基于鲁棒深度强化学习的IRS辅助分散计算网络保密速率和优化

李嘉欣[1],王建萍[1],刘之滨[2],林福宏[1*]
(1. 北京科技大学计算机与通信工程学院,北京 100083;2. 国家电网有限公司华北分部,北京,100053)

摘 要: 为解决恶劣无线通信环境下分散计算网络中高保密传输和高服务质量(Quality of Service,QoS)的协同需求,本文提出了一种智能反射面(Intelligent Reconfigurable Surface,IRS)辅助的分散计算网络保密通信与资源优化方案。首先,由于分散计算网络中的无人机(Unmanned Aerial Vehicle,UAV)节点受到自身能源的限制,本文研究了一种新颖的能量收集(Energy Harvesting,EH)方案,通过在几何空间上对IRS被动反射阵列进行功能划分,使部分反射单元用于信息反射,部分单元用于EH,从而实现信息传输与能量采集的协同进行。其次,构建了IRS辅助分散计算网络中的保密速率和最大化优化模型。该模型通过联合优化用户发射功率、IRS反射元件相移、EH约束以及通信QoS等多个耦合变量,以提升系统整体保密性能和资源利用效率。由于优化问题具有高度非凸性和变量强耦合特性,传统优化方法难以直接获得全局最优解。此外,考虑到分散计算网络中用户移动性强、无线信道动态变化快以及环境状态不确定等特点,本文设计了一种基于鲁棒深度强化学习(Deep Reinforcement Learning,DRL)的动态资源优化算法,以在动态分散计算网络环境中保证QoS。仿真结果表明:所提出的基于鲁棒DRL的IRS辅助分散计算网络方案性能不仅优于现有的其他基于学习的解决方案,还接近穷举法性能,最终验证了所提方案的有效性和优越性。

关键词: 分散计算网络;智能反射面(IRS);能量收集(EH);保密速率和;鲁棒深度强化学习(DRL)
基金项目: 国家自然科学基金(No.62436004);国家重点研发计划基金资助项目(No.2022YFB3104903)
中图分类号: TN929.5 **文献标识码:** A **文章编号:** 0372-2112(XXXX)XX-0001-12
电子学报URL:http://www.ejournal.org.cn **DOI:**10.12263/DZXB.20260219

Secure Sum Rate and Optimization in IRS-Assisted Dispersed Computing Network Based on Robust Deep Reinforcement Learning

LI Jiaxin[1], WANG Jianping[1], LIU Zhibin[2], LIN Fuhong[1*]
(1. *University of Science and Technology Beijing* (*USTB*), *Department of Computer and Communication Engineering*, *Beijing* 100083, *China*;
2. *North China Branch of State Grid Corporation of China*, *Beijing*, 100053, *China*)

Abstract: To address the collaborative requirements of high-security transmission and high quality of service (QoS) in dispersed computing network under harsh wireless communication environments, this paper proposes a secure communication and resource optimization scheme for intelligent reconfigurable surface (IRS)-assisted dispersed computing network. Firstly, due to the energy limitations of unmanned aerial vehicle (UAV) nodes, this paper studies a novel energy harvesting (EH) scheme. By dividing the IRS passive reflection array in geometric space, some reflection elements are used for information reflection, and some elements are used for EH, so as to realize the cooperation of information transmission and EH. Secondly, the secure sum rate maximization optimization model in IRS-assisted dispersed computing network is formulated. The model jointly optimizes multiple coupling variables such as user transmission power, IRS reflection element phase shift, EH constraint and communication QoS, while improve the overall system security performance and resource utilization efficiency. Since the formulated optimization problem is highly non-convex and the variables are strongly coupled, traditional optimization methods are difficult to directly obtain the global optimal solution. Furthermore, considering the characteristics of dispersed computing network, such as high user mobility, rapidly varying wireless channels, and uncertain environmental states, a robust deep reinforcement learning (DRL)-based dynamic resource optimization algorithm is designed to guarantee QoS in dynamic dispersed computing environments. Simulation results show that the performance of the IRS-assisted dispersed computing network scheme based on robust DRL proposed in this paper not only outperforms existing learning-based solutions but also achieves performance close to that of the exhaustive search method, verifying the effectiveness and superiority of the proposed scheme.

收稿日期:2026-03-11;录用日期:2026-05-16;责任编委:姚苏
*通信作者:林福宏

Keywords: dispersed computing network; intelligent reflecting surface (IRS); energy harvesting (EH); secure sum rate; robust deep reinforcement learning (DRL)

Foundation Item(s): National Natural Science Foundation of China (No. No.62436004); National Key Research and Development Program of China (No.2022YFB3104903)

0 引言

在复杂多变的通信环境中,山脉、建筑物等障碍物常常会阻碍地面基站与用户之间的直接通信链路,导致严重的信号衰减甚至通信中断[1]。分散计算作为云计算、雾计算和边缘计算的一种补充范式,拥有更可靠的算力支撑,能够充分利用网络中的计算资源,在军事行动、灾害救援及应急通信等场景中展现出显著优势[2]。然而,在复杂环境下实现高可靠与高安全通信仍然面临严峻挑战。近年来,智能反射面(Intelligent Reflecting Surface,IRS)因其低功耗、可编程信道重构能力等特点,被广泛应用于提升无线系统的频谱效率与物理层安全性能。在分散计算网络中引入IRS[3],可显著增强合法用户信道质量并抑制窃听链路。此外,与传统中继系统相比,IRS具有能耗更低、硬件成本更低、调控自由度更高以及频谱效率更显著等优势。然而,IRS的系统性能在很大程度上依赖于其部署位置,而用户的移动性与网络环境的动态性使得这一依赖尤为关键。目前,地面IRS通常固定部署于建筑物外墙或屋顶,受限于成本与城市规划,其位置难以根据实时需求进行动态调整。为此,将IRS搭载于无人机(Unmanned Aerial Vehicle,UAV)上,构成UAV辅助智能反射面(Unmanned Aerial Vehicle-Intelligent Reflecting Surface,UAV-IRS)分散计算网络,能够在分散计算网络中为通信受限区域提供灵活、普适的覆盖服务[4]。因此,将UAV、传感器网络、移动终端等资源受限的异构设备整合为一个有机的网络整体,通过节点间的计算、存储和网络资源共享机制,能够实现高效的任务处理和实时反馈。

在分散计算网络中引入UAV-IRS不仅能够提供灵活的部署优势,还为增强分散计算网络的物理层安全性开辟了新途径。具体来讲,在分散计算网络中引入IRS,通过智能调控入射信号的相位,能够主动塑造无线信道环境。这不仅增强了合法用户的接收信号质量,还能够有效地抑制窃听者的信道条件,从而从物理层面提升通信安全。在物理层安全方面,已有研究围绕IRS辅助安全传输展开了大量探索。文献[5]提出了基于IRS并利用混合波束成形来增强非正交多址接入网络的安全性。文献[6]研究了IRS辅助的认知无线电非正交多址接入网络的安全传输问题,通过联合优化发射波束成形、IRS的模式选择和相位向量以最大化总可达保密率。文献[7]提出利用IRS提高频谱感知的准确性以及次级用户的保密性能。文献[8]针对无线通信中的数据安全问题,提出一种利用IRS来实现分布式调制的安全通信方案。文献[9]提出在非正交多址接入网络中利用IRS和人工噪声实现安全传输。

尽管UAV-IRS结合了双方的优势,但其实际应用仍面临一个核心挑战:UAV有限的机载电池容量,严重制约了系统的续航时间与通信性能,难以满足所有用户持续的安全通信需求[10]。能量收集(Energy Harvesting,EH)技术,尤其是基于射频信号的无线信息和能量同时传输(Simultaneous Wireless Information and Power Transfer,SWIPT),是缓解该能源瓶颈的关键[11]。其中,“收集-传输-存储”作为一种高效的SWIPT模式,将每个传输时块划分为EH与信息传输两个阶段。然而,仅通过时域分割难以实现资源的高效全局优化,因为UAV-IRS系统的资源分配涉及发射功率、反射相位、传输调度与UAV轨迹的复杂联合优化。此外,当服务用户较少时,启动全部反射单元会造成资源浪费。近期研究提出了一种基于空间分割的EH模型,即利用部分反射单元收集能量,其余单元负责反射信息信号,从而在空间维度上提升了IRS的能效。因此,在空间维度上研究UAV-IRS辅助的安全通信系统的资源分配,有望最大程度地提升系统的整体能效与续航能力。然而,在UAV-IRS辅助的分散计算网络中实现系统的保密安全速率最大化,并同时满足严格的通信服务质量(Quality of Service,QoS)约束是一个复杂的非凸优化问题,其求解极具挑战。针对此类非凸问题,现有研究多采用交替优化、问题分解或基于惩罚的迭代等方法以获得次优解。其中,文献[12]提出利用多个可调相位的IRS来设计节能安全通信,通过开发一种基于逐次凸逼近和惩罚技术的高效交替优化算法以最大化保密能量效率,有效对抗多个窃听者。文献[13]研究利用IRS提高存在被动窃听者的边缘计算系统的安全计算性能,通过开发一种结合泰勒展开法、半定松弛算法、拉格朗日对偶理论和Karush-Kuhn-Tucker条件的迭代优化算法来解决非凸问题。文献[14]研究了IRS辅助的UAV-集成传感与通信网络的安全传输,提出了一种基于交替优化、逐步凸近似和流形优化的迭代算法以获得接近最优解。文献[15]针对一种采用功率分割模型的主动IRS辅助安全集成感知与SWIPT系统,提出两种交替优化算

法来联合优化发射波束形成、人工噪声向量、主动IRS的放大因子和相位偏移等变量以最大化获取到的能量。然而,这些方法通常针对特定场景设计,泛化能力有限。近年来,深度强化学习(Deep Reinforcement Learning,DRL)因其在处理耦合优化与实时决策方面的强大能力,被广泛应用于无线资源分配问题。这为利用DRL解决UAV-IRS系统中的安全通信问题提供了强大动力。例如,文献[16]利用DRL算法对IRS反射系数与用户功率进行联合优化,以提升系统能效或用户可达速率。文献[17]将DRL应用于IRS辅助语义通信或频谱共享网络,通过联合优化子信道分配、波束成形以及IRS反射矩阵来提升系统性能。在动态网络环境下,文献[18]将IRS与UAV相结合,并利用DRL算法实现通信资源与轨迹的联合优化,以适应时变信道环境。尽管上述研究验证了DRL在IRS系统资源分配中的潜力,但现有方法多采用深度确定性策略梯度(Deep Deterministic Policy Gradient,DDPG)或双延迟深度确定性策略梯度(Twin Delayed Deep Deterministic policy gradient,TD3)进行连续控制优化,仍存在一些局限性。具体来讲,DDPG能够处理高维连续动作空间,但其价值函数易产生过估计偏差;TD3通过双Q裁剪机制缓解过估计问题,但在复杂连续动作空间中可能引入系统性低估偏差,从而影响策略更新的稳定性。尤其在IRS辅助安全通信场景中,保密速率目标函数呈现对数差结构,对价值函数估计精度高度敏感,传统DRL框架难以保证稳定收敛。

为满足UAV-IRS辅助的分散计算网络高QoS和高保密需求,本文提出基于空域分割SWIPT机制来增强UAV的续航能力,同时设计一种鲁棒学习算法以满足系统的安全性能,主要贡献如下:

(1)提出一种面向UAV-IRS辅助分散计算网络的空域分割SWIPT安全通信模型。针对UAV能量受限问题和恶劣通信环境中的高保密和高QoS的协同需求,本文突破了传统时域分割资源分配方法的局限性。在UAV-IRS辅助分散计算安全通信场景中引入空域分割EH机制,对IRS反射单元进行功能级划分,实现信息反射与EH的并行运行,从而显著提升了系统的能效与续航能力,为IRS辅助安全通信提供了一种更高效的资源利用方式。

(2)构建混合整数非凸保密速率最大化优化问题,刻画了QoS与能量约束下的资源分配机制。在所提出的空域分割模型基础上,建立了一个以系统可实现保密速率和最大化为目标的混合整数非凸优化问题。在保障QoS和EH需求的同时,通过联合优化用户发射功率、IRS反射相位及反射单元调度变量,实现安全性能的整体提升。该问题高度耦合且非凸,传统优化方法难以直接高效求解。

(3)为了解决所构建的非凸优化问题,设计了一种鲁棒DRL框架。针对传统主流DRL算法如DDPG存在的价值函数过估计问题以及TD3在复杂连续动作空间中引入的低估偏差,本文提出一种基于归一化指数算子的鲁棒DRL算法。该方法能够有效提升策略更新的稳定性,在空域环境下对UAV-IRS系统的动态资源分配进行动态规划。仿真结果表明:本文所提出的空域分割节能安全框架在IRS辅助分散计算网络的安全性能方面是有效的,同时所提算法在不同用户规模的情况下,性能不仅优于现有的主流学习算法,还接近穷举法的最优性能。

1 系统模型与问题描述

如图1所示,本文考虑了在恶劣通信环境中UAV-IRS辅助的分散计算安全通信网络,其中地面部署了一个具有Z个天线的接入点(Access Point,AP)、K个单天线终端用户和E个单天线窃听者,空中部署了配备IRS的UAV,每个IRS具有$\mathcal{L}$个反射单元。此外,本文将整个时间段划分为T个相等的时隙,记为$\mathcal{T}=\{1,2,\cdots,t,t+1,\cdots,T\}$。令$\mathcal{K}=\{1,2,\cdots,K\}$,$\mathcal{E}=\{1,2,\cdots,E\}$,$\mathcal{L}=\{M\times N\}$分别表示终端用户、窃听者及IRS反射元件。每个终端用户k在时隙t时的天线位置表示为$\mathcal{C}^k(t)=\left(x^k(t),y^k(t),H^k(t)\right)$。在以AP为参考原点的笛卡尔坐标系下,终端用户$k$天线的位置由其高度坐标$H^k(t)$和水平坐标$\left(x^k(t),y^k(t)\right)$共同确定。本文假设从基站到用户的直接信号链路因恶劣的通信环境而被阻断,因此考虑由UAV-IRS辅助的两跳通信系统,即AP通过UAV-IRS的反射和中继向用户发送信息,其中UAV-IRS由$\mathcal{L}=\{M\times N\}$个反射单元组成,且用户只能接收经UAV-IRS反射后的信号。将位于第i行、第j列的反射单元表示为$\mathcal{R}_{i,j}$,其在时隙t的空间位置可表示为$\mathcal{C}^r_{i,j}(t)=\left(x^r_{i,j}(t),y^r_{i,j}(t),H^r_{i,j}(t)\right)$,其中$H^r_{i,j}(t)$、$x^r_{i,j}(t)$和$y^r_{i,j}(t)$分别表示该反射单元的垂直位置及其在水平平面内的坐标。此外,反射单元的位置与UAV的轨迹相关联,为了不失一般性,在UAV-IRS系统中,反射阵列表示为$\mathcal{R}=\{\mathcal{R}\}^{M,N}_{i,j=1}$。IRS可以通过所附的智能控制器与AP交换信道状态信息。鉴于在基站到用户的数据传输过程中存在持续的未授权窃听者截取数据的风险,UAV与IRS必须协同工作,优化AP传输功率和IRS相位偏移,以提高用户可用的数据传输速率,同时降低被窃听数据的传输速率。此外,本文创新性地提出了一种空域分割的EH资源分配模型,以增强UAV在传输信号时的续航能力。

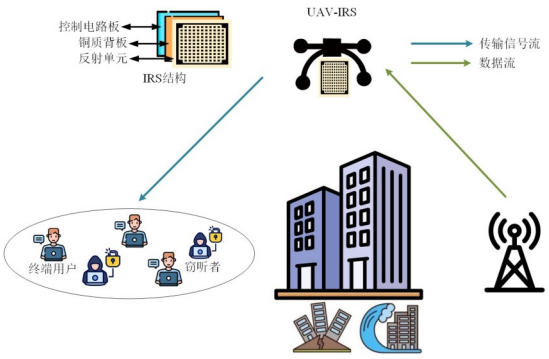

图1 UAV-IRS辅助安全通信系统模型

Figure 1 UAV-IRS assisted secure communication system model

1.1 通信模型

本节建模通信模型。在AP向终端用户传输信息的过程中,IRS的各反射单元用于对入射信号进行反射。参照文献[19],AP的基带发射信号可表示为

$$\boldsymbol{X}=\sum_{k\in\mathcal{K}}\boldsymbol{V}_k S_k \tag{1}$$

其中,S_k与$\boldsymbol{V}_k\in\mathbb{C}^{D\times 1}$分别表示第$k$个终端用户的数据信号和其对应的预编码向量,且$S_k$假设服从零均值、单位方差的环形对称复高斯分布,即$S_k\sim\mathcal{CN}(0,1)$。由式(1)可得,AP的总发射功率可写为[20]

$$\mathbb{E}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})=\sum_{k\in\mathcal{K}}\left\|\boldsymbol{V}_k\right\|^2\leqslant p_{\max} \tag{2}$$

其中,$p_{\max}$表示AP的最大发射功率约束;$\|\cdot\|$表示向量的欧几里得范数;而$p_k=\left\|\boldsymbol{V}_k\right\|^2$表示分配给终端用户$k$的发射功率。在每个时隙$t$进行信息传输时,令$\boldsymbol{G}\in\mathbb{C}^{Z\times\mathcal{L}}$、$\boldsymbol{h}_{r,k}=[h_{1,1}(k),h_{1,2}(k),\cdots,h_{1,N}(k),h_{2,N}(k),\cdots,h_{M,N}(k)]$和$\boldsymbol{h}_{r,e}=[h_{1,1}(e),h_{1,2}(e),\cdots,h_{1,N}(e),h_{2,N}(e),\cdots,h_{M,N}(e)]$分别表示AP到UAV-IRS的信道、UAV-IRS到终端用户k和UAV-IRS到窃听者e的信道。本文假定信道矩阵中的小尺度信道衰落服从瑞利衰落分布。从AP到每个反射单元$\mathcal{R}_{i,j}$的信道向量$\boldsymbol{g}_{i,j}$的路径损耗$\mathrm{PL}_{i,j}$表示为[21]

$$\begin{aligned}\mathrm{PL}_{i,j}=&\left(P_{i,j}(\mathrm{LoS})+\left(1-P_{i,j}(\mathrm{LoS})\right)\varphi\right)\\&\times\left(\sqrt{\left|x_{i,j}^r(t)\right|^2+\left|y_{i,j}^r(t)\right|^2+\left|H_{i,j}^r(t)\right|^2}\right)^{-\alpha}\end{aligned} \tag{3}$$

其中,α表示从反射单元$\mathcal{R}_{i,j}$到AP的路径损耗指数;φ表示由非视距连接引起的额外衰减系数;$P_{i,j}(\mathrm{LoS})$表示AP与反射单元$\mathcal{R}_{i,j}$之间的视距概率。根据文献[22],视距概率$P_{i,j}(\mathrm{LoS})$的计算式为

$$P_{i,j}(\mathrm{LoS})=\frac{1}{1+\mathcal{A}\times\exp\left(-\mathcal{B}(\theta_{i,j}-\mathcal{A})\right)} \tag{4}$$

其中,$\mathcal{A}$和$\mathcal{B}$表示取决于环境条件的常数。AP与反射单元$\mathcal{R}_{i,j}$之间的仰角计算式为[23]

$$\theta_{i,j}=\frac{180}{\pi}\arcsin\left(\frac{H_{i,j}^r(t)}{\sqrt{\left|x_{i,j}^r(t)\right|^2+\left|y_{i,j}^r(t)\right|^2+\left|H_{i,j}^r(t)\right|^2}}\right) \tag{5}$$

此外,UAV-IRS通过控制反射相移被动反射接收到的信息信号,则UAV-IRS的反射系数矩阵可以表示为[24]

$$\boldsymbol{\Phi}=\mathrm{diag}(\varpi_1 e^{\mathrm{j}\theta_1},\varpi_2 e^{\mathrm{j}\theta_2},\ldots,\varpi_L e^{\mathrm{j}\theta_L})\in\mathbb{C}^{\mathcal{L}\times\mathcal{L}} \tag{6}$$

其中,$\mathrm{j}=\sqrt{-1}$表示虚数单位;$\theta_r^l\in(0,2\pi)$表示第l个反射单元的相移;而$\varpi_l\in[0,1]$表示反射幅度系数。此外,为了便于分析,在理想假设下将ϖ_l设为单位值,即认为各反射单元的天线能够实现独立调控,从而充分提升反射效率。基于式(1),经由AP-IRS-终端用户级联信道,第k个终端用户以及第e个窃听者在接收端获得的射频信号可分别表示为[25]:

$$y_k=\hat{\boldsymbol{h}}_{r,k}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{X}+v_k,k\in\mathcal{K}, \tag{7}$$

$$y_e=\hat{\boldsymbol{h}}_{r,e}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{X}+v_e,e\in\mathcal{E}, \tag{8}$$

其中,第k个终端用户和第e个窃听者处的接收噪声分别记为$v_k\sim\mathcal{CN}(0,\sigma_k^2)$与$v_e\sim\mathcal{CN}(0,\sigma_e^2)$,将其统一建模为加性白高斯噪声,噪声功率分别记为$\sigma_k^2$和$\sigma_e^2$。同时,记$\hat{\boldsymbol{h}}_{r,k}$为IRS调度下UAV-IRS-用户$k$的级联信道矩阵,其具体形式可表示为

$$\hat{\boldsymbol{h}}_{r,k}=\begin{bmatrix}\omega_{1,1}^k h_{1,1}(k) & \cdots & \omega_{1,N}^k h_{1,N}(k)\\ \vdots & \vdots & \vdots\\ \omega_{M,1}^k h_{M,1}(k) & \cdots & \omega_{M,N}^k h_{M,N}(k)\end{bmatrix} \tag{9}$$

在信道建模过程中,本文同时引入了$\boldsymbol{h}_{r,k}$链路中的路径损耗和小尺度衰落因素。UAV-IRS与终端用户之间的路径损耗建模为$\kappa\left(\frac{d_{i,j}^k(t)}{d'}\right)^{-\bar{\alpha}}$,其中$\kappa$表示在参考距离$d'=1$ m处的路径损耗系数,$d_{i,j}^k(t)=\left\|\mathcal{C}^k(t)-\mathcal{C}_{i,j}^r(t)\right\|_2$为反射单元$\mathcal{R}_{i,j}$与终端用户$k$之间的欧几里得距离,$\bar{\alpha}$表示UAV-IRS至终端用户链路的路径损耗指数。对于信道$\boldsymbol{h}_{r,k}$中的小尺度衰落,假定其服从瑞森衰落模型,其瑞森因子$K_{\mathrm{rician}}=10$,表达式为[26]

$$\boldsymbol{h}_{r,k}=\sqrt{\frac{K_{\mathrm{rician}}}{1+K_{\mathrm{rician}}}}\boldsymbol{h}_{r,k}^{\mathrm{LoS}}+\sqrt{\frac{1}{1+K_{\mathrm{rician}}}}\boldsymbol{h}_{r,k}^{\mathrm{NLoS}} \tag{10}$$

其中,$\boldsymbol{h}_{r,k}^{\mathrm{NLoS}}$和$\boldsymbol{h}_{r,k}^{\mathrm{LoS}}$分别表示随机的非视距(Non-Line-Of-Sight,NLOS)分量与确定性的视距(Line-Of-Sight,LOS)分量。以此类推,记窃听者e的级联信道矩阵和小尺度衰落为$\hat{\boldsymbol{h}}_{r,e}$和$\boldsymbol{h}_{r,e}$,具体表示如前所述。参考文献[27]中的理想串行干扰消除模型,本文在理论分析阶段假设终端用户在检测目标信号之前能够有效抑制来自其他反射路径的干扰分量。需要指出的是,该假设主要用于获得可解析的闭式表达形式,并刻画

系统的性能上界。在实际系统中,由于信道估计误差与硬件非理想性,可能存在残余干扰项,其将导致可实现保密速率有所下降。然而,该影响不会改变本文所提算法的优化框架与收敛机制,仅影响具体性能数值。因此,本文模型可视为理想上界情形,后续工作将进一步考虑不完全干扰消除条件下的鲁棒设计。由此可得,第k个用户终端和第e个窃听者的接收信噪比(Signal-to-Noise Ratio,SNR)表达式为:

$$\mathrm{SNR}_k=\frac{|\hat{\boldsymbol{h}}_{r,k}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{V}_k|^2}{\sigma_k^2} \tag{11}$$

$$\mathrm{SNR}_e=\frac{|\hat{\boldsymbol{h}}_{r,e}^{\mathrm{H}}\boldsymbol{\Phi}^{\mathrm{H}}\boldsymbol{G}^{\mathrm{H}}\boldsymbol{V}_e|^2}{\sigma_e^2} \tag{12}$$

因此,终端用户的可实现保密速率为[28]

$$R_k=[\log_2(1+\mathrm{SNR}_k)-\max_{\forall e\in\mathcal{E}}\log_2(1+\mathrm{SNR}_e)]^+ \tag{13}$$

其中,$[X]^+=\max\{X,0\}$。系统的可实现保密速率和表示为

$$R=\sum_{k=1}^{K}R_k \tag{14}$$

此外,根据香农公式,在时隙t中第k个终端用户的可达保密速率(bits/second/Hz)表示为[29]

$$\Gamma_k(t)=B\log_2(1+\mathrm{SNR}_k),k\in\mathcal{K},t\in\mathcal{T} \tag{15}$$

其中,系统带宽记为B。此外,在给定的时间区间内,为保证QoS约束,第k个用户终端的可达保密速率应满足不小于$\Gamma_{\min}$的条件,可表示为

$$\Gamma_k(t)\geqslant\Gamma_{\min},\forall k\in\mathcal{K},t\in\mathcal{T} \tag{16}$$

1.2 系统能耗模型

图2展示了UAV-IRS在通信盲区中的典型工作场景。在数据传输过程中,UAV-IRS内的反射单元被划分为信号反射与EH两类功能模块。对应地,在时隙t下,UAV-IRS的能量采集量可表示为[30]

$$E(t)=t\sum_{i=1}^{M}\sum_{j=1}^{N}(1-\sum_{k\in\mathcal{K}}^{K}\omega_{i,j}^{k})\eta\left\|\boldsymbol{g}_{i,j}^{H}\boldsymbol{X}\right\|^2 \tag{17}$$

其中,能量效率$\eta\in(0,1)$;$\boldsymbol{g}_{i,j}=[g_{i,j}^1,\cdots,g_{i,j}^z,\cdots,g_{i,j}^Z]$表示具有$Z$个天线的AP与反射单元$\mathcal{R}_{i,j}$之间的信道向量,其中该信道遵循空气到地面传播模型的路径损耗规律[31]。AP的传输功率为$p=\mathbb{E}(\boldsymbol{X}^H\boldsymbol{X})$。其中,当$\omega_{i,j}^k=0$时,表示IRS中的反射单元$\mathcal{R}_{i,j}$用于收集能量,$\omega_{i,j}^k=1$时表示IRS中的反射单元$\mathcal{R}_{i,j}$向第个终端用户反射信号。此外,表1列出了系统模型中的变量符号并加以说明。

1.3 优化问题描述

本文研究了在满足所需的最小吞吐量约束的同时,在有限的时间范围内最大化UAV-IRS辅助的安全通信网络的可实现保密速率和。因此,优化问题表述如下:

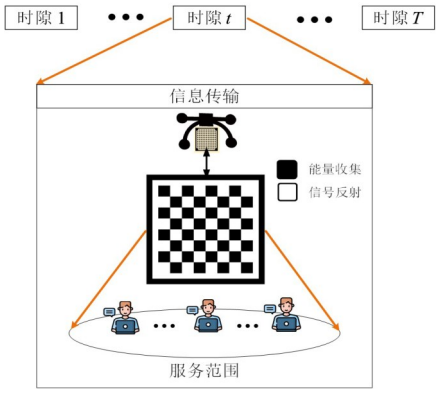

图2 每个时隙EH与信号传输模型

Figure 2 EH and signal transmission model for each time slot

表1 符号说明表

Table 1 Symbol description

符号	说明
$\boldsymbol{X}$	AP发射信号
$\mathbb{E}(\boldsymbol{X}^{\mathrm{H}}\boldsymbol{X})$	AP发射功率
y_k/y_e	用户/窃听者的射频信号
$\hat{\boldsymbol{h}}_{r,k}/\hat{\boldsymbol{h}}_{r,e}$	UAV-IRS-用户/窃听者的级联信道矩阵
$\boldsymbol{h}_{r,k}/\boldsymbol{h}_{r,e}$	UAV-IRS-用户/窃听者的小尺度衰落
$\mathrm{SNR}_k/\mathrm{SNR}_e$	用户/窃听者的SNR
R_k	用户k的可实现SNR
R	可实现保密速率和
$E(t)$	时隙t内收集的能量
$\Gamma_k(t)$	时隙t中第k个用户的可达保密速率

$$(\mathrm{P1}):\ \max_{\boldsymbol{P},\Theta,\omega}\sum_{t=1}^{T}R_k(t) \tag{18}$$

$$\text{s.t. }\Gamma_k(t)\geqslant\Gamma_{\min},\quad\forall k\in\mathcal{K},t\in\mathcal{T} \tag{18a}$$

$$0\leqslant p=\sum_{k=1}^{K}\left\|\boldsymbol{v}_k\right\|^2\leqslant p_{\max} \tag{18b}$$

$$0\leqslant p_k\leqslant p'_{\max},\quad\forall k\in\mathcal{K} \tag{18c}$$

$$\omega_{i,j}^k\in\{0,1\},\forall i\in[0,M],j\in[0,N],k\in\mathcal{K} \tag{18d}$$

$$\sum_{k\in\mathcal{K}}\omega_{i,j}^k\leqslant 1\quad\forall k\in\mathcal{K} \tag{18e}$$

$$\theta_l^r\in[0,2\pi],\quad\forall l\in[0,\mathcal{L}] \tag{18f}$$

$$\left|\mathrm{e}^{\mathrm{j}\theta_l^r}\right|=1 \tag{18g}$$

$$E(t)\geqslant E_{\min},\quad\forall t\in\mathcal{T} \tag{18h}$$

其中,$\boldsymbol{P}=[p_1,p_2,\cdots,p_K]$表示每个终端$K$个用户的发射功率向量;$p'_{\max}$表示每个终端用户的发射功率上限。$\boldsymbol{\Theta}=[\theta_1^r,\theta_2^r,\cdots,\theta_{\mathcal{L}}^r]$表示IRS上所有反射元件的相位偏移向量。$\boldsymbol{\omega}$是IRS的反射单元调度矩阵,表示为

$$\boldsymbol{\omega}=\begin{bmatrix}\omega_{1,1}^1\cdots\omega_{1,N}^1\cdots\omega_{M,N}^1\\ \vdots\quad\vdots\quad\vdots\\ \omega_{1,1}^k\cdots\omega_{1,N}^k\cdots\omega_{M,N}^k\end{bmatrix} \tag{19}$$

约束条件(18a)用于刻画终端用户的最小吞吐量需求,以确保无线通信的QoS。式(18b)和式(18c)分别给出了AP与终端用户k的最大发射功率限制。反射单元调度相关的二进制变量$\omega_{i,j}^{k}$约束由式(18d)和式(18e)给出。此外,式(18f)与式(18g)规定IRS中的反射单元l仅具备相位控制能力,只能引入相位偏移$\theta_l^r\in[0,2\pi]$而不具备信号放大功能。最后,约束条件(18h)对应系统的EH约束条件。问题(P1)为混合整数非凸优化问题,非凸性来源于多个变量之间存在高度耦合、约束条件中的相位偏移向量是单位模约束以及存在二进制调度变量。因此,难以使用标准的凸优化方法来有效地解决问题(P1),采用基于DRL的近似动态规划法来进行求解。

2 DRL优化框架

优化问题(P1)由于非凸约束和多个变量的耦合而具有非凸性,传统的优化方法在实际应用中难以在有限的时间内获得高效的解。因此,本文提出了一种基于DRL的优化框架来解决这个问题。然而,传统的DRL算法往往存在高估和低估的问题,这在复杂的无线通信环境中会降低性能。因此,本文通过提出一种新型IRS能量协同框架,构建混合整数动态保密优化模型,并提出一种鲁棒DRL框架来解决连续控制过/低估问题。此外,本文所提出的鲁棒性主要体现在价值函数估计稳定性与策略收敛稳定性两个方面。传统DDPG算法由于单评价网络结构易产生过估计偏差,而TD3虽然缓解了过估计问题,但其最小值裁剪机制在复杂连续动作空间中可能引入系统性低估偏差。因此,本文通过引入归一化指数算子构造软最小运算,使目标值在统计意义上更加接近真实期望值,从而同时抑制过估计与低估现象。因此,在信道动态变化与用户移动环境下,所提框架能够获得更稳定的策略更新过程和更平滑的收敛轨迹。

2.1 鲁棒DRL框架

在强化学习中构建的马尔可夫决策过程可表示为[32]

$$\mathcal{G}:=\left\langle S,A,\mathcal{P},\mathcal{R},\gamma\right\rangle \tag{20}$$

其中,S和A分别表示状态集和动作集。$\mathcal{R}:S\times A\times S\to\mathbb{R}$表示状态奖励函数,它规定了特定状态间转换的奖励。状态转换概率表示$\mathcal{P}:S\times A\times S\to[0,1]$将当前环境状态与动作交互所产生的概率分布映射到下一个环境状态。折扣因子$\gamma\in[0,1]$决定了未来奖励相对于当前状态的重要性。在每个一致性时间步t中,智能体会根据当前环境状态$s_t\in S$及其策略π_*采取动作$a_t=\pi_*(s_t)$。随后,智能体将获得即时奖励$r_t=\mathcal{R}(s_t,a_t)$及演变后的状态$s_{t+1}\in S$。通常,奖励函数$\mathcal{R}$和转换函数$\mathcal{P}$构成了马尔可夫决策过程模型$\pi_*:S\to A$,该模型用来最大化由所计算的长期奖励所确定的值

$$\max_{\pi_*}J(\pi_*):=\mathbb{E}\left[\sum_{t=0}^{T}\gamma^t r_t\left(s_t,\pi_*(s_t)\right)\right] \tag{21}$$

同样地,动作值函数(Q函数)定义为

$$Q^{\pi_*}(s_t,a_t)=\mathbb{E}\left[\sum_{t=0}^{T}\gamma^t r_t|s_0=s,a_0=a,a_t\sim\pi_*(\cdot|s_t)\right] \tag{22}$$

先前的研究表明:在Q学习中探索连续动作空间将会耗费大量时间[33]。DDPG使用一个确定性策略$\pi_*(s|\delta^\pi)$,其中其函数逼近器由δ^π参数化,以在连续动作空间中最大化Q函数[34]。通过贝尔曼方程学习由δ^π参数化的评价网络$Q(s,a|\delta^Q)$来评价动作网络的性能。动作网络和评价网络的副本$\pi'(s|\delta^{\pi'})$和$Q'(s,a|\delta^{Q'})$被创建作为快速收敛的目标网络。在每一步中,DDPG通过从随机噪声过程$\mathcal{N}$中采样噪声来创建用于连续动作空间学习的探索策略

$$\pi'(s)=\pi(s|\delta^{\pi'})+\mathcal{N} \tag{23}$$

而$\mathcal{N}$可根据具体环境进行选择。综合来看,该代理网络将使用以下近似方法来更新其策略

$$\nabla_{\delta^\pi}J\approx\frac{1}{N_b}\sum_i[\nabla_a Q(s,a|\delta^Q)|_{s_i,a=\pi(s_i)}\nabla_{\delta^\pi}\pi(s|\delta^\pi)|_{s_i}] \tag{24}$$

其中,N_b表示从重放缓冲区$\mathcal{D}$中随机抽取的小批量样本中的转换数量。评价网络根据以下方式更新其策略,以最小化损失

$$L(\delta^Q)=\frac{1}{N_b}\sum_{i=1}^{N_b}\left(y_i-Q(s_i,a_i|\delta^Q)\right)^2 \tag{25}$$

其中,y_i表示为

$$y_i=r(s_i,a_i)+\gamma Q'\left(s_{i+1},\pi'(s_{i+1}|\delta^{\pi'})|\delta^{Q'}\right) \tag{26}$$

随后,DDPG会按照以下方式更新目标网络的权重:

$$\begin{aligned}\delta^{Q'}&\leftarrow\psi\delta^Q+(1-\psi)\delta^{Q'}\\ \delta^{\pi'}&\leftarrow\psi\delta^\pi+(1-\psi)\delta^{\pi'}\end{aligned} \tag{27}$$

其中,$\psi\ll1$表示用于软更新策略型动作网络和评价网络的学习率。

然而,DDPG的一个关键问题在于过度估计现象[35]。针对这一问题,文献[36]指出,TD3算法利用带剪裁的双估计器Q_1和Q_2来对判别器进行改进,显著提高了DDPG的收敛速度和性能。与双Q学习公式类似,这对判别器(Q_1,Q_2)由$(\delta^{Q_1},\delta^{Q_2})$参数化[37]。因此,在TD3算法中,本文通过使用带剪裁的双Q学习方法,从两个判别器中选取最小的估计值

$$y_1,y_2=r+\gamma\min_{i=1,2}Q_i\left(s',\pi'(s'|\delta^{\bar{\pi}})|\delta^{\bar{Q}_i}\right) \tag{28}$$

其中,$\delta^{\bar{Q}_1}$和$\delta^{\bar{Q}_2}$分别表示目标强化学习网络的参数。

因此,通过使用剪枝双Q学习方法,可以减少对价值目标的任何额外高估。然而,TD3仍然存在低估偏差的问题,这显著降低了其性能[38]。

为解决此问题,本文提出在TD3中使用归一化指数函数,以减少连续控制中的任何过度估计和低估偏差。归一化指数函数操作符的定义为

$$\mathrm{softmax}_{\beta}\big(Q(s,\cdot)\big)=\int_{a\in A}\frac{\exp\big(\beta Q(s,a)\big)}{\int_{a'\in A}\exp\big(\beta Q(s,a')\big)\mathrm{d}a'}Q(s,a)\mathrm{d}a \tag{29}$$

其中,β表示归一化指数函数运算的参数。通过使归一化指数函数运算能够表示算法中Q函数的期望值,从而得到了一个无偏的估计值,计算式为

$$\mathrm{softmax}_{\beta}\big(Q(s,\cdot)\big)=\mathbb{E}_{a'\sim p}\Big[\frac{\exp\big(\beta\hat{Q}(s',a')\big)\hat{Q}(s',a')}{p(a')}\Big]\Big/\mathbb{E}_{a'\sim p}\Big[\frac{\exp\big(\beta\hat{Q}(s',a')\big)}{p(a')}\Big] \tag{30}$$

其中,$p(a')$表示遵循高斯分布的概率;$\hat{Q}_i(s',\cdot)$表示所有评价网络的最小估计值,其表达式为

$$\hat{Q}_i(s',a')=\min\Big(Q_i(s',a'|\delta^{\bar{Q}_i}),Q_j(s',a'|\delta^{\bar{Q}_j})\Big) \tag{31}$$

其中,Q_j表示除关键网络Q_i之外的所有关键网络的索引。目标关键网络Q_i的估计值计算式为

$$y_i=r+\gamma\mathcal{T}_{\mathrm{RD}}(s') \tag{32}$$

其中,$\mathcal{T}_{\mathrm{RD}}(s')$表示在连续动作空间中的归一化指数函数运算符,其表达式为

$$\mathcal{T}_{\mathrm{RD}}(s')=\mathrm{softmax}_{\beta}\big(\hat{Q}_i(s',\cdot)\big) \tag{33}$$

此外,所选取的动作通过将一个噪声$\mathcal{N}$加到目标动作$\pi(s',|\delta^{\bar{\pi}})$上得到。由于所选取的每个噪声都被限制在区间$[-c,c]$,因此所选取的动作可以表示为

$$a'=[-c+\pi(s',|\delta^{\bar{\pi}}),c+\pi(s',|\delta^{\bar{\pi}})] \tag{34}$$

该算法的一个实际优势在于其动作空间的有限范围能够确保所采取的动作与原始动作较为接近。因此,本文所提算法能够获得关于归一化指数Q函数的准确且鲁棒的估计值。

算法1展示了本文所提鲁棒学习算法(Robust Deep reinforcement learning,RD)的具体实现细节。本文首先将通信环境状态定义为所提出算法的输入,而一对行为网络$\pi_1(s|\delta^{\pi_1})$和$\pi_2(s|\delta^{\pi_2})$以及一对评价网络$Q_1(s|\delta^{Q_1})$和$Q_2(s|\delta^{Q_2})$则分别以随机参数对$(\delta^{\pi_1},\delta^{\pi_2})$和$(\delta^{Q_1},\delta^{Q_2})$进行初始化。其次,所有行为网络和评价网络的目标网络均使用与其对应网络相同的参数进行初始化。本文为学习过程初始化了一个大小为$N_{\mathcal{D}}$的空回放缓冲区$\mathcal{D}$。在每个时间步,行为网络根据目前策略对(π_1,π_2)以及剪裁后的探索噪声$\mathcal{N}$生成一个动作a_t。再次,该算法在执行相应动作后获取即时奖励r_t。接下来,元组(s,a_t,r_t,s',d)被存储到$\mathcal{D}$中,其中d是完成标志。随后,从重放记忆$\mathcal{D}$中立即抽取一个小型的N_B转换批次,并根据式(30)通过归一化指数函数来计算目标Q值。根据贝尔曼损失函数

$$\frac{1}{N_b}\sum_{s}\Big(y_i-Q_i(s_i,a_i|\delta^{Q_i})\Big)^2 \tag{35}$$

分别更新价值网络Q_i、策略网络π_i以及策略梯度

$$\frac{1}{N_b}\sum_{s}[\nabla_a Q_i(s,a|\delta^{Q_i})|_{a=\pi(s|\delta^{Q_i})}\nabla_{\delta^{\pi_i}}\big(\pi(s|\delta^{\pi_i})\big)] \tag{36}$$

最后,目标网络会以如下方式进行软更新:

$$\begin{aligned}\delta^{\bar{Q}_i}&\leftarrow\psi\delta^{Q_i}+(1-\psi)\delta^{\bar{Q}_i}\\ \delta^{\bar{\pi}_i}&\leftarrow\psi\delta^{\pi_i}+(1-\psi)\delta^{\bar{\pi}_i}\end{aligned} \tag{37}$$

该算法的输出结果为最优动作$\boldsymbol{a}=(\boldsymbol{P},\Theta,\boldsymbol{\omega})$以及UAV-IRS系统可实现保密速率和$\bar{R}$。

算法1 鲁棒学习算法:RD算法

输入: $\boldsymbol{G},\boldsymbol{h}_{r,k},\boldsymbol{h}_{r,e},d_{i,j}^{k},C_{i,j}^{r},\mathcal{C}^{k},N_{\mathcal{D}},N_b$

输出: 最优动作策略$\boldsymbol{a}=(\boldsymbol{P},\Theta,\boldsymbol{\omega})$,$R$

1. 初始化具有随机参数δ^{π_1}和δ^{Q_1}的动作空间$\pi_1(s|\delta^{\pi_1})$和评论家网络$Q_1(s|\delta^{Q_1})$;
2. 初始化具有随机参数δ^{π_2}和δ^{Q_2}的动作空间$\pi_2(s|\delta^{\pi_2})$和评论家网络$Q_2(s|\delta^{Q_2})$;
3. 初始化目标网络$\delta^{\bar{\pi}_1}\leftarrow\delta^{\pi_1},\delta^{\bar{Q}_1}\leftarrow\delta^{Q_1},\delta^{\bar{\pi}_2}\leftarrow\delta^{\pi_2},\delta^{\bar{Q}_2}\leftarrow\delta^{Q_2}$;
4. For 每轮训练 episode do
5. 接收当前的$\boldsymbol{G}$,初始化一个随机噪声过程$\mathcal{N}$,收集每个周期的$\boldsymbol{h}_{r,k}$和$\boldsymbol{h}_{r,e}$;
6. For 每一时刻t do
7. 根据策略(π_1,π_2)选择具有噪声$\mathcal{N}$的动作a_t;
8. 执行动作a_t以观察其对应的奖励r_t、下一个状态s'和完成标志d;将转移元组(s,a_t,r_t,s',d)存储到$\mathcal{D}$中;
9. For $i=1,2$ do
10. 从$\mathcal{D}$中随机抽取一个大小为N_b的小批量转移元组$\{(s,a_t,r_t,s',d)\}$;
11. 根据式(35)更新评论家网络;
12. 根据式(36)使用策略梯度更新动作空间;
13. 根据式(37)执行目标网络软更新;
14. End For
15. End for
16. End for

2.2 状态与动作及奖励函数

在本研究中,DRL环境基于无线网络假设构建,而IRS则作为智能体参与其中。其状态空间、动作空间以及奖励函数定义如下:

(1)状态空间(State):在每个时间步长t,观测值是由当前环境状态s_t构建而成的,该状态s_t包含从AP到UAV-IRS的信道$\boldsymbol{G}$以及从UAV-IRS到第k个终

端用户的信道 $\boldsymbol{h}_{r,k}\in\mathbb{C}^{1\times\mathcal{L}}$；对于所有的用户 $k\in\mathcal{K}$，每个反射单元 $\mathcal{R}_{i,j}$ 与第 k 个终端用户的距离 $d_{i,j}^{k}$；对于所有的用户 $k\in\mathcal{K}$，每个反射单元的位置 $\mathcal{C}_{i,j}^{r}$ 以及每个用户天线的位置 $\mathcal{C}^{k}$。因此，所提出的RD学习算法观测值表示为

$$O(s_t)=\{\boldsymbol{G},\boldsymbol{h}_{r,k},d_{i,j}^{k},\mathcal{C}^{k},\mathcal{C}_{i,j}^{r}\} \tag{38}$$

（2）动作空间（Action）：在第 t 个时间步内，所提出的基于DRL的空域高保密速率方案的动作由三个主要部分组成，即每个终端用户 k 的发射功率范围 $p_k\in[0,p'_{\max}]$；每个反射单元 l 的相位偏移 $\theta_l^r\in[0,2\pi]$；反射单元调度变量 $\omega_{i,j}^{k}\in[0,1],\forall i\in[0,M],j\in[0,N],k\in\mathcal{K}$。此外，$p_k$ 和 θ_l^r 定义在一个连续可行区域，$\omega_{i,j}^{k}$ 被转换为离散变量。

（3）奖励函数（Reward）：积极的奖励代表了所提出框架的目标，即最大限度地提高UAV-IRS系统的整体保密速率和。为兼顾系统整体保密速率最大化目标与QoS约束满足需求，本文构建了带约束惩罚机制的复合奖励函数。具体而言，当用户满足最小可达保密速率要求时，系统获得正向奖励；否则引入惩罚项以抑制违反约束的策略选择。该设计实质上构造了一个近似拉格朗日松弛结构，使强化学习过程在无显式约束优化器的情况下实现软约束满足。此外，在每个时间步长 t 中，即时奖励与保密速率和 R 存在正相关关系，所提出的框架还必须考虑到约束条件（16）中定义的用户最低可达保密速率要求。因此，奖励 r_t 表示为

$$r_t=R(t)\times\rho \tag{39}$$

其中，ρ 表示满足 $\Gamma_{\min}$ 要求的用户终端数量，并定义为

$$\rho=\prod_{k\in\mathcal{K}}\rho_k(t) \tag{40}$$

其中，$\rho_k(t)$ 表示为

$$\rho_k(t)=\begin{cases}0, & \Gamma_k(t)<\Gamma_{\min},\forall k\in\mathcal{K},t\in\mathcal{T}\\ 1, & \Gamma_k(t)\geqslant\Gamma_{\min},\forall k\in\mathcal{K},t\in\mathcal{T}\end{cases} \tag{41}$$

累积奖励的计算公式为 $\max J=\sum_t\gamma^t r_t$。

3 仿真分析

本节通过仿真验证所提出的鲁棒学习RD算法在UAV-IRS辅助的安全通信系统中的有效性。

3.1 仿真参数设置

在训练阶段，UAV-IRS的飞行轨迹依据文献[39]中提出的密度感知部署策略进行配置。具体而言，所提出的方法通过结合密度感知与费马点两种UAV轨迹方案进行对比评估。此外，在训练阶段，终端用户的轨迹与训练阶段不同。针对单用户与多用户场景，终端用户数量分别设定为 $K=1$ 和 $K=3$，且所有终端用户均位于20 m × 20 m的区域内。IRS由16个反射单元组成。图3展示了AP位置以及终端用户的移动轨迹。同时，本文假设所需的QoS约束 $\Gamma_{\min}$ 为70 Mbit/s。表2列出了仿真模拟的部分参数。

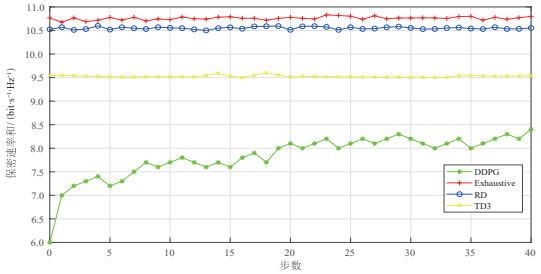

图3 空域EH环境下每步的可实现保密速率和

Figure 3 The secure sum rate for each step in the spatial-domain EH environment

表2 仿真参数

Table 2 Simulation parameter

仿真参数	数值
LoS环境常量($\mathcal{A}$)	9.61
LoS环境常量($\mathcal{B}$)	0.16
参考路径损耗(κ)	−20 dB
终端用户噪声功率(σ_k^2)	−102 dBm
窃听者噪声功率(σ_e^2)	−102 dBm
EH效率(η)	0.7
NLoS衰减因子(φ)	20 dB
AP-IRS路径损耗因子(α)	3
IRS-EU路径损耗因子($\bar{\alpha}$)	2.5
AP最大发射功率($p_{\max}$)	500 W
IRS反射元件数($\mathcal{L}$)	16

3.2 仿真对比设置

为了全面评估所提RD算法的性能表现，本文选取了三种具有代表性的基准算法进行对比。

（1）穷举法（Exhaustive）：代表经典的传统优化方法，广泛应用于无线通信安全领域中，用于解析算法可达的下界性能。

（2）DDPG算法：将深度神经网络应用于连续控制问题，计算开销较小。

（3）TD3算法：通过引入双网络结构、延迟更新策略和目标策略平滑等机制，有效地解决过估计问题。

3.2.1 仿真结果分析

（1）单用户性能分析

图3和图4展示了单个用户在空域EH环境和时域EH环境中不同学习算法的性能。在空域EH环境中，本文提出的RD算法与穷举法极其接近，这表明：本文提出的RD算法几乎能实现最优的资源分配，但

穷举法的成本较高。此外,本文提出的RD算法在所有步骤中的可实现保密速率和均优于TD3算法。然而,DDPG算法在时域EH环境下表现优于TD3算法,这是因为TD3算法存在低估问题。同时,图3和图4中展示的所提算法与基准算法中空域EH方案的可实现保密速率和明显优于时域EH方案。此外,在空域EH方案的所有学习算法中,本文提出的RD算法取得了最好的性能,穷举法由于不确定性多项式时间的复杂性导致其在实际应用中缺乏实用性。综上所述,在权衡有效性和实用性方面,仿真结果表明:本文提出的RD算法在单用户情况下具有优势。

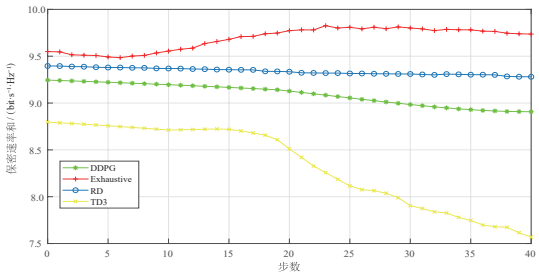

图4 时域EH环境下每步的可实现保密速率和

Figure 4 The secure sum rate for each step in the time-domain EH environment

(2)多用户性能分析

图5和图6展示了多个用户在空域EH环境和时域EH环境中不同学习算法的性能。从整体性能来看,在每个EH方案中,穷举法所用时间始终高于其他学习算法,这是因为穷举法以耗时的方式探索最优解。如图5所示,所提出的RD算法接近穷举法曲线。RD算法在性能上优于其他学习算法。如图6所示,所提出的RD算法和基于DDPG算法的值与穷举法的值接近。TD3算法与穷举法之间的差异大于其他基于学习的算法与穷举法之间的差异。基于DDPG算法可实现保密速率和略高于RD算法,TD3算法达到了最低值。此外,与单个用户的情况一样,空域EH方案也优于时域方案。总体而言,在所有基于学习和穷举法中,空域EH方案的表现均优于时域EH方案。本文提出的RD算法在空域EH方案中表现最佳,因其在有效性与时间消耗之间实现了良好的平衡。

(3)用户数量设置性能分析

图7展示了在空域EH环境中不同算法在不同用户数量下的可实现保密速率和变化趋势。从结果可以观察到,随着用户数量的增加,系统的可实现保密速率和呈现先增加、后趋于饱和的变化趋势。这是因为在用户数量较少的情况下,系统中的通信资源即IRS反射单元、发射功率等相对充足,系统可以同时服务更多合法用户,从而产生更多有效的信息传输链

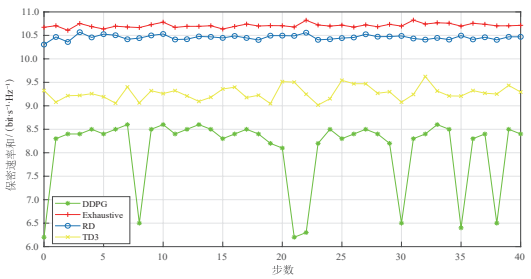

图5 空域EH环境下每步的可实现保密速率和

Figure 5 The secure sum rate for each step in the spatial-domain EH environment

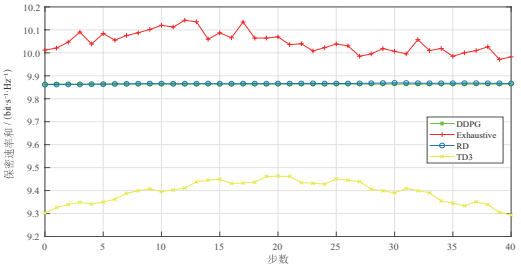

图6 时域EH环境下每步的可实现保密速率和

Figure 6 The secure sum rate for each step in the time-domain EH environment

路。然而,当更多用户接入系统时,每个用户能够获得的反射单元资源与功率分配会相应减少。同时,多用户信号之间的干扰也逐渐增强,这将降低合法用户信道容量,从而削弱单个用户的保密速率,因此系统保密速率和的增长速度会逐渐减缓,并趋于稳定。仿真结果进一步说明:所提算法在多用户场景下仍具有良好的可扩展性。此外,在UAV-IRS辅助的网络场景中,本文提出的算法性能够更接近穷举法,这充分验证了所提算法的优越性。

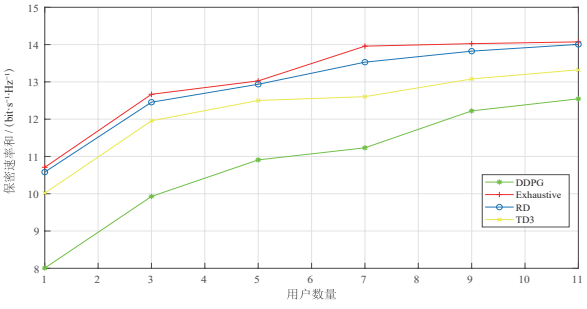

图7 系统性能与用户数量的关系

Figure 7 System performance versus the number of users

(4)IRS设置性能分析

为了进一步验证所提出算法的有效性,本文还研究了IRS设置对系统性能的影响。图8展示了在空域EH环境中不同算法在多用户情况下的可实现保密速

率和变化趋势。可以观察到，当IRS反射元件数量增加时，本文提出的算法性能仍能接近穷举法。系统性能也随着反射元件数量的增加而增加，这是由于红外反射元件数量的增加为相位偏移的设计提供了更多的自由度，从而能够更好地增强合法用户的通信通道，同时抑制窃听者的通信通道。

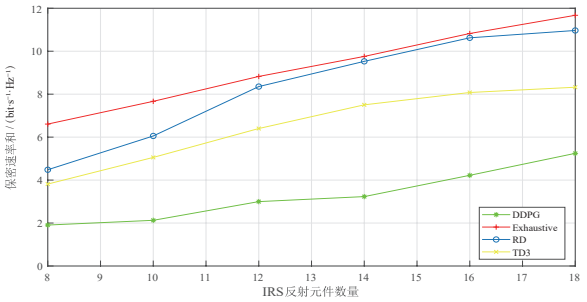

图8 系统性能与IRS反射元件数量的关系

Figure 8 System performance versus the number of IRS reflecting elements

图9展示了在空域EH和时域EH机制下所提算法的性能比较。可以观察到，随着IRS反射元件数量的增加，两种机制下系统的可实现保密速率和均有所提升。此外，由图9可知，空域分割方案始终优于传统时域分割方案，这是因为空域分割机制通过对IRS反射单元进行空间划分，使部分单元用于信息反射，部分单元用于EH，从而实现两种功能的并行运行，提高了IRS资源利用效率，而时域分割机制需要在不同时间段分别执行信息传输与EH，降低了有效通信时间，因此系统保密性能相对较低。

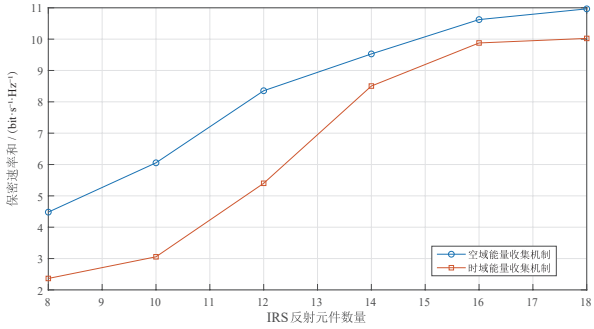

图9 空域EH与时域EH机制的性能比较

Figure 9 Performance comparison between spatial-domain and time-domain EH mechanisms

4 结论

本文针对IRS辅助的分散计算网络，围绕高保密和高QoS需求，提出了一种基于EH的UAV-IRS辅助的保密速率和优化方案。该方案首先构建了基于空域分割的混合传输模型和信道模型，突破了传统时域分割资源分配方法的局限性；其次，通过联合优化用户发射功率、反射元件相移等耦合变量，满足所需的通信QoS和EH约束等，以最大化用户保密速率和；最后，设计了一种基于鲁棒DRL算法，以在动态无线环境中保证无线系统的QoS。仿真评估了在单用户和多用户情况下的算法性能，仿真结果验证了本文所提算法的性能不但优于现有其他基于学习的算法，而且接近穷举法的性能。

尽管本文研究取得了一定进展，但仍存在进一步得拓展空间，未来可从以下几个方向展开：一是考虑不完全干扰消除及信道估计误差条件下的鲁棒优化设计，以提升模型的工程适用性；二是将研究场景拓展到多UAV协同的IRS辅助网络场景，研究多节点协作下的分布式学习与资源分配机制；三是可结合最新的强化学习策略，提升算法在动态环境中的快速适应能力。上述研究将有助于进一步推动IRS辅助分散计算安全通信技术在实际复杂环境中的应用落地。

参考文献

[1] Karaman B, Basturk I, Taskin S, et al. Solutions for sustainable and resilient communication infrastructure in disaster relief and management scenarios[J]. IEEE Communications Surveys & Tutorials, 2026, 28: 716-760.

[2] Wu H J, Zhang J, Cai Z P, et al. Resolving multitask competition for constrained resources in dispersed computing: A bilateral matching game[J]. IEEE Internet of Things Journal, 2021, 8(23): 16972-16983.

[3] Liu Y W, Liu X, Mu X D, et al. Reconfigurable intelligent surfaces: Principles and opportunities[J]. IEEE Communications Surveys & Tutorials, 2021, 23(3): 1546-1577.

[4] Ning Z, Li T, Wu Y, et al. 6G Communication new paradigm: The integration of unmanned aerial vehicles and intelligent reflecting surfaces[J]. IEEE Communications Surveys & Tutorials, 2025,27(6):3382-3416.

[5] Zhang J F, Wang W, Tang J, et al. Robust secure transmission for IRS-aided NOMA networks with hybrid beamforming[J]. IEEE Transactions on Wireless Communications, 2024, 23(4): 3086-3101.

[6] Lv L, Zhao S, Zhou Y N, et al. Secure transmission for dual-function IRS-assisted cognitive radio NOMA networks[J]. IEEE Internet of Things Journal, 2025, 12(8): 10768-10782.

[7] Wang Z, Wu W, Zhou F H, et al. IRS-enhanced spectrum sensing and secure transmission in cognitive radio networks[J]. IEEE Transactions on Wireless Communications, 2024, 23(8): 10271-10286.

[8] Li Z, Zhang L J, Le S W, et al. Distributed modulation ex-

ploiting IRS for secure communications[J]. IEEE Transactions on Mobile Computing, 2025, 24(10): 11193-11208.

[9] Bao J Y, Cao Y, Qin X Q, et al. Secure constructive interference precoding for IRS-aided NOMA networks[J]. IEEE Transactions on Wireless Communications, 2025, 24(12): 10296-10310.

[10] Pandey G K, Gurjar D S, Yadav S, et al. UAV-assisted communications with RF energy harvesting: A comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2025, 27(2): 782-838.

[11] Ponnimbaduge Perera T D, Jayakody D N K, Sharma S K, et al. Simultaneous wireless information and power transfer (SWIPT): Recent advances and future challenges[J]. IEEE Communications Surveys & Tutorials, 2018, 20(1): 264-302.

[12] Song H H, Wen H, Tang J, et al. Secrecy energy efficiency maximization for distributed intelligent-reflecting-surface-assisted MISO secure communications[J]. IEEE Internet of Things Journal, 2023, 10(5): 4462-4474.

[13] Li B G, Liao J, Wu W J, et al. Intelligent reflecting surface assisted secure computation of wireless powered MEC system[J]. IEEE Transactions on Mobile Computing, 2024, 23(4): 3048-3059.

[14] Zhang J F, Xu J L, Lu W D, et al. Secure transmission for IRS-aided UAV-ISAC networks[J]. IEEE Transactions on Wireless Communications, 2024, 23(9): 12256-12269.

[15] Xu J L, Zhang J F, Liu M Q, et al. Secure integrated sensing and SWIPT via active IRS[J]. IEEE Transactions on Wireless Communications, 2025, 24(8): 6997-7011.

[16] Dong J W, Wang F. PIRS and ASTAR-IRS jointly aided wireless communications using RSMA: Deployment design and rate allocations[J]. IEEE Internet of Things Journal, 2025, 12(5): 5575-5588.

[17] Wang H D, Zhao S H, Li L H, et al. Learning joint source-channel encoding in IRS-assisted multi-user semantic communications[C]//2025 IEEE International Symposium on Parallel and Distributed Processing with Applications. Piscataway: IEEE, 2025: 15-20.

[18] Khan A, Hayat B, Ahmad S, et al. AI-empowered multi-UAV and IRS collaboration for spectrum and energy optimization in B5G networks[J]. IEEE Internet of Things Journal, 2026, 13(5): 7972-7988.

[19] Wu Q Q, Zhang R. Joint active and passive beamforming optimization for intelligent reflecting surface assisted SWIPT under QoS constraints[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(8): 1735-1748.

[20] Wan Z X, Jiang W H, Nie J T, et al. Min-max fairness based joint optimal design for IRS-assisted MEC systems[J]. IEEE Transactions on Vehicular Technology, 2024, 73(8): 11949-11963.

[21] Al-Hourani A, Kandeepan S, Jamalipour A. Modeling air-to-ground path loss for low altitude platforms in urban environments[C]//2014 IEEE Global Communications Conference. Piscataway: IEEE, 2014: 2898-2904.

[22] Lei M, Zhang X J, Yu B C, et al. Throughput maximization for UAV-assisted wireless powered D2D communication networks with a hybrid time division duplex/frequency division duplex scheme[J]. Wireless Networks, 2021, 27(3): 2147-2157.

[23] Xiong X, Zheng B X, Swindlehurst A L, et al. A new intelligent reflecting surface-aided electromagnetic stealth strategy[J]. IEEE Wireless Communications Letters, 2024, 13(5): 1498-1502.

[24] Wu Q Q, Zhang R. Beamforming optimization for wireless network aided by intelligent reflecting surface with discrete phase shifts[J]. IEEE Transactions on Communications, 2020, 68(3): 1838-1851.

[25] Hu G J, Wu Q Q, Xu D H, et al. Intelligent reflecting surface-aided wireless communication with movable elements[J]. IEEE Wireless Communications Letters, 2024, 13(4): 1173-1177.

[26] Luo C, Hu J, Xiang L P, et al. Massive wireless energy transfer without channel state information *via* imperfect intelligent reflecting surfaces[J]. IEEE Transactions on Vehicular Technology, 2024, 73(6): 8529-8541.

[27] Tang Y Z, Ma G G, Xie H L, et al. Joint transmit and reflective beamforming design for IRS-assisted multiuser MISO SWIPT systems[C]// 2020 IEEE International Conference on Communications. Piscataway: IEEE, 2020: 1-6.

[28] Yu X H, Xu D F, Sun Y, et al. Robust and secure wireless communications via intelligent reflecting surfaces[J]. IEEE Journal on Selected Areas in Communications, 2020, 38(11): 2637-2652.

[29] Li X L, Huo J H, Huangfu W, et al. Secrecy sum rate maximization in UAV-IRS assisted networks with credit-aware cooperative multi-agent reinforcement learning[J]. IEEE Transactions on Wireless Communications, 2026, 25: 3186-3200.

[30] Peng H R, Wang L C. Energy harvesting reconfigurable intelligent surface for UAV based on robust deep reinforcement learning[J]. IEEE Transactions on Wireless Communications, 2023, 22(10): 6826-6838.

[31] Peng H R, Tsai A H, Wang L C, et al. LEOPARD: Parallel optimal deep echo state network prediction improves service coverage for UAV-assisted outdoor hotspots[J]. IEEE Transactions on Cognitive Communications and Networking, 2022, 8(1): 282-295.
[32] Bellman R. A Markovian decision process[J]. Indiana University Mathematics Journal, 1957, 6(4): 679-684.
[33] Mnih V, Badia A P, Mirza M, et al. Asynchronous methods for deep reinforcement learning[C]//International Conference on Machine Learning. New York: PMLR, 2016: 1928-1937.
[34] Lillicrap T P, Hunt J J, Pritzel A, et al. Continuous control with deep reinforcement learning[PP/OL]. V6. arXiv (2019-07-05)[2026-03-11]. https://doi.org/10.48550/arXiv.1509.02971.
[35] Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
[36] Fujimoto S, Hoof H, Meger D. Addressing function approximation error in actor-critic methods[C]//International conference on machine learning. Stockholm: PMLR, 2018: 1587-1596.
[37] Hasselt H. Double Q-learning[C]//Proceedings of the 24th Annual Conference on Neural Information Processing Systems. Vancouver: Curran Associates, Inc., 2010: 2613-2621.
[38] Pan L, Cai Q, Huang L. Softmax deep double deterministic policy gradients[J]. Advances in neural information processing systems, 2020, 33: 11767-11777.
[39] Lai C C, Chen C T, Wang L C. On-demand density-aware UAV base station 3D placement for arbitrarily distributed users with guaranteed data rates[J]. IEEE Wireless Communications Letters, 2019, 8(3): 913-916.

作者简介

李嘉欣 女，1995年8月出生于山西省朔州市。现为北京科技大学计算机与通信工程学院博士研究生。主要研究方向为多接入边缘计算和人工智能。
E-mail: b20200318@xs.ustb.edu.cn

刘之滨 男，1985年2月出生于北京市。现为中国国家电网公司华北分公司高级工程师。主要研究方向为网络安全、数字化管理和计算机网络。
E-mail: liu.zbin@nc.sgcc.com.cn

王建萍 女，1974年1月出生于河北省保定市。现为北京科技大学计算机与通信工程学院通信工程系教授。主要研究方向为可见光通信、人工智能。
E-mail: jpwang@ustb.edu.cn

林福宏 男，1981年11月出生于山西省朔州市。现为北京科技大学计算机与通信工程学院通信工程系教授。主要研究方向为网络安全、人工智能和边缘/雾计算。
E-mail: fhlin@ustb.edu.cn