有源RIS辅助的半免授权NOMA上行系统安全传输方案

李有明，王宏健，廖昌达，蔡万源
（宁波大学信息科学与工程学院，浙江 宁波 315211）

摘要：针对有源可重构智能反射面（ARIS）辅助半免授权非正交多址（SGF-NOMA）上行系统的安全传输问题，提出一种面向时变信道和窃听信道不确定性的安全资源分配方案。首先，利用ARIS的幅度放大与相位调控能力补偿级联路径损耗，联合优化基站接收波束成形、ARIS幅相系数及免授权用户发射功率，以最大化系统长期保密速率。随后，提出基于集中式训练、分布式执行框架的多智能体双延迟深度确定性策略梯度算法，实现基站与免授权用户的协同安全决策。通过引入双批评家网络、目标策略平滑和延迟更新机制，缓解信道动态性及窃听信道估计误差引起的策略振荡，提升训练稳定性与实时决策能力。仿真结果表明，相较于无源RIS辅助SGF-NOMA方案，所提方案的系统保密速率提升约46%，并在时变和不确定信道条件下表现出更优的保密性能。

关键词：有源可重构智能反射面；半免授权；非正交多址接入；安全传输；多智能体强化学习
中图分类号：TP393.0
文献标志码：A
doi：10.11959/j.issn.1000–436x.TXXB260336

Active RIS-aided secure transmission scheme for semi-grant-free NOMA uplink systems

Li Youming, Wang Hongjian, Liao Changda, Cai Wanyuan
Faculty of Information Science and Engineering, Ningbo University, Ningbo 315211, China

Abstract: To address secure uplink transmission in active reconfigurable intelligent surface (ARIS)-assisted semi-grant-free non-orthogonal multiple access (SGF-NOMA) systems under time-varying channels and eavesdropping channel uncertainty, a secure resource allocation scheme was proposed. First, the amplitude amplification and phase control capabilities of ARIS were exploited to compensate for cascaded path loss. The base station receive beamforming, ARIS amplitude-and-phase coefficients, and grant-free user transmit powers were jointly optimized to maximize the long-term secrecy rate. Then, a multi-agent twin delayed deep deterministic policy gradient algorithm based on centralized training with decentralized execution was developed to enable cooperative secure decision-making between the base station and grant-free users. By incorporating twin critic networks, target policy smoothing, and delayed policy updates, the proposed algorithm mitigated policy oscillations caused by channel dynamics and eavesdropping channel estimation errors, thereby improving training stability and real-time decision-making capability. Simulation results show that the proposed scheme improves the system secrecy rate by approximately 46% compared with the passive RIS-assisted SGF-NOMA scheme and achieves superior secrecy performance under time-varying and uncertain channel conditions.

Key words: active reconfigurable intelligent surface, semi-grant-free, non-orthogonal multiple access, secure transmission, multi-agent reinforcement learning

收稿日期：2026–06–09；**修回日期：**2026–07–28
通信作者：李有明，liyouming@nbu.edu.cn
基金项目：国家重点研发计划基金资助项目（No.2023YFC2809400）；宁波市重点研发计划基金资助项目（No.2023Z030）
Foundation Items: The National Key Research and Development Program of China (No. 2023YFC2809400), The Key Research and Development Program of Ningbo (No.2023Z030)

0　引言

随着第六代移动通信与物联网技术的深度融合，海量终端设备的并发接入需求激增，频谱资源稀缺问题日益严峻。传统授权（grant-based, GB）接入需通过握手机制获取传输资源[1]，其冗长的信令交互过程在大规模短包传输场景下将导致过高的接入时延与信令开销[2]。为解决这些问题，免授权（grant free, GF）传输技术应运而生，其允许物联网设备不需要向基站发送授权申请即可直接传输数据，同时为这些设备预留专用资源用于数据包发送[3]。然而，由于缺乏集中式调度，多个GF用户可能无协调地接入同一时频资源，造成严重的传输碰撞与多用户干扰[4]。半免授权（semi grant-free, SGF）方案的出现为解决免授权传输中的资源竞争与多用户干扰问题提供了有效途径。在半免授权–非正交多址接入（SGF-NOMA）中，GB用户通过授权方式占用正交资源块，而GF用户可在不需要完整授权请求过程的情况下竞争复用为GB用户预留的资源块，并通过功率域非正交叠加与GB用户并发传输；同时，系统可根据信道条件从多个GF用户中选择适合接入的用户，以保障GB用户服务质量并避免过多GF用户同时接入造成碰撞和干扰。因此，SGF-NOMA在降低接入信令开销和传输时延的同时，实现GF用户与GB用户的共存，能够在有限频谱资源下高效地支持大规模物联网设备的连接需求[5-6]。

然而，由于GF用户缺乏集中式调度，多个用户在无协调的情况下并发传输，使SGF-NOMA系统更容易遭受窃听攻击，面临严峻的安全威胁。传统上层加密方案（如高级加密标准、非对称公钥加密算法等）通过密钥协商、分发以及复杂加解密运算实现数据保护[7]，但在大规模物联网中，海量终端的密钥管理将带来沉重的信令与存储开销，令资源受限的物联网设备难以承受，无法满足低时延传输需求。与传统加密机制不同，物理层安全（physical layer security, PLS）技术不依赖计算复杂度假设，而是利用无线信道固有的噪声、衰落与干扰特性，从物理传输层面提升通信安全性[8-9]。文献[10]将PLS引入NOMA系统，通过联合优化功率分配、用户配对与波束成形，利用叠加传输的干扰特性增强合法用户保密速率。但该方案未考虑GF用户的免调度随机接入特性，难以直接适配传输碰撞与动态调度不确定性场景。文献[11]率先从单GF用户保密性能分析入手，逐步拓展至多用户场景并提出了相应的调度方案。文献[12]进一步考虑窃听者信道状态信息（channel state information, CSI）已知与未知两种场景，分别设计了差异化的GF用户调度策略。

与此同时，可重构智能反射面（reconfigurable intelligent surface, RIS）借助大量可编程电磁单元对无线传播环境进行智能调控，在增强合法链路质量的同时恶化窃听信道，从而有效提升系统的抗窃听能力[13-14]。文献[15]研究RIS辅助的NOMA系统，在非理想窃听CSI下，有效提升了多用户保密速率的公平性与系统鲁棒性。然而，传统无源RIS仅被动反射信号，其辅助链路需经历用户至RIS与RIS至基站的双重信道衰落，级联路径损耗严重制约了保密性能的提升。为解决这一问题，研究者提出了有源可重构智能反射面（active reconfigurable intelligent surface, ARIS）[16]。不同于无源RIS仅调控相位，ARIS在每个反射单元集成额外放大器，在消耗少量偏置功率的同时对入射信号实施幅度放大与相位调控[17-18]，从而有效补偿级联损耗并显著提升接收信号强度[19-20]。文献[21]通过实验原型，验证了ARIS的协议设计与盲贪心等控制算法的有效性。文献[22]将ARIS应用于上行多天线NOMA系统，通过ARIS对信号幅度与相位的联合调控，设计了系统和速率最大化方案，显著提升了传输性能。在安全威胁场景下，文献[23]在基站与窃听者均存在范数有界非理想CSI的条件下，提出一种基于交替优化（alternating optimization, AO）的算法，证明了ARIS与NOMA结合在构建安全上行链路方面的优势。文献[24]首次将ARIS引入上行SGF-NOMA隐蔽通信场景，针对GF用户与GB用户共享资源块的传输机制，推导了检测错误概率与隐蔽速率的解析表达式，验证了ARIS在提升隐蔽通信性能方面的优势。

相较于传统NOMA，ARIS辅助的SGF-NOMA系统需同时应对GB与GF用户间的异构业务需求，以及GF用户随机接入引发的信道快速时变问题。文献[23]所提交替优化方法面向传统NOMA场景，未考虑SGF机制下GF用户随机接入带来的动态性。同时，其依赖逐时隙交替迭代求解，难以在信道快速时变场景下及时获得满足实时性要求的资源分配策略。深度强化学习（deep reinforcement

learning, DRL）技术为解决这些问题提供了新的技术路径。其通过与环境的在线交互，以数据驱动方式学习低复杂度策略，能够适配动态变化的无线环境，实现更高效的网络管理与更可靠的通信[25]。文献[26]提出面向SGF-NOMA的多智能体DRL框架，通过联合优化功率控制与接收波束成形提升物理层安全性能。文献[27]提出基于功率池的自主功率控制机制，通过离散功率集合实现GF用户分布式功率选择，在降低信令开销的同时提升系统吞吐性能。上述工作虽已将DRL引入SGF-NOMA系统的资源分配，但未能利用ARIS的主动幅相调控能力增强系统的抗窃听性能。其中，文献[26]未考虑窃听信道的不确定性；文献[27]以系统吞吐量为优化目标，未涉及安全传输问题。此外，ARIS的引入使基站波束成形、ARIS幅相系数与用户发射功率高度耦合，现有逐时隙迭代优化方法难以在信道快速时变条件下直接适用。因此，研究信道时变与窃听信道不确定条件下ARIS辅助SGF-NOMA上行系统的实时安全传输方法具有重要意义。

为此，本文提出了一种基于多智能体双延迟深度确定性策略梯度（multi-agent twin delayed deep deterministic policy gradient, MATD3）的安全资源分配算法，主要贡献概括如下。

（1）构建了面向动态环境的ARIS辅助SGF-NOMA上行安全传输模型。采用一阶自回归模型刻画信道时变特性，以范数有界误差模型刻画窃听CSI不确定性，将基站接收波束成形、ARIS幅相系数与GF用户发射功率的联合优化建模为长期保密和速率最大化问题，并通过GF用户竞争接入约束与基于CSI的SIC解码顺序优先保障GB用户的服务质量。

（2）提出了基于中心化训练去中心化执行（centralized training with decentralized execution, CTDE）框架的MATD3安全资源分配算法。将基站与GF用户建模为协同智能体，训练阶段基于全局信息集中学习、执行阶段基于局部观测分布式决策；通过双Critic网络与目标策略平滑抑制窃听信道估计误差及连续动作扰动引起的Q值过估计，结合延迟策略更新与目标网络软更新缓解信道时变与GF用户竞争接入带来的策略振荡，增强了动态安全环境下的训练稳定性。

（3）通过多组对比仿真验证了所提方案的有效性。结果表明，相较于无源RIS配置，ARIS辅助的系统保密速率提升约46%；与基于DDPG的方案和交替优化方案相比，所提方案在信道时变与窃听信道估计误差条件下均取得更优的保密速率，表现出更强的动态环境适应能力。

1 系统模型

1.1 信号传输模型

本文研究ARIS辅助的SGF-NOMA上行系统，模型如图1所示。系统包含一个配备M根天线的基站（base station, BS）、N个GF用户$U_1,U_2,\cdots,U_N$（记为$\mathcal{N}=\{1,\cdots,N\}$）和1个GB用户$U_0$，全体用户集合记为$\mathcal{I}=\{0,1,\cdots,N\}$。在SGF-NOMA机制下，$N$个GF用户机会式竞争接入与GB用户$U_0$相同的资源块。为控制同一资源块上的多用户干扰并保障GB用户服务质量，系统在每个时隙内最多允许L个GF用户与GB用户共享该资源块，其中$L\leqslant N$。为改善信道环境，系统部署了一个由K个反射单元构成的ARIS，以主动调控入射信号的相位与幅度。此外，系统中存在一个外部窃听者（Eve），试图截获GF用户和GB用户的信息。

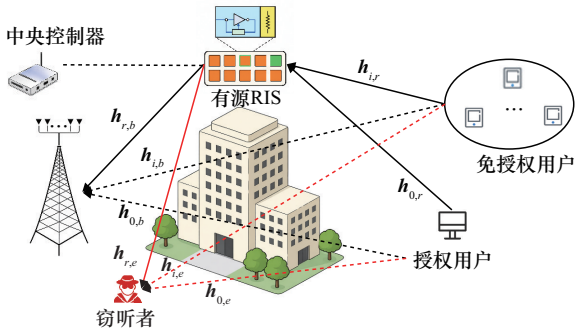


图1 系统模型

为刻画信道时变特性，对于与用户相关联的无线链路，采用一阶自回归模型描述其小尺度快衰落过程。记任意无线链路在第t个时隙下的复信道衰落系数为$h(t)$，则其时域演化可表示为：

$$h(t+1)=\rho h(t)+\sqrt{1-\rho^2}\,e(t) \tag{1}$$

其中，$\rho=J_0\left(2\pi f_{\mathrm{d}}T_{\mathrm{s}}\right)\in[0,1]$为信道时间相关系数，由最大多普勒频移$f_{\mathrm{d}}$和时隙长度$T_{\mathrm{s}}$共同决定，$f_{\mathrm{d}}$与用户移动速度正相关，从而使$\rho$随用户移动速度增加而减小；$e(t)$为与历史状态相互独立的零均值、单位方差复高斯随机变量。

在第t个时隙下，全体用户$U_i(i\in\mathcal{I})$至基站以及窃听者的等效合并信道响应分别表示为：

$$\boldsymbol{G}_{i,b}(t)=\boldsymbol{h}_{i,b}(t)+\boldsymbol{H}_{r,b}^{\mathrm{H}}(t)\boldsymbol{\Theta}(t)\boldsymbol{h}_{i,r}(t) \tag{2}$$

$$G_{i,e}(t)=h_{i,e}(t)+\boldsymbol{h}_{r,e}^{\mathrm{H}}(t)\boldsymbol{\Theta}(t)\boldsymbol{h}_{i,r}(t) \tag{3}$$

其中，$\boldsymbol{h}_{i,b}(t)\in\mathbb{C}^{M\times 1}$为用户$U_i$至基站的直连信道；$\boldsymbol{h}_{i,r}(t)\in\mathbb{C}^{K\times 1}$为用户$U_i$至ARIS的信道；$\boldsymbol{H}_{r,b}(t)\in\mathbb{C}^{K\times M}$为ARIS至基站的信道；$h_{i,e}(t)$为用户$U_i$至窃听者的直连窃听信道；$\boldsymbol{h}_{r,e}(t)\in\mathbb{C}^{K\times 1}$为ARIS至窃听者的信道；$\boldsymbol{\Theta}(t)$为ARIS的反射系数矩阵，定义为$\boldsymbol{\Theta}(t)=\mathrm{diag}\{\beta_1(t)\mathrm{e}^{\mathrm{j}\theta_1(t)},\cdots,\beta_K(t)\mathrm{e}^{\mathrm{j}\theta_K(t)}\}$，$\beta_k(t)\in[0,\beta_{\max}]$和$\theta_k(t)\in[0,2\pi)$分别为第$k$个单元的放大增益和相移。由于ARIS通常部署于开阔位置，其与基站、窃听者之间的链路以及用户至ARIS的链路均具有较强的视距分量，因此$\boldsymbol{H}_{r,b}(t)$、$\boldsymbol{h}_{r,e}(t)$和$\boldsymbol{h}_{i,r}(t)$均采用莱斯衰落模型。用户至基站、用户至窃听者的直连链路更易受到遮挡，因此将$\boldsymbol{h}_{i,b}(t)$和$h_{i,e}(t)$建模为瑞利衰落信道。

设第t个时隙内参与资源竞争的GF用户集合为$C(t)\subseteq\mathcal{N}$，首先，系统根据接入条件得到满足信道要求的GF用户集合$\mathcal{C}_\tau(t)=\{l\in C(t)\big|\boldsymbol{G}_{n,b}(t)|^2\geqslant\tau\}$，其中$\tau$为接入筛选门限。随后，系统从$\mathcal{C}_\tau(t)$中选择等效信道增益最大的$L$个GF用户组成最终接入集合$\mathcal{A}(t)=\mathrm{Top}_N\{|\boldsymbol{G}_{n,b}(t)|^2,n\in\mathcal{C}_\tau(t)\}$，其中$L\leqslant N$，$\mathrm{Top}_L\{\cdot\}$表示选取集合中等效信道增益最大的至多$L$个用户。

基于上述假设，基站接收到的来自$U_0,\cdots,U_N$的叠加信号可表示为：

$$\begin{aligned}\boldsymbol{y}(t)=&\sqrt{P_0(t)}\,\boldsymbol{G}_{0,b}(t)s_0(t)+\\&\sum_{n=1}^{N}\alpha_n(t)\sqrt{P_n(t)}\,\boldsymbol{G}_{n,b}(t)s_n(t)+\\&\boldsymbol{H}_{r,b}^{\mathrm{H}}(t)\boldsymbol{\Theta}(t)\boldsymbol{n}_r(t)+\boldsymbol{n}_b(t)\end{aligned} \tag{4}$$

其中，$P_0(t)$和$P_n(t)$分别表示GB用户U_0和第n个GF用户U_n的发射功率；$s_0(t)$和$s_n(t)$为设备发送的归一化数据符号，满足$\mathbb{E}[|s_0(t)|^2]=\mathbb{E}[|s_n(t)|^2]=1$；$\alpha_n(t)\in\{0,1\}$为GF用户$U_n$的接入控制指示变量，仅当$U_n\in\mathcal{A}(t)$时，$\alpha_n(t)=1$，该用户参与当前时隙的NOMA叠加传输，否则$\alpha_n(t)=0$，该用户保持静默；$\boldsymbol{n}_r(t)\sim\mathcal{CN}(0,\sigma_r^2)$表示ARIS在放大信号过程中引入的额外加性噪声向量；$\boldsymbol{n}_b(t)\sim\mathcal{CN}(0,\sigma^2\mathbf{I})$为基站接收端的加性白高斯噪声向量。

在NOMA系统中，接收端通过串行干扰消除（successive interference cancellation, SIC）技术对合并接收信号进行解码，分离出不同用户的信号。在完成GF用户接入后，基站仅对最终接入集合$\mathcal{A}(t)$中的GF用户与GB用户执行SIC解码。本文采用基于信道状态信息的SIC解码顺序，即在同一NOMA簇内优先解码信道增益较强的用户[28]。具体而言，通过调节ARIS的反射系数矩阵以重构各用户的等效信道增益，可在满足GB用户QoS约束的前提下，使GB用户在SIC过程中最后被解码。该设计能够抑制用户对GB用户U_0解码过程造成的干扰，从而优先保障GB用户的传输可靠性。基于上述SIC解码规则，基站解码GF用户U_n的信干噪比可表示为：

$$\Gamma_{n,b}^{\mathrm{GF}}(t)=\frac{\alpha_n(t)P_n(t)\left|\boldsymbol{w}^{\mathrm{H}}(t)\boldsymbol{G}_{n,b}(t)\right|^2}{P_0(t)\left|\boldsymbol{w}^{\mathrm{H}}(t)\boldsymbol{G}_{0,b}(t)\right|^2+I_n^{\mathrm{GF}}(t)+I_r(t)+\left\|\boldsymbol{w}(t)\right\|^2\sigma^2} \tag{5}$$

基站解码GB用户U_0的信干噪比表达式为：

$$\Gamma_{0,b}^{\mathrm{GF}}(t)=\frac{P_0(t)\left|\boldsymbol{w}^{\mathrm{H}}(t)\boldsymbol{G}_{0,b}(t)\right|^2}{I_r(t)+\left\|\boldsymbol{w}(t)\right\|^2\sigma^2} \tag{6}$$

其中，$\boldsymbol{w}(t)\in\mathbb{C}^{M\times 1}$为第$t$个时隙下的波束成形向量；$I_n^{\mathrm{GF}}(t)=\sum_{m\in\mathcal{I}_n}\alpha_m(t)P_m(t)\left|\boldsymbol{w}^{\mathrm{H}}(t)\boldsymbol{G}_{m,b}(t)\right|^2$为其他未解码的GF用户造成的干扰，$\mathcal{I}_n$是$\mathcal{A}(t)$中其他未解码的GF用户的集合；$I_r(t)\triangleq\left\|\boldsymbol{w}^{\mathrm{H}}(t)\boldsymbol{H}_{r,b}^{\mathrm{H}}(t)\boldsymbol{\Theta}(t)\right\|^2\sigma_r^2$为ARIS引入的等效干扰；$\sigma^2$为基站的高斯白噪声功率。

当窃听者尝试窃听GF用户U_n与GB用户U_0的信号时，其信干噪比分别表示为：

$$\Gamma_{n,e}^{\mathrm{GF}}(t)=\frac{\alpha_n(t)P_n(t)\left|G_{n,e}(t)\right|^2}{P_0(t)\left|G_{0,e}(t)\right|^2+I_{n,e}^{\mathrm{GF}}(t)+I_{r,e}(t)+\sigma_e^2} \tag{7}$$

$$\Gamma_{0,e}^{\mathrm{GB}}(t)=\frac{P_0(t)\left|G_{0,e}(t)\right|^2}{\sum_{n=1}^{N}\alpha_n(t)P_n(t)\left|G_{n,e}(t)\right|^2+I_{r,e}(t)+\sigma_e^2} \tag{8}$$

其中，$I_{n,e}^{\mathrm{GF}}(t)=\sum_{m\neq n}\alpha_m(t)P_m(t)\left|G_{m,e}(t)\right|^2$为其余

GF 用户造成的干扰；$I_{r,e}(t) \triangleq \left\| \boldsymbol{h}_{r,e}^{\mathrm{H}}(t)\boldsymbol{\Theta}(t) \right\|^2 \sigma_r^2$ 表示 ARIS 在窃听者处引入的等效干扰；σ_e^2 为窃听者处的加性白高斯噪声功率。

由式(5)~式(8)，可以得出 GB 用户与 GF 用户的保密速率为：

$$R_{0,s}^{\mathrm{GB}}(t) = \left[R_{0,b}^{\mathrm{GB}}(t) - R_{0,e}^{\mathrm{GB}}(t) \right]^+ \tag{9}$$

$$R_{n,s}^{\mathrm{GF}}(t) = \left[R_{n,b}^{\mathrm{GF}}(t) - R_{n,e}^{\mathrm{GF}}(t) \right]^+ \tag{10}$$

其中，$R_{0,b}^{\mathrm{GB}}(t)$ 与 $R_{n,b}^{\mathrm{GF}}(t)$ 分别为 GB 用户 U_0 与 GF 用户 U_n 在基站处的传输速率；$R_{0,e}^{\mathrm{GB}}(t)$ 与 $R_{n,e}^{\mathrm{GF}}(t)$ 分别为窃听者对 GB 用户 U_0 与 GF 用户 U_n 的信号可解码速率；$[x]^+$ 表示 x 的正部，即 $\max(x,0)$。

鉴于窃听者通常处于被动监听状态，基站难以获取其精确 CSI。为刻画窃听链路的不确定性，本文采用范数有界的 CSI 误差模型[15,23]。实际系统中，受传播损耗与接收动态范围限制，信道估计误差的幅值通常具有有限上界；该模型不依赖于误差的具体统计分布，并可通过调整误差上界覆盖不同程度的 CSI 失配，适用于窃听 CSI 难以准确获取的场景。设 $G_{i,e}(t)$ 为全体用户 $U_i (i \in \mathcal{I})$ 在第 t 个时隙的等效窃听信道，则其真实值与估计值的关系可表示为：

$$G_{i,e}(t) \in \mathcal{G}_{i,e} = \left\{ \hat{G}_{i,e}(t) + \Delta G_{i,e}(t) : \left| \Delta G_{i,e}(t) \right| \leqslant \varepsilon_{i,e} \right\} \tag{11}$$

其中，$\hat{G}_{i,e}(t)$、$\Delta G_{i,e}(t)$ 与 $\varepsilon_{i,e}$ 分别表示等效窃听信道的估计值、估计误差与误差上界。为量化信道向量失配程度，定义信道估计误差（channel estimation error, CEE）表达式为 $\mathrm{CEE} = 10\lg \mathbb{E}\left[\left(\frac{|\Delta G_e|}{|G_e|} \right)^2 \right]$。CEE 用于衡量窃听链路的信道不确定性，降低 CEE 能够帮助基站更精准地估计窃听者信道，提升系统的数据传输安全能力，进而提高用户整体保密速率。

1.2 优化问题建模

基于上述系统模型，本文将基站波束成形向量 $\boldsymbol{w}(t)$、ARIS 反射系数矩阵 $\boldsymbol{\Theta}(t)$ 与各用户发射功率 $\boldsymbol{P}(t)$ 的联合优化，建模为长期保密和速率最大化问题，具体表示为：

$$\max_{\boldsymbol{w}(t),\boldsymbol{P}(t),\boldsymbol{\Theta}(t)} \ \frac{1}{T}\sum_{t=0}^{T-1}\sum_{i=0}^{N} R_{i,s}(t)$$

$$\text{s.t. C1: } R_{0,b}^{\mathrm{GB}}(t) \geqslant R_{\min}^{\mathrm{GB}}$$

$$\text{C2: } \left| \boldsymbol{w}^{\mathrm{H}}(t)\boldsymbol{G}_{0,b}(t) \right|^2 \leqslant \left| \boldsymbol{w}^{\mathrm{H}}(t)\boldsymbol{G}_{n,b}(t) \right|^2, \forall n \in \mathcal{A}(t)$$

$$\text{C3: } \alpha_n(t) = \mathbb{I}\{ n \in \mathcal{A}(t) \}, \sum_{n=1}^{N} \alpha_n(t) \leqslant L, \forall n \in \mathcal{N}$$

$$\text{C4: } P_n(t) \leqslant \alpha_n(t) P_{\max}, \forall n \in \mathcal{N}$$

$$\text{C5: } P_0(t) \left| \boldsymbol{\Theta}(t)\boldsymbol{h}_{0,r}(t) \right|^2 + \sum_{n=1}^{N} \alpha_n(t) \cdot P_n(t) \left| \boldsymbol{\Theta}(t)\boldsymbol{h}_{n,r}(t) \right|^2 + \left\| \boldsymbol{\Theta}(t) \right\|^2 \sigma_r^2 \leqslant P_{\mathrm{RIS}}$$

$$\text{C6: } 0 \leqslant \beta_k(t) \leqslant \beta_{\max}, \forall k \in \mathcal{K}$$

$$\text{C7: } 0 \leqslant \theta_k(t) < 2\pi, \forall k \in \mathcal{K} \tag{12}$$

其中，式(12)将系统长期保密和速率最大化作为目标函数；C1 为 GB 用户的 QoS 约束，要求其可达速率始终不低于最低门限 $R_{\min}^{\mathrm{GB}}$；C2 为基于 CSI 的 SIC 解码顺序约束，确保 GB 用户在簇内最后解码；C3 为 GF 用户竞争接入约束，其中 $\alpha_n(t)$ 由最终接入集合 $\mathcal{A}(t)$ 决定，并限制同一资源块上最多允许 L 个 GF 用户接入；C4 限制各 GF 用户发射功率不超过 $P_{\max}$；C5 为 ARIS 总功耗约束，包含有源放大功耗与放大器引入的噪声功率；C6 与 C7 分别限定 ARIS 各反射单元幅度增益与相移的取值范围。

针对上述多目标耦合关系，本文将系统长期保密和速率最大化与 GB 用户 QoS、SIC 解码顺序、GF 用户接入数量、用户发射功率及 ARIS 功率限制等约束统一建模。由于 ARIS 反射系数矩阵 $\boldsymbol{\Theta}(t)$ 同时影响合法等效信道、窃听等效信道、SIC 解码顺序以及有源放大功耗，本文通过约束 C1 保证 GB 用户 QoS，通过约束 C2 限定基于 CSI 的 SIC 解码顺序，并通过约束 C5~C7 限制 ARIS 总功耗、幅度增益和相移范围，从而在可行域内优化系统长期保密性能。

由于等效信道 $\boldsymbol{G}_{n,b}(t)$ 时变，且 GF 用户的接入状态具有不确定性，传统迭代优化方法对参数扰动敏感，难以满足实时决策需求。同时，各 GF 用户仅能基于局部信道状态并独立决策发射功率，使该问题难以通过传统集中式优化方法求解。因此，本文将基站与 GF 用户的联合资源分配建模为分布式多智能体序列决策过程，通过中心化训练获取全局协同策略，并在执行阶段基于局部观测实现分布式决策，从而学习适应信道时变和信道不确定性的长期安全传输策略。

2　MATD3安全资源分配算法

2.1　MDP建模

在多智能体强化学习框架中，各智能体依据当前观测从动作空间选取动作并与环境交互，环境转移至下一状态，并反馈奖励。通过持续迭代该过程，智能体逐步学习最优联合策略。本文将基站与GF用户的协同资源分配问题建模为马尔可夫决策过程（Markov decision process, MDP），由五元组$\{\mathcal{U},\mathcal{S},\mathcal{A},\mathcal{O},\mathcal{R}\}$刻画，各元素的具体含义如下。

（1）智能体集合$\mathcal{U}$。系统中包含两类智能体，即N个GF智能体$A_n^{\mathrm{GF}}(l=1,\cdots,N)$与基站智能体$A^{\mathrm{BS}}$，其中基站$A^{\mathrm{BS}}$负责联合优化接收波束成形向量$\boldsymbol{w}(t)$与ARIS反射系数矩阵$\boldsymbol{\Theta}(t)$，GF智能体$A_n^{\mathrm{GF}}$负责调节GF用户$U_n$的发射功率$P_n(t)$。

（2）状态空间$\mathcal{S}$与观测空间$\mathcal{O}$。全局状态$\boldsymbol{s}(t)$涵盖各用户的CSI、发射功率、ARIS反射系数矩阵以及基站波束成形向量。受限于局部感知能力，各GF用户智能体A_n^{GF}仅可获取自身信道状态与接入指示，其局部观测为：

$$\boldsymbol{o}_n^{\mathrm{GF}}=\{\boldsymbol{h}_{n,b}(t),\boldsymbol{h}_{n,r}(t),\alpha_n(t)\} \tag{13}$$

基站智能体A^{BS}接收到GF用户的前导信号后，可获取所有GF用户的相关信息与ARIS矩阵作为观测值：

$$\boldsymbol{o}^{\mathrm{BS}}(t)=\left\{\mathcal{H}_b(t),\hat{\mathcal{H}}_e(t),\boldsymbol{P}(t),\boldsymbol{\Theta}(t-1),\boldsymbol{\alpha}(t)\right\} \tag{14}$$

其中，$\mathcal{H}_b(t)$为全体合法接入用户到基站的等效合并信道集合；$\hat{\mathcal{H}}_e(t)$为全体用户到窃听者的估计信道集合；$\boldsymbol{P}(t)=\{P_1(t),\cdots,P_n(t)\}$为当前时隙各GF用户的发射功率向量；$\boldsymbol{\alpha}(t)=[\alpha_1(t),\alpha_2(t),\cdots,\alpha_N(t)]$为当前时隙GF用户接入结果。

（3）动作空间$\mathcal{A}$。GF智能体A_n^{GF}的动作a_n^{GF}为用户U_n的发射功率，基站智能体A^{BS}的动作a^{BS}为接收波束成形向量与ARIS反射系数矩阵：

$$a_n^{\mathrm{GF}}(t)=P_n(t) \tag{15}$$

$$a^{\mathrm{BS}}(t)=\left[\boldsymbol{w}(t),\boldsymbol{\Theta}(t)\right] \tag{16}$$

为实现更精准的动作控制，本文采用连续型动作空间，各动作分量的取值范围由约束C4、C6和C7界定。当$\alpha_n(t)=0$时，GF用户U_n未接入当前资源块，其功率动作被置为0。

（4）奖励函数$\mathcal{R}$。所有智能体共享同一全局奖励，以保证各智能体的局部策略优化与系统长期保密速率最大化目标一致。本文将系统保密和速率作为收益项，并对GB用户QoS约束、SIC解码顺序约束以及ARIS总功耗约束引入惩罚项，奖励函数定义为：

$$r(t)=\sum_{i\in\{0\}\cup A(t)}R_{i,s}(t)-\lambda_1 C_{\mathrm{QoS}}(t)-\lambda_2 C_{\mathrm{SIC}}(t)-\lambda_3 C_{\mathrm{ARIS}}(t) \tag{17}$$

其中，$\lambda_1,\lambda_2,\lambda_3$为惩罚因子；$C_{\mathrm{QoS}}(t)$为GB用户QoS约束惩罚项，表示为$C_{\mathrm{QoS}}(t)=[R_{\min}^{\mathrm{GB}}-R_{0,b}^{\mathrm{GB}}(t)]^+$；$C_{\mathrm{SIC}}(t)$为SIC解码顺序约束惩罚项，表示为$C_{\mathrm{SIC}}(t)=\sum_{n\in A(t)}\left[\left|\boldsymbol{w}^{\mathrm{H}}(t)\boldsymbol{G}_{0,b}(t)\right|^2-\left|\boldsymbol{w}^{\mathrm{H}}(t)\boldsymbol{G}_{n,b}(t)\right|^2\right]^+$；$C_{\mathrm{ARIS}}(t)$为ARIS总功耗约束惩罚项，表示为$C_{\mathrm{ARIS}}(t)=[P_{\mathrm{ARIS}}(t)-P_{\mathrm{RIS}}^{\max}]^+$。当相应约束满足时，对应惩罚项为0；当约束被违反时，惩罚项随违反程度的增大而增大，从而引导智能体在提升系统保密速率的同时满足系统约束。这一机制将奖励信号与约束满足直接绑定，驱动智能体协同学习。从长期视角来看，所有智能体的优化目标是学习联合最优策略$\boldsymbol{\pi}^*=\left\{\pi_1^{\mathrm{GF}*},\cdots,\pi_N^{\mathrm{GF}*},\pi^{\mathrm{BS}*}\right\}$，以最大化期望累积折扣奖励：

$$\pi^*=\underset{\pi=\{\pi_1^{\mathrm{GF}},\cdots,\pi_N^{\mathrm{GF}},\pi^{\mathrm{BS}}\}}{\arg\max}\ \mathbb{E}\left[\sum_{t=1}^{T}\gamma^{t-1}r(t)\,\middle|\,s(1)=s,\pi\right] \tag{18}$$

其中，$0\leqslant\gamma\leqslant1$为折扣因子，用于权衡即时奖励与未来奖励的重要性。

2.2　基于CTDE的MATD3算法

针对多智能体策略并行更新引起的环境非平稳性，本文采用CTDE框架。在中心化训练阶段，各智能体的Critic网络能够获取全局状态与联合动作信息，通过双Critic取较小Q值抑制过估计，为Actor提供稳定策略梯度；在去中心化执行阶段，基站与各GF用户仅基于局部观测独立选择动作。此外，通过延迟策略更新与目标网络软更新，进一步缓解连续动作空间中的策略振荡，学习适应信道动态的长期安全传输策略。

TD3算法是在深度确定性策略梯度（deep deterministic policy gradient, DDPG）基础上发展而来的连续动作空间强化学习算法，通过双Critic网络、目标策略平滑和延迟策略更新有效缓解Q值过估计与策略更新不稳定的问题，已被广泛用于连续动作空间优化任务。对于ARIS辅助SGF-NOMA上行安

全传输系统而言，信道时变会导致等效信道增益、干扰关系和最优资源分配策略随时隙变化，GF 用户的竞争接入会进一步改变参与传输的用户集合和 SIC 解码更新，而窃听信道估计误差会增加保密速率评估的不确定性。为此，MATD3 通过双 Critic 网络取较小 Q 值构造 TD 目标，抑制函数逼近误差和 CSI 不确定性引起的 Q 值过估计；通过目标策略平滑降低连续动作扰动对价值估计的敏感性；利用延迟策略更新与网络软更新，避免 Actor 在 Critic 尚未稳定时频繁更新，从而提高动态安全环境下的稳定性。相较于 MADDPG 算法，MATD3 通过双 Critic 和延迟更新机制进一步降低 Q 值过估计和策略振荡风险；与 MAPPO 等 on-policy 多智能体算法相比，MATD3 通过经验回放机制重复利用历史交互数据，提高样本利用效率。由于 on-policy 方法需要依赖新采样轨迹进行更新，在动态无线资源分配场景中样本复用率较低，而 MATD3 能够在不需要额外环境交互的情况下多次利用已有经验进行训练，因此更适用于 ARIS 辅助的动态无线资源分配问题。与 QMIX 等主要面向离散动作空间的值分解方法相比，MATD3 能够直接处理基站波束成形、ARIS 幅相调控和用户功率分配等连续耦合变量，避免动作离散化带来的控制精度损失。

具体而言，基站智能体 A^{BS} 维护确定性 Actor 网络 $\mu_{\theta^{\mathrm{BS}}}$ 与双 Critic 网络 $Q_{\phi_1^{\mathrm{BS}}},Q_{\phi_2^{\mathrm{BS}}}$；每个 GF 智能体 A_n^{GF} 维护 Actor 网络 $\mu_{\theta_n^{\mathrm{GF}}}$ 与双 Critic 网络 $Q_{\phi_{n,1}^{\mathrm{GF}}},Q_{\phi_{n,2}^{\mathrm{GF}}}$，并配备对应的目标网络 $\mu_{\theta^{\mathrm{BS}\prime}},Q_{\phi_1^{\mathrm{BS}\prime}},Q_{\phi_2^{\mathrm{BS}\prime}}$ 及 $\mu_{\theta_n^{\mathrm{GF}\prime}},Q_{\phi_{n,1}^{\mathrm{GF}\prime}},Q_{\phi_{n,2}^{\mathrm{GF}\prime}}$。在去中心化执行阶段，各智能体 i 仅依据自身局部观测生成动作。为增强训练过程中的探索能力，执行动作时在 Actor 输出的确定性动作上加入探索噪声，并将最终动作裁剪或投影至对应的动作可行域内。

$$a_i(t)=\Pi_{\mathcal{A}_i}\left[\mu_{\theta_i}\left(o_i(t)\right)+\epsilon_i(t)\right] \tag{19}$$

其中，$\epsilon_i(t)$ 表示智能体 i 的探索噪声，服从零均值高斯分布，即 $\epsilon_i(t)\sim\mathcal{N}\left(0,\sigma^2_{\mathrm{explore}}\right)$；$\mathcal{A}_i$ 为智能体 i 的动作可行域，$\Pi_{\mathcal{A}_i}[\cdot]$ 表示将动作裁剪或投影至动作可行域 $\mathcal{A}_i$ 的算子。在训练过程中，向策略输出注入探索噪声以充分探索状态–动作空间，通过动作投影确保生成动作满足系统约束。

在中心化训练阶段，系统将交互样本 $(\boldsymbol{s}(t),\boldsymbol{o}(t),\boldsymbol{a}(t),r(t),\boldsymbol{s}(t+1),\boldsymbol{o}(t+1))$ 存入经验回放池 $\mathcal{D}$。从经验回放池中随机采样小批量样本后，各智能体的目标 Actor 网络依据下一时刻局部观测生成目标动作分量，并组合为联合目标动作 $\boldsymbol{a}'(t+1)$。为降低目标 Q 值对函数逼近误差的敏感性，TD3 在联合目标动作上注入裁剪后的高斯噪声，实现目标策略平滑。

对于任意智能体 i，其双 Critic 网络的 TD 目标值定义为：

$$y_i(t)=r(t)+\gamma\min_{j=1,2}Q_{\phi'_{i,j}}\left(\boldsymbol{s}(t+1),\boldsymbol{a}(t+1)\right) \tag{20}$$

其中，$Q_{\phi'_{i,1}}$ 和 $Q_{\phi'_{i,2}}$ 分别表示智能体 i 的两个目标 Critic 网络，$\boldsymbol{a}'(t+1)$ 为目标策略平滑后的目标动作。通过取两个目标 Critic 输出中的较小值构造 TD 目标，可以有效缓解连续动作空间中常见的 Q 值过估计问题，从而避免 Actor 朝着被高估的动作方向更新。

双 Critic 网络通过最小化均方 TD 误差进行更新：

$$\mathcal{L}(\phi_{i,j})=\frac{1}{B}\sum_{b=1}^{B}\left[Q_{\phi_{i,j}}(s_b,a_b)-y_{i,b}\right]^2,\quad j=1,2 \tag{21}$$

其中，B 为小批量样本大小，$\boldsymbol{s}_b$ 与 $\boldsymbol{a}_b$ 分别表示第 b 个样本的全局状态与联合动作。与去中心化执行阶段不同，Critic 在训练阶段能够访问全局状态和联合动作，从而缓解多智能体策略同时更新所带来的环境非平稳性问题。

为避免 Actor 在 Critic 尚未充分收敛时过早更新策略，MATD3 采用延迟策略更新机制，每经过 d 次 Critic 更新后，再执行一次 Actor 更新。对于智能体 i，其 Actor 以最大化其对应 Critic 网络（通常选取双 Critic 中的第一个）对当前联合策略的期望 Q 值为目标，可表示为：

$$J_i(\theta_i)=\frac{1}{B}\sum_{b=1}^{B}Q_{\phi_{i,1}}\left(s_b,a_{1,b},\cdots,\mu_{\theta_i}(o_{i,b}),\cdots,a_{N+1,b}\right) \tag{22}$$

其中，智能体 i 的动作由其 Actor 网络根据局部观测 $o_{i,b}$ 生成。由于 Actor 仅以自身局部观测作为输入，因此训练完成后能够直接用于去中心化执行；Critic 网络在训练阶段利用全局信息指导策略更新，从而提高策略学习的稳定性。

策略网络与 Critic 网络完成更新后，目标网络采用软更新方式进行参数更新：

$$\theta_{i'} \leftarrow \tau_{\text{soft}}\theta_i + (1-\tau_{\text{soft}})\theta_{i'},$$
$$\phi_{i,j'} \leftarrow \tau_{\text{soft}}\phi_{i,j} + (1-\tau_{\text{soft}})\phi_{i,j'},\quad j=1,2 \tag{23}$$

其中，$\tau_{\text{soft}} \in (0,1)$为软更新系数。通过软更新机制，目标网络参数能够缓慢跟随当前网络变化，从而避免TD目标剧烈波动，进一步提升训练过程的稳定性。

在训练过程中，各智能体基于局部观测独立执行动作，环境反馈全局奖励$r(t)$与下一全局状态$s(t+1)$，系统将全局转移样本存入中央经验回放池$\mathcal{D}$。每训练步从$\mathcal{D}$中随机采样小批量数据，首先由各目标Actor依据下一时刻局部观测生成联合目标动作，经目标策略平滑后计算截断双Q的TD目标值，进而更新各双Critic网络以最小化均方TD误差，并同步对Critic目标网络执行软更新；每隔d个训练步，各Actor网络执行一次延迟策略更新，并同步对Actor目标网络执行软更新。基于CTDE框架的MATD3安全资源分配算法流程如算法1所示。

算法1　基于CTDE框架的MATD3安全资源分配算法

输入　$M,N,K,L,P_{\max},P_{\text{ARIS}},\beta_{\max},\tau_{\text{soft}},R_{\min}^{\text{GB}}$

输出　策略网络参数$\{\theta^{\text{BS}*},\theta_1^{\text{GF}*},\cdots,\theta_N^{\text{GF}*}\}$

1)初始化：Actor/Critic网络及其目标网络、经验回放池与初始状态
2)for $e=1{:}E$ do
3)　for $t=1{:}T$ do
4)　　确定接入集合$A(t)$并更新$\alpha_n(t)$
5)　　按式(19)确定联合动作$\boldsymbol{a}(t)$
6)　　获得$(\boldsymbol{s}(t),\boldsymbol{o}(t),\boldsymbol{a}(t),r(t),\boldsymbol{s}(t+1),\boldsymbol{o}(t+1))$
7)　　将转移样本存入$\mathcal{D}$
8)　　if$|\mathcal{D}| \geqslant B$ then
9)　　从$\mathcal{D}$中随机采样小批量样本
10)　　目标动作，并加入平滑噪声
11)　　按式(20)计算TD目标值
12)　　按式(21)更新双Critic网络
13)　　if $t \bmod d = 0$ then
14)　　按式(22)更新Actor网络
15)　　按式(23)软更新目标网络
16)　　end if
17)　end if
18)　end for
19)end for

2.3　复杂度分析

所提MATD3算法包含N个GF用户Actor网络、1个基站Actor网络，以及对应的双Critic网络。各网络均由输入层、两个隐藏层和输出层组成，隐藏层神经元数量分别为K^{GF}、K^{BS}和K^{C}。其中，GF用户Actor输出发射功率，基站Actor输出接收波束成形向量和ARIS反射系数矩阵。通过统计前向传播过程中的乘法运算次数，单个GF用户Actor网络复杂度为$\mathcal{O}_{\text{GF}}=\mathcal{O}((M+K)K^{\text{GF}}+(K^{\text{GF}})^2)$。基站Actor网络需要处理全体用户相关信道信息，输出接收波束成形向量和ARIS反射系数矩阵，计算复杂度为$\mathcal{O}_{\text{BS}}=\mathcal{O}((N(M+K)+K)K^{\text{BS}}+(K^{\text{BS}})^2)$。在中心化训练阶段，Critic网络以全局状态和联合动作为输入，因此单个Critic网络的计算复杂度为$\mathcal{O}_{\text{C}}=\mathcal{O}((N(M+K)+K)K^{\text{C}}+(K^{\text{C}})^2)$。由于MATD3采用双Critic结构，每个智能体均需要维护两个Critic网络，并结合延迟策略更新机制减少Actor更新频率，因此训练阶段的计算复杂度为$\mathcal{O}\left(ETB\left[\frac{1}{d}(NO_{\text{GF}}+O_{\text{BS}})+2(N+1)O_{\text{C}}\right]\right)$，其中，$E$表示训练回合数，$T$表示每个训练回合包含的时隙数，$B$表示小批量训练样本数量，$d$表示Actor延迟更新间隔。训练完成后，执行阶段仅涉及Actor网络前向推理，不再更新Critic网络，因此每个时隙的在线执行复杂度为$\mathcal{O}\left(N\mathcal{O}_{\text{GF}}+\mathcal{O}_{\text{BS}}\right)$。此外，延迟更新间隔$d$通过调节Actor与Critic的更新频率影响收敛时间，若d过小，Actor在Critic价值估计收敛前频繁更新，易引起策略振荡、延长收敛；若d过大，则策略改进缓慢。

相较于传统的逐时隙迭代优化方法，MATD3利用预训练策略网络实现一次推理决策，降低了在线计算复杂度。然而，CTDE训练阶段依赖全局信息交互，其开销主要来源于Critic网络对全局状态与联合动作的输入需求。每个训练时隙涉及的信息包括合法信道与窃听信道估计值、GF用户接入状态、发射功率、基站接收波束成形向量与ARIS反射系数，维度分别为$(N+1)(M+K)+MK+N+1$、N、N、M与$2K$。由于信道状态信息的维度远高于其余各项，单训练的信息交互维度为$\mathcal{O}((N+1)(M+K)+MK)$。经过$E$个训练回合（每个回合包含$T$个时隙）后，训练阶段总信息交互开销为$\mathcal{O}(ET[(N+1)(M+K)+MK])$，该开销随GF用

户数、基站天线数与 ARIS 反射单元数的增加而增长。训练完成后，Actor 网络参数也仅需一次性部署至各智能体。在分布式执行阶段，各智能体仅依据局部观测独立决策，不需要交换全局状态与联合动作信息，不引入额外的在线信息交互开销。因此，所提 MATD3 算法将计算与信令负担集中于离线训练阶段，保障了动态资源分配场景下的实时决策能力。

3　仿真结果

本节对所提 ARIS 辅助的 SGF-NOMA 安全传输方案进行仿真验证，参数设置如下[23]。基站配备 $M = 4$ 根天线，服务 $N = 4$ 个 GF 用户、1 个 GB 用户以及 1 个窃听者；ARIS 配备 $K = 16$ 个反射单元，最大放大增益 $\beta_{\max} = 10$，功率预算 $P_{\text{ARIS}} = 10$ dBm，GB 用户发射功率固定为 $P_0 = 30$ dBm，合法链路与窃听链路噪声功率均为−90 dBm；信道建模方面，ARIS 相关链路采用莱斯衰落 $K_{br} = 5, K_{nr} = 4$，直连与窃听链路采用瑞利衰落，路径损耗参考距离 $d_0 = 1$ m 处损耗 30 dB，LoS 和 NLoS 损耗指数分别为 2.2 和 3.4，用户移动相关系数 $\rho = 0.95$；传输约束设置 $R_{\min} = 1$ bit/(s·Hz)，GF 用户接入筛选门限 $\tau = 10^{-11}$；训练采用 CTDE-MATD3，经验回放池容量为 10^6，批量大小 256，折扣因子 $\gamma = 0.99$，Actor 和 Critic 学习率分别为 5×10^{-6} 与 1×10^{-5}，策略延迟更新步数 $d = 2$，软更新系数 $\tau_s = 0.005$，共训练 500 回合，每回合 100 时隙。

为评估所提 MATD3 安全资源分配算法在 ARIS 辅助 SGF-NOMA 上行系统中的性能，本节将其与典型深度强化学习算法、传统优化算法及不同 RIS 配置进行对比。所有对比均在相同的系统模型、信道生成方式与仿真参数设置下进行，以保证对比结果的公平性。鉴于目前针对 ARIS 辅助 SGF-NOMA 上行安全传输场景的相关研究较少，本文通过多类基准算法验证所提算法在保密速率提升、动态环境适应性和训练稳定性方面的优势。

（1）基于 DDPG 方案。该算法同样采用 CTDE 框架处理连续动作空间下的协调决策问题。本文将其作为 DRL 基准算法，并在相同状态、动作和奖励设置下进行对比，以验证所提 MATD3 通过引入双 Critic 网络、目标策略平滑和延迟策略更新机制所带来的性能提升。

（2）交替优化方案。将文献[23]中面向 ARIS 辅助上行 NOMA 安全传输的交替优化算法适配至本文 SGF-NOMA 场景，用于求解基站接收波束成形、ARIS 反射系数和用户发射功率的联合优化问题。通过与 AO 算法的对比，可验证所提 MATD3 算法在时变信道和 GF 用户动态接入条件下的性能优势。

（3）无源 RIS 方案。将 ARIS 替换为无源 RIS，各反射单元仅保留相位调控功能，不提供信号放大且不引入有源噪声。该方案用于评估 ARIS 主动放大相较于传统无源 RIS 的性能增益。

（4）随机相位方案。在该方案中，ARIS 反射系数矩阵随机，仅优化用户功率与基站波束成形，以单独评估 ARIS 系数联合优化对系统安全传输的影响。

图 2 展示了 MATD3 算法在不同基站天线数 M 下的训练收敛性能。从图 2 观察到，3 种配置下的平均奖励均随训练轮次的增加而快速上升，并在约 50 轮后趋于平稳，验证了所提算法具有良好的收敛性。对比不同 M 的收敛结果可见，随着基站天线数从 $M = 1$ 增加到 $M = 16$，收敛后的平均奖励显著提升。这是由于更多天线为基站提供了更高的空间自由度，使其能够形成更精确的接收波束成形向量，从而提升了系统的长期保密和速率。与此同时，基站天线数的增加也会提高波束成形变量的维度，使策略搜索空间更加复杂，但不同配置下的奖励曲线仍能保持平稳收敛，表明 MATD3 中的双 Critic 网络、目标策略平滑、延迟策略更新和软更新机制能够在一定程度上抑制连续动作空间中的价值估计偏差和策略振荡，从而提升训练过程的稳定性。

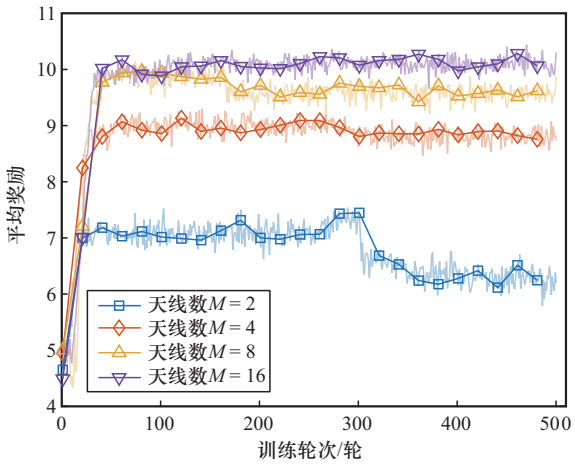


图 2　MATD3 算法在不同基站天线数 M 下的训练收敛性能

图3展示了不同算法下系统保密速率随GF用户最大发射功率变化的曲线，其中实线和虚线分别对应信道时间相关系数ρ为0.95和0.85的场景。可以看出，所有算法的保密速率均随最大发射功率的增加而提升，这是由于较高的发射功率增强了合法链路接收信干噪比。在相同信道时间相关系数及相同仿真参数设置下，本文方案在所考察的最大发射功率范围内整体优于对比方案。相较交替优化方案，MATD3通过多智能体协同学习持续优化动态传输策略，能够更好地适应信道时变和GF用户随机接入带来的不确定性；在与本文方案采用相同CTDE框架及相同状态、动作和奖励设置的条件下，双Critic网络、延迟策略更新和目标策略平滑机制有效缓解了Q值过估计与策略振荡问题，因此本文方案取得了优于DDPG方案的保密速率。当信道时间相关系数由0.95降至0.85时，各算法保密速率均下降，原因在于信道时变性增强，使资源分配策略难以持续匹配瞬时信道状态。尽管如此，本文方案仍保持领先地位，展现出较强的动态环境适应能力。此外，ARIS辅助SGF-NOMA整体性能优于无源RIS，说明ARIS可通过联合调控相移和幅度增强合法链路并抑制窃听链路，从而提升系统安全性能。

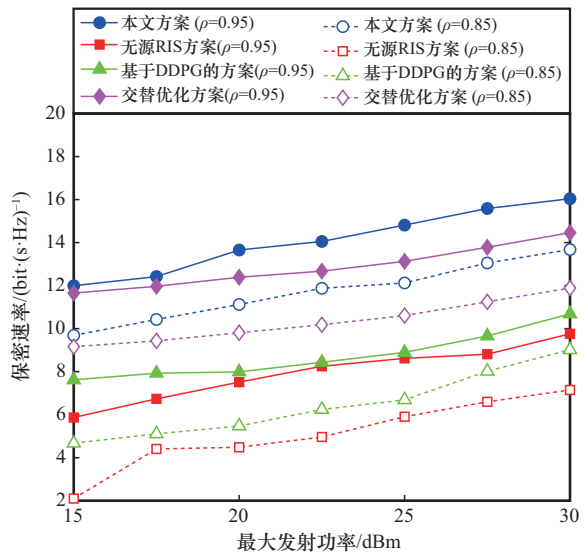


图3　不同算法下系统保密速率随GF用户最大发射功率变化的曲线

图4展示了不同算法下系统和保密速率随窃听信道估计误差CEE的变化曲线。从图4观察到，所有算法的保密速率均随CEE的增加而逐渐下降，这是因为窃听信道估计精度降低导致基站难以精确抑制窃听链路。在低CEE区域，交替优化方案基于较精确的CSI通过迭代优化可获得较优策略，与本文方案差距较小；随着CEE增大，交替优化方案基于静态CSI交替迭代，难以在信道时变过程中及时调整，所提MATD3算法通过在线交互持续适应信道动态，在CSI失配条件下仍保持较优性能。此外，ARIS的主动放大能力有效补偿了级联路径损耗，进一步拉大了与无源RIS的性能差距。

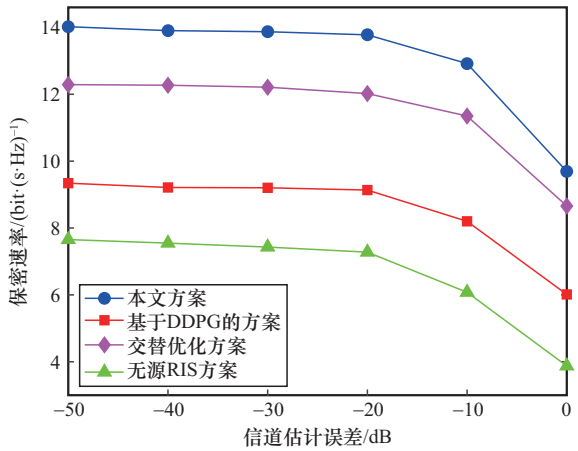


图4　不同算法下系统和保密速率随窃听信道估计误差CEE的变化曲线

图5展示了不同RIS元件数下的保密速率。结果表明，系统性能随RIS元件数的增加而提升，这是因为RIS元件数量的增加能有效提高保密速率的总增益。在相同RIS元件数下，ARIS辅助方案的保密速率均高于无源RIS方案；同时，$\beta = 10$时的保密速率高于$\beta = 5$，这表明更大的放大增益能够带来更强的合法信号增强与窃听链路抑制能力。

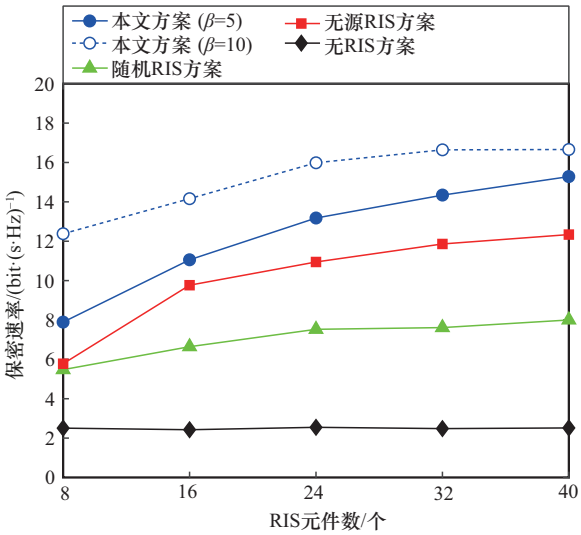


图5　不同RIS元件数下的保密速率

图 6 展示了不同 ARIS 功率预算下的保密速率。由图 6 可知，由 ARIS 辅助的保密速率随 ARIS 功率预算的增加而单调上升，但当预算高于一定门限后逐渐趋于饱和。这是因为在功率预算充足时，ARIS 放大功耗与有源噪声功率不再成为约束瓶颈，继续增加预算对合法链路的边际增益递减。无源 RIS 方案不消耗额外功率预算，其性能不受功率预算变化影响。

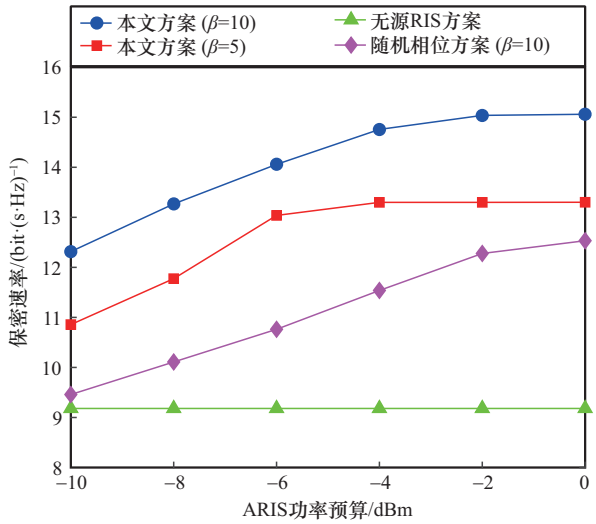


图 6　不同 ARIS 功率预算下的保密速率

图 7 给出了不同算法在不同信道时间相关系数条件下的系统保密速率变化情况。从图 7 可以看出，随着信道时间相关系数的减小，各算法的系统保密速率整体呈下降趋势。这表明在信道快速时变条件下，资源分配策略难以及时适应信道状态变化，从而导致系统安全传输性能下降。在各信道时间相关系数取值下，本文方案的系统保密速率均高于其他对比方案。这主要是由于其多智能体协同在线学习机制能够有效应对信道时变性带来的动态环境变化，并结合双 Critic 结构与延迟策略更新机制，有效抑制 Q 值过估计与策略振荡，从而提升了系统在动态信道条件下的鲁棒性与保密速率性能。

图 8 给出了系统和保密速率随接入 GF 用户数变化的结果。从图 8 可以看出，随着 GF 用户数的增加，各算法的系统和保密速率整体提升。这是因为当更多 GF 用户参与 NOMA 叠加传输时，系统可通过联合功率控制、接收波束成形和 ARIS 幅相调控，更充分地利用多用户传输增益。在相同的系统模型与仿真参数设置下，本文方案在所考察的 GF 用户数范围内均优于其余对比方案，表明其在本文场景下具有较强的协同优化能力。在相同算法框架与训练设置下，基于 DDPG 的方案受 Q 值过估计和策略振荡影响，其保密速率低于本文方案；交替优化方案受逐时隙迭代和动态环境影响，性能提升相对有限。此外，无源 RIS 方案性能最低，表明 ARIS 的主动放大能力能有效补偿级联路径损耗，从而提升系统安全传输性能。

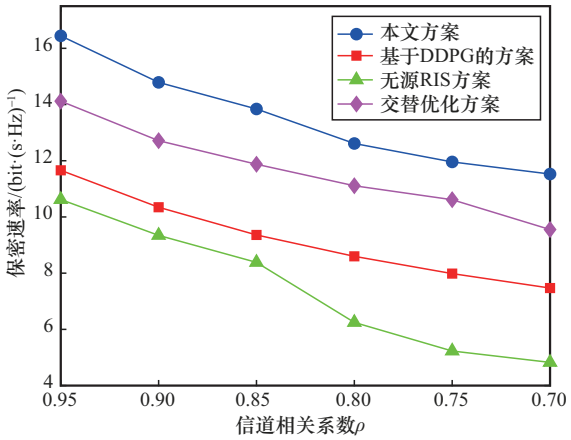


图 7　不同算法在不同信道时间相关系数条件下的系统保密速率变化情况

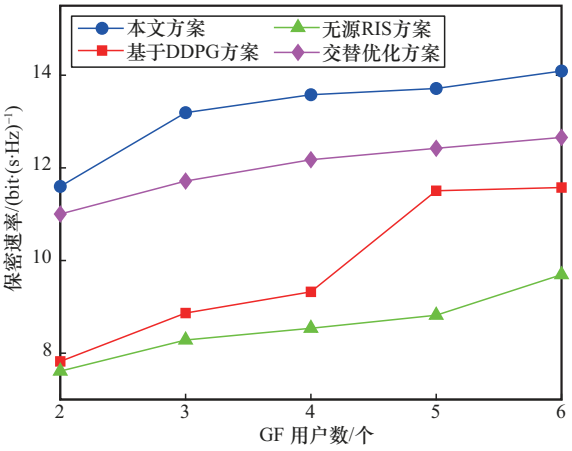


图 8　不同 GF 用户数下的保密速率

4　结束语

本文研究了信道时变与窃听 CSI 不确定条件下 ARIS 辅助 SGF-NOMA 上行系统的安全传输问题，建立了基站波束成形、ARIS 幅相系数与 GF 用户发射功率的联合优化模型，并提出了基于 CTDE 框架的 MATD3 安全资源分配算法，通过双 Critic 网络与延迟策略更新机制实现了基站与 GF

用户的协同安全决策。仿真结果表明，所提算法在训练约50轮后稳定收敛，各对比场景下保密速率均高于对比算法，其中较无源RIS方案提升约46%，验证了方案在动态不确定环境下的有效性。本文采用的范数有界误差模型仅考虑单窃听者，且算法训练复杂度随用户数与反射单元数增加而上升，未来工作将围绕多窃听者协同攻击与多ARIS部署场景下的安全资源分配展开深入研究，并引入优先经验回放机制提高样本利用效率，缩短算法收敛所需的训练回合数，从而降低大规模网络场景下的训练开销。

参考文献：

[1] Shashin A, Belogaev A, Krasilov A, et al. Adaptive parameters selection for uplink grant-free URLLC transmission in 5G systems[J]. Computer Networks, 2023, 222: 109527.

[2] Lu H B, Xie X Z, Shi Z Y, et al. Advanced NOMA assisted semi-grant-free transmission schemes for randomly distributed users[J]. IEEE Transactions on Wireless Communications, 2023, 22(7): 4638-4653.

[3] Nadif S, Sabir E, Elbiaze H, et al. Green grant-free power allocation for ultra-dense Internet of things: a mean-field perspective[J]. Journal of Network and Computer Applications, 2024, 229: 103908.

[4] Balcı A, Sokullu R. Fairness aware deep reinforcement learning for grant-free NOMA-IoT networks[J]. Internet of Things, 2024, 25: 101079.

[5] Ding Z, Schober R, Fan P, et al. Simple semi-grant-free transmission strategies assisted by non-orthogonal multiple access[J]. IEEE Transactions on Communications, 2019, 67(6): 4464-4478.

[6] Zhang C, Liu Y W, Ding Z G. Semi-grant-free NOMA: a stochastic geometry model[J]. IEEE Transactions on Wireless Communications, 2022, 21(2): 1197-1213.

[7] Ji Y H, Yu J G, Yao Y, et al. Securing wireless communications from the perspective of physical layer: a survey[J]. Internet of Things, 2022, 19: 100524.

[8] Lv L, Xu D Y, Hu R Q, et al. Safeguarding next-generation multiple access using physical layer security techniques: a tutorial[J]. Proceedings of the IEEE, 2024, 112(9): 1421-1466.

[9] Devi P, Bharti M R, Gautam D. A survey on physical layer security for 5G/6G communications over different fading channels: Approaches, challenges, and future directions[J]. Vehicular Communications, 2025, 53: 100891.

[10] Li Y Q, Jiang M, Zhang Q, et al. Secure beamforming in downlink MISO nonorthogonal multiple access systems[J]. IEEE Transactions on Vehicular Technology, 2017, 66(8): 7563-7567.

[11] Lei H J, Yang F T, Liu H W, et al. On secure NOMA-aided semi-grant-free systems[J]. IEEE Transactions on Wireless Communications, 2024, 23(1): 74-90.

[12] Cao K R, Ding H Y, Wang B H, et al. Enhancing physical-layer security for IoT with nonorthogonal multiple access assisted semi-grant-free transmission[J]. IEEE Internet of Things Journal, 2022, 9(24): 24669-24681.

[13] Alexandropoulos G C, Katsanos K D, Wen M W, et al. Counteracting eavesdropper attacks through reconfigurable intelligent surfaces: a new threat model and secrecy rate optimization[J]. IEEE Open Journal of the Communications Society, 2023, 4: 1285-1302.

[14] Singh R, Kaushik A, Shin W, et al. Indexed multiple access with reconfigurable intelligent surfaces: the reflection tuning potential[J]. IEEE Communications Magazine, 2024, 62(4): 120-126.

[15] Lou Y L, Zou Y L, Wang H, et al. Minimum multiuser secrecy rate maximization in RIS-aided NOMA systems with imperfect eve CSI[J]. IEEE Transactions on Cognitive Communications and Networking, 2026, 12: 2486-2498.

[16] Ahmed M, Raza S, Soofi A A, et al. Active reconfigurable intelligent surfaces: expanding the frontiers of wireless communication-a survey [J]. IEEE Communications Surveys & Tutorials, 2025, 27(2): 839-869.

[17] Zhang Z J, Dai L L, Chen X B, et al. Active RIS vs. passive RIS: which will prevail in 6G?[J]. IEEE Transactions on Communications, 2023, 71(3): 1707-1725.

[18] Zhi K D, Pan C H, Ren H, et al. Active RIS versus passive RIS: which is superior with the same power budget? [J]. IEEE Communications Letters, 2022, 26(5): 1150-1154.

[19] Kaur R, Bansal B, Majhi S, et al. A survey on reconfigurable intelligent surface for physical layer security of next-generation wireless communications[J]. IEEE Open Journal of Vehicular Technology, 2024, 5: 172-199.

[20] Sharma V, Allu R, Singh S K, et al. Robust transmission design in multiobjective RIS-aided SWIPT IoT communications[J]. IEEE Internet of Things Journal, 2024, 11(10): 18605-18618.

[21] Chian D M, Chen F J, Chang Y C, et al. Active RIS-assisted MIMO-OFDM system: analyses and prototype measurements[J]. IEEE Communications Letters, 2024, 28(1): 208-212.

[22] Yang X L, Wang H, Feng Y H. Sum rate maximization for active RIS-aided uplink multi-antenna NOMA systems[J]. IEEE Wireless Communications Letters, 2023, 12(7): 1149-1153.

[23] Singh S, Raviteja A, Singh K, et al. Secrecy rate maximization for active RIS-aided robust uplink NOMA communications[J]. IEEE Wireless Communications Letters, 2024, 13(11): 2960-2964.

[24] Yin Y Y, Yang L. Covert communication of active RIS-assisted uplink SGF-NOMA systems[J]. IEEE Wireless Communications Letters, 2025, 14(6): 1836-1840.

[25] Astiti N M E P, Lee B M. A survey of quantum deep reinforcement

learning for resource allocation in future wireless networks[J]. IEEE Communications Surveys & Tutorials, 2026, 28: 109-148.

[26] Luo M F, Zhou Y C, Yang L, et al. Joint optimization of power control and receive beamforming for security enhancement in SGF-NOMA systems: a distributed MADRL approach[J]. IEEE Transactions on Vehicular Technology, 2025, 74(10): 16494-16499.

[27] Fayaz M, Yi W Q, Liu Y W, et al. Toward autonomous power control in semi-grant-free NOMA systems: a power pool-based approach[J]. IEEE Transactions on Communications, 2024, 72(6): 3273-3289.

[28] Ding Z G, Schober R, Poor H V. Unveiling the importance of SIC in NOMA systems: part 1: state of the art and recent findings[J]. IEEE Communications Letters, 2020, 24(11): 2373-2377.

[作者简介]



李有明（1963– ），男，博士，宁波大学教授，主要研究方向为无线宽带通信、电力线通信、协作中继、认知无线电等。



王宏健（2002– ），男，宁波大学硕士生，主要研究方向为智能反射表面辅助无线通信、物理层安全、信号处理等。



廖昌达（2002– ），男，宁波大学硕士生，主要研究方向为无线通信、物理层安全等。



蔡万源（1998– ），男，宁波大学博士生，主要研究方向为毫米波 MIMO 信道估计、脉冲噪声估计、图信号处理理论与应用。