深度强化学习驱动的低空安全通信：感通一体化设计

魏志强[①] 张家烁[①] 刘 凡[②] 杨 在[*①] 费泽松[③]

[①](西安交通大学 西安 710049)

[②](东南大学 南京 211189)

[③](北京理工大学 北京 100081)

摘要：针对低空无人机通信中的物理层安全挑战，该文提出了一种感通一体化(ISAC)方案，并据此基于深度强化学习(DRL)方法在线优化通信无人机的航迹和通信资源分配策略。所提方案通过复用通信无人机传输的人工噪声，同时实现对窃听无人机的感知与干扰，保障地面用户的安全通信服务。基于对窃听无人机的状态估计和预测，该文将在线无人机航迹和通信资源分配联合设计建模为马尔可夫决策过程，基于深度确定性策略梯度(DDPG)方法，逐步学习最优策略，动态优化通信无人机的航迹与通信资源分配策略，最大化系统的长期感知和安全通信性能。仿真结果表明，该文所提方案和优化方法在感知性能不损失的前提下，安全通信性能上优于基线方案，在感知和安全通信性能之间实现更好的折中，验证了感知和在线航迹规划的增益，也验证了深度强化学习优化方法在感知、通信和航迹规划联合设计问题中的可行性和先进性。

关键词：感通一体化；在线航迹优化；无人机通信；物理层安全；深度确定性策略梯度

中图分类号：TN958 **文献标识码**：A **文章编号**：2095-283X(2025)04-1005-14

DOI: 10.12000/JR25025 **CSTR**: 32380.14.JR25025

引用格式：魏志强, 张家烁, 刘凡, 等. 深度强化学习驱动的低空安全通信：感通一体化设计[J]. 雷达学报(中英文), 2025, 14(4): 1005–1018. doi: 10.12000/JR25025.

Reference format: WEI Zhiqiang, ZHANG Jiashuo, LIU Fan, *et al*. Low-altitude secure communication driven by deep reinforcement learning: An integrated sensing and communication design[J]. *Journal of Radars*, 2025, 14(4): 1005–1018. doi: 10.12000/JR25025.

Low-altitude Secure Communication Driven by Deep Reinforcement Learning: An Integrated Sensing and Communication Design

WEI Zhiqiang[①] ZHANG Jiashuo[①] LIU Fan[②] YANG Zai[*①] FEI Zesong[③]

[①](*Xi'an Jiaotong University*, *Xi'an* 710049, *China*)

[②](*Southeast University*, *Nanjing* 211189, *China*)

[③](*Beijing Institute of Technology*, *Beijing* 100081, *China*)

Abstract: To address the physical layer security challenges in low-altitude Unmanned Aerial Vehicle (UAV) communications, this paper proposes an Integrated Sensing And Communication (ISAC) scheme. For the proposed ISAC scheme, an online optimization framework for UAV trajectory and communication resource allocation is developed using Deep Reinforcement Learning (DRL). In the proposed scheme, artificial noise

收稿日期：2025-02-04；改回日期：2025-06-30；网络出版：2025-07-09

*通信作者： 杨在 yangzai@xjtu.edu.cn *Corresponding Author: YANG Zai, yangzai@xjtu.edu.cn

基金项目：国家自然科学基金(12371464, 62331023, 62471039)，陕西省秦创原引用高层次创新创业人才项目(QCYRCXM-2023-094)，广东省基础与应用基础研究面上项目(2024A1515011218)

Foundation Items: The National Natural Science Foundation of China (12371464, 62331023, 62471039), Qin Chuang Yuan High-Level Innovation and Entrepreneurship Talent Program (QCYRCXM-2023-094), Guangdong Basic and Applied Basic Research Foundation (2024A1515011218)

责任主编：张海君 Corresponding Editor: ZHANG Haijun

©The Author(s) 2025. This is an open access article under the CC-BY 4.0 License (https://creativecommons.org/licenses/by/4.0/)

transmitted by a communication UAV is reused to simultaneously sense and jam a potential eavesdropping UAV, thereby enhancing secure communications for ground users. By estimating and predicting the state of the eavesdropping UAV, the trajectory and resource allocation design problem is reformulated as a Markov decision process. Using the Deep Deterministic Policy Gradient (DDPG) algorithm, the optimal framework is learned over time, dynamically optimizing the communication UAV's trajectory and resource allocation to maximize long-term sensing and secure communication performance. Simulation results demonstrate that the proposed scheme achieves a superior trade-off between sensing and security without degrading sensing performance and outperforms baseline methods in terms of secure communication performance. This validates the performance gains achieved through sensing and online trajectory design, as well as the potential and superior performance of applying DRL to the integrated design of sensing, communication, and trajectory.

Key words: Integrated Sensing And Communications (ISAC); Online trajectory optimization; Unmanned Aerial Vehicle (UAV) communications; Physical layer security; Deep Deterministic Policy Gradient (DDPG)

1 引言

低空经济是指利用低空空域，以无人机、直升机等低空飞行器为工具开展的一系列经济活动。无人机通信利用无人驾驶航空器(Unmanned Aerial Vehicle, UAV)体积小、操作灵活、适应复杂环境等特性，可独立为特定区域(如农村、灾区或热点区域)提供通信支持，也可与地面通信设施互补。尤其在与现有蜂窝网络协同设计时，无人机通信能够显著提升网络覆盖范围与服务质量，为构建综合性的现代化通信网络提供重要支撑。

无人机通信系统的安全性正面临日益严峻的挑战。首先，以广阔空域视距传播为主的无人机通信信道特性显著增加了信息传输过程中被窃听的风险。其次，当网络中部署具有高机动性的窃听无人机时，其灵活的移动能力进一步增大了信息被窃取的可能性，而窃听信道的高度不确定性导致传统物理层安全(Physical Layer Security)技术[1,2]的性能显著下降甚至失效。针对这一问题，现有研究[3–5]通过结合合法链路与窃听链路的信道特性，提出了基于无人机轨迹与功率分配的离线联合优化策略，然而这些方法通常假设发射端能够获取完美窃听信道状态信息，在实际应用场景中往往难以满足。尽管一些鲁棒设计方法[6,7]可以应对窃听信道的不确定性，对于高机动窃听无人机而言，窃听信道快速时变，传统方法难以适用。人工噪声(Artificial Noise)干扰技术[2]是指发射端在与合法用户通信的同时，发射人工噪声干扰信号以防止信息泄露，但其干扰效果同样依赖于窃听信道的信道状态信息获取质量。因此，窃听信道状态信息的获取是无人机安全通信的瓶颈问题。

感知与通信一体化技术(Integrated Sensing And Communications, ISAC)为窃听信道状态信息的获取提供了新的技术途径。该技术通过共享频谱与硬件资源，利用电磁波复用同时实现感知与通信双重功能，不仅显著提升了资源利用效率，还实现了感知与通信之间的深度协同，近年来已成为学术界和工业界广泛关注的研究热点[8,9]。文献[10–12]初步探索了感知辅助的物理层安全，但尚未充分考虑机动窃听者的动态特性，不适用于无人机窃听场景。文献[13]提出了一种基于贝叶斯学习的低复杂度多用户感知和通信一体化算法，显著提升了毫米波通信应用场景中的时频资源利用效率。

本研究将感通一体化理念引入低空无人机通信领域，提出了一种基于人工噪声复用的协同感知与干扰机制。该机制使通信无人机在发射人工噪声干扰窃听无人机的同时，能够利用后向散射回波实现对窃听无人机状态的实时感知。通过在广域空域范围内对窃听无人机的状态进行实时感知，显著提高了对窃听无人机信道的估计精度，从而为通信无人机后续的航迹规划和通信资源分配设计提供依据。这一创新方法不仅提高了系统的安全性，还为低空无人机通信网络的智能化管理提供了新的技术路径。

无人机的高机动性为提升低空无人机通信系统的传输速率和安全性能提供了额外的优化自由度。具体而言，通信无人机可基于实时感知与预测的窃听无人机状态信息，联合优化其飞行轨迹与通信资源分配策略，实现窃听无人机感知性能与地面用户安全通信性能的协同提升，从而保障对窃听无人机的持续感知和对地面用户的持续安全通信服务。然而，在线联合优化无人机航迹与通信资源分配是一个具有挑战性的问题，长期以来一直是无人机通信领域的研究难点[14,15]。首先，无人机在线航迹规划是一个典型的动态规划(Dynamic Programming)问题，当前位置的优化结果将作为后续无人机航迹设计的起点，影响后续感知和安全通信性能。其次，当前的通信资源分配决定了下一时刻的通信信号和

人工噪声信号，进而影响下一时刻的窃听无人机感知性能。最后，传统无人机航迹规划和通信资源优化的耦合关系是单向的，即航迹规划会通过信道的变化影响通信资源分配的设计，反之则不会有影响。然而，感知的引入使得两个问题的耦合是双向的，即不仅航迹规划会影响通信资源分配策略，通信资源分配策略决定在线感知性能，进而间接影响航迹规划的结果。这种双向耦合使得系统的感知和安全通信性能难以分析，在线联合无人机航迹和通信资源分配优化更具挑战。作者的前期工作[16]，将航迹规划与通信资源分配解耦为两个独立的问题进行求解，且航迹规划是一种基于距离的启发式航迹规划策略，往往无法兼顾长期的感知精度与安全通信性能。

深度强化学习(Deep Reinforcement Learning, DRL)是一种结合深度学习与强化学习的智能优化方法，能够通过与环境的交互学习最优策略，尤其适用于解决复杂动态系统中的决策问题。经典的无模型DRL方法包括深度Q网络(Deep Q-Network)[17]、深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)[18]、近端策略优化(Proximal Policy Optimization)[19]和软演员-评论家算法(Soft Actor-Critic)[20]等。深度强化学习优化方法被广泛用于解决轨迹规划、任务分配和通信资源管理等复杂决策问题。文献[21]针对智能反射面辅助的复杂通信环境，提出了一种基于深度强化学习的联合优化方法，通过协同设计基站端的数字波束成形矩阵和智能反射面端的模拟波束成形矩阵，显著扩大了通信覆盖范围，同时有效降低了信号传输损耗。文献[22]利用深度强化学习优化无人机网络中的资源分配方案，同时提升无人机的感知与通信性能。文献[23]则针对智能反射面辅助的毫米波网络，基于深度强化学习设计了一种基站与用户关联及波束成形策略。针对无人机安全通信，文献[24]通过深度强化学习优化无人机的轨迹和资源分配，增强了毫米波通信网络的物理层安全性。但是，上述工作没有利用感知手段辅助窃听信道获取，同时没有考虑感知与安全通信之间的折中优化问题。本文利用深度强化学习高效的建模和学习能力，在通信无人机与环境的决策交互过程中，逐步学习出最大化系统累积长期回报(感知与安全通信性能的权衡)的策略函数，为复杂的在线联合无人机航迹和通信资源分配优化问题提供高质量可行解。

综上，本文针对低空无人机通信中的物理层安全挑战，提出了一种感通一体化框架，该框架复用人工噪声感知和干扰窃听无人机，并提出了一种基于深度强化学习的通信无人机航迹与通信资源分配在线优化方法，为保障低空安全通信提供一种可行的技术路线。主要贡献如下：

(1) 本文构建了一个如图1所示的感通一体化问题框架。与作者前期工作[16]中提出的基于短期回报最大化的感通一体化设计问题不同，本文创新性地提出了一个基于长期回报最大化的感通一体化设计数学模型，联合优化无人机航迹和通信预编码策略，在感知性能约束下最大化系统的累积安全通信性能。所提方案突破了传统航迹设计-通信策略解耦优化方法的局限性，为低空无人机通信系统的性能优化提供了新的研究思路。

(2) 基于所提出的ISAC问题框架，本文将无人机航迹和通信资源分配联合优化问题建模为马尔可夫决策过程(Markov Decision Process)。然后，本文利用深度确定性策略梯度方法求解该问题，在智能体与环境的动态交互过程中，利用深度神经网络分别学习系统的动作价值函数和动作策略函数，在感知性能约束下最大化系统的累积安全通信性能，为复杂的在线联合无人机航迹和通信资源分配联合设计问题提供高质量可行解。

(3) 本文所提出的ISAC方案和优化方法能够在感知和通信性能之间实现更好的权衡。仿真结果表明，过度追求感知精度反而会损失系统的安全通信性能，所提出的方案和优化方法在感知性能不损失的前提下，安全通信性能上优于基线方案，验证了深度强化学习优化方法在感通一体化设计问题中的可行性和先进性。

2　系统模型及所提出的方案

2.1　系统模型

我们考虑一个下行无人机通信系统，假设环境中存在K个单天线地面合法用户，通信无人机向K个用户广播数据流，同时向窃听无人机发射人工

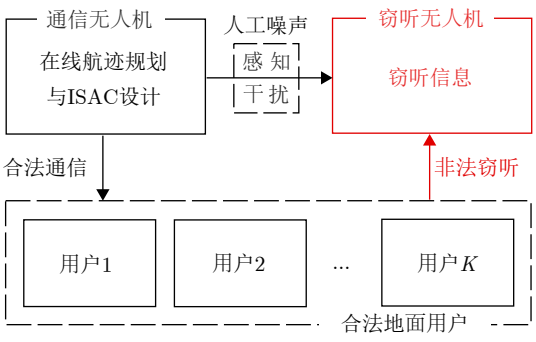

图 1 所提ISAC问题框架

Fig. 1 The proposed ISAC problem framework

噪声以干扰窃听无人机对合法信息的窃听，并利用人工噪声信号的回波感知窃听无人机的位置及状态。通信无人机配备有两个均匀平面阵列(Uniform Planar Arrays, UPA)，分别由$M_{\mathrm{b}}^{\mathrm{t}}$($M_{\mathrm{b}}^{\mathrm{t}x}$行和$M_{\mathrm{b}}^{\mathrm{t}y}$列)与$M_{\mathrm{b}}^{\mathrm{r}}$($M_{\mathrm{b}}^{\mathrm{r}x}$行和$M_{\mathrm{b}}^{\mathrm{r}y}$列)根天线组成，用于三维信号发射和接收波束成形。窃听无人机亦配备了一个用于接收的均匀平面阵列，由M_{e}(M_{e}^{x}行和M_{e}^{y}列)根天线构成，如图2所示。

将总服务时间T分为N个等长时隙，每个时隙的长度为δ，即$T=N\delta$。在第n个时隙，通信无人机通过其自身的惯性导航系统(Inertial Navigation System)获得自身位置与速度，通信无人机的位置记为$\boldsymbol{q}_{\mathrm{b}}[n]=[x_{\mathrm{b}}[n],y_{\mathrm{b}}[n],z_{\mathrm{b}}[n]]^{\mathrm{T}}$，速度记为$\dot{\boldsymbol{q}}_{\mathrm{b}}[n]=[\dot{x}_{\mathrm{b}}[n],\dot{y}_{\mathrm{b}}[n],\dot{z}_{\mathrm{b}}[n]]^{\mathrm{T}}$，窃听无人机的位置和速度分别表示为$\boldsymbol{q}_{\mathrm{e}}[n]=[x_{\mathrm{e}}[n],y_{\mathrm{e}}[n],z_{\mathrm{e}}[n]]^{\mathrm{T}}$与$\dot{\boldsymbol{q}}_{\mathrm{e}}[n]=[\dot{x}_{\mathrm{e}}[n],\dot{y}_{\mathrm{e}}[n],\dot{z}_{\mathrm{e}}[n]]^{\mathrm{T}}$。本文假设窃听无人机沿着既定的轨道巡航。为便于后续表述，将通信无人机与窃听无人机的状态分别定义为$\boldsymbol{a}_{\mathrm{b}}[n]=[\boldsymbol{q}_{\mathrm{b}}^{\mathrm{T}}[n],\dot{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{T}}[n]]^{\mathrm{T}}$与$\boldsymbol{a}_{\mathrm{e}}[n]=[\boldsymbol{q}_{\mathrm{e}}^{\mathrm{T}}[n],\dot{\boldsymbol{q}}_{\mathrm{e}}^{\mathrm{T}}[n]]^{\mathrm{T}}$。假定所有地面用户的位置都是固定的，他们的位置记为$\boldsymbol{q}_k=[x_k,y_k,0]^{\mathrm{T}}$，$\forall k\in\{1,2,\cdots,K\}$。依据文献[6,25,26]，本文假定通信无人机与地面用户的位置完全已知，并通过独立的反馈链路在节点间共享。但是，窃听无人机的状态，即$\boldsymbol{a}_{\mathrm{e}}[n]$，是未知的，在本文中将由通信无人机在线感知。相反，通信无人机与地面用户的位置与状态对于窃听无人机是完全已知的，这是安全通信问题背景下的最差情况。

本文所提感通一体化方案的时序结构如图3所示，通信无人机的在线航迹优化通过设计各时隙的加速度实现。在第n时隙开始时，通信无人机首先根据窃听无人机在第$n-1$时刻的状态估计来预测其在第n时刻的状态，即$\hat{\alpha}_{\mathrm{e}}[n|\,n-1]$，然后根据该预测状态设计第$n$时隙的加速度与通信资源分配策略，并发射机密信息流与人工噪声。在接收到人工噪声的回波后，提取第n时刻的测量值$\hat{\beta}_{\mathrm{e}}[n]$，并更新窃听无人机第$n$时刻的状态估计$\hat{\alpha}_{\mathrm{e}}[n]$，其中$\hat{\beta}_{\mathrm{e}}[n]$代表通信无人机处可观测参数的集合，具体含义将在后续部分中说明。本文所考虑的无人机通信系统存在固有的因果关系，第n时隙的加速度与通信资源分配方案设计取决于第$n-1$时隙的状态预测，同时影响后续时隙的感知性能、航迹优化策略和资源分配性能，因此无人机航迹和通信资源分配联合设计必须是在线的，且需要考虑系统的长期回报。

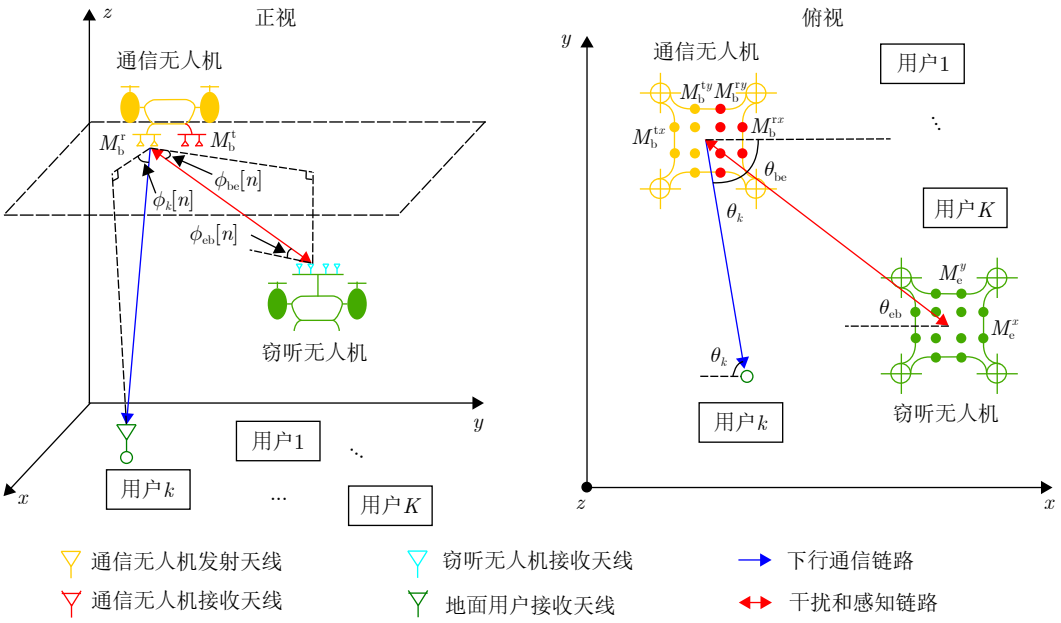

图 2 系统的视距信道模型

Fig. 2 LoS channel model of the considered system

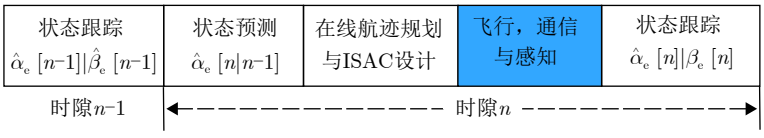

图 3 面向低空安全通信的ISAC框架

Fig. 3 The proposed ISAC framework for low-altitude secure communications

2.2 通信模型

在第n时隙中，通信无人机的下行发射信号由式(1)给出：

$$\boldsymbol{x}\left(n,t\right)=\sum_{k=1}^{K}\boldsymbol{w}_k\left[n\right]s_k(n,t)+\boldsymbol{w}_{\mathrm{e}}\left[n\right]a_{\mathrm{e}}\left(n,t\right)\tag{1}$$

其中，$t\in(0,\delta)$是时隙n内的时刻，$s_k(n,t)$表示通信无人机向第k个用户传输的符号信息，$s_k(n,t)\sim\mathcal{CN}\left(0,1\right)$，$a_{\mathrm{e}}(n,t)$表示通信无人机向窃听无人机发射的人工噪声，$a_{\mathrm{e}}(n,t)\sim\mathcal{CN}\left(0,1\right)$。向量$\boldsymbol{w}_k\left[n\right]\in\mathbb{C}^{M_{\mathrm{b}}^{\mathrm{t}}\times1}$和$\boldsymbol{w}_{\mathrm{e}}\left[n\right]\in\mathbb{C}^{M_{\mathrm{b}}^{\mathrm{t}}\times1}$分别表示通信无人机对第$k$个地面用户的通信预编码和对窃听无人机的人工噪声预编码向量。

假设完美时频同步以补偿所有地面用户的时间延迟和多普勒频移，第k个地面用户在时隙n中接收到的信号如式(2)所示：

$$\begin{aligned}\tilde{y}_k\left(n,t\right)=&\boldsymbol{h}_k^{\mathrm{H}}\left[n\right]\boldsymbol{w}_k\left[n\right]s_k\left(n,t\right)\\&+\sum_{k'\neq k}\boldsymbol{h}_k^{\mathrm{H}}\left[n\right]\boldsymbol{w}_{k'}\left[n\right]s_{k'}\left(n,t\right)\\&+\boldsymbol{h}_k^{\mathrm{H}}\left[n\right]\boldsymbol{w}_{\mathrm{e}}\left[n\right]a_{\mathrm{e}}\left(n,t\right)+\tilde{v}_k\left(n,t\right)\end{aligned}\tag{2}$$

其中，$v_k\left(n,t\right)\sim\mathcal{CN}\left(0,\sigma_k^2\right)$为第$k$个地面用户的加性高斯白噪声(Additive White Gaussian Noise, AWGN)，方差为σ_k^2，该噪声不受时频同步操作的影响[27]。时隙n中通信无人机和用户k之间的信道向量为

$$\boldsymbol{h}_k\left[n\right]=\frac{\beta_0}{d_k\left[n\right]}\boldsymbol{a}_{M_{\mathrm{b}}^{\mathrm{t}x},M_{\mathrm{b}}^{\mathrm{t}y}}\left(\theta_k\left[n\right],\phi_k\left[n\right]\right)\tag{3}$$

其中，β_0^2表示在参考距离下的信道功率增益，$d_k\left[n\right]=\left\|q_{\mathrm{b}}\left[n\right]-q_k\left[n\right]\right\|$表示通信无人机与第$k$个地面用户之间的距离，$\theta_k\left[n\right]\in[-\pi/2,\pi/2]$和$\phi_k\left[n\right]\in[-\pi/2,0]$分别表示通信无人机到第$k$个地面用户的方位角和俯仰角，$\boldsymbol{a}_{M^x,M^y}(\theta,\phi)\in\mathbb{C}^{M^xM^y\times1}$表示大小为$M^x\times M^y$的UPA导向向量，由式(4)给出：

$$\begin{aligned}&\boldsymbol{a}_{M^x,M^y}\left(\theta,\phi\right)\\&=\left(1,\mathrm{e}^{-\frac{\mathrm{j}2\pi d\cos\phi\sin\theta}{\lambda_{\mathrm{c}}}},\cdots,\mathrm{e}^{-\frac{\mathrm{j}2(M^x-1)\pi d\cos\phi\sin\theta}{\lambda_{\mathrm{c}}}}\right)^{\mathrm{T}}\\&\otimes\left(1,\mathrm{e}^{-\frac{\mathrm{j}2\pi d\cos\phi\cos\theta}{\lambda_{\mathrm{c}}}},\cdots,\mathrm{e}^{-\frac{\mathrm{j}2(M^y-1)\pi d\cos\phi\cos\theta}{\lambda_{\mathrm{c}}}}\right)^{\mathrm{T}}\end{aligned}\tag{4}$$

除了AWGN外，第k个地面用户的接收信号还会受到用户间干扰与人工噪声的干扰，即式(2)的第2项与第3项。因此，用户k在时隙n中的可达通信速率由式(5)给出：

$$R_k\left[n\right]=\log_2\left(1+\frac{\left|\boldsymbol{h}_k^{\mathrm{H}}\left[n\right]\boldsymbol{w}_k\left[n\right]\right|^2}{\sum\limits_{k'\neq k}\left|\boldsymbol{h}_k^{\mathrm{H}}\left[n\right]\boldsymbol{w}_{k'}\left[n\right]\right|^2+\left|\boldsymbol{h}_k^{\mathrm{H}}\left[n\right]\boldsymbol{w}_{\mathrm{e}}\left[n\right]\right|^2+\sigma_k^2}\right)\tag{5}$$

本文假设窃听无人机执行最大比合并(Maximum Ratio Combining)[16]以最大化接收信号的功率。假设窃听无人机同样实现完美的时频同步，其在时隙n中的接收波束成形输出信号为

$$\begin{aligned}y_{\mathrm{e}}\left(n,t\right)=\boldsymbol{u}_{\mathrm{e}}^{\mathrm{H}}[n]\,\boldsymbol{y}_{\mathrm{e}}=&\sum_{k=1}^{K}\sqrt{M_{\mathrm{e}}}\boldsymbol{h}_{\mathrm{be}}^{\mathrm{H}}\left[n\right]\boldsymbol{w}_k\left[n\right]s_k\left(n,t\right)\\&+\sqrt{M_{\mathrm{e}}}\boldsymbol{h}_{\mathrm{be}}^{\mathrm{H}}\left[n\right]\boldsymbol{w}_{\mathrm{e}}\left[n\right]a_{\mathrm{e}}\left(n,t\right)+v_{\mathrm{e}}\left(n,t\right)\end{aligned}\tag{6}$$

其中，第1项为下行机密信息，第2项为人工噪声，第3项$v_{\mathrm{e}}(n,t)\sim\mathcal{CN}\left(0,\sigma_{\mathrm{e}}^2\right)$表示经过最大比组合后窃听无人机处的背景噪声。通信无人机与窃听无人机之间的等效窃听信道$\boldsymbol{h}_{\mathrm{be}}\left[n\right]$由式(7)表示：

$$\boldsymbol{h}_{\mathrm{be}}\left[n\right]=\frac{\beta_0}{d_{\mathrm{e}}\left[n\right]}\boldsymbol{a}_{M_{\mathrm{b}}^{\mathrm{t}x},M_{\mathrm{b}}^{\mathrm{t}y}}\left(\theta_{\mathrm{be}}\left[n\right],\phi_{\mathrm{be}}\left[n\right]\right)\tag{7}$$

$\theta_{\mathrm{be}}\left[n\right]\in[-\pi/2,\pi/2]$和$\phi_{\mathrm{be}}\left[n\right]\in[-\pi/2,\pi/2]$分别表示从通信无人机到窃听无人机的出发方位角和俯仰角，$\theta_{\mathrm{eb}}\left[n\right]\in[-\pi/2,\pi/2]$和$\phi_{\mathrm{eb}}\left[n\right]\in[-\pi/2,\pi/2]$分别表示从通信无人机到窃听无人机的到达方位角和俯仰角。值得注意的是，除三维空间上的位置和角度外，通信无人机与窃听无人机间的信道还受其自身的姿态影响，如偏航角、俯仰角和滚转角。在这里，我们假设通信无人机和窃听无人机的姿态可以通过其内部的传感器获取并设计对应的接收波束成形矢量进行补偿。因此，可以假设UAV的自身姿态在本文中不影响预编码设计，在后文中未予提及。

为使所提方案即使在最差情况下也能可靠运行，本文假设窃听无人机在窃取用户k的信息时，可以消除其他地面用户引起的用户间干扰，窃听功能仅受人工噪声的干扰[1]，即时隙n中与地面用户k相关的信息泄露速率为

$$R_{\mathrm{e}}^k\left[n\right]=\log_2\left(1+\frac{M_{\mathrm{e}}\left|\boldsymbol{h}_{\mathrm{be}}^{\mathrm{H}}\left[n\right]\boldsymbol{w}_k\left[n\right]\right|^2}{M_{\mathrm{e}}\left|\boldsymbol{h}_{\mathrm{be}}^{\mathrm{H}}\left[n\right]\boldsymbol{w}_{\mathrm{e}}\left[n\right]\right|^2+\sigma_{\mathrm{e}}^2}\right)\tag{8}$$

2.3 跟踪模型

由于在现实场景中，窃听无人机通常不会主动向通信无人机发送其位置信息。为了实现有效的通信资源分配和在线航迹规划，通信无人机需要利用各时隙人工噪声的回波信号来跟踪窃听无人机的状

态。本节介绍本文所采用的跟踪模型，包括2.3.1节的回波信号模型、2.3.2节的窃听无人机参数估计模型、2.3.3节的窃听无人机观测模型和2.3.4节的窃听无人机状态演化模型。其中，2.3.1节和2.3.2节是2.3.3节的底层模型，2.3.3节和2.3.4节构成了窃听无人机跟踪模型。跟踪模型为通信资源分配和航迹规划提供了信道预测；航路规划确认了通信无人机下一时隙的位置，为通信资源分配和感知提供了空间位置信息；通信资源分配确定了下一时隙的人工噪声预编码向量和功率，为跟踪模型提供了感知信号。

2.3.1 回波信号模型

通信无人机发射的信号部分被窃听无人机接收，部分被窃听无人机的机体散射。假设通信无人机以全双工模式工作，允许在同一频段内同时发送和接收信号，但接收到的信号可能存在残留的自干扰。

在时隙n时，通信无人机接收到的回波信号如式(9)：

$$\begin{aligned}\boldsymbol{r}_{\mathrm{b}}(n,t)=&\mathrm{e}^{\mathrm{j}2\pi\nu_{\mathrm{e}}[n]t}\boldsymbol{H}_{\mathrm{be}}^{\mathrm{r}}[n]\sum_{k=1}^{K}\boldsymbol{w}_k[n]s_k(n,t-\tau_{\mathrm{e}}[n])\\&+\mathrm{e}^{\mathrm{j}2\pi\nu_{\mathrm{e}}[n]t}\boldsymbol{H}_{\mathrm{be}}^{\mathrm{r}}[n]\boldsymbol{w}_{\mathrm{e}}[n]a_{\mathrm{e}}(n,t-\tau_{\mathrm{e}}[n])\\&+\boldsymbol{v}_{\mathrm{b}}(n,t)\end{aligned}\tag{9}$$

其中，第1项为下行链路信号回波，第2项为人工噪声回波，$\boldsymbol{H}_{\mathrm{be}}^{\mathrm{r}}[n]\in\mathbb{C}^{M_{\mathrm{b}}^{\mathrm{t}}\times M_{\mathrm{b}}^{\mathrm{t}}}$为往返信道矩阵，如式(10)所示：

$$\begin{aligned}\boldsymbol{H}_{\mathrm{be}}^{\mathrm{r}}[n]=&\frac{\epsilon_{\mathrm{e}}[n]\beta_0}{2d_{\mathrm{e}}[n]}\boldsymbol{a}_{M_{\mathrm{b}}^{\mathrm{tx}},M_{\mathrm{b}}^{\mathrm{ty}}}(\theta_{\mathrm{be}}[n],\phi_{\mathrm{be}}[n])\boldsymbol{a}_{M_{\mathrm{b}}^{\mathrm{tx}},M_{\mathrm{b}}^{\mathrm{ty}}}^{\mathrm{H}}\\&\cdot(\theta_{\mathrm{be}}[n],\phi_{\mathrm{be}}[n])\end{aligned}\tag{10}$$

其中，$\tau_{\mathrm{e}}[n]$和$\nu_{\mathrm{e}}[n]$分别表示往返时间和多普勒频移，本文定义$\epsilon_{\mathrm{e}}[n]=\sqrt{\dfrac{v_{\mathrm{e}}}{4\pi d_{\mathrm{e}}^2[n]}}$为窃听无人机在$n$时隙的散射系数，$v_{\mathrm{e}}$代表窃听无人机的雷达散射截面。向量$\boldsymbol{v}_{\mathrm{b}}(n,t)\sim\mathcal{CN}(0,\sigma_{\mathrm{b}}^2\boldsymbol{I}_{M_{\mathrm{b}}^{\mathrm{r}}})$表示通信无人机处接收信号的背景噪声和剩余自干扰。本文省略了环境中由地面用户和其他散射体散射的杂波信号。事实上，与窃听无人机处后向散射的回波相比，杂波具有不同的回波角度和多普勒频移，因此本文假设地面用户和其他散射体的杂波可以被现有的杂波抑制技术有效抑制[28]。假设从窃听无人机到通信无人机的方位角与从通信无人机到窃听无人机的方位角相同，这种假设对于点目标模型和广域视距传播信道场景是合理的[29]。令$\theta_{\mathrm{e}}[n]=\theta_{\mathrm{be}}[n]$和$\phi_{\mathrm{e}}[n]=\phi_{\mathrm{be}}[n]$。由于窃听无人机的位置未知，$\tau_{\mathrm{e}}[n]$，$\nu_{\mathrm{e}}[n]$，$\theta_{\mathrm{e}}[n]$，$\phi_{\mathrm{e}}[n]$均未知，但可由接收到的回波信号进行估计。

2.3.2 窃听无人机参数估计模型

根据式(9)中的回波，本文采用匹配滤波法对$\tau_{\mathrm{e}}[n]$，$\nu_{\mathrm{e}}[n]$，$\theta_{\mathrm{e}}[n]$，$\phi_{\mathrm{e}}[n]$进行估计。而根据文献[29,30]，在人工噪声信号和用户信号相独立的情况下，$\tau_{\mathrm{e}}[n]$，$\nu_{\mathrm{e}}[n]$，$\theta_{\mathrm{e}}[n]$，$\phi_{\mathrm{e}}[n]$的估计方差用式(11)表示：

$$\begin{aligned}\sigma_{\tau_{\mathrm{e}}[n]}^2=\frac{c_{\tau_{\mathrm{e}}}}{\mathrm{SNR}},\quad\sigma_{\nu_{\mathrm{e}}[n]}^2=\frac{c_{\nu_{\mathrm{e}}}}{\mathrm{SNR}}\\\sigma_{\phi_{\mathrm{e}}[n]}^2=\frac{c_{\phi_{\mathrm{e}}}}{\mathrm{SNR}},\quad\sigma_{\theta_{\mathrm{e}}[n]}^2=\frac{c_{\theta_{\mathrm{e}}}}{\mathrm{SNR}}\end{aligned}\tag{11}$$

其中，匹配滤波器的输出信噪比可以表示为[27]

$$\mathrm{SNR}=\frac{v_{\mathrm{e}}\beta_0^2G_{\mathrm{MF}}M_{\mathrm{b}}^{\mathrm{r}}\left|\boldsymbol{A}_{M_{\mathrm{b}}^{\mathrm{tx}},M_{\mathrm{b}}^{\mathrm{ty}}}^{\mathrm{H}}(\theta_{\mathrm{e}}[n],\phi_{\mathrm{e}}[n])\boldsymbol{w}_{\mathrm{e}}[n]\right|^2}{16\pi\sigma_{\mathrm{b}}^2d_{\mathrm{e}}^4[n]}\tag{12}$$

参数$c_{\tau_{\mathrm{e}}},c_{\nu_{\mathrm{e}}},c_{\phi_{\mathrm{e}}},c_{\theta_{\mathrm{e}}}>0$是已知的，且取决于具体的估计方法[28]，$G_{\mathrm{MF}}$为匹配滤波器的增益，与该时隙内发射符号的数量成正比。

值得注意的是，本文聚焦基于回波参数估计的窃听无人机跟踪。因此，本文假设参数估计误差与人工噪声的回波功率近似成反比，这一假设适用于一般的回波信号处理方法。基于回波参数估计和参数估计误差，推导跟踪算法的后验协方差矩阵，并据此联合设计通信无人机的航迹与通信资源分配策略。

2.3.3 窃听无人机观测模型

观测模型表征了可观测窃听信道参数与窃听无人机隐状态之间的关系，是推断窃听无人机状态的关键。基于图2的几何关系，给出$\tau_{\mathrm{e}}[n]$，$\nu_{\mathrm{e}}[n]$，$\theta_{\mathrm{e}}[n]$，$\phi_{\mathrm{e}}[n]$的测量模型

$$\begin{aligned}\tau_{\mathrm{e}}[n]&=\frac{2\|\boldsymbol{q}_{\mathrm{e}}[n]-\boldsymbol{q}_{\mathrm{b}}[n]\|}{\mathrm{c}}+v_{\tau_{\mathrm{e}}[n]}\\\nu_{\mathrm{e}}[n]&=\frac{2\left(\dot{\boldsymbol{q}}_{\mathrm{e}}^{\mathrm{T}}[n]-\dot{\boldsymbol{q}}_{\mathrm{b}}^{\mathrm{T}}[n]\right)(\boldsymbol{q}_{\mathrm{e}}[n]-\boldsymbol{q}_{\mathrm{b}}[n])f_{\mathrm{c}}}{\mathrm{c}\|\boldsymbol{q}_{\mathrm{e}}[n]-\boldsymbol{q}_{\mathrm{b}}[n]\|}\\&\quad+v_{\nu_{\mathrm{e}}[n]}\\\sin\theta_{\mathrm{e}}[n]&=\frac{x_{\mathrm{e}}[n]-x_{\mathrm{b}}[n]}{\sqrt{|x_{\mathrm{e}}[n]-x_{\mathrm{b}}[n]|^2+|y_{\mathrm{e}}[n]-y_{\mathrm{b}}[n]|^2}}\\&\quad+v_{\sin\theta_{\mathrm{e}}[n]}\\\sin\phi_{\mathrm{e}}[n]&=\frac{z_{\mathrm{e}}[n]-z_{\mathrm{b}}[n]}{\|\boldsymbol{q}_{\mathrm{e}}[n]-\boldsymbol{q}_{\mathrm{b}}[n]\|}+v_{\sin\phi_{\mathrm{e}}[n]}\end{aligned}\tag{13}$$

其中，$f_{\mathrm{c}}=\mathrm{c}/\lambda_{\mathrm{c}}$为载频，c为光速，高斯随机变量$v_{\tau_{\mathrm{e}}[n]},v_{\nu_{\mathrm{e}}[n]},v_{\sin\theta_{\mathrm{e}}[n]},v_{\sin\phi_{\mathrm{e}}[n]}$分别建模具有$\sigma_{\tau_{\mathrm{e}}[n]}^2,\sigma_{\nu_{\mathrm{e}}[n]}^2,\sigma_{\sin\theta_{\mathrm{e}}[n]}^2,\sigma_{\sin\phi_{\mathrm{e}}[n]}^2$方差的测量噪声。由于在通常情况下，角度$\theta_{\mathrm{e}}[n]$和$\phi_{\mathrm{e}}[n]$的测量误差绝对值数量级都相对较小，因此可以近似得到：

$$\begin{aligned}\sigma_{\sin\theta_{\mathrm{e}}[n]} &\approx \cos\theta_{\mathrm{e}}[n]\,\sigma_{\theta_{\mathrm{e}}[n]}\\ \sigma_{\sin\phi_{\mathrm{e}}[n]} &\approx \cos\phi_{\mathrm{e}}[n]\,\sigma_{\phi_{\mathrm{e}}[n]}\end{aligned} \tag{14}$$

令 $\boldsymbol{\beta}_{\mathrm{e}}[n]=[\tau_{\mathrm{e}}[n],\nu_{\mathrm{e}}[n],\sin\theta_{\mathrm{e}}[n],\sin\phi_{\mathrm{e}}[n]]^{\mathrm{T}}\in\mathbb{R}^{4\times1}$ 包含所有可观测的信道参数。测量模型可以改写为

$$\boldsymbol{\beta}_{\mathrm{e}}[n]=g_n(\boldsymbol{\alpha}_{\mathrm{e}}[n])+\boldsymbol{v}_{\boldsymbol{\beta}_{\mathrm{e}}[n]} \tag{15}$$

其中，$\boldsymbol{v}_{\boldsymbol{\beta}_{\mathrm{e}}}[n]$为观测噪声，$\boldsymbol{v}_{\boldsymbol{\beta}_{\mathrm{e}}}[n]=\left[v_{\beta_{\mathrm{e}}[n]},v_{\nu_{\mathrm{e}}[n]},v_{\sin\theta_{\mathrm{e}}[n]},v_{\sin\phi_{\mathrm{e}}[n]}\right]^{\mathrm{T}}\sim\mathcal{N}\left(\mathbf{0},\boldsymbol{Q}_{\boldsymbol{\beta}_{\mathrm{e}}[n]}\right)$，噪声协方差阵 $\boldsymbol{Q}_{\boldsymbol{\beta}_{\mathrm{e}}[n]}$ 可由式(16)给出：

$$\boldsymbol{Q}_{\boldsymbol{\beta}_{\mathrm{e}}[n]}=\mathrm{diag}\left\{\sigma^2_{\tau_{\mathrm{e}}[n]},\sigma^2_{\nu_{\mathrm{e}}[n]},\sigma^2_{\sin\theta_{\mathrm{e}}[n]},\sigma^2_{\sin\phi_{\mathrm{e}}[n]}\right\} \tag{16}$$

非线性函数$g_n:\mathbb{R}^{6\times1}\to\mathbb{R}^{4\times1}$为观测函数，由式(13)的观测模型构成。观测函数$g_{\mathrm{n}}$取决于通信无人机的位置和速度，因而是随时间变化的。由式(11)，式(12)可知，测量协方差矩阵$\boldsymbol{Q}_{\boldsymbol{\beta}_{\mathrm{e}}[n]}$取决于通信无人机和窃听无人机的位置以及通信无人机的预编码向量，因而也是随时间变化的，并且只能在线获得。换句话说，通信无人机的资源分配和航迹设计会影响感知性能，这是与传统无人机航迹和通信资源分配联合优化的重要不同点。

2.3.4 窃听无人机状态演化模型

为实现准确的跟踪，除了测量模型外，还需要假设合理的窃听无人机状态演化模型。假设窃听无人机短时间内匀速直线运动，其状态演化模型为

$$\boldsymbol{\alpha}_{\mathrm{e}}[n]=\boldsymbol{F}\boldsymbol{\alpha}_{\mathrm{e}}[n-1]+\boldsymbol{v}_{\alpha_{\mathrm{e}}}[n] \tag{17}$$

其中，$\boldsymbol{F}\in\mathbb{R}^{6\times6}$为状态转换矩阵，由式(18)给出：

$$\boldsymbol{F}=\begin{bmatrix}\boldsymbol{I}_3 & \boldsymbol{\delta}\boldsymbol{I}_3\\ \mathbf{0} & \boldsymbol{I}_3\end{bmatrix} \tag{18}$$

其中，$\mathbf{0}$代表零矩阵。实数向量$\boldsymbol{v}_{\alpha_{\mathrm{e}}}=[v_{x_{\mathrm{e}}},v_{y_{\mathrm{e}}},v_{z_{\mathrm{e}}},v_{\dot{x}_{\mathrm{e}}},v_{\dot{y}_{\mathrm{e}}},v_{\dot{z}_{\mathrm{e}}}]^{\mathrm{T}}$表示状态演化噪声。假设状态演化噪声相互独立，我们得到$\boldsymbol{v}_{\alpha_{\mathrm{e}}}\sim\mathcal{N}(\mathbf{0},\boldsymbol{Q}_{\alpha_{\mathrm{e}}})$，其中

$$\boldsymbol{Q}_{\alpha_{\mathrm{e}}}=\mathrm{diag}\left\{\sigma^2_{x_{\mathrm{e}}},\sigma^2_{y_{\mathrm{e}}},\sigma^2_{z_{\mathrm{e}}},\sigma^2_{\dot{x}_{\mathrm{e}}},\sigma^2_{\dot{y}_{\mathrm{e}}},\sigma^2_{\dot{z}_{\mathrm{e}}}\right\} \tag{19}$$

包含了相应状态变量的状态演化噪声方差，其值可以通过通信无人机根据长期测量得到。

在建立窃听无人机的观测模型和状态演化模型后，即可基于扩展卡尔曼滤波算法实现对其状态的实时跟踪。鉴于该算法的实现流程已属成熟技术，此处不再详细展开说明。

3 在线式航迹与通信资源联合优化

针对本文提出的ISAC框架，本节基于窃听无人机状态估计和预测的通信无人机在线式航迹与通信资源联合优化问题。如第2节所述，无人机航迹规划和通信资源优化是双向耦合的，联合设计复杂度较高但可以显著提升系统感知和安全通信性能。此外，无人机通信系统存在固有的因果关系，第n时隙的飞行策略与通信资源分配方案设计取决于第$n-1$时隙的状态预测，同时影响后续时隙的整体安全通信性能，因此无人机航迹和通信资源分配的联合优化必须是在线的，且需要考虑系统的长期回报。

本节首先建模基于长期回报的航迹与通信资源联合优化问题，然后将其建模为马尔可夫决策过程，最后基于深度确定性策略梯度，联合优化在线通信无人机航迹与通信资源分配策略，最大化系统的长期安全通信性能。

3.1 在线式航迹与通信资源联合优化问题建模

低空安全通信需要同时保障合法用户的可达通信速率与信息泄露速率。因此，我们定义信息泄露速率上界R_{Leakage}与可达通信速率下界$R_{\min}$。通信无人机在时隙n的位置和通信资源分配策略需要满足对用户k的可达通信速率约束

$$\mathrm{C}_1:R_k[n]\geq R_{\min} \tag{20}$$

和信息泄露速率约束

$$\mathrm{C}_2:R^k_{\mathrm{e}}[n]\leq R_{\mathrm{Leakage}} \tag{21}$$

则可认为在第n时隙用户k实现了安全通信。

为了能够量化系统在第n时隙的安全通信性能，本文定义二元变量$u_k[n]$，$k\in\{1,2,\cdots,K\}$。如果在第n时隙用户k的可达通信速率和信息泄露速率能够同时满足约束式(20)与式(21)，则令$u_k[n]=1$，反之$u_k[n]=0$。借助二元变量$u_k[n]$，我们将系统在第n时隙的安全通信性能定义为在第n时隙可以实现安全通信的用户数，即

$$\sum_{k=1}^{K}u_k[n] \tag{22}$$

由于通信无人机可以在线感知窃听无人机，并且在线优化其飞行策略，第n时隙的安全通信性能与前序时隙的感知和安全通信性能存在强相关关系。因此，本文将系统的长期安全通信性能作为优化目标，即

$$\sum_{n=1}^{N}\sum_{k=1}^{K}u_k[n] \tag{23}$$

此外，通信无人机的飞行和通信策略需要满足飞行空域的环境约束与通信无人机的硬件约束。具体来说，可分为以下3方面：

(1) 运动学约束

通信无人机的速度需小于通信无人机可达到的最大速度，即

$$\mathrm{C}_3:\left\|\dot{\boldsymbol{q}}_{\mathrm{b}}[n]\right\|\le V_{\max} \tag{24}$$

通信无人机的运动需在允许空间内进行，即

$$\mathrm{C}_4:\boldsymbol{q}_{\mathrm{b}}[n]\le\boldsymbol{q}_{\max},\boldsymbol{q}_{\mathrm{b}}[n]\ge\boldsymbol{q}_{\min} \tag{25}$$

其中，$V_{\max}$代表通信无人机的最大速度，$\boldsymbol{q}_{\max}$和$\boldsymbol{q}_{\min}$分别表示通信无人机在三维笛卡儿坐标系中各维度的边界位置。

(2) 功率约束

单一时隙内通信无人机对用户的发射功率和对窃听无人机的干扰功率总和需小于通信无人机硬件限制的总发射功率$p_{\max}$，即

$$\mathrm{C}_5:\sum_{k=1}^{K}\boldsymbol{w}_k^{\mathrm{H}}[n]\boldsymbol{w}_k[n]+\boldsymbol{w}_{\mathrm{e}}^{\mathrm{H}}[n]\boldsymbol{w}_{\mathrm{e}}[n]\le p_{\max} \tag{26}$$

(3) 跟踪精度约束

为保证通信无人机对窃听无人机状态的连续可靠感知，本文要求后验协方差矩阵的迹小于等于最大观测误差限制，即

$$\mathrm{C}_6:\mathrm{Tr}\left\{\boldsymbol{C}_{\mathrm{e}}[n]\right\}\le\mathrm{MSE}_{\max} \tag{27}$$

其中，$\mathrm{MSE}_{\max}$为可对窃听无人机进行有效跟踪的最大允许误差。

为了能适应复杂多变的环境，针对不同的通信无人机状态给出相应的策略，本文采取在线式的航迹规划方案，即通过控制每一时隙n通信无人机的加速度$\ddot{\boldsymbol{q}}_{\mathrm{b}}[n]$来控制其飞行轨迹。而为了能得到最大的长期安全通信性能，我们需要同时优化每一时隙通信无人机的加速度与预编码策略，即$\boldsymbol{w}_{\mathrm{e}}[n]$，$\boldsymbol{w}_k[n],\ddot{\boldsymbol{q}}_{\mathrm{b}}[n],\forall k\in\{1,2,\cdots,K\},\forall n\in\{1,2,\cdots,N\}$。我们将基于长期回报的感通一体化设计问题建模为如下优化问题：

$$\begin{aligned}&\max_{\substack{u_k[n],\boldsymbol{w}_{\mathrm{e}}[n],\boldsymbol{w}_k[n],\ddot{\boldsymbol{q}}_{\mathrm{b}}[n]\\ \forall k\in\{1,2,\cdots,K\},\forall n\in\{1,2,\cdots,N\}}}\sum_{n=1}^{N}\sum_{k=1}^{K}u_k[n]\\&\text{s.t. }\mathrm{C}_1,\mathrm{C}_2,\mathrm{C}_3,\mathrm{C}_4,\mathrm{C}_5,\mathrm{C}_6\end{aligned} \tag{28}$$

3.2 马尔可夫决策过程建模

基于长期回报的感通一体化设计问题(28)最大化系统的长期安全通信性能。显然，式(28)的目标函数与优化变量之间的关系难以显式得到。另外，由于二元变量$u_k[n]$的存在，式(28)是一个混合整数规划问题。此外，由于通信无人机无法事先获得窃听无人机的航迹，需要在飞行过程中利用人工噪声的回波进行在线感知，并据此在线优化下一时隙的飞行策略。因此，我们更期待一个在线的感通一体化设计方法，能够根据系统实时获取的环境状态给出最大化系统长期回报的决策。因此，本文将原问题(28)建模为马尔可夫决策过程，然后利用深度强化学习求解该问题。

3.2.1 状态空间$\mathcal{S}$

在马尔可夫决策过程中，状态空间代表了智能体(通信无人机)所感知到的环境信息。状态信息是智能体制定决策和评估其长期回报的依据，直接决定了深度强化学习算法的收敛性能。在本文中，第n–1时隙的环境状态由第n–1时隙窃听无人机的后验状态估计、后验协方差矩阵和通信无人机的状态构成，即

$$\boldsymbol{s}_{n-1}\triangleq\left\{\hat{\boldsymbol{\alpha}}_{\mathrm{e}}[n-1],\boldsymbol{C}_{\mathrm{e}}[n-1],\boldsymbol{\alpha}_{\mathrm{b}}[n-1]\right\}\in\mathcal{S} \tag{29}$$

值得注意的是，马尔可夫决策过程建模中的环境状态与第2节窃听无人机跟踪中的状态定义不同。事实上，除上述参数以外，环境中还包含地面用户的位置$\boldsymbol{q}_k,k=1,2,\cdots,K$。但是，由于系统模型假设地面用户静止不动，且不会随着通信无人机的飞行和通信资源分配策略发生变化，因而没有列入每时隙的环境状态向量中。

3.2.2 动作空间$\mathcal{A}$

在马尔可夫决策过程中，动作空间为智能体与环境交互时可采取的动作组成的集合。正如第2节所述，在本问题中，第n时隙通信无人机的动作包含其加速度和预编码策略，即

$$\boldsymbol{a}_n\triangleq\left\{\ddot{\boldsymbol{q}}_{\mathrm{b}}[n],\boldsymbol{\mathcal{W}}[n]\right\}\in\mathcal{A} \tag{30}$$

其中，$\ddot{\boldsymbol{q}}_{\mathrm{b}}[n]$是通信无人机在时隙$n$的加速度，$\boldsymbol{\mathcal{W}}[n]$为通信无人机的预编码集合，包括对地面用户的通信预编码和对窃听无人机的人工噪声预编码，即

$$\boldsymbol{\mathcal{W}}[n]=\left\{\boldsymbol{w}_1[n],\boldsymbol{w}_2[n],\cdots,\boldsymbol{w}_K[n],\boldsymbol{w}_{\mathrm{e}}[n]\right\} \tag{31}$$

3.2.3 状态转移$\mathcal{T}$

本文采用确定性状态转移关系。通信无人机的位置与速度变化由其加速度决定，即

$$\begin{aligned}\dot{\boldsymbol{q}}_{\mathrm{b}}[n]&=\dot{\boldsymbol{q}}_{\mathrm{b}}[n-1]+\ddot{\boldsymbol{q}}_{\mathrm{b}}[n]\delta\\\boldsymbol{q}_{\mathrm{b}}[n]&=\boldsymbol{q}_{\mathrm{b}}[n-1]+\dot{\boldsymbol{q}}_{\mathrm{b}}[n]\delta+\frac{1}{2}\ddot{\boldsymbol{q}}_{\mathrm{b}}[n]\delta^2\end{aligned} \tag{32}$$

另一方面，在智能体给出n时隙的加速度与预编码方案后，窃听无人机的后验状态估计与估计协方差矩阵更新由扩展卡尔曼滤波器给出。

3.2.4 回报函数$\mathcal{R}$

回报函数是环境针对智能体输出的动作给出的评分，其设计方式直接决定了智能体的学习方向，对最终深度强化学习的性能有举足轻重的作用。

3.1节已完整构建了感通一体化设计的数学模型。据此，可定义可安全通信用户数$\sum_{k=1}^{K}\hat{u}_k[n]$作为时隙$n$的瞬时回报函数。同时，为了能确保智能体的动作策略满足式(28)中的约束条件，引入一个二元惩罚因子$\varsigma[n]$，当约束条件$\mathrm{C}_1,\mathrm{C}_2,\mathrm{C}_3,\mathrm{C}_4,\mathrm{C}_5,\mathrm{C}_6$均能得到满足时，$\varsigma[n]=1$，反之$\varsigma[n]=0$。最终瞬时回报函数为

$$r_n \triangleq \varsigma[n]\sum_{k=1}^{K}\hat{u}_k[n] \tag{33}$$

当且仅当智能体给出的动作策略能够满足所有约束时，才会产生正向得分，激励智能体在满足$\mathrm{C}_1-\mathrm{C}_6$的可行域内积极探索。因此，上述瞬时回报函数指引智能体在满足感知性能约束(27)的前提下，最大化系统的安全通信性能，以获得感知-安全通信性能的折中。

3.3 基于深度确定性策略梯度的在线式航迹与通信资源联合优化设计

在完成了原优化问题(28)的马尔可夫决策过程建模后，原问题转变为一个序列化的决策问题。深度强化学习方法通过智能体与环境的交互，获得每一时隙的$\{\boldsymbol{s}_n,\boldsymbol{a}_n,\boldsymbol{s}_{n+1},r_n\}$四元组，逐渐学习在线感通一体化设计问题的最优决策。深度确定性策略梯度(DDPG)是一种基于策略-价值(Actor-Critic)框架的无模型深度强化学习算法[18]。DDPG专为具有连续动作空间的规划任务设计，其核心是通过一个确定性策略网络直接输出连续的动作，而不需要事先对动作空间进行离散化。DDPG结构简单，易于实现，非常适合用于在线感通一体化设计问题。

3.3.1 价值网络

价值网络(Critic Network)通过深度神经网络对动作价值函数$Q(\boldsymbol{s}_n,\boldsymbol{a}_n)$进行近似建模，可表示为$Q(\boldsymbol{s}_n,\boldsymbol{a}_n|\boldsymbol{\theta}_Q)$，其中，$\boldsymbol{\theta}_Q$代表待优化的网络参数。该函数用于刻画在状态$\boldsymbol{s}_n$下执行动作$\boldsymbol{a}_n$所带来的期望长期回报。

3.3.2 策略网络

策略网络(Actor Network)利用深度神经网络对动作策略函数$\mu(\boldsymbol{s}_n)$进行参数化建模，记为$\mu(\boldsymbol{s}_n|\boldsymbol{\theta}_\mu)$，其中$\boldsymbol{\theta}_\mu$代表待优化的网络参数。该函数通过确定性映射，将输入状态$\boldsymbol{s}_n$映射至对应的动作$\boldsymbol{a}_n$，从而实现策略生成。

3.3.3 网络参数优化

深度确定性策略梯度方法的最终目标是训练得到最优的动作策略函数，使其可以在任意状态下获得最大化的长期回报[18]，即

$$\mu\left(\boldsymbol{s}_n|\boldsymbol{\theta}_\mu^*\right)=\underset{\boldsymbol{a}_n}{\operatorname{argmax}}\,Q\left(\boldsymbol{s}_n,\boldsymbol{a}_n|\boldsymbol{\theta}_Q^*\right) \tag{34}$$

其中，$\boldsymbol{\theta}_\mu^*$与$\boldsymbol{\theta}_Q^*$代表训练后达到最优的网络参数。

在每次智能体与环境的交互中，系统均生成四元组样本$(\boldsymbol{s}_n,\boldsymbol{a}_n,r_n,\boldsymbol{s}_{n+1})$，储存于大小为$|\mathcal{M}|$的回放缓存区。为了提升训练的稳定性并减弱样本间的时间相关性，深度确定性策略梯度方法引入了小批量训练(minibatch)机制[31]。具体而言，每次从回放缓冲区取出大小为$|\mathcal{N}|$的样本集合，用于计算梯度并更新网络参数，从而实现训练过程的平滑化与稳定化。

价值网络的训练基于时序差分方法[31]，定义损失函数如下：

$$\begin{aligned}L(\boldsymbol{\theta}_Q)=\frac{1}{|\mathcal{N}|}\sum_{i=1}^{|\mathcal{N}|}&\left(Q(\boldsymbol{s}_i,\boldsymbol{a}_i|\boldsymbol{\theta}_Q)-[r_i\right.\\&\left.+Q\left(\boldsymbol{s}_{i+1},\mu(\boldsymbol{s}_{i+1}|\boldsymbol{\theta}_\mu)|\boldsymbol{\theta}_Q\right)]\right)^2\end{aligned} \tag{35}$$

其中，瞬时回报r_i由式(33)给出。为提高训练的稳定性，深度确定性策略梯度方法引入了目标网络(Target Network)，通过延迟更新的方式缓解训练中的梯度波动和收敛不稳定的问题，从而提高了模型的鲁棒性。

策略网络的训练可通过链式法则，计算长期回报对策略网络参数的梯度：

$$\begin{aligned}\nabla_{\boldsymbol{\theta}_\mu}Q\left(\boldsymbol{s}_n,\mu(\boldsymbol{s}_n|\boldsymbol{\theta}_\mu)|\boldsymbol{\theta}_Q\right)=&\nabla_a Q(\boldsymbol{s}_n,\boldsymbol{a}_n|\boldsymbol{\theta}_Q)\\&\cdot\nabla_{\theta_\mu}\mu(\boldsymbol{s}_n|\boldsymbol{\theta}_\mu)\end{aligned} \tag{36}$$

在小批量样本上取均值后，利用Adam优化器更新策略网络参数[29]。为增加探索能力，在确定性动作中加入噪声$\boldsymbol{n}_a$，即

$$\mu(\boldsymbol{s}_n)=\mu(\boldsymbol{s}_n|\boldsymbol{\theta}_\mu)+\boldsymbol{n}_a \tag{37}$$

4 仿真实验

本节将通过仿真实验评估本文所提基于深度强化学习的在线式航迹与通信资源联合优化方案。在本实验中通信无人机的初始位置为$[200,200,80]$，初始速度和初始加速度均为0。窃听无人机的初始位置为$[0,100,50]$，窃听无人机飞行高度不变且经过每个地面用户的上空。通信无人机位置，窃听无

人机位置，窃听无人机轨迹，地面用户位置状态如图4所示。其余参数除有特殊说明，均为表1所示。

4.1 感知精度约束对安全通信性能的影响

跟踪误差上限参数$\mathrm{MSE}_{\max}$刻画了感知精度约束的程度，其取值将决定感知和安全通信性能的折中。为确定合理的$\mathrm{MSE}_{\max}$取值，本节探究跟踪误差上限参数$\mathrm{MSE}_{\max}$对感知和安全通信性能的影响。

本节分别设置$\mathrm{MSE}_{\max}=50,\mathrm{MSE}_{\max}=100,\mathrm{MSE}_{\max}=200$这3组参数，对比通信无人机对窃听无人机的瞬时归一化跟踪误差$\lg(\dfrac{\left|\boldsymbol{\alpha}_e\left[n\right]-\hat{\boldsymbol{\alpha}}_e\left[n\right]\right|^2}{100})$。为评估无人机安全通信性能，我们定义了两个经验互补累积分布函数。

(1) 估计CCDF：由于通信无人机无法获取窃听无人机的真实状态，训练时只能基于扩展卡尔曼滤波的跟踪状态$\hat{\boldsymbol{\alpha}}_e\left[n\right]$来设计航迹和预编码策略。估计CCDF，记为$\hat{F}\left(m\right)$，反映了所提DDPG框架的学习能力，其定义为：假设估计状态为真实状态时，系统可获得的安全通信用户数的CCDF。该指标同时也是MDP设计中回报函数的体现。

$$\hat{F}\left(m\right)=\frac{\sum_{n=1}^{N}I\left\{\sum_{k=1}^{K}\hat{u}_k\left[n\right]>m\right\}}{N}$$

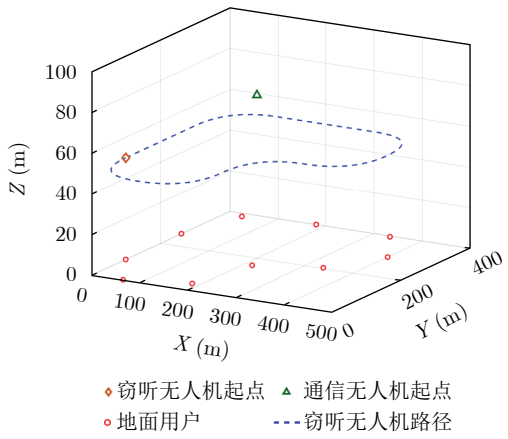

图 4 系统初始状态示意图

Fig. 4 The initial system state

(2) 真实CCDF：在DDPG框架生成时隙n的预编码策略后，我们基于3.1节提出的计算流程，利用真实的窃听无人机状态计算$u_k\left[n\right]$，进而得到该时隙的安全通信用户数$\sum_{k=1}^{K}u_k\left[n\right]$。由此获得的真实安全通信用户数CCDF，记为$F\left(m\right)$。$F\left(m\right)$反映了所提框架的实际安全通信性能。通过比较$F\left(m\right)$与$\hat{F}\left(m\right)$的差异，不仅能够评估状态估计误差对系统性能的影响，同时也凸显了感知环节在整个系统中的重要性。$F\left(m\right)$的计算方式如下：

$$F\left(m\right)=\frac{\sum_{n=1}^{N}I\left\{\sum_{k=1}^{K}u_k\left[n\right]>m\right\}}{N}$$

由图5可见，窃听无人机跟踪误差随着$\mathrm{MSE}_{\max}$的变小而减小。然而，由图6可见，随着$\mathrm{MSE}_{\max}$数值减小，感知约束逐渐增强，通信无人机需要将更多的资源用于发射人工噪声，安全服务通信用户数明显下降。由此可见，过度追求感知精度反而会损失系统的安全通信性能。有趣的是，$\mathrm{MSE}_{\max}$数值过大($\mathrm{MSE}_{\max}=200$)导致对窃听无人机的状态估计准确性降低，实际可实现的安全通信性能会出现大幅下降。反映在图6中，$\mathrm{MSE}_{\max}=200$时估计CCDF与真实CCDF之间存在一定间距。

总之，对比图5和图6，当$\mathrm{MSE}_{\max}=50$时，通信无人机将大量资源用于发射人工噪声，安全通信性能降低；当$\mathrm{MSE}_{\max}=100$和$\mathrm{MSE}_{\max}=200$时，感知性能略微损失，但安全通信性能较高；当$\mathrm{MSE}_{\max}=200$时，由于感知精度下降，通信无人机的真实安全服务用户数不及$\mathrm{MSE}_{\max}=100$。值得注意的是，在低空安全通信系统中，感知的目的是辅助航迹优化和通信策略设计，安全通信性能才是决定性的评价指标。因此，在后续实验中，我们选择$\mathrm{MSE}_{\max}=100$。

4.2 与基线方案的对比

本文采取文献[16]中所提方案作为基线方案，比较对窃听无人机的感知性能和对地面用户的安全通信性能。在文献[16]中，作者提出了一种基于传

表 1 仿真参数

Tab. 1 Simulation parameters

参数	数值	参数	数值	参数	数值	参数	数值
K	10	$\left[M_{\mathrm{e}}^{x},M_{\mathrm{e}}^{y}\right]$	$[4,4]$	$c_{\theta_{\mathrm{be}}}$	0.1	$V_{\max}$	30 m/s
δ	1 s	$\left[M_{\mathrm{b}}^{\mathrm{t}x},M_{\mathrm{b}}^{\mathrm{t}y}\right]$	$[4,4]$	$c_{\phi_{\mathrm{be}}}$	0.1	$p_{\max}$	30 dBm
β_0^2	–40 dBW	$\left[M_{\mathrm{b}}^{\mathrm{r}x},M_{\mathrm{b}}^{\mathrm{r}y}\right]$	$[4,4]$	$c_{\tau_{\mathrm{e}}}$	10^{-7}	N	200
$R_{\min}$	3 bit/(s·Hz)	R_{Leakage}	0.1 bit/(s·Hz)	$c_{\nu_{\mathrm{e}}}$	1	f_{c}	3 GHz
$\sigma_{\mathrm{e}}^2,\sigma_k^2$	–90 dBm	$\boldsymbol{q}_{\min}$	$[0,0,50]^{\mathrm{T}}$	c	3×10^8 m/s	G_{MF}	10^4
σ_{b}^2	–50 dBm	$\boldsymbol{q}_{\max}$	$[500,500,100]^{\mathrm{T}}$	υ_{e}	0.1 m^2	$\left[A_{\mathrm{cc}}^{x},A_{\mathrm{cc}}^{y},A_{\mathrm{cc}}^{z}\right]$	$[4,4,2]$

统优化方法的在线式航迹设计与通信资源分配方案。该方案解耦地处理航迹设计与通信资源分配策略优化问题，在环境中设置了一个静态“通信点”，并用平衡系数$\lambda \in (0,1)$控制通信无人机偏向于靠近窃听无人机和静态“通信点”的程度，λ越大，通信无人机的飞行轨迹越靠近窃听无人机，与之对应，该航迹设计方案就更加有利于感知，λ越小，通信无人机的飞行轨迹越靠近静态“通信点”，与之对应，该航迹设计方案就越利于通信。在文献[16]中，作者深入研究了平衡系数λ对系统感知与通信性能的权衡影响。研究表明，当$\lambda = 0.9$时，系统能够在保证通信性能不明显下降的前提下获得最优的感知性能；当$\lambda = 0.5$时，系统可以在维持有效感知能力的同时实现最优的通信性能。为充分凸显本文所提方案在感知-通信性能协同优化方面的先进性，本研究将文献[16]中的两个典型参数设置作为对比基线：以通信性能最优的$\lambda = 0.5$作为基线方案1，以感知性能最优的$\lambda = 0.9$作为基线方案2。

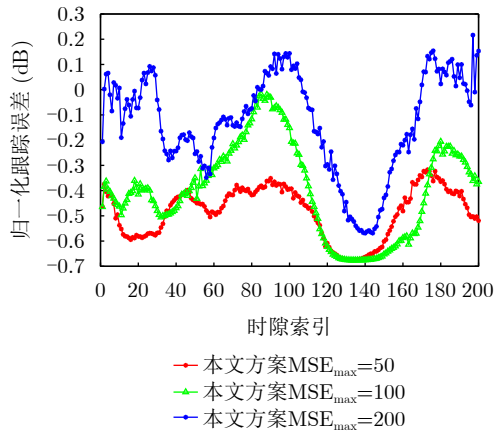

图 5 不同MSE$_{max}$下窃听无人机瞬时归一化跟踪误差

Fig. 5 Instant normalized tracking error of eavesdropping UAV under different MSE$_{max}$

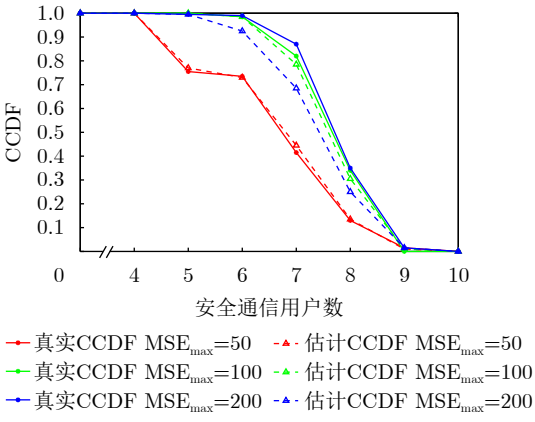

图 6 不同MSE$_{max}$下安全通信用户数CCDF

Fig. 6 The CCDF of the number of securely served GUs under different MSE$_{max}$

此外，在理想感知场景下，即窃听无人机航迹精确已知，假设通信无人机保持静止，不进行航迹规划而只优化通信资源分配，作为基线方案3。与基线方案3相比，本文所提方案可以展示航迹规划带来的性能增益。同样，在理想感知场景下，假设通信无人机运动，且联合优化航迹和通信资源分配策略，作为基线方案4。基线方案4在理想感知场景中联合设计航迹与通信策略，可以为本文所提方案提供性能上界。

本文提出方案、基线方案1和基线方案2所设计的通信无人机轨迹如图7所示，相应的窃听无人机瞬时归一化跟踪误差$\lg\left(\dfrac{\left|\boldsymbol{\alpha}_e[n]-\hat{\boldsymbol{\alpha}}_e[n]\right|^2}{\mathrm{MSE}_{\max}}\right)$展示在图8中。图9展示了本文所提方案和4种基线方案得到的N个时隙内真实安全服务用户数的互补累计分布函数。

由图7可知，本文所提方案的通信无人机规划航迹相较于基线方案1和基线方案2[16]有较大区别：

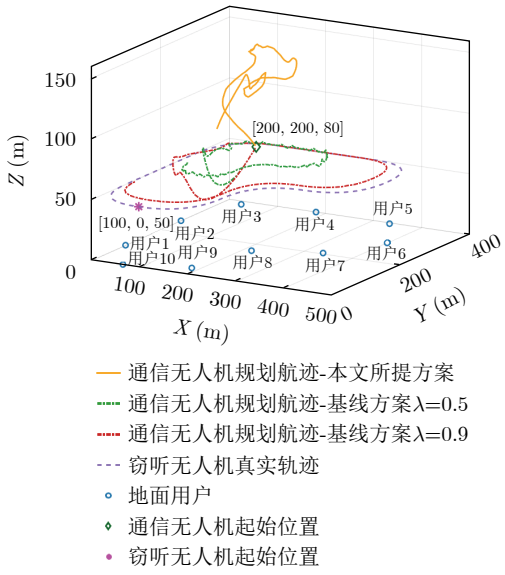

图 7 通信无人机规划航迹图

Fig. 7 The designed trajectory of the communication UAV

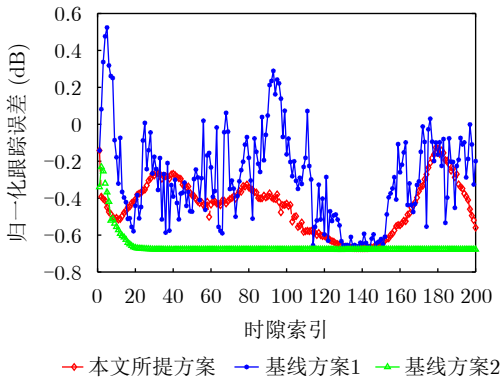

图 8 窃听无人机跟踪的归一化状态估计误差

Fig. 8 The NMSE of eavesdropping UAV tracking

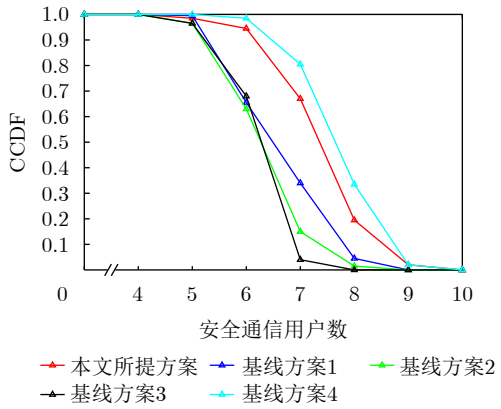

图 9 安全通信用户数CCDF

Fig. 9 The CCDF of the number of securely served GUs

(1) 本文所提方案的通信无人机规划航迹较少受到窃听无人机航迹的影响。相反，基线方案1和基线方案2的通信无人机规划航迹依赖于窃听无人机的航迹。事实上，基线方案1和基线方案2均考虑了一种基于距离的启发式航迹规划策略，更加注重优化窃听无人机的短期感知性能。然而，本文所提方案更加注重窃听无人机的长期感知性能，故其设计的通信无人机航迹与窃听无人机航迹无明显关系，从而充分平衡感知和安全通信性能。

(2) 本文所提方案相较于两组基线方案，更加明显地利用了飞行高度这一维度。传统方案受限于基于距离的航迹优化目标，只能在某一预设高度进行飞行，无法充分利用通信无人机的机动特性。然而，本文所提方案以长期感知和安全通信性能作为航迹优化目标，可以更好地利用通信无人机的飞行高度，从而引导通信无人机更加灵活地飞行，寻找能带来更大增益的空间位置。

由图8可见，在大多数时隙内，本文所提方案的感知误差低于基线方案1，但略高于基线方案2。这是由于基线方案2选择$\lambda=0.9$，通信无人机的航迹始终保持在窃听无人机的附近，如图7所示，较高的人工噪声回波功率促进了感知性能的提升。但是，由图9可知，相较于基线方案1和基线方案2，本文所提方案显著提高了N个时隙内的总安全通信用户数，获得了更好的安全通信性能。因此，本文所提方案可以更好地平衡感知与安全通信性能，避免了基线方案2过分追求感知精度而导致的安全通信性能下降。此外，在图9中，对比基线方案3和本文所提方案可知，将航迹规划引入感通一体化系统并进行联合设计，可以带来非常可观的安全通信性能增益。尽管基线方案2的感知精度很高，但是通过和基线方案3对比可知，其对高感知精度的追求损失了航迹规划的自由度，甚至获得了与静止通信无人机相近的安全通信性能。最后，在图9中，对比基线方案4和本文所提方案可知，本文所提方案获得了接近其上界的安全通信性能，验证了本文所提基于深度强化学习的联合设计方法可以为感通一体化设计问题提供高质量可行解。

总之，本文所提基于深度强化学习的在线式航迹与通信资源联合优化方案考虑长期感知和安全通信性能作为优化目标，可以充分利用通信无人机的飞行高度，在感知和通信之间达到更好的平衡。

5 结语

面向低空无人机安全通信需求，本研究提出了一种在线式航迹与通信资源联合优化方案及基于深度强化学习的感通一体化设计方法。具体而言，本研究首先构建了ISAC与航迹优化一体化框架，利用人工噪声复用技术同时实现窃听无人机感知与干扰功能；其次，感通一体化设计问题建模通过在线联合优化通信无人机航迹和通信资源分配策略，最大化系统的长期感知与安全通信性能；最后，本研究将在线式航迹与通信资源联合优化问题建模为马尔可夫决策过程，并基于DDPG算法进行求解，得到感通一体化设计问题的高质量可行解。相较于基线方案，本文所提方案通过更加灵活的飞行策略，充分利用了通信无人机的飞行高度，在感知和通信之间实现了更高效的平衡，显著提升了系统的长期安全通信性能。本研究验证了感通一体化联合设计可以为低空通信带来非常可观的安全通信性能增益，也验证了深度强化学习方法在感通一体化设计问题中的可行性与先进性。

未来的研究方向将聚焦于窃听无人机与通信无人机博弈对抗场景下的安全通信问题。通过引入多智能体强化学习框架，进一步探索通信无人机与具有自主决策能力的窃听无人机之间的博弈机制，从而丰富ISAC理论体系并拓展其应用边界。

利益冲突 所有作者均声明不存在利益冲突

Conflict of Interests The authors declare that there is no conflict of interests

参 考 文 献

[1] MUKHERJEE A, FAKOORIAN S A A, HUANG Jing, *et al*. Principles of physical layer security in multiuser wireless networks: A survey[J]. *IEEE Communications Surveys & Tutorials*, 2014, 16(3): 1550–1573. doi: 10.1109/SURV.2014.012314.00178.

[2] LIU Chenxi, LEE J, and QUEK T Q S. Safeguarding UAV communications against full-duplex active eavesdropper[J].

IEEE Transactions on Wireless Communications, 2019, 18(6): 2919–2931. doi: 10.1109/TWC.2019.2906177.

[3] WANG Qian, CHEN Zhi, MEI Weidong, *et al*. Improving physical layer security using UAV-enabled mobile relaying[J]. *IEEE Wireless Communications Letters*, 2017, 6(3): 310–313. doi: 10.1109/LWC.2017.2680449.

[4] ZHANG Guangchi, WU Qingqing, CUI Miao, *et al*. Securing UAV communications via joint trajectory and power control[J]. *IEEE Transactions on Wireless Communications*, 2019, 18(2): 1376–1389. doi: 10.1109/TWC.2019.2892461.

[5] ZHOU Xiaobo, WU Qingqing, YAN Shihao, *et al*. UAV-enabled secure communications: Joint trajectory and transmit power optimization[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(4): 4069–4073. doi: 10.1109/TVT.2019.2900157.

[6] CUI Miao, ZHANG Guangchi, WU Qingqing, *et al*. Robust trajectory and transmit power design for secure UAV communications[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(9): 9042–9046. doi: 10.1109/TVT.2018.2849644.

[7] CAI Yuanxin, WEI Zhiqiang, LI Ruide, *et al*. Joint trajectory and resource allocation design for energy-efficient secure UAV communication systems[J]. *IEEE Transactions on Communications*, 2020, 68(7): 4536–4553. doi: 10.1109/TCOMM.2020.2982152.

[8] LIU Fan, MASOUROS C, PETROPULU A P, *et al*. Joint radar and communication design: Applications, state-of-the-art, and the road ahead[J]. *IEEE Transactions on Communications*, 2020, 68(6): 3834–3862. doi: 10.1109/TCOMM.2020.2973976.

[9] ZHANG J A, RAHMAN M L, WU Kai, *et al*. Enabling joint communication and radar sensing in mobile networks-A survey[J]. *IEEE Communications Surveys & Tutorials*, 2022, 24(1): 306–345. doi: 10.1109/COMST.2021.3122519.

[10] WEI Zhongxiang, LIU Fan, MASOUROS C, *et al*. Toward multi-functional 6G wireless networks: Integrating sensing, communication, and security[J]. *IEEE Communications Magazine*, 2022, 60(4): 65–71. doi: 10.1109/MCOM.002.2100972.

[11] SU Nanchi, LIU Fan, WEI Zhongxiang, *et al*. Secure dual-functional radar-communication transmission: Exploiting interference for resilience against target eavesdropping[J]. *IEEE Transactions on Wireless Communications*, 2022, 21(9): 7238–7252. doi: 10.1109/TWC.2022.3156893.

[12] SU Nanchi, LIU Fan, and MASOUROS C. Secure radar-communication systems with malicious targets: Integrating radar, communications and jamming functionalities[J]. *IEEE Transactions on Wireless Communications*, 2021, 20(1): 83–95. doi: 10.1109/TWC.2020.3023164.

[13] HUANG Chongwen, YANG Zhaohui, ALEXANDROPOULOS G C, *et al*. Multi-Hop RIS-empowered terahertz communications: A DRL-based hybrid beamforming design[J]. *IEEE Journal on Selected Areas in Communications*, 2021, 39(6): 1663–1677. doi: 10.1109/JSAC.2021.3071836.

[14] SUN Yan, XU Dongfang, NG D W K, *et al*. Optimal 3D-trajectory design and resource allocation for solar-powered UAV communication systems[J]. *IEEE Transactions on Communications*, 2019, 67(6): 4281–4298. doi: 10.1109/TCOMM.2019.2900630.

[15] YOU Changsheng and ZHANG Rui. Hybrid offline-online design for UAV-enabled data harvesting in probabilistic LoS channels[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(6): 3753–3768. doi: 10.1109/TWC.2020.2978073.

[16] WEI Zhiqiang, LIU Fan, LIU Chang, *et al*. Integrated sensing, navigation, and communication for secure UAV networks with a mobile eavesdropper[J]. *IEEE Transactions on Wireless Communications*, 2024, 23(7): 7060–7078. doi: 10.1109/TWC.2023.3337148.

[17] MNIH V, KAVUKCUOGLU K, SILVER D, *et al*. Playing Atari with deep reinforcement learning[J]. *arXiv preprint arXiv*: 1312.5602, 2013. doi: 10.48550/arXiv.1312.5602.

[18] LILLICRAP T P, HUNT J J, PRITZEL A, *et al*. Continuous control with deep reinforcement learning[J]. *arXiv preprint arXiv*: 1509.02971, 2015. doi: 10.48550/arXiv.1509.02971.

[19] SCHULMAN J, WOLSKI F, DHARIWAL P, *et al*. Proximal policy optimization algorithms[J]. *arXiv preprint arXiv*: 1707.06347, 2017. doi: 10.48550/arXiv.1707.06347.

[20] HAARNOJA T, ZHOU A, ABBEEL P, *et al*. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor[J]. *arXiv preprint arXiv*: 1801.01290, 2018. doi: 10.48550/arXiv.1801.01290.

[21] GAN Xu, HUANG Chongwen, YANG Zhaohui, *et al*. Bayesian learning for double-RIS aided ISAC systems with superimposed pilots and data[J]. *IEEE Journal of Selected Topics in Signal Processing*, 2024, 18(5): 766–781. doi: 10.1109/JSTSP.2024.3408071.

[22] QIN Yunhui, ZHANG Zhongshan, LI Xulong, *et al*. Deep reinforcement learning based resource allocation and trajectory planning in integrated sensing and communications UAV network[J]. *IEEE Transactions on Wireless Communications*, 2023, 22(11): 8158–8169. doi: 10.1109/TWC.2023.3260304.

[23] ZHU Yuqian, LI Ming, LIU Yang, *et al*. DRL-based joint beamforming and BS-RIS-UE association design for RIS-assisted mmWave networks[C]. 2022 IEEE Wireless

Communications and Networking Conference (WCNC), Austin, USA, 2022: 345–350. doi: 10.1109/WCNC51071.2022.9771607.

[24] DONG Runze, WANG Buhong, TIAN Jiwei, *et al*. Deep reinforcement learning based UAV for securing mmWave communications[J]. *IEEE Transactions on Vehicular Technology*, 2023, 72(4): 5429–5434. doi: 10.1109/TVT.2022.3224959.

[25] ZENG Yong, WU Qingqing, and ZHANG Rui. Accessing from the sky: A tutorial on UAV communications for 5G and beyond[J]. *Proceedings of the IEEE*, 2019, 107(12): 2327–2375. doi: 10.1109/JPROC.2019.2952892.

[26] YOU Changsheng and ZHANG Rui. 3D trajectory optimization in Rician fading for UAV-enabled data harvesting[J]. *IEEE Transactions on Wireless Communications*, 2019, 18(6): 3192–3207. doi: 10.1109/TWC.2019.2911939.

[27] NASIR A A, DURRANI S, MEHRPOUYAN H, *et al*. Timing and carrier synchronization in wireless communication systems: A survey and classification of research in the last 5 years[J]. *EURASIP Journal on Wireless Communications and Networking*, 2016, 180(2016): 1–38. doi: 10.1186/s13638-016-0670-9.

[28] SKOLNIK M I. Introduction to Radar Systems[M]. New York: McGraw-Hill, 1980: 581.

[29] LIU Fan, YUAN Weijie, MASOUROS C, *et al*. Radar-assisted predictive beamforming for vehicular links: Communication served by sensing[J]. *IEEE Transactions on Wireless Communications*, 2020, 19(11): 7704–7719. doi: 10.1109/TWC.2020.3015735.

[30] KAY S M. Fundamentals of Statistical Signal Processing[M]. Englewood Cliffs, NJ: PTR Prentice-Hall, 1993.

[31] MONTAVON G, ORR G B, and MÜLLER K R. Neural Networks: Tricks of the Trade[M]. 2nd ed. Berlin: Springer, 2012: 769. doi: 10.1007/978-3-642-35289-8.

作 者 简 介

魏志强，博士，教授，主要研究方向为高移动信道传输方法、高移动网络优化方法、感通一体化技术等。

张家烁，硕士生，主要研究方向为深度强化学习在高移动网络优化中的应用。

刘　凡，博士，教授，主要研究方向为雷达通信一体化、车联网、毫米波通信等。

杨　在，博士，教授，主要研究方向为信息处理与无线通信的数学理论与方法。

费泽松，博士，教授，主要研究方向为5G/6G移动通信关键技术、通信感知一体化、星地融合通信、智能通信、无线短距离通信等。

(责任编辑：于青)