



XXXX

STARS 辅助 ISCAP 系统的鲁棒资源分配优化

赵伟, 宋翱, 陶瑞鑫, 赵建立, 张智娟

(华北电力大学 电气与电子工程学院, 保定 071003)

摘 要: 针对可同时透射与反射智能表面(simultaneously transmitting and reflecting surface, STARS)辅助的感知、通信与能量传输一体化(integrated sensing, communication, and power transfer, ISCAP)系统中的感知精度、通信安全与能量捕获之间的性能权衡问题, 提出一种基于 Stackelberg 博弈的元软演员-评论家 (stackelberg game-based meta soft actor-critic, ST-Meta-SAC) 算法的鲁棒资源分配方案。首先, 基于感知信道状态信息(channel state information, CSI) 误差的约束下, 推导了安全传输速率与能量捕获效率的闭式表达式。其次, 在最小捕获能量的约束下, 构建了一个最大化通信安全传输速率并最小化感知克拉美-罗界(Cramer-Rao bound, CRB) 矩阵行列式的多目标优化问题, 通过联合优化 STARS 透射与反射系数矩阵、感知信号协方差矩阵及信号波束成形矢量进行求解。最后, 考虑到所提优化问题具有多级非凸性, 传统算法难以求解, 为此提出 ST-Meta-SAC 算法进行求解。仿真结果表明, 所提方法相比基准方案具有更快的收敛速度, 验证了该方法的有效性 with 优越性。

关键词: 同时透射与反射智能表面; 克拉美-罗界; 安全传输速率; 能量捕获效率

中图分类号: TP929

文献标志码: A

doi: 10.11959/j.issn.1000-0801.

Robust Resource Allocation for STARS-Assisted Secure ISCAP Systems

ZHAO Wei, SONG Ao, TAO Ruixin, ZHAO Jianli, ZHANG Zhijuan

School of Electrical and Electronic Engineering, North China Electric Power University, Baoding 071003, China

Abstract: To address the trade-off among communication security, sensing accuracy, and harvested energy in a simultaneously transmitting and reflecting surface (STARS)-assisted integrated sensing, communication, and power transfer (ISCAP) system, a robust resource allocation scheme based on a stackelberg game-based meta soft actor-critic (ST-Meta-SAC) algorithm is proposed. First, closed-form expressions for the secrecy rate (SR) and harvested energy are derived under the sensing channel state information (CSI) errors constraints. Subsequently, we formulate a multi-

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

通信作者: 赵伟, andyzaoster@163.com

基金项目: 国家自然科学基金项目 (62471181); 河北省自然科学基金项目 (F2022502020)

Foundation Items: The National Natural Science Foundation of China (62471181), Natural Science Foundation of Hebei Province (F2022502020)



objective optimization problem that maximizes the communication secrecy rate while minimizing the determinant of the sensing Cramer-Rao bound (CRB) matrix subject to the minimum harvested energy constraint, by jointly optimizing the STARS reflection and transmission coefficient matrices, the sensing signal covariance matrix, and the signal beamforming vector. Finally, since the high-dimensional, multi-level, and non-convex nature of the problem makes it intractable for conventional methods, we propose a ST-Meta-SAC algorithm to obtain an efficient solution. Simulation results show that our proposed method outperforms benchmark schemes in terms of convergence speed, demonstrating its effectiveness.

Key words: simultaneously transmitting and reflecting surface, Cramér - Rao bound, secrecy rate, harvested energy

1 引言

随着第6代无线网络的发展,感知、通信与能量传输一体化(integrated sensing, communication, and power transfer, ISCAP)技术因能够同时支持高可靠通信、高精度感知与高效能量捕获而受到广泛关注^[1]。然而,受无线信道广播特性与严重路径损耗影响,ISCAP面临着信号泄露、环境依赖性强、能量传输效率低等挑战^[2]。值得注意的是,同时透射与反射表面(simultaneously transmitting and reflecting surface, STARS)可实现信号全空间覆盖,为上述难题提供了创新性的解决方案。理论研究表明,优化STARS的波束成形与单元数量,能够提高安全传输速率、提高感知精度^{[3],[4]}和降低发射功率需求^[5]。因此,将STARS集成至ISCAP系统,对增强通信安全性、感知精度与能量传输效率至关重要。

此外,现有的ISCAP研究主要集中于传感、通信或能量传输的孤立分析,并未考虑到这些子系统之间的协调交互机制^{[6],[7]}。具体而言,文献[6]针对RIS辅助ISCAP系统的联合资源优化问题,在通信和感知性能约束下,通过联合设计基站发射波束与RIS反射波束,以最大化能量采集性能。文献[7]进一步面向近场安全ISCAP系统,在满足感知精度和能量传输需求的条件下,通过联合波束设计提升安全传输速率(secrecy rate, SR),从而兼顾通信安全、感知和能量传输需求。文献[8]针对多用户ISCAP系统开展联合波束设

计,并分析通信、感知与能量传输之间的性能关系,指出随着通信信噪比和能量阈值增加,感知性能会下降,表明这三者之间存在明显的性能竞争。但上述工作总体仍局限于在完美信道状态信息下,以单一指标为主导的分析和优化,缺乏对通信、感知、能量三者间相互辅助协同机制的系统化分析与联合设计^[9],并且上述研究的结果均是基于完美信道状态信息(channel state information, CSI)展开的,缺乏工程的可实现性。

在ISCAP系统中获取完美的信道状态信息并不现实,由于大规模阵列特性与复杂信道环境,传统方式获取完美CSI将导致巨大的导频、反馈信令等通信开销^[2]。为了解决这个问题,文献[9]针对ISAC系统中的上行CSI估计问题,利用感知获得的到达角信息作为先验,提出感知辅助卡尔曼滤波CSI估计方法,通过对最小二乘CSI估计结果进行修正,在提高信道估计性能的同时降低计算复杂度。文献[10]针对雷达通信协作系统中的隐蔽通信问题,利用雷达回波跟踪敌方目标,并采用扩展卡尔曼滤波预测其运动轨迹及相应CSI,在此基础上联合设计雷达波形与通信发射波束成形,以提高隐蔽传输速率。文献[11]针对无线供能通信网络中的CSI获取开销问题,利用可重构智能感知表面进行到达角估计,以感知获得的角度信息替代传统信道估计,并进一步分析到达角估计误差对用户接收能量的影响。上述工作表明,利用感知辅助CSI获取能够降低资源开销,并为通信或能量性能提升提供新的路径,

但是利用感知精度或 CSI 误差对通信和能量捕获的影响尚未被研究。

针对以上研究的不足,提出了一种基于感知的 CSI 误差模型,并推导了存在传播误差时通信安全传输速率和捕获能量的闭合表达式。在最小收获能量约束下,建立联合优化框架以最大化通信 SR 和感知克拉美-罗界(Cramer-Rao bound, CRB)行列式。为解决该多目标优化的动态性和非凸性问题,提出了一种基于 Stackelberg 博弈的元软演员-评论家(stackelberg game-based meta soft actor-critic, ST-Meta-SAC)算法,该算法保留了 SAC 探索的同时增强了跨任务适应性和样本效率,利用 stackelberg 交互来协调代理参数更新,减轻局部最优并加速收敛。仿真结果证明了该算法的优越性。

2 系统模型

定义了一个由 STARS 辅助的安全 ISCAP 系统,如图1所示^[12]。

配备 N_B 根天线的基站(BS, base station)在由 M 个单元元件的 STARS 辅助下,有 K 个单天线透射用户、 J 个单天线反射用户以及一个单天线能量捕获者(energy harvester, EH)提供服务。系统

中存在窃听者1(eavesdropper 1, Eve1)和窃听者2(eavesdropper 2, Eve2)两个单天线窃听者, Eve1 位于反射区域, Eve2 位于折射区域,其中窃听者 $i, i \in \{1, 2\}$ 表示为 ei 。 $e1$ 和 $e2$ 通过最大比合并(maximum ratio combining, MRC)协作进行窃听。假设从 BS 到用户、EH 以及 Eve 的所有直达链路均被阻塞。

2.1 感知模型

BS 可发射专用感知信号 $\mathbf{x}_c \in \mathbb{C}^{N_B \times 1} \sim \mathcal{CN}(0, \mathbf{R}_c)$ 用于用户、EH 和窃听者的位置与信道估计,其中 $\mathbf{R}_c = \mathbf{E}[\mathbf{x}_c \mathbf{x}_c^H] \in \mathbb{C}^{N_B \times N_B}$, $\mathbf{E}[\cdot]$ 表示期望算子。BS 通过接收感知回波信号进行位置及信道估计,BS 与 STARS 之间的信道建模为 $\mathbf{G} \in \mathbb{C}^{M \times N_B}$, 其表达式为:

$$\mathbf{G} \triangleq \alpha_G \mathbf{a}_R^H(\theta_G^a, \phi_G^a) \mathbf{a}_M(\theta_G^d, \phi_G^d) \quad (1)$$

其中, α_G 表示复路径增益,由路径损耗和复反射因子决定, $\mathbf{a}_R \in \mathbb{C}^{N_B \times 1}$ 为基站的导向矢量, $\mathbf{a}_M \in \mathbb{C}^{M \times 1}$ 为 STARS 的导向矢量。 θ_G^a, ϕ_G^a 为从基站到 STARS 的方位角和仰角即基站的到达角度(angle of arrival, AOA)。 θ_G^d, ϕ_G^d 为从 STARS 到基站的方位角和仰角即 STARS 的离开角度(angle of departure, AOD)。类似地, STARS 到用户

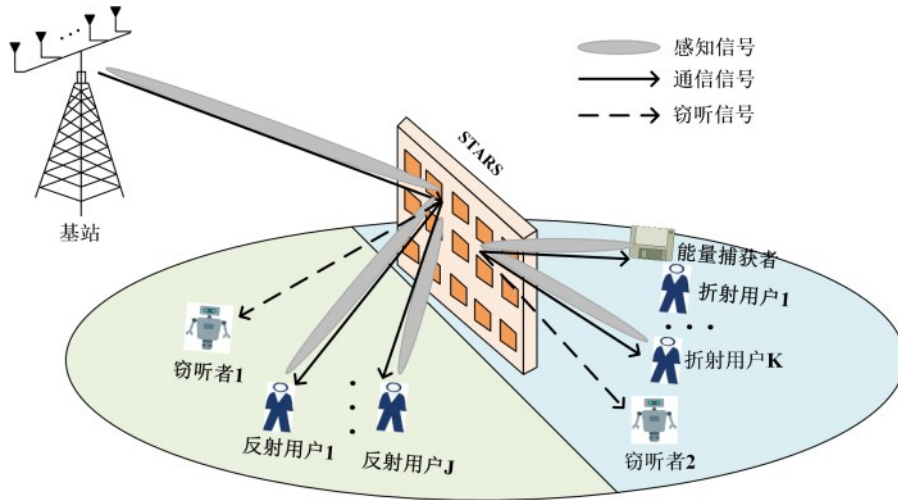


图1 STAR 辅助的 ISCAP 系统模型



$j \in J$ 、用户 $k \in K$ 、EH 和 ei 的信道分别记为 $\mathbf{h}_j \in \mathbb{C}^{M \times 1}$ 、 $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$ 、 $\mathbf{h}_{EH} \in \mathbb{C}^{M \times 1}$ 、 $\mathbf{h}_{ei} \in \mathbb{C}^{M \times 1}$ 。可建模为： $\mathbf{h}_u \triangleq \alpha_u \mathbf{a}_M(\theta_u^d, \varphi_u^d)$ ， $\mathbf{h}_u \in \mathbb{C}^{M \times 1}$ ， $u \in \{j, k, EH, ei\}$ 。其中 θ_u^d 、 φ_u^d 为离开 STARS 到 u 的方位角和仰角即 STARS 的 AOD， α_u 表示由路径损耗和复反射因子决定的系数。因此，BS 接收到的感知回波信号可表示为：

$$\mathbf{y}_s = \mathbf{H}_{bu} \mathbf{x}_c + n_s \quad (2)$$

其中， $\mathbf{H}_{bu} = \mathbf{h}_u^H \Theta_o \mathbf{G}$ ， $\mathbf{H}_{bu} \in \mathbb{C}^{1 \times N_B}$ 表示从基站 BS 到用户 $u \in \{j, k, EH, ei\}$ 的级联信道，其中， $o \in \{t, r\}$ 当 $u \in \{j, ei\}$ 时， $o = r$ ；当 $u \in \{k, EH, ei\}$ 时， $o = t$ 。由于 $u \in \{j, k, EH, ei\}$ ，因此 $\mathbf{H}_{bu}$ 可以具体表示为： $\mathbf{H}_{bj} = \mathbf{h}_j^H \Theta_o \mathbf{G}$ ， $\mathbf{H}_{bk} = \mathbf{h}_k^H \Theta_o \mathbf{G}$ ， $\mathbf{H}_{bEH} = \mathbf{h}_{EH}^H \Theta_o \mathbf{G}$ ， $\mathbf{H}_{bei} = \mathbf{h}_{ei}^H \Theta_o \mathbf{G}$ ， $\mathbf{H}_{bj}, \mathbf{H}_{bk}, \mathbf{H}_{bEH}, \mathbf{H}_{bei} \in \mathbb{C}^{1 \times N_B}$ ，分别表示基站到用户 k 、EH、用户 j 以及窃听者的级联信道， $n_s \sim \mathcal{CN}(0, \sigma^2 I)$ 为加性高斯白噪声 (additive white gaussian noise, AWGN)， I 为单位矩阵。 $\Theta_r \in \mathbb{C}^{M \times M}$ ， $\Theta_t \in \mathbb{C}^{M \times M}$ 为 STARS 反射和折射系数矩阵，表示为：

$$\Theta_r = \text{diag}(\sqrt{\beta_{r,1}} e^{j\theta_{r,1}}, \sqrt{\beta_{r,2}} e^{j\theta_{r,2}}, \dots, \sqrt{\beta_{r,M}} e^{j\theta_{r,M}}) \quad (3)$$

$$\Theta_t = \text{diag}(\sqrt{\beta_{t,1}} e^{j\theta_{t,1}}, \sqrt{\beta_{t,2}} e^{j\theta_{t,2}}, \dots, \sqrt{\beta_{t,M}} e^{j\theta_{t,M}}) \quad (4)$$

其中，STARS 采用能量分割模式同时支持透射反射， $\sqrt{\beta_{o,m}} \in [0, 1]$ 和 $\theta_{o,m} \in [0, 2\pi)$ ， $m \in \{1, 2, 3, \dots, M\}$ 分别为第 m 个 STARS 单元的幅度和相位。且幅度满足能量守恒定律： $\beta_{r,M} + \beta_{t,M} = 1$ 。

为了利用感知回波信号确定出目标的位置信息，假设位置估计的过程中基站位置已知，STARS 位置已知，用户 j 、用户 k 、EH 以及窃听者的位置均未知。设 $\mathbf{r}_\rho = [r_{\rho x}, r_{\rho y}, r_{\rho z}]^T$ 为位置坐

标， $\rho \in \{ST, j, k, EH, ei\}$ 。其中，ST 表示 STARS 反射元件。根据位置坐标可以计算离开方位角 θ_u^d 和离开仰角 φ_u^d ，分别表示为：

$$\theta_u^d = \arctan\left(\frac{r_{yu} - r_{yST}}{r_{xu} - r_{xST}}\right) - \pi - \theta_S \quad (5)$$

$$\varphi_u^d = \arctan\frac{\sqrt{(r_{xu} - r_{xST})^2 + (r_{yu} - r_{yST})^2}}{|r_{zu} - r_{zST}|} - \pi - \varphi_S \quad (6)$$

其中， θ_S 、 φ_S 分别表示 STARS 天线与正 x 轴和正 z 轴之间的夹角。

根据公式(2)的感知回波信号，令等效的级联信道为 $\mathbf{H}_s(\xi) = \mathbf{H}_{bk}(\xi) + \mathbf{H}_{bj}(\xi) + \mathbf{H}_{bei}(\xi) + \mathbf{H}_{bEH}(\xi)$ 。其中， $\mathbf{H}_s(\xi) \in \mathbb{C}^{1 \times N_B}$ ， ξ 为感知目标的位置向量。因此公式(2)可以重新表述为：

$$\mathbf{y}_s = \mathbf{H}_s \mathbf{x}_c + n_s \quad (7)$$

为了评估位置估计的感知精度，通过计算 CRB 来衡量，定义位置向量 $\xi = [\xi_k, \xi_j, \xi_{ei}, \xi_{EH}]$ ，其中 ξ_k 、 ξ_j 、 ξ_{ei} 、 ξ_{EH} 分别表示用户 k 、用户 j 、Eve 和 EH 的位置向量。用于估计 $\xi \in \mathbb{C}^{3 \times 4}$ 的费舍尔信息矩阵 (fisher information matrix, FIM) 可写为^[12]：

$$\mathbf{J}_{\xi\xi} = \frac{2}{\sigma^2} \text{Re} \left\{ \left(\frac{\partial \mathbf{u}_s}{\partial \xi} \right)^H \left(\frac{\partial \mathbf{u}_s}{\partial \xi} \right) \right\} = \begin{bmatrix} \mathbf{J}_{\xi_k, \xi_k} & \mathbf{J}_{\xi_k, \xi_j} & \mathbf{J}_{\xi_k, \xi_{ei}} & \mathbf{J}_{\xi_k, \xi_{EH}} \\ \mathbf{J}_{\xi_j, \xi_k} & \mathbf{J}_{\xi_j, \xi_j} & \mathbf{J}_{\xi_j, \xi_{ei}} & \mathbf{J}_{\xi_j, \xi_{EH}} \\ \mathbf{J}_{\xi_{ei}, \xi_k} & \mathbf{J}_{\xi_{ei}, \xi_j} & \mathbf{J}_{\xi_{ei}, \xi_{ei}} & \mathbf{J}_{\xi_{ei}, \xi_{EH}} \\ \mathbf{J}_{\xi_{EH}, \xi_k} & \mathbf{J}_{\xi_{EH}, \xi_j} & \mathbf{J}_{\xi_{EH}, \xi_{ei}} & \mathbf{J}_{\xi_{EH}, \xi_{EH}} \end{bmatrix} \quad (8)$$

其中， $\mathbf{u}_s \triangleq \mathbf{H}_s(\xi) \mathbf{x}_c$ ， $\mathbf{J}_{\xi\xi} \in \mathbb{C}^{12 \times 12}$ ， ξ 的 CRB 矩阵表示为 $\text{CRB} = \mathbf{J}_{\xi\xi}^{-1}$ 。 $\mathbf{J}_{\xi\xi}$ 中元素 $\mathbf{J}_{\xi_k, \xi_k}$ 具体表示为：

$$\mathbf{J}_{\xi_k, \xi_k} = \frac{2}{\sigma^2} \Re \left(\text{tr} \left(\left(\frac{\partial \mathbf{H}_s}{\partial \xi_k} \right)^H \mathbf{R}_c \left(\frac{\partial \mathbf{H}_s}{\partial \xi_k} \right) \right) \right) \quad (9)$$

其中， $\frac{\partial \mathbf{H}_s}{\partial \xi_k}$ 可表示为：

$$\begin{aligned}\frac{\partial \mathbf{H}_s}{\partial \xi_k} &= \frac{\partial \mathbf{h}_k^H}{\partial \xi_k} \Theta_t \mathbf{G} \\ &= \left(\left(\frac{\partial \mathbf{h}_k}{\partial r_{x_k}} \right)^H + \left(\frac{\partial \mathbf{h}_k}{\partial r_{y_k}} \right)^H + \left(\frac{\partial \mathbf{h}_k}{\partial r_{z_k}} \right)^H \right) \Theta_t \mathbf{G}\end{aligned}\quad (10)$$

进一步地, $\mathbf{h}_k$ 对各位置坐标的偏导数分别为:

$$\frac{\partial \mathbf{h}_k}{\partial r_{x_k}} = \alpha_k \left(\mathbf{a}_{M,\theta}^H \frac{\Delta r_{y_k}}{D_k} + \mathbf{a}_{M,\varphi}^H \left(-\frac{\Delta r_{z_k} \Delta r_{x_k}}{\sqrt{D_k} Q_k} \right) \right) \quad (11)$$

$$\frac{\partial \mathbf{h}_k}{\partial r_{y_k}} = \alpha_k \left(-\mathbf{a}_{M,\theta}^H \frac{\Delta r_{x_k}}{D_k} + \mathbf{a}_{M,\varphi}^H \left(-\frac{\Delta r_{z_k} \Delta r_{x_k}}{\sqrt{D_k} Q_k} \right) \right) \quad (12)$$

$$\frac{\partial \mathbf{h}_k}{\partial r_{z_k}} = \alpha_k \mathbf{a}_{M,\varphi}^H \frac{\sqrt{D_k}}{Q_k} \quad (13)$$

其中, 参数 Q_k , D_k 分别表示为: $D_k = (\Delta r_{y_k})^2 + (\Delta r_{x_k})^2$, $Q_k = D_k^2 + (\Delta r_{z_k})^2$, $\Delta r_{yu} = r_{yu} - r_{yST}$, $\Delta r_{xu} = r_{xu} - r_{xST}$ 因此, $J_{\xi,\xi}$ 的表达式即可确定。类似地, $\mathbf{J}_{\xi_j, \xi_j}$ 、 $\mathbf{J}_{\xi_j, \xi_k}$ 、 $\mathbf{J}_{\xi_{ei}, \xi_{ei}}$ 、 $\mathbf{J}_{\xi_{EH}, \xi_{EH}}$ 均可采用相同方法推导得到。

因此, 感知过程得到估计位置信息 $\hat{\xi}_u$ 估计信道记为 $\hat{\mathbf{h}}_u = \mathbf{h}_u(\hat{\xi}_u) = \mathbf{h}_u + \Delta \mathbf{h}_u$, $\hat{\mathbf{h}}_u \in \mathbb{C}^{M \times 1}$ 。其中, 信道估计误差 $\Delta \mathbf{h}_u$ 的下界可以近似可表示为:

$$\text{cov}(\Delta \mathbf{h}_u) \approx \frac{\partial \mathbf{h}_u}{\partial \xi_u} \text{CRB}(\xi_u) \frac{\partial \mathbf{h}_u^H}{\partial \xi_u} \quad (14)$$

其中, $\text{cov}(\cdot)$ 为协方差。此外, 由于 BS 和 STARS 的位置已知, θ_G^a , φ_G^a 为确定的值, 此时信道 $\mathbf{G}$ 的不完美性并非来源于 STARS 的位置误差, 于是将其建模为 $\hat{\mathbf{G}} = \mathbf{G} + \Delta \mathbf{G}$, $\hat{\mathbf{G}} \in \mathbb{C}^{N_B \times M}$, $\|\Delta \mathbf{G}\|_2 \leq \sigma_G$, 其中 σ_G^2 为误差的二范数有界值。因此, 从 BS 到用户 $u \in \{j, k, E, ei\}$ 的不完美级联信道可以进一步表示为 $\mathbf{H}_{bu} = \hat{\mathbf{h}}_u^H \Theta_o \hat{\mathbf{G}}$, $\mathbf{H}_{bu} \in \mathbb{C}^{1 \times N_B}$ 。

2.2 通信模型

基于通过感知过程估计得到的 CSI, BS 向用户 $u' \in \{j, k\}$ 发送通信信号 $\mathbf{x}_k$ 和 $\mathbf{x}_j$ 。发射信号则可以表示为 $\mathbf{x} = \sum_{k=1}^K \omega_k \mathbf{x}_k + \sum_{j=1}^J \omega_j \mathbf{x}_j$ 。其中, $\omega_k \in \mathbb{C}^{N_B \times 1}$, $\omega_j \in \mathbb{C}^{N_B \times 1}$ 分别表示面向用户 k 和用

户 j 的波束成形向量, 其协方差矩阵为 $\mathbf{R}_x = \mathbb{E}[\mathbf{x}\mathbf{x}^H]$, $\mathbf{R}_x \in \mathbb{C}^{N_B \times N_B}$ 。

用户 u' 处的 SINR 表示为:

$$\gamma_{u'} = \frac{|\hat{\mathbf{H}}_{bu'} \omega_{u'}|^2}{\sum_{t \neq u'}^{U'} |\hat{\mathbf{H}}_{bu'} \omega_t|^2 + \sum_{\varepsilon=1}^{\Xi} |\hat{\mathbf{H}}_{bu'} \omega_{\varepsilon}|^2 + \sigma_{u'}^2} \quad (15)$$

其中, 若 $u'=k$, $U'=K$, 则 $\varepsilon=j$, $\Xi=J$; 若 $u'=j$, $U'=J$, 则 $\varepsilon=k$, $\Xi=K$, $\sigma_{u'}^2$ 为噪声功率。与此同时, ei 窃听信号时的 SINR 表示为:

$$\gamma_{ei,u'} = \frac{|\hat{\mathbf{H}}_{bei} \omega_{u'}|^2}{\sum_{t \neq u'}^{U'} |\hat{\mathbf{H}}_{bei} \omega_t|^2 + \sum_{\varepsilon=1}^{\Xi} |\hat{\mathbf{H}}_{bei} \omega_{\varepsilon}|^2 + \sigma_{ei,u'}^2} \quad (16)$$

其中, $\sigma_{ei,u'}^2$ 为噪声功率。由于窃听者之间采用最大比合并方式合作窃听信号, 则窃听者的总 SINR 可表示为 $\gamma_{e,u'} = \gamma_{e1,u'} + \gamma_{e2,u'}$ 。此时系统的安全速率可表示为:

$$R_{\text{sec}} = \min_u \left[\ln(1 + \gamma_{u'}) - \ln(1 + \gamma_{e,u'}) \right]^+ \quad (17)$$

其中, $[a]^+ = \max(a, 0)$ 。

2.3 能量捕获模型

这里采用基于 sigmoidal 函数的非线性能量采集模型^[13]。相比传统线性能量采集模型, 该模型能够更准确地刻画实际 RF 能量采集电路的非线性输入—输出特性及高输入功率下的饱和效应, 并与实际电路测量数据具有较好的一致性^[13]。因此, 该模型更适合描述本文 EH 的实际能量采集过程。在不完美感知 CSI 条件下, EH 从通信信号中采集到的能量记为:

$$\hat{E}_e = \left(1 + e^{-a(\hat{P}_{er} - b)} \right)^{-1} \mathfrak{I} - \mathfrak{I} \mathcal{W} 1 - \mathcal{W} \quad (18)$$

其中, $\hat{P}_{er} = \sum_{k=1}^K |\omega_k \hat{\mathbf{H}}_{bEH}|^2 + \sum_{j=1}^J |\omega_j \hat{\mathbf{H}}_{bEH}|^2$, $\mathfrak{I}$ 为常数表示当能量捕获电路饱和时 EH 处捕获的最大能量, $\mathcal{W} = 1/1 + \exp(ab)$, a 和 b 是 EH 的详细电路规范相关的常数。在实际的应用中常数 a ,



b 以及 $\mathfrak{I}$ 可以通过标准的拟合曲线工具得到具体的数。常数 $\mathcal{W}$ 确保能量捕获电路的零输入零输出响应。

2.4 多目标优化问题与性能权衡分析

基于上述模型, 通过优化 Θ_t 、 Θ_r 、 ω_k 、 ω_j 和 $\mathbf{R}_c$, 在最大化 SR 的同时最小化 CRB 矩阵的行列式。由于 CRB 是 FIM 的逆矩阵, 最小化 $|\mathbf{J}_{\xi\xi}^{-1}|$ 等价于最大化 $|\mathbf{J}_{\xi\xi}|$, 因此多目标优化问题表述为:

$$P1: \max_{\omega_k, \omega_j, \mathbf{R}_c, \Theta_t, \Theta_r} [R_{\text{sec}}, |\mathbf{J}_{\xi\xi}|] \quad (19a)$$

$$s.t. \quad \hat{E}_e \geq E_{\text{ref}} \quad (19b)$$

$$\text{tr}(\mathbf{R}_c) \leq P_c, \text{tr}(\mathbf{R}_x) \leq P_s \quad (19c)$$

$$\beta_{r,M} + \beta_{t,M} = 1 \quad (19d)$$

$$\sqrt{\beta_{o,M}} \in [0, 1], \theta_{o,M} \in [0, 2\pi] \quad (19e)$$

其中, $o \in \{t, r\}$, $h \in [0, 1]$ 表示加权系数, E_{ref} 表示为最小采集能量阈值, P_c 和 P_s 为 BS 发送信号的最大功率。(19b) 为能量采集约束, (19c) 为发射功率约束, (19d)、(19e) 为 STARS 单元的幅度与相位条件。为进一步说明安全通信性能与感知性能之间的权衡关系, 需要区分感知辅助通信产生的协同增益与共享资源条件下的性能竞争。一方面, 由式 (14) 可知, 位置估计 CRB 通过感知辅助 CSI 获取过程影响级联信道估计误差, 结合式 (15) 一式 (17) 可知, 较小的 CRB 能够降低信道不确定性, 从而改善合法用户的接收 SINR 和安全传输性能。另一方面, 安全传输速率与感知 FIM 均由 STARS 透射/反射系数及 BS 波束资源共同决定, 而 STARS 单元受到 $\beta_{r,M} + \beta_{t,M} = 1$ 的能量分割约束。因此, 安全通信所需的合法链路增强和窃听链路抑制, 与提高目标参数估计精度所需的感知波束配置通常不能同时达到各自的单目标最优状态, 从而在共享资源约束下形成通信与感知性能之间的权衡。

为刻画上述性能权衡, 第 l 个子任务采用权

重向量 $\lambda^l = [\lambda_1^l, \lambda_2^l]^T$, 其中, λ_1^l 和 λ_2^l 分别表示安全通信目标与感知目标的权重, 且满足 $\lambda_1^l + \lambda_2^l = 1$, $\lambda_1^l, \lambda_2^l \in [0, 1]$ 。当 λ_1^l 增大时, 优化过程更加侧重安全传输速率; 当 λ_2^l 增大时, 则更加侧重感知精度。若某一可行解不存在另一可行解能够在不降低其中一个目标性能的条件下进一步提高另一目标, 则该解为 Pareto 有效解。因此, 通过设置不同的权重向量, 可以表征安全通信与感知性能之间不同的折中偏好; 当相应加权子问题获得全局最优解时, 其解对应多目标问题的受支持 Pareto 有效解。

问题 P1 的多目标优化要求在 CSI 变化条件下进行性能权衡, 而传统凸优化方法通常需要固定权重来简化这种权衡, 无法实时捕捉目标之间的动态平衡。元强化学习可通过多任务预训练提取可迁移知识^[14], 其动态参数调整能力特别适用于系统不断变化的无线通信场景。因此, 采用一种基于元强化学习的算法来求解该问题。

3 问题解决方案

ST-Meta-SAC 算法结合了 Stackelberg 博弈和元学习, 通过分层决策结构协调多智能体交互。它分两个阶段运行: 基于 Stackelberg 博弈的全局博弈阶段和 Meta-SAC 本地更新阶段。在 Stackelberg 博弈阶段, 各智能体依据当前策略进行领导者一跟随者交互, 并获得相应的效用及环境反馈; 在本地更新阶段, 各智能体利用交互样本通过 Meta-SAC 更新策略网络、价值网络及元参数。更新后的策略参数继续用于下一轮 Stackelberg 博弈, 两个阶段交替执行直至策略趋于稳定。具体过程如算法 1 所示。

算法引入了 L 个权重向量 $\lambda^l = [\lambda_1^l, \lambda_2^l]^T$ 并满足 $\lambda_1^l + \lambda_2^l = 1, l \in L$, 从而将多目标优化问题分解为多个子问题, 每个子问题对应一个特定的权重

向量。不同权重向量对应第2.4节所述的不同通信—感知性能偏好，使智能体能够针对不同的Pareto折中工作点学习相应的资源分配策略。该问题可表述为一个马尔可夫决策过程，其重要参数可以由元组 $(\mathcal{S}, \mathcal{A}, \pi, \mathcal{R}, \gamma, \mathcal{P})$ 来刻画。其中， π 为采取行动的策略；折扣因子 $\gamma \in [0, 1]$ 用于量化未来奖励相对于即时奖励的重要程度； $\mathcal{P}$ 表示系统从状态 $\mathcal{S}^t$ 转移到状态 $\mathcal{S}^{t+1}$ 的概率。具体而言， $\mathcal{S}$ 、 $\mathcal{A}$ 、 $\mathcal{R}$ 分别表示如下：

状态 $\mathcal{S}$ ：BS及用户 $u \in \{j, k, EH, ei\}$ 处元智能体，第 t 个时隙的状态空间分别为： $s_{BS}^t = \{\hat{\mathbf{G}}, \hat{\mathbf{h}}_u\}$ ， $u \in \{j, k, EH, ei\}$ ， $s_u^t = \{\gamma_k, \gamma_j, \hat{E}_e\}$ 。智能体的状态不仅取决于当前的动作而且与之前的状态有关。

动作 $\mathcal{A}$ ：BS处元智能体的动作空间： $a_{BS}^t = \{\omega_k, \omega_j, \mathbf{R}_c\}$ ，用户 u 处元智能体的动作空间为： $a_u^t = \{\Theta_r, \Theta_t\}$ 。

奖励 $\mathcal{R}$ ：在 t 时隙BS的奖励函数为： $r_{BS}^t = |J_{\xi\xi}|$ ，

用户 u 的奖励函数为 $r_u^t = R_{\text{sec}}$ 。定义第 l 个子任务的联合奖励函数为： $R_l^t = \lambda_1^t R_{\text{sec}} / R_{\text{ref}} + \lambda_2^t |J_{\xi\xi}| / |J_{\max}|$ 其中， R_{ref} 为理论最大速率， $|J_{\max}|$ 为CRB的理论最大值。该联合奖励用于评价全局博弈阶段产生的联合策略，并作为本地更新阶段的奖励反馈。

对于第 l 个子任务，将BS与跟随者之间的策略交互建模为Stackelberg博弈，其中BS作为领导者先制定策略，跟随者在获得领导者策略后进行响应，双方分别通过最大化各自效用完成策略更新。在全局博弈阶段，BS处的元智能体作为

算法1 ST-Meta-SAC算法

```

1  初始化本地数据集 $C_l$ ，领导者策略参数 $\varphi_{BS}$ ，跟随者策略参数 $\varphi_u$ ，Actor参数 $\phi^i$ ，Critic参数 $\delta_1^i, \delta_2^i$ ，熵参数 $\eta^i$ ，子任务数 $L$ ，最大训练迭代次数 $e_m$ 
2  for  $e = 1:e_m$  do
3    for  $l = 1:L$  do
4      全局博弈阶段
5      BS领导者根据当前策略进行决策，并根据式(20)计算领导者效用 $u_b(\tilde{\mathcal{R}})$ 
6      按照式(21)更新领导者策略参数 $\varphi_{BS}$ 
7      跟随者根据领导者策略进行响应，并根据式(22)计算跟随者效用 $u(\mathcal{C}_u, \mathcal{R}_l^t)$ 
8      按照式(23)更新跟随者策略参数 $\varphi_u \leftarrow \varphi_u + \varepsilon \nabla_{\varphi_u} u(\mathcal{C}_u, \mathcal{R}_l^t)$ 
9    end for
10   本地更新阶段
11   将交互样本 $(s_i^{t-1}, s_i^t, a_i^t, s_i^t, r_i^t)$ 存入本地数据集 $C_l$ ，并随机采样批量数据；
12   由式(27)和式(28)计算Critic损失及目标软价值，更新双Critic网络 $\delta^i$ ；
13   根据式(29)更新Actor网络参数 $\phi^i$ 
14   由式(30)更新熵参数 $\eta^i$ ；
15   按照式(31)~式(33)更新元参数 $\phi_m^i, \delta_{zm}^i, \eta_m^i$ ；
16   按照式(34)和式(35)软更新目标网络 $\tilde{\delta}^i, \tilde{\phi}^i$ ；
17 end for

```

领导者，参数为 φ_{BS} ，策略网络 $\pi_{BS}(\cdot | s_u^t; \varphi_{BS})$ ，而用户 $u \in \{j, k, EH, ei\}$ 作为跟随者，参数为 φ_u ，策略网络 $\pi_u(\cdot | s_u^t; \varphi_u)$ 。在第 t 个时隙开始时，基站基于先前的状态 s_{BS}^{t-1} 制定策略，预测出波束成形以及感知信号协方差矩阵的动作 a_{BS}^t 。用户 u 根据先前状态 s_u^{t-1} 以及与基站的交互来预测STARS幅度相位的调整动作 a_u^t ，以此最大化交互效用。

基站领导者的效用函数，领导者制定一项旨在最大化其效用的激励策略，追随者收益取决于追随者可以提供的回报、信誉以及其运算时间，因此其效用函数表示为：

$$u_b(\mathcal{R}_l^t) = \beta \sum_i \ln(T_{\max}^i - T_{\text{comm}}^i - T_{\text{comp}}^i) \mathcal{H}(\mathcal{R}_l^t) \left(1 - e^{-\sum_i C_i}\right) - \sum_i \mathcal{H}(\mathcal{R}_l^t) \mathcal{R}_l^t / |i| \quad (20)$$

其中， $\mathcal{H}(\mathcal{R}_l^t) = (1 - \bar{\omega}) \log(\mathcal{R}_l^t) + \bar{\omega}$ 为追随者的信誉

函数， $\bar{\omega}$ 是默认信誉值， β 为收益加权参数，



$T_{\max}^i, T_{\text{comm}}^i$ 和 T_{comp}^i 分别为领导者所允许追随者完成一轮任务循环的最大时间、追随者数据传输时间、追随者数据处理时间, $|i|$ 为智能体追随者的个数, $C_i = (s_i^{t-1}, s_i^t, a_i^t, s_i^t, r_i^t)$, $\ln(T_{\max}^i - T_{\text{comm}}^i - T_{\text{comp}}^i)$ 表示其收益与每个追随者完成时间呈现递减的关系。此外, 领导者的目标是最大化其效用函数 $u_b(\mathcal{R}_l^t)$ 。领导者的参数更新为:

$$\varphi_{BS} \leftarrow \varphi_{BS} + \varepsilon \nabla_{\varphi_{BS}} \mathbb{E}[u_b] \quad (21)$$

其中, $\nabla_{\varphi_{BS}} \mathbb{E}[u_b] = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \nabla_{\varphi_{BS}} \log \pi_{BS}(a_{BS}^t | s_{BS}^t) (u_b^t - b_{BS})$, $\mathcal{B} \in C_i$, b_{BS} 为基线用于降方差, ε 为学习率。

用户 u 作为追随者, 其对应智能体 i 通过提供本地数据 $(s_i^{t-1}, s_i^t, a_i^t, r_i^t)$ 参与学习任务, 在提高全局收益的同时获得相应回报。其效用函数由两部分组成: 一部分为参与学习任务获得的收益, 另一部分为计算和传输产生的成本, 因此可表示为:

$$u_u(C_i, \mathcal{R}_l^t) = \frac{C_u}{\sum_{n=1}^{|u|} |C_n|} \frac{\sum_i \mathcal{H}(\mathcal{R}_l^t) \mathcal{R}_l^t}{|i|} - C_i(c_{\text{comm}}^i + c_{\text{comp}}^i) \quad (22)$$

其中, c_{comm}^i 和 c_{comp}^i 分别表示数据传输成本和计算成本, $\sum_{n=1}^{|u|} |C_n|$ 为用于训练数据集的总数。此外追随者的目标也是最大化其效用函数 $u_u(C_i, \mathcal{R}_l^t)$, 追随者的参数更新为:

$$\varphi_u \leftarrow \varphi_u + \varepsilon \nabla_{\varphi_u} \mathbb{E}[u(C_u, \mathcal{R}_l^t)] \quad (23)$$

其中, $\nabla_{\varphi_u} \mathbb{E}[u_u] = \frac{1}{|\mathcal{B}|} \sum_{i \in \mathcal{B}} \nabla_{\varphi_u} \log \pi_u(a_u^t | s_u^t, a_{BS}^t) (u_u^t - b_u)$, $\mathcal{B} \in C_i$, b_u 为基线用于降方差, ε 为学习率。

完成第 l 个子任务的全局博弈后, 将领导者与跟随者产生的联合动作作用于当前系统环境, 并获得对应的联合奖励 R_l^t 及下一状态。由当前状态、联合动作、奖励和下一状态构成的交互样本

存入本地数据集 C_i , 用于后续 Meta-SAC 参数更新。因此, 全局博弈阶段负责策略交互及样本生成, 本地更新阶段利用上述交互样本完成网络参数更新, 更新后的策略继续用于下一轮全局博弈。

在本地更新阶段, 每个智能体 i 利用全局博弈阶段产生的本地数据集 C_i 进行 Meta-SAC 训练。为统一表示不同智能体的策略网络参数, 记智能体 i 的 Actor 参数为 φ^i , 当 $i=BS$ 时, $\varphi^i = \varphi_{BS}$; 当 $i=u$ 时, $\varphi^i = \varphi_u$ 。Meta-SAC 算法采用一个 Actor 和两个 Critic 网络, 参数分别为 $\varphi^i, \delta_1^i, \delta_2^i$ 。此外, 为了增强收敛性, 引入了一个目标 Actor 网络 ($\tilde{\varphi}^i$) 和两个目标 Critic 网络 (δ_1^i 和 δ_2^i)。Meta-SAC 算法的基本过程如下:

$$V^\pi(s_i^t) = \mathbb{E}_{a_i^t \sim \pi} [Q(s_i^t, a_i^t) + \eta H(\pi(\cdot | s_i^t))] \quad (24)$$

$$Q(s_i^t, a_i^t) = \mathbb{E}_{s_i^{t+1} \sim \mathcal{P}} [r_i^t + \gamma V^\pi(s_i^{t+1})] \quad (25)$$

$$\pi = \arg \min_{\pi} D_{\text{KL}} \left[\pi(a_i^t | s_i^t) // e^{\frac{1}{\eta} Q(s_i^t, a_i^t)} / Z(s_i^t) \right] \quad (26)$$

其中, $H(\pi(\cdot | s_i^t)) = -\mathbb{E}_{a_i^t \sim \pi} \log \pi(a_i^t | s_i^t)$ 表示策略 π 下的熵, η 是温度参数。式(24)表示引入熵项后的软价值函数; 式(25)表示软 Bellman 方程(通过将熵与对应于下一状态的值函数相结合来综合评估策略); 式(26)通过最小化 KL 散度, 以寻求一种新的更新策略, 该策略不仅旨在实现更大的值, 而且保持更高的熵。是归一化分布函数 $Z(s_i^t)$ 。

每个智能体通过最小化软 Bellman 残差来更新 Critic 网络, 因此 Critic 网络的损失函数可写为:

$$L(\delta_z^i, C_i) = \mathbb{E}_{C_i} \left[\left(Q_{\delta_z^i}(s_i^{t-1}, a_i^t) - f^t \right)^2 \right], z = \{1, 2\} \quad (27)$$

其中, 每个智能体的目标软价值表示为:

$$f^i = \gamma \left[\min_{z=1,2} Q_{\delta_z^i}(s_i^t, \hat{a}_i^{t+1}) - \eta^i \log \pi_{\phi^i}(\hat{a}_i^t | s_i^{t+1}) \right] + r_i^t \quad (28)$$

其中, $\hat{a}_i^t$ 并不是数据集中存储的动作 a_i^t , 而是根据策略采取的动作, 定义为 $\hat{a}_i^t \sim \pi(\cdot | s_i^{t-1})$ 。其中, 每个 Actor 网络的损失函数可以表示为:

$$L(\phi^i, C_i) = \mathbb{E}_{C_i} [\eta^i \log \pi_{\phi^i}(\hat{a}_i^t | s_i^{t+1}) - \min_{z=1,2} Q_{\delta_z^i}(s_i^t, \hat{a}_i^{t+1})] \quad (29)$$

为了增强策略在不同训练阶段之间的适应能力, 引入了一种动态熵调节机制, 以实现对学习过程的自适应控制。目标熵定义为 $\bar{W}^i = \dim(a_i^t)$, 相应的损失函数表示为:

$$L(\eta^i, C_i) = \mathbb{E}_{C_i} [\eta^i \log \pi_{\phi^i}(\hat{a}_i^t | s_i^{t+1}) - \eta^i \bar{W}^i] \quad (30)$$

参数与熵的更新可如下面的公式表示为 $\phi^i = \arg \min_{\phi^i} \sum_{l \in L} L(\phi_l^i, C_i)$, $\delta_z^i = \arg \min_{\delta_z^i} \sum_{l \in L} L(\delta_{z,l}^i, C_i)$, $\eta^i = \arg \min_{\eta^i} \sum_{l \in L} L(\eta_l^i, C_i)$ 。其中, $\phi_l^i, \delta_{z,l}^i, \eta_l^i$ 表示第 l 个任务的参数。随后, 在元适应阶段, 智能体随机采样一批 C_i , 并按如下方式更新演员网络和评论家网络的参数:

$$\phi_m^i = \phi_m^i - \alpha_{\phi_m^i} (\phi^i - \phi_m^i), \quad (31)$$

$$\delta_{zm}^i = \delta_{zm}^i - \alpha_{\delta_{zm}^i} (\delta_z^i - \delta_{zm}^i), \quad (32)$$

$$\eta_m^i = \eta_m^i - \alpha_{\eta_m^i} (\eta^i - \eta_m^i), \quad (33)$$

其中, $\alpha_{\phi_m^i}, \alpha_{\delta_{zm}^i}, \alpha_{\eta_m^i}$ 表示相应的学习率。为了在保持学习稳定性的同时保证策略演化的平稳性, 采用软更新来更新目标评论家网络和目标演员网络的参数:

$$\tilde{\delta}_z^i = \zeta_{\delta_z^i} \tilde{\delta}_z^i + (1 - \zeta_{\delta_z^i}) \delta_z^i \quad (34)$$

$$\tilde{\phi}^i = \zeta_{\phi^i} \tilde{\phi}^i + (1 - \zeta_{\phi^i}) \phi^i \quad (35)$$

其中, ζ_{ϕ^i} 和 $\zeta_{\delta_z^i}$ 表示软更新的学习率。上述全局博弈阶段和本地更新阶段在各训练轮次中重复执行, 直至达到预设的最大训练轮次 e_m 。本文通过训练过程中平均联合奖励的变化评价算法的数值

收敛性, 当平均奖励随训练轮次增加逐渐趋于稳定时, 表明策略更新进入稳定状态。由于原优化问题具有非凸性, 本文不对全局最优收敛作理论保证, 而通过训练奖励的稳定性评价所提算法的数值收敛性能。

ST-Meta-SAC 算法的计算复杂度分析由两部分组成, 即全局博弈阶段和本地更新阶段。其中, 全局博弈的计算复杂度表示为 $\mathcal{O}(Le_m(K+J)^2 DMN_B)$ 。其中, D 表示神经网络参数的维度。本地更新的复杂度可表示为 $\mathcal{O}(Le_m(K+J)B2D)$, 其中, B 是数据批次 C_i 的大小。因此, 算法的总复杂度可表示为 $\mathcal{O}(Le_m(K+J)D((K+J)MN_B+2B))$ 。

基于 Stackelberg 博弈与元学习的 ST-Meta-SAC 分布式优化算法能够在不同系统规模和环境下快速适应, 并且随着智能体数量增加及其交互复杂度提升, 仍然能够实现高效扩展。

4 仿真结果

假设 STARS 中心元件位于坐标原点处 $(0,0,0)$, STAR-RIS 距离 BS 50 米坐标为 $(30,40,0)$, EH 随机分布在距离 STARS 50 米的范围内, 用户随机分布在距离 STARS 两侧 30 至 60 米的范围内, Eve 在距离 STARS 30 米处。带宽为 50MHz, 载波频率为 28GHz, STAR-RIS 的元件数量 $M=64$, BS 天线数 $N_B=10$, EH 的天线数 $N_E=10$, 透射用户与反射用户数分别为 K 和 J , 最大发射功率 $P_c=P_s=30\text{dBm}$, 噪声功率 $\sigma_{su}^2 = \sigma_k^2 = \sigma_s^2 = \sigma_{e1}^2 = \sigma_{e2}^2 = -110\text{dBm}$, 算法批量大小为 32。其中, BS 天线数、最大发射功率和噪声功率等物理层参数参考已有 STARS 辅助 ISAC 研究 [12] 进行设置; BS、STARS 及用户的部署距离参考文献 [12] 中 STARS 辅助 ISAC 系统采用的数十米级部署尺度, 并根据本文透射/反射双区域用户



分布进行调整；EH和Eve的部署位置则结合本文能量采集和安全通信场景进行设置。非线性能量采集模型及相关电路参数参考文献[13]设置。ST-Meta-SAC的学习率、折扣因子和批量大小等算法超参数通过预实验确定，以兼顾训练过程的收敛速度与稳定性，其余参数见表1^[13]。

表1 仿真参数

仿真参数	物理含义	数值
a	非线性能量采集模型电路参数	1500
b	非线性能量采集模型电路参数	0.0022
$\mathfrak{S}$	EH电路饱和时的最大采集能量	20mW
ε	Stackelberg 博弈策略更新学习率	5×10^{-4}
ϖ	默认信誉值	0.5
Υ	折扣因子	0.99
$\alpha_{\varphi^i} = \alpha_{\delta_l^i} = \alpha_{\eta^i}$	元学习率	5×10^{-4}
L	子任务数量	100
$\xi_{\varphi^i} = \xi_{\delta_z^i}$	目标网络软更新系数	0.05

为了从算法求解性能和系统结构性能两个层面验证所提方案的有效性，在算法层面选取传统的深度确定性策略梯度(deep deterministic policy gradient, DDPG)^[15]、半定松弛(semidefinite relaxation, SDR)^[2]作为现有方法的对比基准。同时，为分析ST-Meta-SAC中不同组成机制对性能的贡献，进一步设置Meta-SAC和SAC两种消融方案，其中Meta-SAC去除Stackelberg博弈机制而保留元学习机制，SAC进一步去除元学习机制，仅保留基础SAC学习框架。在系统结构层面，设置传统RIS辅助方案作为对比基准。

图2展示了不同表面结构及系统配置下感知CRB对合法用户接收SINR的影响。随着CRB增大，各配置下的SINR均呈下降趋势，表明感知误差增大会加剧级联信道不确定性，从而降低合法链路的有效信号增益。在STARS方案中，增大表面元件数量能够提供更多空间自由度和阵列增益，因此在相同CRB下可获得更高的SINR。为保证RIS与STARS比较的公平性，本文仅比较元

件数和用户数量均相同的两组配置，即RIS的 $M=100$ ， $K=J=10$ 与STARS的 $M=100$ ， $K=J=10$ ，其余STARS曲线用于分析元件数量和用户数量变化对SINR的影响。由图2可见，在相同的 M 、 K 和 J 条件下，STARS获得了高于传统RIS的SINR，这是由于STARS能够同时调控透射和反射链路，为两侧用户提供更灵活的空间域资源配置。

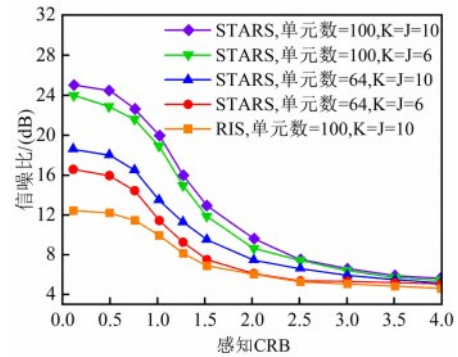


图2 克拉美-罗界和信噪比之间的关系

图3展示了不同优化算法在训练过程中的奖励值变化，来评估算法在优化问题中的学习效率与稳定性。可以观察到，所提ST-Meta-SAC在训练初期即能实现更快的回报提升，并在较少迭代轮次内达到更高的稳定值，同时曲线后期振荡幅度更小，表明该方法不仅具有更强的收敛速度，也具备更好的训练稳定性与鲁棒性。相比之下，基线方法往往表现为上升更慢或在一定阶段出现明显波动，这通常意味着在强非凸耦合约束下，策略更新更容易受到估计偏差与探索不足的影响，从而导致收敛效率受限。因此图3从训练过程的角度验证了所提方法能够在稳定收敛后维持更高的长期回报，这为性能指标的提升提供了支撑。

图4展示了不同优化算法下感知CRB与捕获能量关系对比，分析了所提方法与多种基线算法在性能优化上的表现。整体上可以看到，随着CRB变小即感知精度提升，系统可实现更高的能

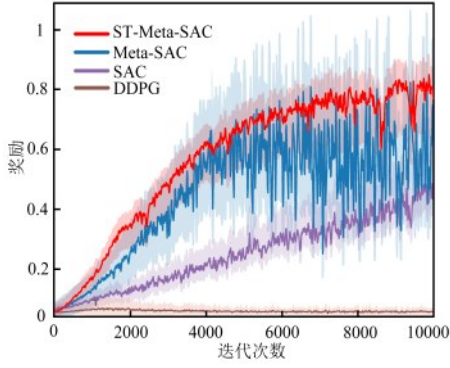


图3 算法的收敛性比较

量捕获水平,而当CRB增大时,由于感知误差加剧导致级联信道不确定性增强,能量波束难以持续聚焦到EH,从而使平均捕获能量下降并呈现更强的误差敏感性。在相同CRB水平下,ST-Meta-SAC算法曲线通常位于其他方法之上,意味着它能在给定感知精度约束下获得更高的平均捕获能量,等价地,在相同能量需求下,它可以实现更小的CRB,从而体现出更优的感知和能量的协同优化能力。

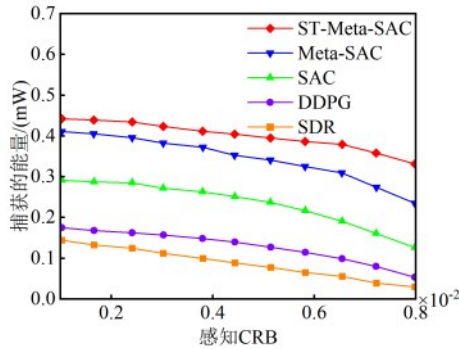


图4 不同算法下克拉美-罗界与采集能量之间的关系

图5展示了随着STARS元件数增加,系统安全速率整体提升的趋势,并揭示了不同算法在安全速率优化上的差异。随着元件数增大,使系统具备更强的波束成形与干扰抑制能力,从而带来安全速率的持续增长。并且所提ST-Meta-SAC在不同STARS元件数下均表现出更高的安全速率曲线,且在中大规模数量下优势更明显,而部分基线方法虽然也随STARS元件数量增大而上升,但

提升幅度相对有限,难以有效挖掘STARS阵列规模扩展带来的潜在增益。另外分析了参数 α 的敏感性,可以看出 $\alpha=5 \times 10^{-4}$ 时获得了最好的安全速率性能。

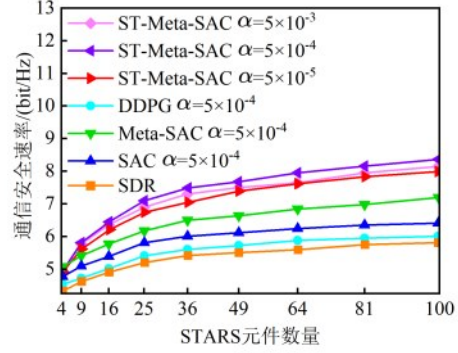


图5 不同算法下STARS单元数与安全传输速率之间的关系

5 结论

研究了STARS辅助ISCAP系统中的感知辅助通信与能量采集问题。首先推导了用于位置估计的CRB,以量化感知精度;随后在考虑感知误差项的基础上,推导了通信安全传输速率和采集能量的闭式表达式。针对多目标优化问题,提出了ST-Meta-SAC算法。结果表明,与基准方案相比,所提方法能够获得显著的性能增益和更快的收敛速度。

参考文献:

- [1] CHEN Y, REN Z, XU J, et al. Integrated sensing, communication, and powering: Toward multi-functional 6G wireless networks[J]. IEEE Communications Magazine, 2025, 63(8): 146-153.
- [2] DAI J, LIU C, PAN C, et al. Secure downlink transmission for integrated sensing, energy, and communication systems[J]. IEEE Communications Letters, 2024, 28(9): 1996-2000.
- [3] LIU Z, LI X, JI H, et al. Exploiting STAR-RIS for physical layer security in integrated sensing and communication networks[C]//Proceedings of the 2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC). Piscataway: IEEE Press, 2023: 1-6.
- [4] KUMAR V, CHAFII M. Green integration of sensing, commu-



- nication, and power transfer via STAR-RIS[C]//2025 IEEE International Conference on Communications (ICC). Montreal, Canada: IEEE, 2025: 2894-2899.
- [5] YASWANTH J, KATWE M, SINGH K, et al. Toward green communication: Power-efficient beamforming for STAR-RIS-aided SWIPT[J]. IEEE Transactions on Green Communications and Networking, 2024, 8(4): 1545-1563.
- [6] YANG Y, GAO H, YANG X, et al. Joint beamforming for RIS-assisted integrated communication, sensing and power transfer systems[J]. IEEE Wireless Communications Letters, 2024, 13(2): 288-292.
- [7] REN Z, ZHANG S, LI X, et al. Secure communications in near-field ISCAP systems with extremely large-scale antenna arrays [C]//Proceedings of the 2024 19th International Symposium on Wireless Communication Systems (ISWCS). Piscataway: IEEE Press, 2024: 1-6.
- [8] ZHOU Z, LI X, ZHU G, et al. Integrating sensing, communication, and power transfer: Multiuser beamforming design[J]. IEEE Journal on Selected Areas in Communications, 2024, 42(9): 2228-2242.
- [9] CHEN X, FENG Z, ZHANG J A, et al. Sensing-aided uplink channel estimation for joint communication and sensing[J]. IEEE Wireless Communications Letters, 2023, 12(3): 441-445.
- [10] WANG X, FEI Z, LIU P, et al. Sensing-aided covert communications: Turning interference into allies[J]. IEEE Transactions on Wireless Communications, 2024, 23(9): 10726-10739.
- [11] LUO C, HU J, XIANG L, et al. Reconfigurable intelligent sensing surface aided wireless powered communication networks: A sensing-then-reflecting approach[J]. IEEE Transactions on Communications, 2024, 72(3): 1835-1848.
- [12] WANG Z, MU X, LIU Y. STARS enabled integrated sensing and communications[J]. IEEE Transactions on Wireless Communications, 2023, 22(10): 6750-6765.
- [13] BOSHKOVSKA E, NG D W K, ZLATANOV N, et al. Robust resource allocation for MIMO wireless powered communication networks based on a non-linear EH model[J]. IEEE Transactions on Communications, 2017, 65(5): 1984-1999.
- [14] ZHANG Z, WU Z, ZHANG H, et al. Meta-learning-based deep reinforcement learning for multiobjective optimization problems[J]. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(10): 7978-7991.
- [15] MOON S, LIU H, HWANG I. Joint beamforming for RIS-assisted integrated sensing and secure communication in UAV networks[J]. Journal of Communications and Networks, 2024, 26(5): 502-508.

赵伟（1980-），男，华北电力大学电气与电子工程学院教师，主要研究方向通信、感知、计算与控制一体化技术、控一体化技术、新型电力系统数智化技术、工业互联网战略。

宋翱（2004-），男，华北电力大学电气与电子工程学院研究生，主要研究方向通信、感知、计算



一体化的隐蔽通信。

赵建立（1973-），男，华北电力大学电气与电子工程学院讲师、研究生导师，主要研究方向电力通信网综合网络管理，智能电网通信技术，认知无线电通信技术等。

张智娟（1971-），女，华北电力大学电气与电子工程学院副教授、研究生导师，主要研究方向功率变换器控制及负载拓扑结构、无限功率传输能量采集。

陶瑞鑫（2003-），女，华北电力大学电气与电子工程学院研究生，主要研究方向通信、感知、计算与控制一体化技术。

