的网络中具有更强的适应性。文献[15]提出了一种基于深度强化学习的地理数据包路由优化算法, 其基于近端策略优化 (PPO, proximal policy optimization) 算法, 通过针对性设计奖励函数, 有效避免了路由环路问题, 但忽略了时延以及能耗指标, 且缺乏对机间链路稳定性的考虑, 导致其在拓扑动态变化的大规模无人机集群场景中路由性能不佳。

以上研究均是单独针对拓扑控制或路由协议开展的, 忽略了拓扑控制与路由协议间的耦合性问题。然而, 在无人机集群这种动态网络中, 拓扑控制算法通过自适应调整网络中各节点的运动状态, 使节点间链路保持强连接性, 避免了频繁的链路中断, 为数据路由提供稳定的通信环境。因此, 在动态网络中, 要进一步对拓扑控制与路由进行耦合研究。文献[16]针对地面移动传感器网络节点动态性、能量受限以及能耗不均衡等问题, 根据节点移动模型, 提出了一种基于节点移动预测的能量均衡拓扑控制算法。进一步地, 基于蚁群优化算法设计路由协议, 提升了移动传感器网络的数据路由性能。文献[17]面向资源受限环境下的卫星通信网络场景, 基于时变图模型, 提出了一种以最短时延为目标的拓扑与路由联合规划算法。该算法面向业务的时延保障需求, 计算出最优路径, 根据路径进行拓扑动态规划, 并进一步对卫星链路进行动态调度与路由。

然而, 相比于地面移动传感器网络中节点的慢速移动和卫星通信网络中节点运行轨道的确定性,

无人机的高速移动特性导致无人机集群网络拓扑具有高动态、不确定性等特点。现有研究成果难以解决区域监视场景下无人机集群数据高效路由的难题。

基于此，本文面向无人机集群区域监视场景，考虑无人机集群拓扑控制与路由的耦合特性，提出一种联合拓扑控制的智能路由规划方法，以提升无人机集群的数据路由性能。本文主要的研究工作概况如下。

1) 考虑无人机集群拓扑控制与路由的耦合特性，提出了一种联合拓扑控制的智能路由规划 (IRPJTC, intelligent route planning with jointing topology control) 方法。IRPJTC 由基于虚拟力的自适应拓扑控制 (VFATC, virtual force-based adaptive topology control) 与基于近端策略优化的地理路由规划 (PPO-GRP, PPO-based geographic routing planning) 组成。其中，拓扑控制自适应调整机间距离以维持链路稳定连接，为路由规划提供稳定的通信环境；路由规划联合拓扑控制，综合多种通信指标进行路由决策。

2) 在路由规划阶段，本文基于深度强化学习理论设计了 PPO-GRP。在 VFATC 构建的稳定拓扑基础上，PPO-GRP 通过引入 VFATC 中的拓扑链路稳定性指标，并结合端到端时延和能耗指标设计多目标奖励函数，以提升无人机与地面站的多跳中继通信性能。

3) 仿真实验结果表明，本文所提出的 IRPJTC 相比于现有路由方法，在保证分组传输成功率的同时，端到端时延降低 12.11%，无人机集群能耗降低 4.56%，且具备更强的能耗均衡能力。

1 系统建模

1.1 系统模型

考虑如图 1 所示的无人机集群区域监视场景，该场景系统模型参数如表 1 所示。其中 k 架四旋翼无人机 $U = \{u_1, u_2, \dots, u_k\}$ 配备 GPS、惯性测量单元与无线通信模块。无人机部署在特定的任务区域内，将监视任务持续时间 T 划分为 m 个时隙，则有 $T = \{t_1, t_2, \dots, t_{m-1}, t_m\}$ ，每个时隙的长度为常数 τ 。因此，在时隙 t_n ，无人机集群拓扑结构可表示为一个无向图 $G(t_n) = (V(t_n), E(t_n))$ 。 $V(t_n) \in \{U(t_n) \cup BS\}$ 表示无人机集群 $U(t_n)$ 和单个固定位置的地面站 BS。

$E(t_n) \subseteq \{\{u_i, u_j\} | u_i, u_j \in V\}$ 表示无人机节点之间或无人机与地面站间是否建立双向链路，即在拓扑图中，两节点是否存在边。

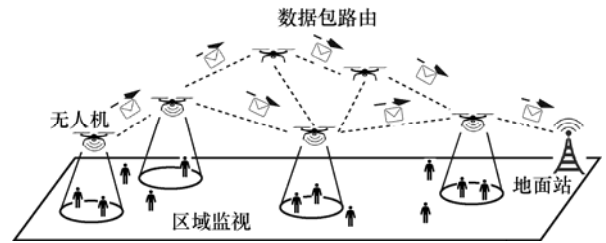


图 1 无人机集群区域监视场景

表 1 系统模型参数

参数	含义
$T = \{t_1, t_2, \dots, t_{m-1}, t_m\}$	区域监视总任务时间与时隙划分
τ	每个时隙的长度
$U = \{u_1, u_2, \dots, u_k\}$	无人机集合
$\vec{p}_i(t_n)$	无人机 u_i 在时隙 t_n 的位置向量
\vec{p}_{BS}	地面站的位置坐标
$\vec{v}_i(t_n)$	无人机 u_i 在时隙 t_n 的速度向量
$\vec{a}_i(t_n)$	无人机 u_i 在时隙 t_n 的加速度向量
R_r	安全距离约束范围 (斥力范围)
R_a	无人机最大传输距离 (引力范围)
R_{BS}	无人机与地面站的通信范围
N_i	u_i 的单跳邻居集
N_i^a	u_i 引力范围内的单跳邻居集
N_i^r	u_i 斥力范围内的单跳邻居集
$d_{ij}(t_n)$	u_i 与 u_j 在时隙 t_n 的欧氏距离
f_{ij}^{Time}	单跳数据包传输时间
f_{ij}^{Speed}	单跳数据包传输速度
LD_{ij}	链路寿命

在区域监视任务过程中，无人机 u_i 使用机载 GPS 定位自身位置 $\vec{P}_i(t_n) = [x_i, y_i, z_i]$ 。无人机的最大传输范围为 R_a ， u_i 与 u_j 两无人机的欧氏距离为 d_{ij} 。若 $d_{ij} \leq R_a$ ，则表示两无人机在各自的最大传输范围内能够建立双向通信链路，故在拓扑图中，两无人机节点之间存在边。为了满足无人机间的安全距离和传输距离约束，定义安全距离约束范围为 R_r 。若 $d_{ij} \leq R_r$ ，则表示两无人机距离过近，不仅无法对监视区域提供有效的覆盖，且具有相互碰撞的风险。故无人机间距 d_{ij} 需保持在

$R_r < d_{ij} \leq R_a$ ，才能有效执行区域监视任务。各无人机通过机载传感器监视地面，并定期将监视数据传输给地面站进行处理。无人机有限的传输距离导致其仅能和自身邻居进行直接数据交换，故各无人机需通过多跳路由转发方式传输监视数据至地面站。地面站与无人机间的通信范围为 R_{BS} ，当无人机与地面站之间距离在通信范围内时，则两者能够建立双向通信链路，故可视为无人机与地面站间存在边。

1.2 信道模型

区域监视场景下无人机集群飞行高度较高，几乎不受到障碍物遮挡，故本文使用自由空间损耗模型。源无人机 u_i 与接收无人机 u_j 间的信道功率增益定义为 $G_{ij}(t_n) = \rho_o d_{ij}(t_n)^{-\psi}$ ，其中， ρ_o 表示参考距离为 1 m 时的信道功率增益， ψ 表示路径损耗指数， d_{ij} 表示两架无人机间的距离。根据文献[18]，若 u_i 发射功率为 P_i^{tx} ，则无人机 u_j 处的 SINR 为

$$\text{SINR}_{ij}(t_n) = \frac{G_{ij}(t_n)P_i^{tx}}{I_{ij}(t_n) + \sigma^2(t_n)} \quad (1)$$

其中， $I_{ij}(t_n)$ 表示 u_j 的邻居无人机对 u_j 产生的传输干扰， $\sigma^2(t_n)$ 表示环境噪声。

本文假设无人机集群为避免数据冲突，MAC 协议基于时分多址 (TDMA, time division multiple access) 方法来保证每架无人机都拥有各自的广播时隙。假设 u_j 的邻居无人机在每个时隙内都以共同概率 β 传输数据，则传输干扰 $I_{ij}(t_n)$ 可近似表示为 $I_{ij}(t_n) = \sum_{c \in U, c \neq i, j} \beta G_{cj}(t_n) P_c^{tx}$ ，其中 $c \in U$ 表示同步传输数据的邻居节点集合。故在干扰 $I_{ij}(t_n)$ 下，SINR (单位为 dB) 可表示为

$$\text{SINR}_{ij}(t_n) = 10 \log \frac{G_{ij}(t_n)P_i^{tx}}{\sum_{c \neq i, j} \beta G_{cj}(t_n)P_c^{tx} + \sigma^2(t_n)} \quad (2)$$

若 $\text{SINR}_{ij}(t_n)$ 大于预定义的 SINR 阈值 (SINR_{th})，表示链路成功建立。因此，无人机最大传输范围为

$$d_{ij}^{th} = \left[\frac{\rho_o P_i^{tx}}{(\bar{I}_{ij} + \sigma^2(t_n)) 10^{\frac{\text{SINR}_{th}}{10}}} \right]^{\frac{1}{\psi}} \quad (3)$$

假设每架无人机均配备全向天线，则其最大传输范围可表示为半径为 $R_a = d_{ij}^{th}$ 的球体。

对于给定的带宽 B ，无人机在时隙 t_n 的数据传输速率可表示为 $C_{ij}(t_n) = B \log[1 + \text{SINR}_{ij}(t_n)]$ 。令 l_{ij} 表示无人机 u_i 与 u_j 之间建立的通信链路，根据文献[18]，链路 l_{ij} 的误包率 (PER, packet error rate) 可近似表示为

$$\text{PER}_{ij}(\text{SINR}_{ij}(t_n)) \approx \begin{cases} a_{id} \exp(-g_{id} \text{SINR}_{ij}(t_n)), & \text{SINR}_{ij}(t_n) \geq \text{SINR}_{th} \\ 1, & \text{SINR}_{ij}(t_n) < \text{SINR}_{th} \end{cases} \quad (4)$$

其中，id 为不同传输模式下的索引， a_{id} 与 g_{id} 为传输模式相关参数，其参数值在采用不同调制方式的传输模式中存在差异，各传输模式与相关参数的对应值来自文献[19]。

1.3 时延模型

链路 l_{ij} 的数据传输时延 D_{ij}^{total} 由单跳 MAC 时延 D_{ij}^{mac} 、排队时延 D_{ij}^{que} 、传播时延 D_{ij}^{pg} 和传输时延 D_{ij}^{tx} 组成，故有 $D_{ij}^{total} = D_{ij}^{mac} + D_{ij}^{que} + D_{ij}^{pg} + D_{ij}^{tx}$ 。其中， $D_{ij}^{mac} = t_{ACK} - t_{send}$ ， t_{send} 为 u_i 发送数据包的时刻， t_{ACK} 为 u_i 收到 u_j 回传的 ACK 包的时刻。在 MAC 层采用 M/M/1 排队模型， D_{ij}^{que} 取决于 u_j 的到达速率 A_j 和服务速率 F_j ，且到达速率 A_j 服从泊松分布。故 D_{ij}^{que} 可表示为 $D_{ij}^{que} = \frac{1}{F_j - A_j}$ 。 D_{ij}^{pg} 可表示为 $D_{ij}^{pg} = \frac{d_{ij}}{v_p}$ ，其中， v_p 为电磁波传播速度。 D_{ij}^{tx} 可表示为 $D_{ij}^{tx} = \frac{P_{size}}{C_{ij}}$ ，其中， P_{size} 为数据包大小， C_{ij} 为数据传输速率。

由于无人机无法准确获取当前时隙的机间链路传输时延，故需要基于过往时隙的链路时延估算当前时隙的链路时延。本文采用窗口均值-指数加权移动平均 (WM-EWMA) 法估算时隙 t_n 两无人机间的精确时延 $\text{delay}_{ij}(t_n)$ [12]，计算式为

$$\text{delay}_{ij}(t_n) = (1 - \rho) \frac{\sum_{k=1}^{n-1} \text{delay}_{ij}(t_k)}{W} + \rho D_{ij}^{total} \quad (5)$$

其中， W 为纳入计算的过往时延数据数量， $\text{delay}_{ij}(t_n)$ 为时隙 t_n 的指数加权移动平均估计值， t_{u_j} 为当前时隙的时延观测值， $\rho \in [0, 1]$ 为权重系数。

考虑传输过程中由于单跳传输误包产生的重传，链路 l_{ij} 的单跳包传输时间 (PTT, packet travel

time) 可表示为

$$f_{ij}^{\text{Time}} = \frac{\text{delay}_{ij}(t_n)}{1 - \text{PER}_{ij}(\text{SINR}_{ij}(t_n))} \quad (6)$$

f_{ij}^{Time} 考虑数据路由过程中单跳误包导致的包重传, 精确量化了数据包从 u_i 传输至 u_j 所需要的时间。

1.4 能耗模型

无人机通信能耗取决于当前时隙 t_n 发送与接收数据的大小。给定发送数据速率 $C_{ij}^{\text{tx}}(t_n)$ 、传输功率 P_i^{tx} 、数据包大小 P_{size} , 则无人机 u_i 的数据传输能耗可定义为 $E_i^{\text{tx}} = \frac{P_i^{\text{tx}} P_{\text{size}}}{C_i^{\text{tx}}(t_n)}$ 。同理可计算出无人机 u_i

的数据接收能耗 $E_i^{\text{rx}} = \frac{P_i^{\text{rx}} P_{\text{size}}}{C_i^{\text{rx}}(t_n)}$ 。

综上所述, 对于一个总能量为 E_{total} 的无人机 u_i , 在时隙 t_n 的剩余能量可表示为

$$\text{RE}_i(t_n) = E_{\text{total}} - \sum_{t_1}^{t_{n-1}} [E_i^{\text{tx}} + E_i^{\text{rx}}] \quad (7)$$

2 算法设计

本节将介绍 IRPJTC 方法的架构, 以及 2 个子模块的算法设计。IRPJTC 方法的架构如图 2 所示。其中, IRPJTC 由 VFATC 与 PPO-GRP 两部分组成, VFATC 通过虚拟力控制各无人机移动, 以动态更新集群拓扑。PPO-GRP 引入 VFATC 中的链路稳定性指标, 结合端到端时延与能耗指标, 针对性设置奖励函数, 并采用 PPO 算法训练路由策略, 以实现高效稳定的数据路由。

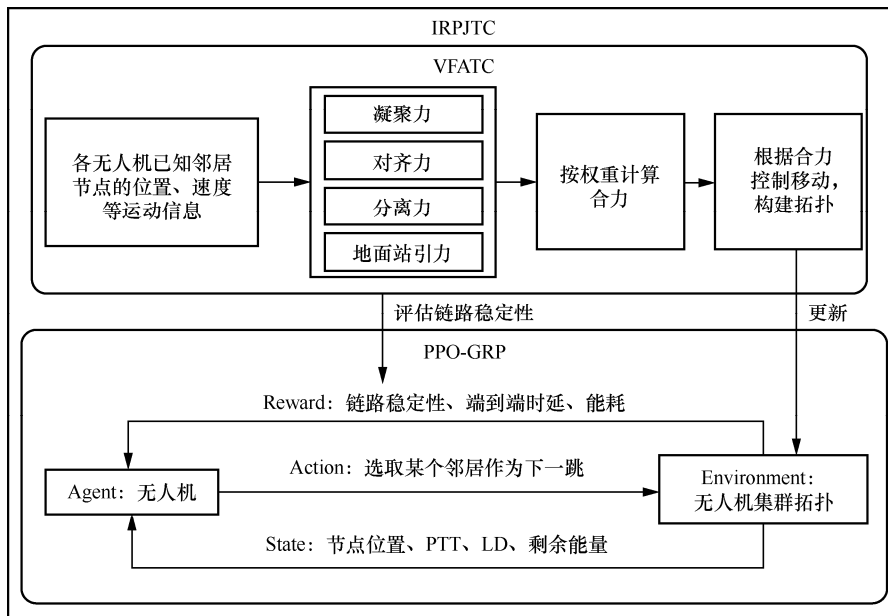


图 2 IRPJTC 方法的架构

2.1 基于虚拟力拓扑控制 VFATC

VFATC 中, 各无人机通过广播 Hello 报文以发现、建立和维护邻居关系。Hello 报文中包含无人机运动信息 (MI, mobility information), MI 包括无人机的位置、速度和加速度信息。在每个时隙中, 各无人机与邻居相互通信以获取 MI, 并计算虚拟力, 各节点根据自身受到的虚拟力合力计算自身在下一时隙的运动状态。VFATC 中存在 4 种虚拟力分量: 凝聚力、对齐力、分离力以及地面站引力, 虚拟力分区如图 3 所示, 各虚拟力分量的定义如下。

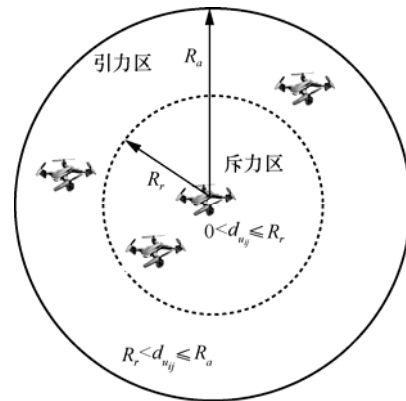


图 3 虚拟力分区

凝聚力使每架无人机被邻居无人机位置的质心所吸引，其目的是使各无人机彼此靠近，以避免频繁的链路中断。由图 3 可知，若无人机间距离满足 $R_r < d_{ij} \leq R_a$ ，位于引力范围内，凝聚力的计算式为

$$\overline{\mathbf{CF}}_i(t_n) = \lambda_{\text{CF}} \frac{\sum_{u_j \in N_i^a(t_n)} [\bar{\mathbf{p}}_j(t_n) - \bar{\mathbf{p}}_i(t_n)]}{|N_i^a(t_n)|} \quad (8)$$

其中， $N_i^a(t_n)$ 为 u_i 的邻居节点中，满足 $R_r < d_{ij} \leq R_a$ 的邻居节点集合； λ_{CF} 为凝聚力常量； $|N_i^a(t_n)|$ 为 $N_i^a(t_n)$ 集合中的节点数量。

对齐力使每架无人机根据邻居速度向量的均值调整移动方向，使集群整体朝着统一方向移动，对齐力的计算式为

$$\overline{\mathbf{AF}}_i(t_n) = \lambda_{\text{AF}} \frac{\sum_{u_j \in N_i^a(t_n)} [\bar{\mathbf{v}}_j(t_n) - \bar{\mathbf{v}}_i(t_n)]}{|N_i^a(t_n)|} \quad (9)$$

其中， λ_{AF} 为对齐力常量。

分离力使每架无人机与邻居无人机保持一定距离，在避免相互碰撞的同时，减少各无人机对地面监视范围的重叠。当无人机间距离满足 $0 < d_{ij} \leq R_r$ 时，位于分离力范围内，分离力的计算式为

$$\overline{\mathbf{SF}}_i(t_n) = -\lambda_{\text{SF}} \frac{\sum_{u_j \in N_i^r(t_n)} \frac{R_r [\bar{\mathbf{p}}_j(t_n) - \bar{\mathbf{p}}_i(t_n)]}{\|[\bar{\mathbf{p}}_j(t_n) - \bar{\mathbf{p}}_i(t_n)]\|^2}}{|N_i^r(t_n)|} \quad (10)$$

其中， $N_i^r(t_n)$ 为 u_i 的邻居节点中，满足 $0 < d_{ij} \leq R_r$ 的节点集合； λ_{SF} 为分离力常量； $|N_i^r(t_n)|$ 为 $N_i^r(t_n)$ 集合中的节点数量。

地面站引力使无人机集群中的边缘节点能时刻保持与地面站的连接，其计算式为

$$\overline{\mathbf{BSF}}_i(t_n) = \lambda_{\text{BSF}} [\bar{\mathbf{p}}_{\text{BS}} - \bar{\mathbf{p}}_i(t_n)] \quad (11)$$

其中， λ_{BSF} 为地面站引力常量。

对上述分量进行加权计算，得到合力为

$$\overline{\mathbf{TF}}_i(t_n) = [\delta_1 \overline{\mathbf{CF}}_i(t_n) + \delta_2 \overline{\mathbf{AF}}_i(t_n) + \delta_3 \overline{\mathbf{SF}}_i(t_n)] + \overline{\mathbf{BSF}}_i(t_n) \quad (12)$$

其中， $\delta_1 \sim \delta_3$ 为各虚拟力分量的权重，且 $\delta_1 + \delta_2 + \delta_3 = 1$ 。

VFATC 通过感知无人机与邻居的最小距离与最大距离以自适应调节虚拟力权重。当无人机与邻

居的最小距离 $\min_{u_j \in N_i^r(t_n)} d_{ij}(t_n) \leq R_r$ 时，增大分离力权重，使 $\delta_3 > \delta_1 + \delta_2$ ；当无人机与邻居的最大距离

$\max_{u_j \in N_i^a(t_n)} d_{ij}(t_n) \geq 0.9R_a$ 时，增大凝聚力权重，使 $\delta_1 > \delta_2 + \delta_3$ 。虚拟力权重的自适应调整，有助于稳定机间链路的连接。

各无人机在当前时隙 t_n ，根据自身 $\overline{\mathbf{TF}}_i(t_n)$ 计算加速度 $\bar{\mathbf{a}}_i(t_n)$ ，并根据 $\bar{\mathbf{a}}_i(t_n)$ 计算自身在下一时隙 t_{n+1} 的速度 $\bar{\mathbf{v}}_i(t_{n+1})$ 和位置 $\bar{\mathbf{p}}_i(t_{n+1})$ ，计算式为

$$\bar{\mathbf{a}}_i(t_n) = \left(\frac{\overline{\mathbf{TF}}_i(t_n)}{\|\overline{\mathbf{TF}}_i(t_n)\|} \right) \tan^{-1} \left(\|\overline{\mathbf{TF}}_i(t_n)\| \right) a_{\text{max}} \frac{2}{\pi} \quad (13)$$

$$\bar{\mathbf{v}}_i(t_{n+1}) = \bar{\mathbf{v}}_i(t_n) + \bar{\mathbf{a}}_i(t_n) \tau \quad (14)$$

$$\bar{\mathbf{v}}_i(t_{n+1}) = \begin{cases} \frac{\bar{\mathbf{v}}_i(t_{n+1})}{\|\bar{\mathbf{v}}_i(t_{n+1})\|} v_{\text{max}} \exp(\text{syn}_i - 1), \|\bar{\mathbf{v}}_i(t_{n+1})\| \geq v_{\text{max}} \\ \bar{\mathbf{v}}_i(t_{n+1}) \exp(\text{syn}_i - 1), \|\bar{\mathbf{v}}_i(t_{n+1})\| < v_{\text{max}} \end{cases} \quad (15)$$

$$\text{syn}_i = \frac{\left\| \sum_{u_j \in N_i^a(t_n)} \bar{\mathbf{v}}_{u_j}(t_n) \right\|}{\sum_{u_j \in N_i^a(t_n)} \|\bar{\mathbf{v}}_{u_j}(t_n)\|} \quad (16)$$

$$\bar{\mathbf{p}}_i(t_{n+1}) = \bar{\mathbf{p}}_i(t_n) + \bar{\mathbf{v}}_i(t_n) \tau + \frac{1}{2} \bar{\mathbf{a}}_i(t_n) \tau^2 \quad (17)$$

其中， syn_i 为速度限制参数，其限制各无人机只有在与邻居速度方向一致时才能达到最大速度，用于使无人机平稳飞行，避免碰撞。

由于 Hello 报文的广播是固定间隔的，对于 2 个相互远离的无人机，只有在两无人机还处于各自通信范围内时相互交换 Hello 报文才能及时调整各自运动状态，在下一时隙相互靠近，以保证链路不中断。因此，VFATC 通过各无人机的 MI 估算链路寿命 (LD, link duration)。LD 表示无人机与邻居飞离各自通信范围所需的时间，LD 越大，证明该链路越不容易中断，具有更高的稳定性。LD 具体计算方法如下。

在时隙 t_n 与 $t_{n'}$ (其中 $t_{n'} > t_n$)， u_i 通过与 u_j 交换 Hello 报文的方式获取两时隙的位置信息 $\bar{\mathbf{p}}_j(t_n)$ 与 $\bar{\mathbf{p}}_j(t_{n'})$ ，时隙 t_n 时两节点间距离为 $d_{ij}(t_n)$ ，故时隙 t_n 和 $t_{n'}$ 时两节点的相对距离差可表示为 $\Delta d = d_{ij}(t_{n'}) - d_{ij}(t_n)$ 。当 $\Delta d > 0$ 时，代表两节点正互相远离，两节点的相对速度可表示为 $\frac{\Delta d}{t_{n'} - t_n}$ ；当

$\Delta d \leq 0$ 时, 代表两节点正在靠近, 证明链路无断连风险。则 LD 计算式为

$$LD_{ij} = \begin{cases} \frac{R_a - d_{ij}(t_n)}{\Delta d}, & \Delta d > 0 \\ \frac{d_{ij}(t_n)}{v_{\min}}, & \Delta d \leq 0 \end{cases} \quad (18)$$

综上所述, VFATC 算法流程如算法 1 所示。

算法 1 VFATC 算法

阶段 1 广播 Hello 报文

- 1) for each UAV do
- 2) 与邻居 UAV 交换 Hello 报文;
- 3) 根据 Hello 报文更新邻居表;
- 4) end for

阶段 2 根据虚拟力更新网络拓扑

- 1) for each UAV do
- 2) 根据式(8)~式(11)分别计算各虚拟力分量
- 3) if $\min_{u_j \in N_i^r(t_n)} d_{ij}(t_n) \leq R_r$ then
- 4) 增大分离力权重, 使 $\delta_3 > \delta_1 + \delta_2$;
- 5) else if $\max_{u_j \in N_i^a(t_n)} d_{ij}(t_n) \geq 0.9R_a$ then
- 6) 增大凝聚力权重, 使 $\delta_1 > \delta_2 + \delta_3$;
- 7) end if
- 8) 根据式(12)计算合力;
- 9) 根据式(13)~式(15)计算无人机速度;
- 10) 根据式(17)计算无人机位置;
- 11) 根据各无人机位置更新网络拓扑;
- 12) 根据式(18)计算链路寿命;
- 13) end for

2.2 基于 PPO 的地理路由规划 PPO-GRP

2.2.1 PPO 算法

PPO-GRP 基于 PPO 算法。PPO 是一种属于标准策略梯度 (PG, policy gradient) 的算法。不同于 PG 算法每次更新都需要重新采样数据, PPO 算法可以对采样的数据重复使用, 以此提高训练速度。相比于 DDPG、Actor-Critic 等算法, PPO 算法更易于实现, 更具有通用性, 且更加稳定。

PPO 算法基于 Actor-Critic 架构, 其包含策略网络 (Actor) 和评估网络 (Critic)。策略网络用于输出动作值的均值, 评估网络用于输出状态价值。在任意一个回合中, 首先利用现有策略 $\pi(s; \theta_{\text{old}})$ 与环境交互 N 个

时间步长, 获得 N 组数据, 即 $\{s_{t_n}, a_{t_n}, P(\pi_{\theta_{\text{old}}}(a_{t_n} | s_{t_n})), r_{t_n}\}$, $n = 0, 1, 2, \dots, N$ 。其中, s_{t_n} 表示时隙 t_n 的状态, a_{t_n} 表示时隙 t_n 的动作。PPO 中的策略网络输出动作的均值, 使动作值服从高斯分布, 高斯分布的标准差随网络更新频率衰减。然后随机采样出一个动作 a_{t_n} , $P(\pi_{\theta_{\text{old}}}(a_{t_n} | s_{t_n}))$ 表示在服从高斯分布的动作值中选取动作 a_{t_n} 的概率。执行动作后, 环境状态更新为 $s_{t_{n+1}}$, 并且产生相应的奖励 r_{t_n} 。在此过程中不会对网络参数进行更新。获取一定批量的数据后, 按照折扣率 γ 计算这批数据中每个时隙对应的奖励期望和优势估计分别为

$$\hat{R}_n = r_{t_n} + \gamma r_{t_{n+1}} + \dots + \gamma^{N-n+1} r_{t_{N-1}} + \gamma^{N-n} V(s_{t_N}) \quad (19)$$

$$\hat{A}_n = \hat{R}_n - V(s_{t_n}) \quad (20)$$

其中, $V(s_{t_n})$ 是利用评估网络获得的 s_{t_n} 状态的价值。网络的目标函数为

$$L_{t_n}^{\text{clip+vf+S}_e}(\theta) = \hat{\mathbb{E}}_{t_n} [L_{t_n}^{\text{clip}}(\theta) - c_1 L_{t_n}^{\text{vf}}(\theta) + c_2 S_e[\pi_{\theta}](s_{t_n})] \quad (21)$$

其中, S_e 是策略模型的熵; $L_{t_n}^{\text{clip}}(\theta)$ 是策略梯度目标函数; $L_{t_n}^{\text{vf}}(\theta)$ 是评估网络目标函数; c_1 和 c_2 是常系数, 用于调整网络目标函数中各部分的权重。 $L_{t_n}^{\text{clip}}(\theta)$ 和 $L_{t_n}^{\text{vf}}(\theta)$ 分别为

$$L_{t_n}^{\text{clip}}(\theta) = \hat{\mathbb{E}}_{t_n} [\min(P_{\theta}^{\text{ratio}} \hat{A}_n, \text{clip}(P_{\theta}^{\text{ratio}}, 1 - \varepsilon, 1 + \varepsilon) \hat{A}_n)] \quad (22)$$

$$P_{\theta}^{\text{ratio}} = \frac{\pi_{\theta}(a_{t_n} | s_{t_n})}{\pi_{\theta_{\text{old}}}(a_{t_n} | s_{t_n})} \quad (23)$$

$$L_{t_n}^{\text{vf}}(\theta) = (V_{\theta}(s_{t_n}) - \hat{R}_n)^2 \quad (24)$$

式(22)中为 $P_{\theta}^{\text{ratio}}$ 设定了上下界约束, 可以限制策略更新幅度, 防止策略过度更新。

最后, 通过最大化 $L_{t_n}^{\text{clip+vf+S}_e}$ 更新网络参数 θ , 使用采集的 N 组数据对网络参数 θ 连续更新 n_{update} 次后, 将参数 θ_{old} 更新为 θ 。

2.2.2 状态空间与动作空间

区域监视场景下, 集群中的无人机仅知道邻居无人机的状态信息与地面站的位置信息, 故需要在未知全局网络拓扑的情况下, 基于局部邻居信息进行分布式路由决策。

由于各无人机定期交换 Hello 报文，无人机与邻居交换 Hello 报文后，获得了各邻居的 MI 以及剩余能量数据，基于式(6)与式(18)估算各邻居链路的 PTT 与 LD，并将此作为邻居无人机的状态。故 PPO-GRP 的状态空间由当前无人机的三维位置信息，以及邻居的 MI、剩余能量、LD 与 PTT，以及地面站的三维位置信息组成。

基于此，PPO-GRP 的状态空间可表示为 $s_{i_n} = \{\bar{p}_i, \bar{p}_1 E_1, LD_{i1}, f_{i1}^{Time}, \bar{p}_2 E_2, \dots, \bar{p}_{BS}\}$ 。由于 PPO 算法中，输入神经网络的状态空间维度在训练过程中必须保持一致，而在无人机集群路由场景中，状态空间维度取决于当前无人机的邻居数量，故状态空间维度在训练过程中是变化的。针对此问题，PPO-GRP 取当前拓扑中的最大邻居数量计算最大状态空间维度，其中最大邻居数量指集群拓扑状态稳定后，取邻居数量最多的节点的邻居数量作为拓扑的最大邻居数量。训练过程中，若当前状态维度小于最大状态空间维度，则用特殊值（如-1）填充。由此可得，若最大邻居数量为 K ，则状态空间维度为 $6K+6$ 。

由于当前无人机需要从邻居中选择数据包的下一跳转发节点，故动作空间由当前无人机的邻居集合组成，可表示为 $a_{i_n} = \{u_1, u_2, \dots, u_k\}$ 。最大动作空间维度为最大邻居数量 K 。同样地，若当前无人机动作空间维度小于 K ，则用特殊值填充。

2.2.3 奖励函数

PPO-GRP 通过链路稳定性、端到端时延以及能耗 3 个评估指标进行多目标奖励函数设置，具体如下。

1) 链路稳定性

链路稳定性表示无人机间链路发生中断的可能性大小，PPO-GRP 引入 VFATC 的 LD 指标来评估各无人机间链路的稳定性。基于此，多目标奖励函数的第一部分可表示为

$$r_{ij}^{LD} = \frac{LD_{ij}}{\max_{u_k \in N_i} [LD_{ik}]} \quad (25)$$

2) 端到端时延

端到端时延表示一个数据包从源无人机发送到地面站所需要的总时间。1.3 节中定义了单跳包传输时间 f_{ij}^{Time} ，进一步地，定义单跳包传输速度 (PTS, packet travel speed) 为

$$f_{ij}^{Speed} = \frac{d(u_i, BS) - d(u_j, BS)}{f_{ij}^{Time}} \quad (26)$$

其中， $d(u_i, BS)$ 为当前无人机 u_i 与地面站的距离， $d(u_j, BS)$ 为下一跳无人机 u_j 与地面站的距离。若 f_{ij}^{Speed} 为正值且数值越大，代表 u_j 距离地面站越近，且 f_{ij}^{Time} 越小，故选取 f_{ij}^{Speed} 更大的下一跳，能降低总传输跳数以及单跳时延，进而实现减小端到端时延的优化目标。

基于此，多目标奖励函数的第二部分可表示为

$$r_{ij}^{Delay} = \begin{cases} \frac{f_{ij}^{Speed}}{\max_{u_k \in N_i} [f_{ij}^{Speed}]}, & f_{ij}^{Speed} > 0 \\ -1, & f_{ij}^{Speed} \leq 0 \end{cases} \quad (27)$$

3) 能耗

在数据路由过程中，若只关注链路稳定性以及端到端时延，会导致某些无人机承担过多的数据转发任务，能量过早耗尽，这降低了区域监视的任务效率，故需要将能耗作为优化指标之一。

基于此，多目标奖励函数的第三部分可表示为

$$r_{ij}^{Energy} = \frac{RE_j}{E_{total}} \quad (28)$$

综上所述，多目标奖励函数定义为

$$r_{u_j} = \begin{cases} r_{max} = 5, & u_j \text{ 为 BS} \\ r_{min} = -1, & \text{错误动作} \\ r = (\omega_1 r_{ij}^{LD} + \omega_2 r_{ij}^{Delay} + \omega_3 r_{ij}^{Energy}), & \text{其他} \end{cases} \quad (29)$$

其中， ω_1 、 ω_2 和 ω_3 为奖励权重因子，且 $\omega_1 + \omega_2 + \omega_3 = 1$ 。由 2.2.2 节可知，PPO-GRP 的动作空间中存在填充的特殊值，故为特殊值设置惩罚 r_{min} ，以避免选取错误动作。

2.2.4 训练过程

训练开始前，初始化缓冲区以及策略网络与评估网络的参数，VFATC 首先进行初始拓扑构建。PPO-GRP 基于此拓扑，随机选取一个无人机作为源节点开始路由策略训练。在训练过程中，VFATC 定期调整各无人机的 MI 以进行拓扑更新，而 PPO-GRP 则在每个时隙中进行一次下一跳转发节点的选择，直到数据包到达地面站后，再随机选取新的源节点开始新一轮数据路由过程。当路由转发

次数到一定数量时，结束当前回合，并重置各无人机的位置、速度、加速度与能量，进入下一回合。

在时隙 t_n ，PPO-GRP 的策略网络根据当前状态 s_{t_n} 输出动作值，即被选取作为下一跳的邻居节点，然后根据所选下一跳节点评估奖励。若所选下一跳节点为地面站，则代表数据包传输成功，给出最大奖励。其余情况下，则根据链路稳定性、端到端时延和能耗 3 个指标综合计算奖励。PPO-GRP 执行完此次动作后，进入下一个状态 $s_{t_{n+1}}$ ，并将时隙 t_n 的状态观测值、所选的动作、选择该动作的概率以及获得的奖励值保存至缓冲区中，存储一定数量之后，对网络参数进行 n_{update} 次更新，并清空缓存数据。综上所述，PPO-GRP 算法的详细步骤如算法 2 所示。

算法 2 PPO-GRP 算法

- 1) 初始化网络参数 θ ，缓冲区 D ;
- 2) for each episode do
- 3) 初始化各无人机位置、速度、加速度与能量，随机选取源无人机开始数据路由。
- 4) for each time slot t_n do
- 5) 获取 t_n 的状态观测值 s_{t_n} ;
- 6) 执行 VFATC 更新各无人机运动状态;
- 7) 选择下一跳转发节点 $a_{t_n} = \pi_{\theta_{\text{old}}}(s_{t_n})$ ，保存动作概率 $P(\pi_{\theta_{\text{old}}}(a_{t_n} | s_{t_n}))$;
- 8) if 下一跳为地面站，then
- 9) 随机选取源节点生成新的数据包;
- 10) else
- 11) 执行动作，数据转发至下一跳节点进入新的状态 $s_{t_{n+1}}$;
- 12) end if
- 13) 将 $\{s_{t_n}, a_{t_n}, P(\pi_{\theta_{\text{old}}}(a_{t_n} | s_{t_n})), r_{t_n}\}$ 保存至缓冲区 D ;
- 14) if D 中数据已经足够，then
- 15) 根据式(19)计算奖励期望;
- 16) 根据式(20)计算优势估计;
- 17) end if
- 18) for each update-time=1, n_{update} do
- 19) 由评估网络获取状态价值;
- 20) 根据式(21)计算目标函数 $L_{t_n}^{\text{clip+vf+S}_e}$;
- 21) 通过最大化 $L_{t_n}^{\text{clip+vf+S}_e}$ 更新网络参数 θ ;

- 22) end for
- 23) 更新状态 $s_{t_n} \leftarrow s_{t_{n+1}}$;
- 24) end for
- 25) end for

3 仿真实验与性能分析

本节基于深度学习框架 PyTorch，结合 OpenAI Gym 来实现深度强化学习仿真环境的搭建。所用硬件平台为 PC，CPU 为 Intel Core i9 13900K 5.8 GHz，GPU 为 NVIDIA RTX 4090，内存为 64 GB。使用 OpenAI Gym 构建无人机集群仿真环境。使用 PyTorch 进行 IRPJTC 算法构建，具体包括 VFATC 的实现，以及在 PPO-GRP 中定义 Actor-Critic 神经网络模型、选择优化器、定义损失函数等。

3.1 仿真环境设置

本文设置一个 $1\ 000\ \text{m} \times 1\ 000\ \text{m} \times 3\ 00\ \text{m}$ 的任务区域， M 个无人机在此区域内进行监视任务。系统仿真参数如表 2 所示，PPO-GRP 的网络参数如表 3 所示。仿真起始阶段，无人机随机部署在任务区域内，地面站位置保持固定，地面站位置 $\vec{p}_{\text{BS}} = [1\ 000, 1\ 000, 0]$ 。无人机飞行高度为 $200 \sim 300\ \text{m}$ 。无人机最大速度 $v_{\text{max}} = 15\ \text{m/s}$ ，最小速度 $v_{\text{min}} = 2\ \text{m/s}$ ，最大加速度 $a_{\text{max}} = 3\ \text{m/s}^2$ 。初始能量 $E_{\text{total}} = 2 \times 10^5\ \text{J}$ 。仿真开始后，IRPJTC 启动 VFATC 进行初始拓扑构建，无人机集群受到虚拟力作用，将自主形成集群，并向地面站靠近。当各无人机间均建立连接，且集群内有若干边缘无人机与地面站建立连接后，拓扑构建阶段完成。基于构建的网络拓扑，IRPJTC 基于 PPO-GRP 算法开始路由决策训练。训练过程中，VFATC 定期进行无人机运动状态调整，以满足机间距离约束。

表 2 系统仿真参数

参数	含义	值
M	无人机数量	30~80
τ/s	每个时隙的长度	0.1
σ/s	Hello 包广播间隔	0.5
ψ	路径损耗指数	3
SINR_{th}	SINR 阈值	1
R_a/m	最大距离（吸引距离）	150
R_r/m	最小距离（排斥距离）	100
B/MHz	带宽	20
$E_{\text{total}}/\text{J}$	无人机初始能量	20 000

表 3 PPO-GRP 的网络参数

参数	含义	值
c_1	评估网络目标函数所占权重值	-0.5
c_2	策略模型的熵所占权重值	-0.01
σ_{\max}	关于动作值概率分布的标准差的最大值	0.6
σ_{\min}	关于动作值概率分布的标准差的最小值	0.1
α_{σ}	关于动作值概率分布标准差的衰减因子	0.999 5
Episodes	仿真回合数	500
Buffer-Size	缓冲区 D 的大小	2 000
n_{update}	网络连续更新次数	16
ϵ	PPO 的裁剪参数	0.1
γ	计算奖励期望的折扣系数	0.95
策略网络学习率		0.000 03
评估网络学习率		0.000 1

3.2 性能评估指标

用于评估 IRPJTC 算法性能的性能指标详细定义如下。

1) 平均最大、最小邻居距离。在仿真初始阶段, 无人机随机分布在任务区域, 各无人机与邻居的最大、最小距离是随机的。在拓扑构建过程中, IRPJTC 采用 VFATC 对各无人机进行移动控制, 以满足与邻居间的最大、最小距离约束。因此, 通过观察平均最大、最小邻居距离可有效评估 IRPJTC 通过约束邻居距离保证链路稳定的能力。

2) 分组传输成功率。分组传输成功率是指成功发送至地面站的数据包数量与源节点产生的数据包总数之比。分组传输成功率直观地体现了 IRPJTC 的路由性能。

3) 平均端到端时延。端到端时延指数据包从源无人机发送至地面站过程中所需要消耗的时间。端到端时延越低, 则代表此路由策略在下一跳节点的选取过程中越倾向于选择链路时延低的链路, 且实现了更少的转发跳数。

4) 归一化剩余能量。归一化剩余能量为无人机剩余能量与总能量的比值, 其直接反映了无人机集群的能量消耗大小, 各无人机的归一化剩余能量能评估 IRPJTC 是否能降低能耗, 且实现能耗均衡。

3.3 仿真结果分析

图 4 表示由 50 个无人机组成的集群拓扑初始状态。在初始状态, 各无人机随机分布在任务区域内, 未形成集群。此时 IRPJTC 启动 VFATC, 在拓扑构

建过程中, 无人机集群向地面站靠近, 并相互建立连接, 经过时长为 200 s 的拓扑构建后, 无人机集群拓扑状态如图 5 所示。从图 5 可以看出, 此时各无人机均与相邻无人机建立通信链路, 且集群中靠近地面站的边缘无人机也与地面站建立通信链路。

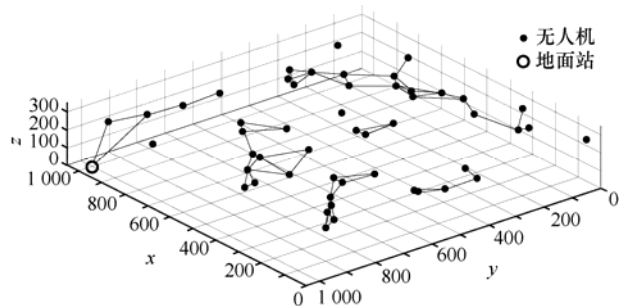


图 4 拓扑初始状态

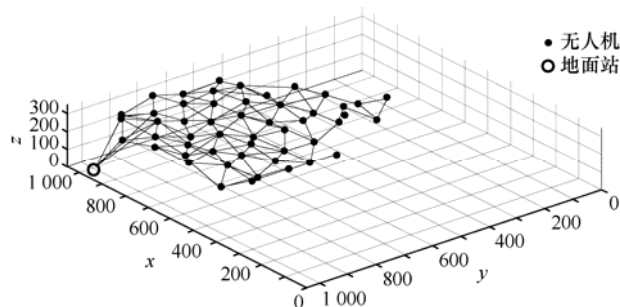


图 5 拓扑构建后的无人机集群拓扑状态

在 IRPJTC 进行拓扑构建过程中平均邻居距离的变化趋势如图 6 所示。由图 6 可以看出, 在初始状态, 平均最小邻居距离为 74 m, 小于预定义的最小距离 100 m。拓扑构建过程中, 最小平均邻居距离收敛至 95 m 以上, 同时最大平均邻居距离收敛至 150 m 以下, 证明 IRPJTC 能够自适应调节虚拟力权值, 使各无人机与邻居满足预设的距离约束。

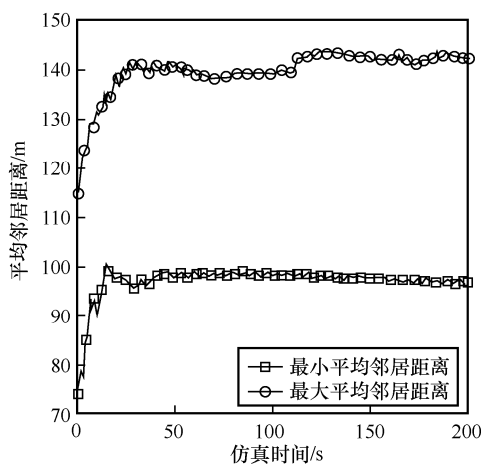


图 6 拓扑构建过程中平均邻居距离的变化趋势

在不同无人机数量的拓扑环境下，采用 PPO-GRP 进行训练的过程中，IRPJTC 每回合累计奖励如图 7 所示。由图 7 可以看出，随着迭代次数的增加，累计奖励逐渐收敛。无人机数量越少，累计奖励的收敛值越高，这是由于每个回合中的时隙数量为固定值，当无人机数量越少、拓扑结构越简单时，数据包路由至地面站所需的跳数越少，单个训练回合内传输成功的数据包数量越多，故累计奖励越高。

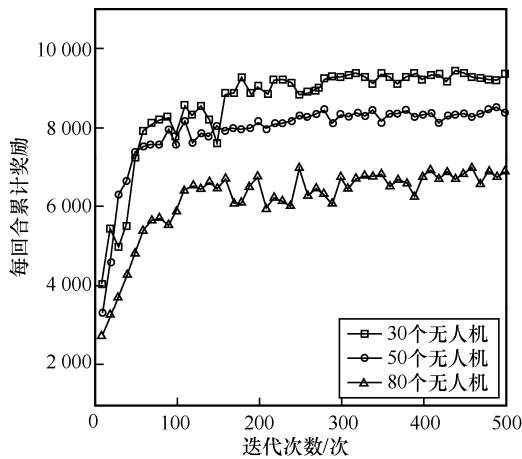


图 7 IRPJTC 每回合累计奖励

IRPJTC 每回合平均分组传输成功率如图 8 所示。随着训练回合的增加，分组传输成功率逐渐收敛。在 30 个无人机的拓扑下，分组传输成功率收敛至 95%左右；在 50 个无人机的拓扑下，分组传输成功率收敛至 92%左右；在 80 个无人机的拓扑下，分组传输成功率收敛至 88%左右。

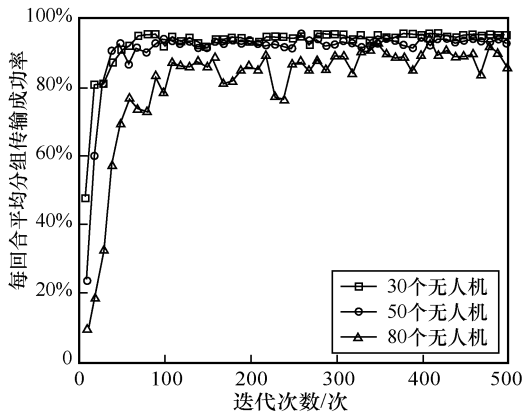


图 8 IRPJTC 每回合平均分组传输成功率

为了全方位评估本文所提 IRPJTC 的性能，本文将 IRPJTC 与 GPSR^[10]、基于 Q 学习的地理路由 (QGeo)^[11]以及基于深度强化学习的地理数据包路

由优化 (DRL-GPR) 方法^[15]进行性能对比，为了保证客观性，纳入对比的路由方法均采用 VFATC 进行拓扑构建与调整。对比指标包括分组传输成功率、端到端时延以及归一化剩余能耗。

在不同拓扑规模下，各路由方法数据分组传输成功率如图 9 所示。IRPJTC 分组传输成功率与 QGeo 基本一致，且相比于 GPSR 和 DRL-GPR 分别提升了 5.84%和 1.34%。GPSR 在选取下一跳节点时采取贪婪策略，倾向于选择距离目的节点更近的下一跳，缺乏对链路稳定性的考虑，且在遇到路由空洞时，会采取周长转发策略，导致其需要更多的转发次数才能到达目的节点，故其分组传输成功率显著降低。DRL-GPR 在奖励函数中，针对路由空洞问题，设计了奖惩机制，故其分组传输成功率显著高于 GPSR；但其未考虑链路稳定性指标，导致分组传输成功率略低于 IRPJTC。

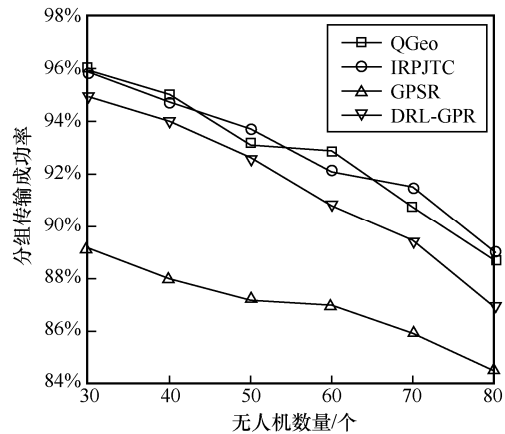


图 9 各路由方法数据分组传输成功率

在不同拓扑规模下，各路由方法端到端时延如图 10 所示。IRPJTC 的端到端时延相比于 QGeo、GPSR 和 DRL-GPR 分别降低了 12.11%、28.43%和 16.71%。这是由于 IRPJTC 与 QGeo 均考虑包传输速率 PTS 作为下一跳链路的选取指标，但 IRPJTC 额外考虑了链路稳定性指标 LD，减轻了链路中断引发的包重传问题，有效降低了端到端时延。GPSR 遇到路由空洞后会采用周长转发方式绕过空洞区域，导致数据包传输所需跳数显著增加，在集群拓扑规模增大时，会显著增加端到端时延。DRL-GPR 对路由空洞问题针对性设计了奖励函数，故端到端时延优于 GPSR。但未考虑链路稳定性指标 LD 以及包传输速率 PTS，故端到端时延弱于 IRPJTC 与 QGeo。

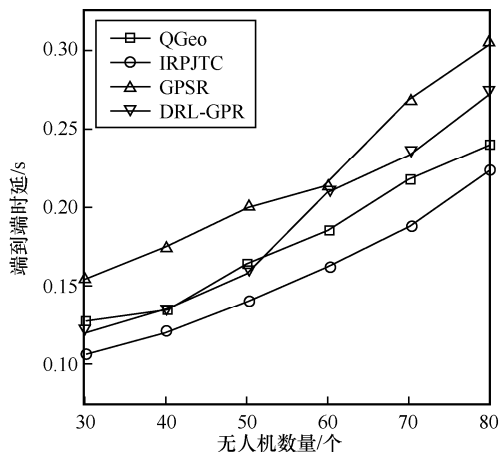


图 10 各路由方法端到端时延

在 50 个无人机组成的拓扑中, 各路由方法的归一化剩余能量如图 11 所示。用箱型图来表示当前集群中无人机的最高归一化剩余能量、平均归一化剩余能量与最低归一化剩余能量。IRPJTC 的平均归一化剩余能量高于其他路由方法, 相比于 QGeo、GPSR 和 DRL-GPR, 分别提升了 4.56%、10.42% 和 5.16%。且其最大、最小归一化剩余能量也更接近平均值, 这表示 IRPJTC 相比于其他路由方法, 具有更强的能耗均衡能力。

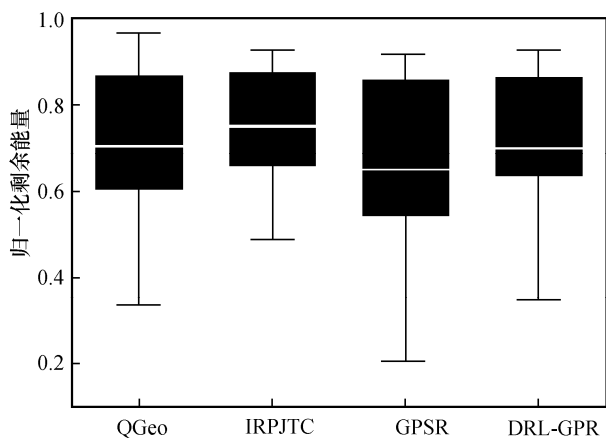


图 11 各路由方法的归一化剩余能量

4 结束语

本文面向无人机集群区域监视场景, 提出了一种联合拓扑控制的智能路由规划方法 IRPJTC。IRPJTC 由基于虚拟力的自适应拓扑控制 VFATC 与基于近端策略优化的地理路由规划 PPO-GRP 两部分组成。VFATC 可实现拓扑构建与自适应调整, 为数据路由提供稳定通信环境。在此基础上, PPO-GRP 通过引入 VFATC 中的链路稳定性指标,

并结合端到端时延以及能耗指标, 针对性设计奖励函数, 提升了区域监视场景下无人机集群的数据路由性能。仿真实验结果表明, IRPJTC 能保证高数据分组传输成功率的同时, 在端到端时延以及能耗方面优于现有路由方法, 有利于区域监视场景下的大规模无人机集群数据路由性能的提升。

参考文献:

- [1] CHENG X, DONG C, DAI H P, et al. MOOC: a mobility control based clustering scheme for area coverage in FANETs[C]//Proceedings of the 2018 IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks". Piscataway: IEEE Press, 2018: 14-22.
- [2] LIN Z J, LIU H H T, WOTTON M. Kalman filter-based large-scale wildfire monitoring with a system of UAVs[J]. IEEE Transactions on Industrial Electronics, 2019, 66(1): 606-615.
- [3] WANG H J, ZHAO H T, WU W Y, et al. Deployment algorithms of flying base stations: 5G and beyond with UAVs[J]. IEEE Internet of Things Journal, 2019, 6(6): 10009-10027.
- [4] 许文俊, 吴思雷, 王凤玉, 等. 基于多智能体强化学习的大规模灾后用户分布式覆盖优化[J]. 通信学报, 2022, 43(8): 1-16.
- [5] XU W J, WU S L, WANG F Y, et al. Large-scale post-disaster user distributed coverage optimization based on multi-agent reinforcement learning[J]. Journal on Communications, 2022, 43(8): 1-16.
- [6] 余雪勇, 邱礼翔, 宋家宁, 等. 无人机辅助边缘计算中安全通信与能效优化策略[J]. 通信学报, 2023, 44(3): 45-54.
- [7] YU X Y, QIU L X, SONG J N, et al. Security communication and energy efficiency optimization strategy in UAV-aided edge computing[J]. Journal on Communications, 2023, 44(3): 45-54.
- [8] ALAM M M, ARAFAT M Y, MOH S, et al. Topology control algorithms in multi-unmanned aerial vehicle networks: an extensive survey[J]. Journal of Network and Computer Applications, 2022, 207: 103495.
- [9] KIM D Y, LEE J W. Joint mission assignment and topology management in the mission-critical FANET[J]. IEEE Internet of Things Journal, 2020, 7(3): 2368-2385.
- [10] TROTTA A, MONTECCHIARI L, FELICE M D, et al. A GPS-free flocking model for aerial mesh deployments in disaster-recovery scenarios[J]. IEEE Access, 2020, 8: 91558-91573.
- [11] ZHAO H T, WEI J B, HUANG S C, et al. Regular topology formation based on artificial forces for distributed mobile robotic networks[J]. IEEE Transactions on Mobile Computing, 2019, 18(10): 2415-2429.
- [12] KARP B, KUNG H T. GPSR: greedy perimeter stateless routing for wireless networks[C]//Proceedings of the 6th Annual International Conference on Mobile Computing and Networking. New York: ACM Press, 2000: 243-254.
- [13] JUNG W S, YIM J, KO Y B. QGeo: Q-learning-based geographic ad hoc routing protocol for unmanned robotic networks[J]. IEEE Communications Letters, 2017, 21(10): 2258-2261.

- [12] LIU J M, WANG Q, HE C T, et al. QMR: Q-learning based multi-objective optimization routing protocol for flying ad hoc networks[J]. *Computer Communications*, 2020, 150: 304-316.
- [13] 张雅楠, 仇洪冰. 基于深度强化学习的无人机可信地理位置路由协议[J]. *电子与信息学报*, 2022, 44(12): 4211-4217.
ZHANG Y N, QIU H B. Trusted geographic routing protocol based on deep reinforcement learning for unmanned aerial vehicle network[J]. *Journal of Electronics & Information Technology*, 2022, 44(12): 4211-4217.
- [14] LIN D P, PENG T, ZUO P L, et al. Deep-reinforcement-learning-based intelligent routing strategy for FANETs[J]. *Symmetry*, 2022, 14(9): 1787.
- [15] BAI Y J, ZHANG X, YU D J, et al. A deep reinforcement learning-based geographic packet routing optimization[J]. *IEEE Access*, 2022, 10: 108785-108796.
- [16] 陈辉. 无线 AdHoc 路由算法和拓扑控制算法研究[D]. 西安: 长安大学, 2014.
CHEN H. Research on wireless AdHoc routing algorithm and topology control algorithm[D]. Xi'an: Changan University, 2014.
- [17] 刘航. 资源受限卫星网络拓扑与路由规划方法研究[D]. 西安: 西安电子科技大学, 2022.
LIU H. Research on topology and routing planning of resource-constrained satellite networks[D]. Xi'an: Xidian University, 2022.
- [18] WU Y L, ZHANG B, YANG S S, et al. Energy-efficient joint communication-motion planning for relay-assisted wireless robot surveillance[C]//*Proceedings of the IEEE Conference on Computer Communications*. Piscataway: IEEE Press, 2017: 1-9.
- [19] LIU Q W, ZHOU S L, GIANNAKIS G B. Cross-layer combining of queuing with adaptive modulation and coding over wireless links[C]//*Proceedings of the IEEE Military Communications Conference*. Piscataway: IEEE Press, 2004: 717-722.

[作者简介]



颜志 (1986-), 男, 湖南衡东人, 博士, 湖南大学副教授、博士生导师, 主要研究方向为机器人通信与组网。



易正伦 (1998-), 男, 广东惠州人, 湖南大学硕士生, 主要研究方向为无人机集群通信与组网。



欧阳博 (1987-), 男, 湖南耒阳人, 博士, 湖南大学副教授、博士生导师, 主要研究方向为机器学习、多机器人系统、复杂系统分析与控制、无线通信技术。



王耀南 (1957-), 男, 云南龙陵人, 博士, 湖南大学教授、博士生导师, 主要研究方向为机器人感知与控制技术。