doi: 10.7690/bgzdh.2024.04.007

基于“电磁陷阱”的自适应卫星通信干扰策略

颜培杰，储飞黄，王梦阳
（航天工程大学航天信息学院，北京 101416）

摘要：针对基于 Q-Learning 的自适应卫星节点，结合传统与智能干扰方式，提出一种“电磁陷阱”的诱骗干扰策略。通过设置诱骗模块、热启动模块、反智模块，形成干扰闭环，达到“诱骗反智”的作战效果。仿真结果表明：相比其他算法，该方法可有效干扰自适应卫星节点，提升干扰精确度，降低通信吞吐量。

关键词：电磁陷阱；诱骗干扰；卫星通信；Q-Learning；信道选择
中图分类号：TN927　　**文献标志码：**A

Adaptive Satellite Communication Jamming Strategy Based on “Electromagnetic Trap”

Yan Peijie, Chu Feihuang, Wang Mengyang
(*College of Space Information*, *Space Engineering University*, *Beijing* 101416, *China*)

Abstract: Aiming at the adaptive satellite node based on Q-Learning, a decoy jamming strategy of “electromagnetic trap” is proposed by combining traditional and intelligent jamming methods. By setting up decoy module, hot start module and anti-intelligence module, a closed loop of jamming is formed to achieve the operational effect of “decoy and anti-intelligence”. The simulation results show that compared with other algorithms, the proposed method can effectively jam the adaptive satellite nodes, improve the jamming accuracy and reduce the communication throughput.

Keywords: electromagnetic trap; deception and interference; satellite communication; Q-Learning; channel selection

0　引言

近年来，各国大力发展卫星事业，以 Starlink[1]、Oneweb[2]、鸿雁系统[3-4]为代表的巨型星座迅速崛起，卫星数量呈现爆炸式增长。然而，频谱资源有限，传统的卫星通信方式逐渐无法满足资源分配的要求[5]，自适应技术[6]可以有效提高频谱利用率；因此，具有学习能力的自适应卫星通信是一个重要的发展趋势。

面对复杂的电磁空间态势，卫星通信对抗日益激烈；然而，传统的卫星通信干扰样式[7]无法有效打击自适应卫星节点。在通信对抗领域，人工智能[8-10]提供了很好的解决思路。干扰和抗干扰双方想要智能对抗，需要与环境实时交互，学习状态和行为的映射关系，即学到对手策略，做出相应的动作，以获取最大累计回报。强化学习作为人工智能的研究热点，具有与环境在线交互与学习的能力，刚好可以满足通信对抗的需求，被广泛应用到通信对抗领域。

在通信干扰领域，Q-Learning[11]作为强化学习的典型算法，被应用于求解最佳干扰频率选择问题。此外，文献[12]基于 Actor-Critic（AC）提出了 2 种信道攻击策略：1）基于前馈神经网络（feedforward neural network，FNN）；2）基于深度强化学习（deep reinforcement learning，DRL）策略，使得基于 DRL 用户执行的动态多信道访问的准确性最小化。强化学习还可以应用于通信功率控制，文献[13]基于 DQN 算法设计了一种干扰方式，在相同功率约束情况下，可以大大降低用户的总传输速率。在波束选择方面，文献[14]提出了一种基于 MAB 的干扰策略，在不知道对手网络拓扑和信道信息的情况下，干扰机找到了最佳的干扰波束宽度和方向。

上述基于强化学习的干扰策略可以有效打击传统抗干扰用户；然而，面对具备学习能力的自适应抗干扰智能用户[15-17]，单一的智能干扰样式仍然存在效果不佳的问题。此外，目前大部分研究主要以地面通信对抗[18-21]为背景，卫星通信对抗的研究关注度较少。

受到雷达拖引干扰[22]的启发，笔者以低轨卫星为背景，针对基于 Q-Learning 的自适应卫星节点，结合传统与智能干扰方式，提出一种“电磁陷阱”的诱骗干扰策略，基于强化学习算法，最终实现“诱骗反智”的作战效果。

收稿日期：2023-12-05；**修回日期：**2024-01-07
第一作者：颜培杰（1998—），男，山东人，硕士。

1 系统模型与问题描述

1.1 系统模型

笔者考虑低轨卫星通信对抗场景，如图 1 所示。系统模型由 1 组卫星收发对 T_s 和 R_s，以及 1 颗干扰卫星 J_s 3 部分构成。笔者假设 3 颗卫星均匀分布在同一条轨道上，即卫星之间的相对位置基本不变。为方便描述，笔者对系统模型的正交信道做离散化处理，即 $C=\{c_1, c_2, c_3, \cdots, c_n\}$。卫星用户和干扰卫星的可用信道集分别被定义为 C^U，$C^U=\{c_1^u, c_2^u, \cdots, c_n^u\}$ 和 C^J，$C^J=\{c_1^j, c_2^j, \cdots, c_n^j\}$，其中 c_n^u 和 c_n^j 分别表示用户和干扰的信道索引。假设用户和干扰卫星的通信频带相同，即 $C^U=C^J=C$。在每一时隙，发射机可选择一个信道向接收机传输数据，而干扰卫星将发动“电磁陷阱”攻击。为更好地说明这一对抗过程，对模型中干扰的可行性进行分析。

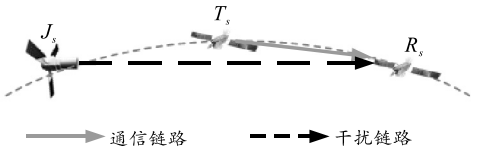

图 1 系统模型

1.1.1 星间干扰可行性分析

从空间位置以及天线角度方面分析星间干扰的可行性。如图 2 所示，O 表示地心，r_e 和 H 分别表示地球平均半径干扰以及轨道高度。

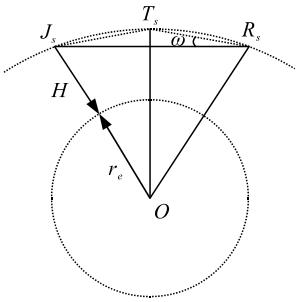

图 2 星间干扰可行性分析

以 Starlink 星座规划分布[23-24]为参考，笔者假设卫星均匀分布在 550 km 轨道高度上，星间距离为 100 km。卫星 J_S 与用户传输信号主瓣夹角被定义为 ω：

$$\omega = 90° - \arccos(d_{TR}/(2\bullet(H+r_e))) \approx 0.5° 。 \quad (1)$$

参考 ITU-R S.1528 建议书[25-26]提供的卫星天线方向图，低轨卫星参考 3 dB 波束宽度为 1.6°。显然，夹角 ω＜3 dB 波束宽度 ms 波通信下的卫星天线主瓣波束宽度；因此，该场景下，星间干扰是可行的。

1.1.2 基于“电磁陷阱”的诱骗干扰模型

基于“电磁陷阱”的诱骗干扰策略可以划分为 2 类干扰样式和 3 种对抗模块。其中，2 类干扰样式是指传统干扰和智能干扰样式。此外，在一个对抗回合中包含诱骗、热启动以及反智 3 模块。笔者以扫频干扰和基于 Q-Learning 干扰为例，分析诱骗干扰过程。

如图 3 所示，在诱骗模块，干扰机首先发动扫频干扰作为“陷阱”，诱骗基于 Q-Learning 的抗干扰用户。随后，具有学习能力的自适应卫星节点将学习扫频干扰策略，避开受扰信道，选择安全信道传输数据；随着不断地训练和学习，最终形成稳定的抗干扰策略。

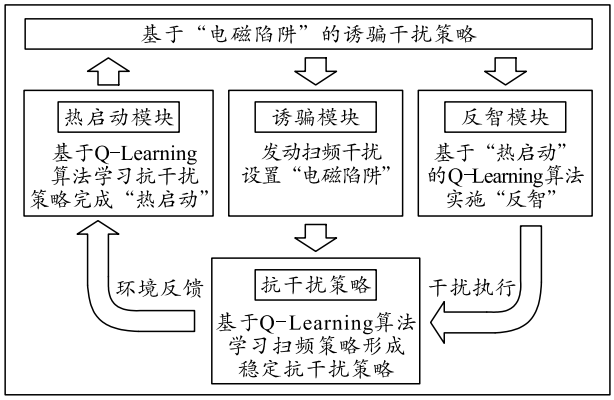

图 3 基于“电磁陷阱”的诱骗干扰框架

在热启动模块，基于 Q-Learning 的干扰卫星将根据环境反馈，学习用户的抗扫频策略，离线训练“反智”干扰策略。当基于 Q-Learning 的干扰卫星完成“热启动”，即形成较为稳定的干扰策略。扫频干扰将切换为基于 Q-Learning 干扰样式。

在反智模块，新切换的 Q-Learning 干扰样式不再需要初始化，而是直接利用热启动模块学到的干扰策略，有针对性地打击自适应抗干扰卫星用户。在一段时间内，基于“热启动”的 Q-Learning 干扰可以有效攻击落入“电磁陷阱”的自适应抗干扰卫星用户，从而实现“诱骗反智”的作战效果。

由于自适应抗干扰用户的智能性，干扰效果将逐渐降低。当干扰效果不再明显，为防止自适应抗干扰卫星形成新的抗干扰策略。基于 Q-Learning 干扰样式将再次切换为扫频干扰，从而“牵引”用户回归抗扫频干扰策略，至此，一个回合的对抗结束。

1.2 问题描述

笔者从抗干扰的角度，描述自适应卫星用户的

吞吐量，引出本文中的优化目标。

在干扰存在的情况下，自适应卫星用户的吞吐量为：

$$T(c_i^u)=B\bullet\log_2(1+\mathrm{SINR})。\tag{2}$$

式中 B 为单个信道的固定带宽。此外，自适应卫星用户的信道干扰比 SINR 为：

$$\mathrm{SINR}(c_i^u)=P_u\bullet\overline{H}^{c_i^u}_{T_s,R_s}\Big/(P_j\bullet\overline{H}^{c_i^j}_{J_s,R_s}\bullet f(c_i^u,c_i^j)+\sigma)。\tag{3}$$

式中：P_u 和 P_j 分别为用户和干扰卫星的功率；$\overline{H}^{c_i^u}_{T_s,R_s}$ 和 $\overline{H}^{c_i^j}_{J_s,R_s}$ 为用户和干扰传输过程的平均路径损耗；σ 为高斯噪声。此外，$f(c_i^u,c_i^j)$ 为指示函数，用来定义用户和干扰是否同时选择相同的信道，即用户是否受扰。

$$f(c_i^u,c_i^j)=\begin{cases}1, & c_i^u=c_i^j\\ 0, & c_i^u\neq c_i^j\end{cases}。\tag{4}$$

与文献[27−28]类似，卫星通信的路径损耗可以用自由空间损耗模型来描述。节点 x 和 y 之间的平均传输损耗为：

$$\overline{H}^{c_i}_{x,y}=E[H^{c_i}_{x,y}]=\left(4\pi\bullet d_{x,y}\bullet f/c\right)^{-\alpha}\bullet\overline{\beta}^{c_i}_{x,y}。\tag{5}$$

式中：$E[\bullet]$ 为期望；$d_{x,y}$ 为 x 和 y 节点的距离；f 为频率；c 为光速；α 为路径损耗因子；$\beta^{c_i}_{x,y}$ 为衰落系数。

由式(3)可知，自适应卫星用户的信道选择决定着吞吐量的大小；因此，笔者将用户的平均累加信道选择奖励 P_1 作为优化目标，从而最大化用户的吞吐量。

$$r_k^u=\begin{cases}0, & c_i^u=c_i^j\\ 1, & c_i^u\neq c_i^j\end{cases};\ P1:\max\sum_k r_k^u\Big/k。\tag{6}$$

自适应卫星用户的目的是通过选择最优信道，从而最大化吞吐量。而干扰卫星的优化目标可以表述为 P_2，即最大化平均干扰精确度，从而降低用户吞吐量。

$$r_k^j=\begin{cases}1, & c_i^j=c_i^u\\ 0, & c_i^j\neq c_i^u\end{cases};\ P2:\max\sum_k r_k^j\Big/k。\tag{7}$$

2　算法描述

1.1.2 节所描述的基于“电磁陷阱”诱骗干扰策略包含 3 种对抗模块。笔者从算法层面详细描述每种模式的对抗过程，引出基于“电磁陷阱”诱骗干扰算法。

2.1　“诱骗模块”对抗过程

在“诱骗模块”，干扰卫星发动扫频干扰作为“陷阱”，诱骗自适应卫星通信用户学习扫频规律。用户的信道选择过程可以被建模为马尔科夫决策过程(Markov decision process，MDP)[29−30]。该过程可以被定义为一个四元组$\{s^u,\ a^u,\ p^u,\ r^u\}$。其中，状态 $s_k^u=\{c_i^u(k),c_i^j(k-1)\}$ 由当前时隙用户的信道选择和上一时隙的干扰信道组成。此外，当前动作 $a_k^u\in\{c_1^u,c_2^u,\cdots,c_n^u\}$ 是指用户在当前状态下选择下一时隙的信道。根据 1.2 节中分析的用户优化目标，奖励 r_k^u 可定义为：

$$r_k^u=\begin{cases}0, & c_i^u(k)=c_i^j(k)\\ 1, & c_i^u(k)\neq c_i^j(k)\end{cases}。\tag{8}$$

针对马尔科夫决策问题，通常通过值迭代和策略迭代解决，但是由于环境先验信息，即状态转移概率 p^u 未知，故采用值迭代方式更新策略。任意 k 时隙，用户 $Q(s_k^u,a_k^u)$ 值表的更新公式为：

$$\begin{aligned}&Q(s_{k+1}^u,a_{k+1}^u)=Q(s_k^u,a_k^u)+\\ &\quad\alpha[r_k^u+\lambda\bullet\max Q(s_{k+1}^u,a_{k+1}^u)-Q(s_k^u,a_k^u)]。\end{aligned}\tag{9}$$

式中：α 为学习率；λ 为折扣因子。诱骗模块具体的对抗过程如下：

算法 1：扫频干扰 vs 基于 Q−Learning 抗干扰。

输入：扫频干扰，干扰时隙 t_j；

初始化：设置学习率 α，折扣因子 λ，可用信道集 C，传输时隙 t_u；迭代次数 iteration，初始化 Q 值表为全零矩阵；设置初始状态 $s_0^u=\{c_i^u(0),c_i^j(0)\}$；

For　1：iteration

Step1：当前用户选择信道 $c_i^u(k)$，干扰信道 $c_i^j(k-1)$，当前状态 $s_k^u=\{c_i^u(k),c_i^j(k-1)\}$；

Step2：自适应卫星节点根据当前策略选择动作 a_k；

Step3：根据 (s_k^u,a_k^u) 计算当前奖励 r_k^u；

Step4：根据 (s_k^u,a_k^u,r_k^u) 更新 $Q(s_k^u,a_k^u)$ 及策略 π^u。

Step5：记录信道选择 c_k^u 和奖励 r_k^u。

End

输出：卫星用户信道选择 c_k^u、奖励值 r_k^u 以及抗干扰信道选择策略 π^u。

2.2　“热启动模块”干扰离线训练过程

在“热启动模块”，虽然干扰卫星仍处于在线扫

频干扰模式，自适应卫星用户也逐渐学习到较为稳定的抗扫频策略。然而，干扰卫星将收集并记录自适应卫星用户的信道选择，并且基于 Q-Learning 进行离线训练，从而实现“热启动”，为“反智模块”做准备。

干扰卫星离线训练“热启动模块”的学习过程，与 2.1 节中自适应卫星用户学习扫频干扰的过程类似。详细的训练过程如下：

算法 2：基于 Q-Learning 干扰 vs 抗扫频抗干扰。

输入：自适应卫星用户的抗扫频信道选择策略 π^u 和卫星用户信道选择 c_k^u；

初始化：设置学习率 α'，折扣因子 λ'，可用信道集 C，传输时隙 t_j；迭代次数 iteration，初始化 Q 值表为全零矩阵；设置初始状态 $s_0^j=\{c_i^j(0),c_i^u(0)\}$；

For 1：iteration

Step1：当前干扰卫星信道选择 $c_i^j(k)$，用户信道选择 $c_i^u(k-1)$，当前状态 $s_k^j=\{c_i^j(k),c_i^u(k-1)\}$；

Step2：干扰卫星根据当前策略 π^j 选择动作 a_k^j；

Step3：根据 (s_k^j,a_k^j) 计算当前奖励 r_k^j；

Step4：根据 (s_k^j,a_k^j,r_k^j) 更新 $Q(s_k^j,a_k^j)$ 及策略 π^j；

Step5：记录干扰信道选择 c_k^j 和奖励 r_k^j。

End

$Q(s_{k+1}^j,\ a_{k+1}^j)=Q(s_k^j,\ a_k^j)+\alpha'[r_k^j+\lambda'\cdot\max Q(s_{k+1}^j,\ a_{k+1}^j)-Q(s_k^j,\ a_k^j)]$ 输出：干扰信道选择奖励 r_k^j 和 Q 值表。

基于 Q-Learning 的干扰离线训练“热启动”过程的马尔科夫四元组被定义为 $\{s^j,a^j,p^j,r^j\}$。其中，当前状态 $s_k^j=\{c_i^j(k),c_i^u(k-1)\}$ 由当前时隙干扰信道选择和上一时隙的用户数据传输信道组成。此外，当前动作 $a_k^j\in\{c_1^j,c_2^j,\cdots,c_n^j\}$ 是指干扰卫星在当前状态下选择下一时隙的干扰信道。根据干扰的优化目标，奖励 r_k^j 可定义为：

$$r_k^j=\begin{cases}1, c_i^j=c_i^u\\ 0, c_i^j\neq c_i^u\end{cases}。\tag{10}$$

此外，干扰卫星根据式(11)的 Q 更新 Q 值表。

$$Q(s_{k+1}^j,a_{k+1}^j)=Q(s_k^j,a_k^j)+\alpha'[r_k^j+\lambda'\cdot\max Q(s_{k+1}^j,a_{k+1}^j)-Q(s_k^j,a_k^j)]。\tag{11}$$

2.3 “反智模块”对抗过程

不同于“诱骗模块”和“热启动模块”，“反智模块”的干扰样式由扫频干扰切换为智能干扰。此模块是基于“热启动”的 Q-Learning 干扰卫星与 Q-Learning 抗干扰卫星用户之间的对抗。具体对抗过程如下：

算法 3：基于“热启动”的 Q-Learning 干扰 vs 基于 Q-Learning 抗干扰。

初始化：迭代次数 iteration；初始化用户信道选择 $c_i^u(0)$；

For 1：iteration

Step1：干扰卫星根据 $c_i^u(k)$ 运行算法 2；

Step2：干扰卫星信道选择 $c_i^j(k)$；

Step3：抗干扰用户根据 $c_i^j(k)$ 运行算法 1

Step4：用户选择信道 $c_i^u(k)$；

End

输出：干扰信道选择奖励 r_k^j。

2.4 基于“电磁陷阱”诱骗干扰算法

在 2.1—2.3 节的基础上，综合描述基于“电磁陷阱”诱骗干扰策略。具体对抗过程如下：

算法 4：基于“电磁陷阱”诱骗干扰算法。

初始化：对抗结束时间 T；固定步长 M；

For 1：T

Step1：发动扫频干扰，基于算法 1 实现诱骗干扰模块；

Step2：保持在线扫频干扰，基于算法 2 实现离线训练，完成热启动；

Step3：发动基于热启动的 Q-Learning 干扰，基于算法 3 实现反智模块；

Step4：记录干扰信道选择奖励 r_k^j，每隔 M 步取一次平均值。

End

输出：干扰平均精确度。

所提干扰策略就是利用自适应卫星的学习能力，结合扫频干扰和智能干扰 2 种干扰样式。将扫频干扰作为“陷阱”引诱自适应用户落入“陷阱”；而后，用完成离线训练的基于热启动的 Q-Learning 干扰有针对性的发起攻击，从而实现“诱骗反智”的作战效果。

可见，“电磁陷阱”有 3 个作用：1) 诱骗用户学习扫频干扰规律，形成稳定的抗干扰策略。2) 为基于 Q-Learning 干扰争取训练时间。切换干扰模式后，完成“热启动”即离线训练的智能干扰机，可

以直接准确攻击用户的抗扫频策略。3) 当智能干扰再次切换为扫频干扰时，“牵引”用户回归抗扫频干扰策略，推动下一回合的对抗。

3　仿真结果与分析

笔者考虑低轨卫星局部对抗场景，假设干扰卫星、用户发射卫星以及接收卫星均匀分布在 550 km 的倾斜轨道上，且相邻卫星之间的星间距离为 100 km。在该场景下，以 6 个可用信道为例，通过仿真时频瀑布图、平均干扰精确度和用户平均吞吐量，详细分析基于“电磁陷阱”诱骗干扰卫星的学习训练过程，及其与自适应卫星用户的对抗过程。

3.1　学习与训练过程分析

图 4 展示了基于 Q-Learning 抗干扰用户学习扫频干扰的时频瀑布图。其中，虚线方框表示干扰频段，实现方框表示用户通信频段。通过不断学习，自适应卫星用户可以有效躲避干扰攻击。从基于“电磁陷阱”诱骗干扰策略的角度分析，用户已经落入干扰“陷阱”。

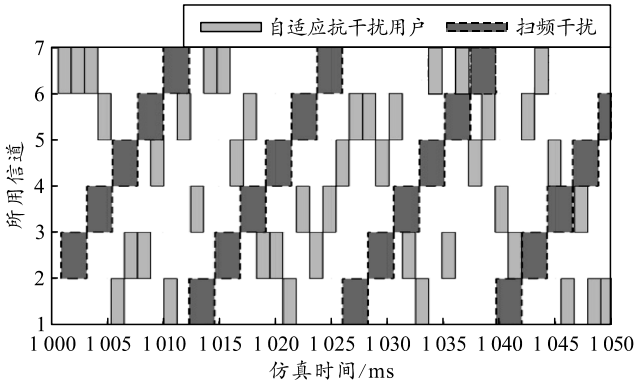

图 4　基于 Q-Learning 抗干扰用户的时频瀑布

图 5 展示了基于 Q-Learning 干扰的离线训练过程。可以观察到，干扰方通过不断地训练与学习，基本可以学习到自适应用户的抗扫频策略，实现“热启动”。

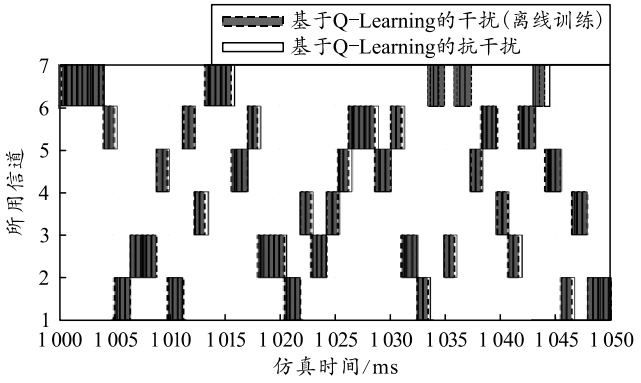

图 5　基于 Q-Learning 离线训练的干扰时频瀑布

在仿真中，设置每 200 次迭代取一次平均，从平均干扰精确度的角度，分析第 3 节中算法 1 和算法 2 的仿真效果。如图 6 所示，倒三角实线表示基于 Q-Learning 抗干扰用户对抗扫频干扰的学习过程；实心圆曲线表示基于 Q-Learning 干扰的离线训练过程。自适应卫星抗干扰用户可以很快学习到扫频干扰规律，并且完美规避干扰。此外，基于 Q-Learning 抗干扰用户通过离线训练，也可以基本学习到卫星用户的抗扫频策略，从而实现较高的干扰精确度。

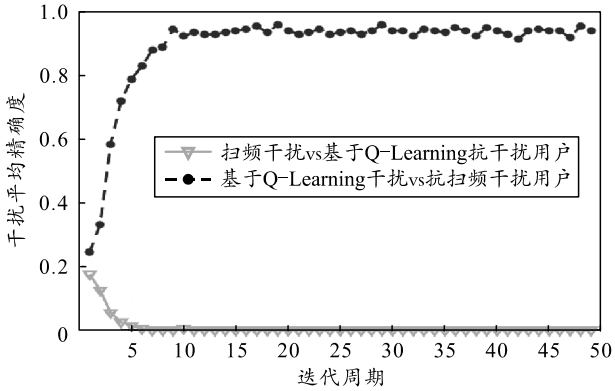

图 6　平均干扰精确度

为证明平均干扰精确度评价指标的有效性，从用户平均吞吐量的角度验证干扰和抗干扰效果，如图 7 所示。可以观察到图 6 和 7 具有相反的变化趋势，该现象验证了干扰平均准确度与用户平均吞吐量的反比关系。

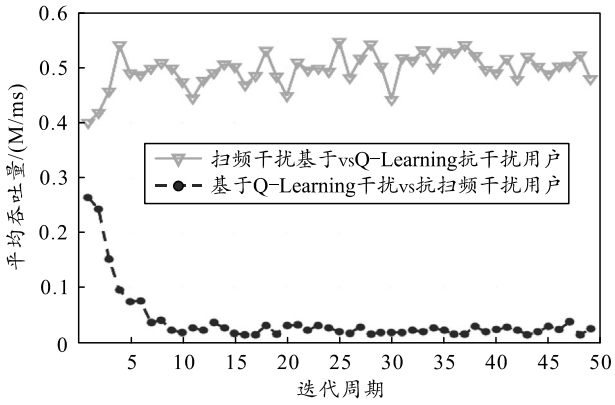

图 7　用户平均吞吐量

3.2　对抗过程分析

笔者从整个对抗过程的角度，分析基于“电磁陷阱”诱骗策略的干扰效果。分别将单一扫频干扰(倒三角虚线)和单一基于 Q-Learning 干扰策略(空心圆点虚线)作为对比算法；针对基于 Q-Learning 抗干扰卫星用户，基于所提算法(正方形实线)干扰对抗过程如图 8 所示。

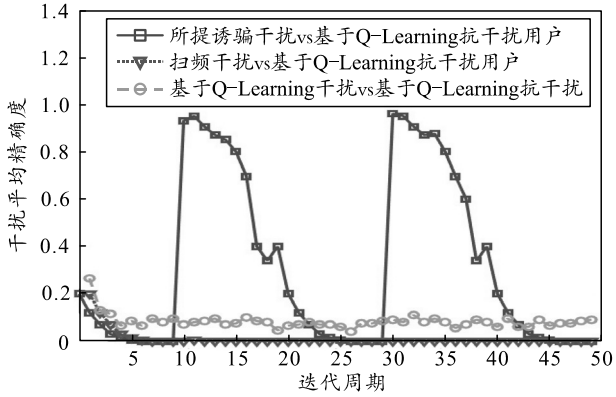

图 8 干扰精确度

观察整个对抗过程，可以得到以下规律：1) 相比于单一扫频和单一基于 Q-Learning 干扰，所提算法具有更高的平均干扰精确度，体现了所提干扰策略的优越性。2) 所提算法的平均干扰精确度曲线波动很大。下面具体分析波动原因。

笔者设置干扰样式每隔 10 个迭代周期切换一次。观察图 8 中的正方形实线，首先，在 0～5 迭代周期范围内，干扰方处于“诱骗模块”。干扰卫星执行扫频干扰，引诱基于 Q-Learning 的抗干扰卫星用户学习扫频规律。显然，随着平均干扰精确度的逐渐降低，抗干扰方逐渐学习到扫频规律。

其次，在 5～10 迭代周期内，干扰方处于“热启动模块”。此时，平均干扰精确度稳定在 0，说明自适应卫星用户已经形成较为稳定的抗扫频干扰方案。干扰卫星虽然仍执行在线扫频干扰，但也开始基于 Q-Learning 进行离线训练。

第 10 个迭代周期内，干扰方切换干扰策略，开始执行基于“热启动”的 Q-Learning 干扰。可以观察到，此时，平均干扰精确度由 0 陡然提升至 95% 左右。说明“热启动模块”中离线训练的干扰策略可以精准打击抗扫频卫星用户，体现了所提算法的优越性。

然而，在 10～20 迭代周期范围内，平均干扰精确度逐渐降低。这是由于信道数目对于卫星用户有优势，用户只需在 6 个信道中，躲避 1 个干扰信道即可正常通信。此外，由于基于 Q-Learning 的抗干扰卫星用户具有实时学习能力，卫星用户发现抗扫频策略效果不佳；因此，开始寻找新的抗干扰策略。

在 20～25 迭代周期内，为避免自适应卫星用户找到新的抗干扰策略，干扰方再次切换为扫频干扰，“牵引”自适应卫星用户再次落入扫频“陷阱”，从而推动下一回合的对抗。

用户平均吞吐量如图 9 所示。

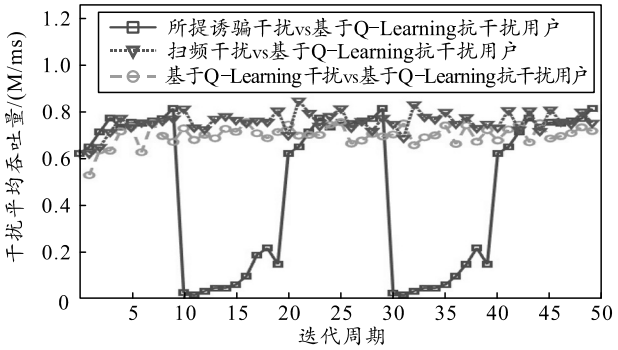

图 9 用户平均吞吐量

对比图 8 和 9，平均用户吞吐量的变化趋势印证了平均干扰精确度的曲线走向。从物理意义的角度再次验证了干扰精确度对用户吞吐量的决定性作用。此外，相比于其他干扰策略，可观察到所提方法下，用户表现出更低的平均吞吐量，体现了所提算法的优越性。

图 10 对比了 1 个对抗回合的平均干扰精确度和平均用户吞吐量。从图中可以看出，相比于其他 2 种干扰样式，所提方法具有更高的干扰精确度和更低的用户吞吐量。相比单一扫频干扰，所提干扰策略的平均干扰精确度高出 37.4%，用户吞吐量降低 0.264 M/ms；相比基于 Q-Learning 干扰策略，所提方法的平均干扰精确度高出 29.5%，用户吞吐量降低 0.206 M/ms。

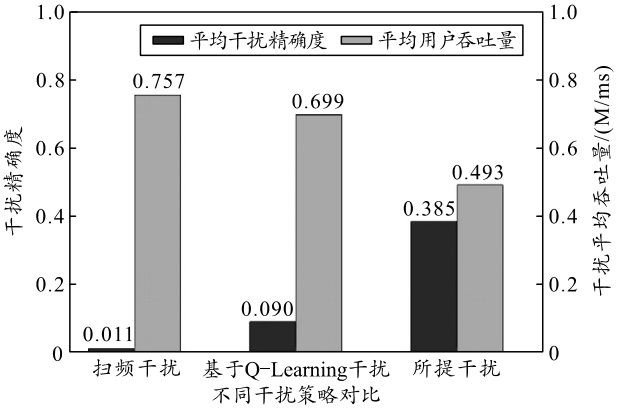

图 10 不同策略干扰性能对比

4 结束语

笔者针对基于 Q-Learning 的自适应卫星节点，结合传统与智能干扰方式，提出一种“电磁陷阱”的诱骗干扰策略，通过设置诱骗模块，热启动模块、反智模块，形成干扰闭环，最终实现“诱骗反智”的作战效果。仿真结果表明：相比单一干扰样式，即扫频干扰或基于 Q-Learning 干扰，该方法可有效地打击自适应卫星节点，提升干扰精确度，降低用户吞吐量。

参考文献：

[1] MCDOWELL J C. The Low Earth Orbit Satellite Population and Impacts of the SpaceX Starlink Constellation[J]. The Astrophysical Journal, 2020(4): 1−12.

[2] 李倬, 周一鸣. 美国 OneWeb 空间互联网星座的发展分析[J]. 卫星应用, 2018(10): 52−55.

[3] 吴巍. 天地一体化信息网络发展综述[J]. 天地一体化信息网络, 2020, 1(1): 1−16.

[4] 徐晓帆, 王妮炜, 高璎园, 等. 陆海空天一体化信息网络发展研究[J]. 中国工程科学, 2021, 23(2): 39−45.

[5] 刘瑞, 朱诗兵, 李长青, 等. 认知卫星通信频谱感知及资源分配技术综述[J]. 电讯技术, 2021, 61(8): 1048−1058.

[6] 李杨. 自适应卫星通信系统方案及相关技术研究[D]. 上海: 中国科学院研究生院(上海微系统与信息技术研究所), 2005.

[7] 郝东方. 卫星通信干扰技术的研究[D]. 西安: 西安电子科技大学, 2012.

[8] LI Y, XU Y, XU Y, et al. Dynamic Spectrum Anti-Jamming in Broadband Communications: A Hierarchical Deep Reinforcement Learning Approach[J]. IEEE Wireless Communication Letters, 2020, 9(10): 1616−1619.

[9] LI Y. Secure OFDM System Design and Capacity Analysis Under Disguised Jamming[J]. 2020, 15: 738−752.

[10] CHEN Z, FENG W, GURSOY M C, et al. Adversarial Jamming Attacks on Deep Reinforcement Learning Based Dynamic Multichannel Access[C]//2020 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2020.

[11] WANG L, PENG J, XIE Z, et al. Optimal Jamming Frequency Selection for Cognitive Jammer based on Reinforcement Learning[C]//2019 IEEE 2nd International Conference on Information Communication and Signal Processing (ICICSP). IEEE, 2019.

[12] CHEN Z, FENG W, GURSOY M C, et al. Adversarial Jamming Attacks on Deep Reinforcement Learning Based Dynamic Multichannel Access[C]//2020 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2020.

[13] LU Z, GURSOY M C. Dynamic Channel Access and Power Control via Deep Reinforcement Learning[C]//2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall). IEEE, 2019.

[14] KIM G, LIM H. Reinforcement Learning Based Beamforming Jammer for Unknown Wireless Networks[J]. IEEE Access, 2020, 8: 210127−210139.

[15] SLIMENI F, SCHEERS B, CHTOUROU Z, et al. Jamming mitigation in cognitive radio networks using a modified Q-learning algorithm[C]. International Conference on Military Communications and Information Systems (ICMCIS). IEEE, 2015: 101−107.

[16] SLIMENI F, CHTOUROU Z, SCHEERS B, et al. Cooperative Q-learning based channel selection for cognitive radio networks[J]. Wireless Networks, 2019, 25(7): 4161−4171.

[17] HAND Q, LI A, ZHANG L L, et al. Deep learning-guided jamming for cross-technology wireless networks: Attack and defense[J]. IEEE/ACM Trans. Networks, 2021, 29(5): 1922−1932.

[18] PELECHRINIS K, ILIOFOTOU M, KRISHNAMURTHY S V. Denial of service attacks in wireless networks: The case of jammers[J]. IEEE Commun. Surveys & Tutorials, 2011, 13(2): 245−257.

[19] GROVER K, LIM A, YANG Q. Jamming and anti-jamming techniques in wireless networks: a survey[J]. International Journal of Ad Hoc and Ubiquitous Computing, 2014, 17(4): 197−215.

[20] AREF M A, JAYAWEERA S K, YEPEZ E. A Survey on cognitive anti-jamming communications[J]. IET Communications, 2020, 14(4): 3110−3127.

[21] MPITZIOPOULOS A, GAVALAS D, KONSTANTOPOULOS C, et al. A survey on jamming attacks and countermeasures in WSNs[J]. IET Communications, 2009, 11(4): 42−56.

[22] 李迎春, 王国宏, 关成斌, 等. 速度拖引干扰和杂波背景下脉冲多普勒雷达目标跟踪算法[J]. 电子与信息学报, 2015, 37(4): 989−994.

[23] ITU-R. Satellite antenna radiation patterns for non-geostationary orbit satellite antennas operating in the fixed-satellite service below 30 GHz: ITU-R S.1528[S]. 2001.

[24] 张晟宇. 本质、技术与竞争: 漫谈 Starlink 星座[J]. 卫星与网络, 2018(3): 16−20.

[25] 韩锐, 杨夏青, 石会鹏, 等. COMPASS 系统与 CTDRS 系统 Ka 频段星间链路干扰仿真研究[J]. 南京邮电大学学报(自然科学版), 2017, 37(2): 33−37.

[26] 李伟, 严康, 耿静茹, 等. NGSO 通信星座系统间同频干扰场景与建模研究[J]. 天地一体化信息网络, 2021, 2(1): 20−27.

[27] HAN C, HUO L, TONG X, et al. Spatial Anti-jamming Scheme for Internet of Satellites Based on the Deep Reinforcement Learning and Stackelberg Game[J]. IEEE Transactions on Vehicular Technology, 2020, 69(5): 5331−5342.

[28] WU Q, XU Y, WANG J, et al. Distributed Channel Selection in Time-Varying Radio Environment: Interference Mitigation Game With Uncoupled Stochastic Learning[J]. IEEE Transactions on Vehicular Technology, 2013, 62(9): 4524−4538.

[29] 周志华. 机器学习[J]. 中国民商, 2016, 3(21): 93−93.

[30] ALAGOZ O, HSU H, SCHAEFER A J, et al. Markov Decision Processes: A Tool for Sequential Decision Making under Uncertainty[J]. Medical Decision Making, 2010, 30(4): 474−483.