doi:10.11887/j.cn.202004003 http://journal.nudt.edu.cn

在线学习的主用户仿冒攻击策略*

盛 响,王少尉
(南京大学 电子科学与工程学院,江苏 南京 210023)

摘 要:在认知无线网络中,次用户通过频谱感知来学习频谱环境,从而接入那些没有被主用户占用的频谱空隙。事实上,多种恶意攻击的存在会影响次用户频谱感知的可靠性。只有深入研究恶意攻击策略,才能确保认知无线网络的安全。基于此,研究了一种认知无线网络中的欺骗性干扰策略,即主用户仿冒攻击策略,该攻击策略通过在信道上传输伪造的主用户信号来降低次用户频谱感知的性能。具体来说,将攻击策略问题建模为在线学习问题,并提出基于汤普森采样的攻击策略以实现在探索不确定信道和利用高性能信道间的权衡。仿真结果表明,与现有的攻击策略相比,提出的攻击策略能更好地通过在线学习优化攻击决策以适应非平稳的认知无线网络。

关键词:认知无线电;在线学习;主用户仿冒攻击;频谱感知;汤普森采样
中图分类号:TN92 **文献标志码**:A **开放科学(资源服务)标识码(OSID)**:
文章编号:1001-2486(2020)04-012-06

Online learning for primary user emulation attack strategy

SHENG Xiang, *WANG Shaowei*
(School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China)

Abstract: In cognitive radio system, the secondary users learn radio spectrum environment through spectrum sensing to get the spectrum holes that the primary users do not occupy. In practice, the existence of various malicious attacks can seriously affect the reliability of spectrum sensing of the secondary users. Only in-depth study of the malicious attack strategies can ensure the security of cognitive radio networks. Based on this, a spoofing jamming strategy in cognitive wireless network, called as primary user emulation attack strategy, was studied. The strategy deteriorates the spectrum sensing performance of the secondary users by transmitting the forged primary user signals over channels. More concretely, the attack strategy was modeled as an online learning problem, and a Thompson sampling based attack strategy was proposed to find an efficient tradeoff between the exploitation of high-performance channels and the exploration of uncertain channels. The simulation results show that compared with the existing attack strategy, the proposed attack strategy can better adapt to the non-stationary cognitive wireless network by optimizing the attack decision through online learning.

Keywords: cognitive radio; online learning; primary user emulation attack; spectrum sensing; Thompson sampling

在认知无线网络[1-2]中,用户被分成主用户和次用户。主用户持有使用授权频段的牌照,可以随时接入该频段,而次用户只能通过频谱感知[3-5]在不影响主用户的情况下机会性地接入该频段。然而,频谱感知技术并不是绝对安全的[6-7]。主用户仿冒攻击是一种针对频谱感知技术的欺骗性干扰,它通过传输伪造的主用户信号来修改频谱环境以阻碍次用户的频谱感知[8]。只有深入研究主用户仿冒攻击策略即在每个时隙如何选定攻击信道,才能进一步确保认知无线网络的安全性。主用户仿冒攻击者(Primary User Emulation Attacker, PUEA)的目的在于阻止次用户接入频谱,这要求能在次用户感知的频谱空隙上传输伪造的主用户信号。然而实际上,主用户仿冒攻击者并没有任何关于频谱环境和次用户行为的先验知识,很可能会攻击那些被主用户占据或未被次用户感知的频段,这严重影响了攻击的有效性。因此,智能的攻击者需要使其攻击策略适应于非平稳的频谱环境和次用户频谱感知策略,其主要难点有以下两个方面:一是攻击者无法判断攻击是否有效,这是因为被攻击的信道永远不会被次用户接入;二是攻击者对非平稳的认知

* **收稿日期**:2019-12-25
基金项目:国家自然科学基金资助项目(61671233, 61801208, 61931023)
作者简介:盛响(1998—),男,江苏连云港人,博士研究生,E-mail:shengxiang1998@foxmail.com;
王少尉(通信作者),男,教授,博士,博士生导师,E-mail:wangsw@nju.edu.cn

无线网络没有任何先验知识,只能通过观察信道状态来学习。对攻击策略的充分理解可以帮助认知无线网络量化主用户仿冒攻击对次用户吞吐量的影响,这有助于相应的检测和防御策略的评估。此外,智能的主用户仿冒攻击者为其他现有干扰者的策略设计提供了一种新思路,并可以指导认知无线网络中安全机制的设计。这是本文的动机和意义所在。

文献[9]提出了基于部分可观测马尔可夫决策过程的攻击策略。这种攻击策略的部署依赖于攻击者获得攻击结果的能力。也就是说,攻击者需要知道是否有次用户曾感知被攻击的信道。文献[10]提出了三种攻击策略(均匀随机攻击、最大拦截攻击和选择性攻击)来检验次用户防御策略的有效性,其假设次用户算法的参数对攻击者是已知的。次用户算法的参数是其对信道的度量,该值越高的信道被感知的概率越大。文献[11]提出了基于在线学习算法 EXP3. G 的攻击策略,该策略不依赖于任何关于认知无线网络的先验知识,其考虑的系统模型是平稳的认知无线网络,也就是该网络中频谱环境和次用户频谱感知策略的统计特性是固定的。

本文研究了在非平稳的认知无线网络中的主用户仿冒攻击策略问题,将攻击策略问题归约为在线学习问题,并提出了基于汤普森采样的在线攻击策略。提出的在线攻击策略根据攻击者观察到的频谱环境信息和次用户行为不断更新以最大化攻击效果,可以有效地适应非平稳的认知无线网络。仿真结果表明,相较于现有的攻击策略,提出的在线攻击策略在稳态信道和非稳态信道场景下的两项性能指标(主用户仿冒累积攻击和次用户累积接入)都表现优越。

1　系统模型和问题归约

1.1　系统模型

考虑一个典型的认知无线网络,该网络包括 K 个授权信道、N 个次用户和 1 个主用户仿冒攻击者。该网络以时隙方式运行,即在一个时隙内授权信道的状态保持不变。

认知无线网络中主用户的行为(数据传输或空闲)决定了授权信道的主用户使用状态。本文只考虑信道的主用户使用状态,而不指定主用户的行为模式。记 K 个信道为 $\mathcal{K}=\{1,2,\cdots,K\}$,主用户使用状态共计有 2^K 个可能。在时隙 t 的主用户使用状态表示为 $D_t=[d_t(1),\cdots,d_t(K)]$,其中 $d_t(i)\in\{0,1\}$(当信道 i 被主用户占据时等于 0,当信道 i 处于空闲状态时等于 1)。值得注意的是,信道主用户使用状态的统计特性不是固定的,而是时变的甚至是任意的。

记认知无线网络中 N 个次用户的集合为 $\mathcal{N}=\{1,2,\cdots,N\}$。在每个时隙,每个次用户依次进行频谱感知和数据传输。在时隙 t 的频谱感知阶段,次用户 j 受限于其有限的采样率只能感知一个信道 $q_{t,j}$。在数据传输阶段,如果该信道处于空闲状态,次用户 j 接入该信道并传输数据,否则次用户 j 保持静默。考虑多个次用户的情形,可以利用现有的控制信道来避免因两个次用户感知同一个信道而产生的碰撞,这样的话就可以将 N 个次用户看作 1 个可以同时感知和接入 N 个信道的次用户。记该次用户群感知的信道为 $Q_t=\{q_{t,1},\cdots,q_{t,N}\}$。

在每个时隙,主用户仿冒攻击者依次进行攻击、观察和学习,如图 1 所示。在时隙 t 的攻击阶段,攻击者选择在一个信道 I_t 上传输伪造的主用户信号,这会使次用户误认为该信道被主用户占用。攻击者在时隙 t 成功攻击的次数为 $d_t(I_t)\cdot 1[I_t\in Q_t]$($1[x]$ 是示性函数,若 x 发生,$1[x]=1$;若 x 不发生,$1[x]=0$)。该值对攻击者来说是未知的,因为他无法确定次用户是否感知过信道 I_t,这意味着他无法通过攻击行为获取任何关于认知无线网络的信息。在观察阶段,攻击者保持静默并感知 λ 个信道的状态,记被感知的信道集合为 J_t,记感知信道 $i\in J_t$ 的结果为 $O_t(i)\in\{-1,0,1\}$(当信道 i 被主用户占用时为 -1,当信道 i 处于空闲状态时为 0,当信道 i 被次用户接入时为 1)。最后在学习阶段,攻击者根据观察结果更新攻击策略,使其适应非平稳的网络。定义在总时隙数 T 下攻击者成功攻击的总次数为主用户仿冒累积攻击,记为 $A_T=\sum_{t=1}^{T}d_t(I_t)\cdot 1[I_t\in Q_t]$。

图 1　主用户仿冒攻击者的帧结构

Fig. 1　Time slot structure of PUEA

1.2　问题归约

攻击者的目标是最大化 A_T。为此,在每个时隙,攻击者不仅需要攻击最可能被次用户接入的信道 I_t,还需要观察能给之后决策带来最大帮助的信道 J_t。这是一个典型的多臂赌博机问题[12],一种特殊的在线学习问题[13],它可以被描述为:

在一系列实验中,赌徒通过在每个回合选择 K 个摇臂中的一个来最大化总奖赏,每个摇臂都服从赌徒不知道的奖赏分布,该分布的特性只能通过过去的奖赏得到部分反映。具体来说,在时隙 t,定义信道 $k\in\mathcal{K}$ 的奖赏 $r_t(k)$ 为攻击者攻击该信道的成功次数 $d_t(I_t)\cdot 1[I_t\in Q_t]$。该问题的攻击策略可以表示为 $\varphi=\{\varphi(t)\}_{t\geqslant 1}$,其中 $\varphi(t):\mathcal{K}\rightarrow(I_t,J_t)$ 只依赖于过去 $t-1$ 个时隙观察到的信道状态信息 $\{O_m(n)\}_{t>m\geqslant 1,n\in J_t}$。攻击者的最优攻击策略 φ^* 可表示为:

$$\varphi^* = \arg\max_{\varphi}\sum_{t=1}^{T} r_t(I_t) \tag{1}$$

上述优化问题是很难解决的,一方面信道奖赏分布的统计特性只能通过观察结果部分反映,另一方面信道奖赏分布的统计特性不是固定的。

2 在线攻击策略

汤普森采样[14-15]是一种旨在解决多臂赌博机问题的启发式算法。该算法的核心思想是在每个回合根据每个摇臂是最优摇臂的后验概率随机选择信道,其中后验概率按照贝叶斯规则根据观察结果进行更新。提出的在线攻击策略分为两个阶段——攻击和观察。

在最开始,需要对认知无线网络建模。相较于文献[11]中直接对信道奖赏分布建模,利用攻击者对网络的认识,即信道状态由主用户和次用户共同决定,分别对信道的主用户占用行为和次用户感知行为进行建模将能更充分地利用观察结果。对于信道 k,用伯努利分布分别对主用户占用行为 Z_k(0 表示空闲,1 表示主用户占用信道 k)和次用户感知行为 S_k(0 表示空闲,1 表示次用户感知信道 k)建模,两个估计分布 $\hat{Z}_k$ 和 $\hat{S}_k$ 的参数服从贝塔分布。具体来说,主用户占用行为的估计 $\hat{Z}_k$ 服从下述条件分布:

$$P(\hat{Z}_k\mid\theta_{1,K})=\begin{cases}\theta_{1,k} & \hat{Z}_k=1\\ 1-\theta_{1,k} & \hat{Z}_k=0\end{cases} \tag{2}$$

式中,$\theta_{1,k}$ 的值在 0 到 1 之间且服从贝塔分布。$\theta_{1,k}$ 的分布 $\mathrm{B}(S_{1,k},F_{1,k})$ 可以表示为:

$$P(\theta_{1,k})=\frac{\Gamma(S_{1,k}+F_{1,k})}{\Gamma(S_{1,k})\Gamma(F_{1,k})}\theta_{1,k}^{S_{1,k}-1}(1-\theta_{1,k})^{F_{1,k}-1} \tag{3}$$

式中,Γ 是伽马函数,$S_{1,k}$ 和 $F_{1,k}$ 是该贝塔分布的参数。相应的次用户感知行为的估计 $\hat{S}_k$ 的参数为 $\theta_{2,k}$、$S_{2,k}$、$F_{2,k}$。

攻击阶段的主要任务是在时隙 t 根据估计分布 $\hat{Z}_k$ 和 $\hat{S}_k$ 的后验参数 $S_{1,k}$、$F_{1,k}$ 和 $S_{2,k}$、$F_{2,k}$ 选定攻击信道。基于上述分布,信道奖赏分布的估计 $\hat{r}_t(k)$ 可以导出:

$$\begin{aligned}P(\hat{r}_t(k)=1)&=P(\hat{Z}_k=0)\cdot P(\hat{S}_k=1)\\&=(1-\theta_{1,k})\cdot\theta_{2,k}\end{aligned} \tag{4}$$

攻击者根据每个信道是最优的概率随机选择一个信道攻击。值得注意的是,不需要将上述概率精确计算出:在每一轮对 $\theta_{1,k}$ 和 $\theta_{2,k}$ 进行一次抽样,然后选择有着最大奖赏期望的信道就足够了。记 $\theta_{1,k}$ 和 $\theta_{2,k}$ 的抽样值为 $\theta_{1,k}(t)$ 和 $\theta_{2,k}(t)$,有着最大奖赏期望的信道也就是

$$\begin{aligned}I_t&=\arg\max_k E(\hat{r}_t(k)\mid\theta_{1,k}(t),\theta_{2,k}(t))\\&=\arg\max_k[(1-\theta_{1,k}(t))\cdot\theta_{2,k}]\end{aligned} \tag{5}$$

观察阶段的主要任务是根据时隙 t 的观察结果 $O_t(i)$ $(i\in J_t)$ 对分布 $\hat{Z}_k$ 和 $\hat{S}_k$ 中的参数进行更新。值得注意的是,被攻击信道 I_t 的观察结果 $O_t(I_t)$ 只可能为 -1 或 0,而未被攻击信道的观察结果可能为 -1 或 0 或 1,这是因为次用户不可能接入被攻击的信道。显然,观察未被攻击信道的状态能带来更多关于网络的信息,因此在 $\lambda<K$ 也就是不能观察所有信道的情况下应优先观察未被攻击信道的状态。理想情况下,攻击者希望观察那些可能是最优的信道。但在实际网络中,信道奖赏分布的统计特性是时变的,即便在过去看来是最差的信道也可能在之后变成最优。因此,攻击者在 $\lambda<K$ 的情况下均匀随机选择 λ 个未被攻击的信道进行观察。在观察到 $O_t(k)$ 后,$\theta_{1,k}(t)$ 和 $\theta_{2,k}(t)$ 的后验分布可根据贝叶斯规则导出:

$$P(\theta_{1,k},\theta_{2,k}\mid O_t(k))\propto P(O_t(k)\mid\theta_{1,k},\theta_{2,k})P(\theta_{1,k},\theta_{2,k}) \tag{6}$$

其中,

$$P(O_t(k)\mid\theta_{1,k},\theta_{2,k})=\begin{cases}(1-\theta_{1,k})\cdot\theta_{2,k} & O_t(k)=1\\ (1-\theta_{1,k})\cdot(1-\theta_{2,k}) & O_t(k)=0\\ \theta_{1,k} & O_t(k)=-1\end{cases} \tag{7}$$

将 $\theta_{1,k}$ 和 $\theta_{2,k}$ 的分布代入上述贝叶斯规则,可得到在观察到 $O_t(k)$ 后的参数更新规则为:

$$\begin{cases}S_{1,k}\leftarrow S_{1,k}+1[O_t(k)=-1]\\ F_{1,k}\leftarrow F_{1,k}+1-1[O_t(k)=-1]\\ S_{2,k}\leftarrow S_{2,k}+1[O_t(k)=1]\\ F_{2,k}\leftarrow F_{2,k}+1[O_t(k)=0]\end{cases} \tag{8}$$

考虑实际网络中主用户占用行为和次用户感知行为的非平稳性,引入遗忘因子来减少过去观

察的影响以适应时变的环境。在上述贝叶斯更新后,再根据遗忘因子 γ 更新所有信道的参数,即

$$(S_{1,k},F_{1,k},S_{2,k},F_{2,k})\leftarrow\gamma(S_{1,k},F_{1,k},S_{2,k},F_{2,k})+(1-\gamma)(\overline{S}_{1,k},\overline{F}_{1,k},\overline{S}_{2,k},\overline{F}_{2,k}) \tag{9}$$

式中,$\overline{S}_{1,k}$、$\overline{F}_{1,k}$、$\overline{S}_{2,k}$、$\overline{F}_{2,k}$是信道 k 的先验参数。在没有任何先验知识的情况下,令 $\theta_{1,k}$和 $\theta_{2,k}$服从均匀分布,即 $\overline{S}_{1,k}$、$\overline{F}_{1,k}$、$\overline{S}_{2,k}$、$\overline{F}_{2,k}$为 1。在线攻击策略的时间复杂度为 $O(KT)$,其整体流程如算法 1 所示。

算法 1　在线攻击策略

Alg. 1　Online attacking strategy

参数:$\gamma\in(0,1]$; $\overline{S}_{1,k},\overline{F}_{1,k},\overline{S}_{2,k},\overline{F}_{2,k},\forall k\in\mathcal{K}$

初始化:$(S_{1,k},F_{1,k},S_{2,k},F_{2,k})\leftarrow(\overline{S}_{1,k},\overline{F}_{1,k},\overline{S}_{2,k},\overline{F}_{2,k})$, $\forall k\in\mathcal{K}$

for $t=1,2,\cdots,T$ **do**
  for $k=1,2,\cdots,K$ **do**
    $\theta_{1,k}(t)\sim\mathrm{B}(S_{1,k},F_{1,k})$
    $\theta_{2,k}(t)\sim\mathrm{B}(S_{2,k},F_{2,k})$
  end for
  攻击信道 $I_t=\arg\max_k[(1-\theta_{1,k}(t))\cdot\theta_{2,k}(t)]$
  随机选择 λ 个未被攻击信道并记为 J_t
  观察到 $O_t(k)$, $\forall k\in J_t$
  对于 $\forall k\in J_t$,根据式(8)更新参数 $S_{1,k}$、$F_{1,k}$和 $S_{2,k}$、$F_{2,k}$
  对于 $\forall k\in\mathcal{K}$,根据式(9)更新参数 $S_{1,k}$、$F_{1,k}$和 $S_{2,k}$、$F_{2,k}$
end for

3　仿真结果与分析

本节给出了提出的在线攻击策略在稳态信道和非稳态信道场景下的仿真结果。与文献[11,16]中的设置相同,假定每个信道的主用户使用状态服从独立的两状态马尔可夫链,信道 k 的转移概率矩阵可表示为:

$$\boldsymbol{P}_k=\begin{bmatrix}p_{00}(k) & p_{01}(k)\\ p_{10}(k) & p_{11}(k)\end{bmatrix} \tag{10}$$

式中,$p_{ij}(k)$是信道 k 从状态 i 到状态 j 的转移概率且 $p_{i0}(k)+p_{i1}(k)=1$。在稳态信道场景,主用户使用状态 D_t 的统计特性保持不变,也就是说在每次独立实验的开始随机生成 k 个独立的 $\boldsymbol{P}_k$ 并在实验阶段保持不变。而在非稳态信道场景下,主用户使用状态 D_t 的统计特性是时变的。在仿真中,$\boldsymbol{P}_k$ 每隔时间 ΔT 就以概率 $P_\Delta\in[0,1]$ 重新随机生成一次。数值仿真比较了在线攻击策略和两个现有策略的性能,其中现有策略包括:①攻击和随机观察学习算法[11](Play and Random Observe Learning Algorithm, PROLA),该攻击策略基于在线学习算法 EXP3. G 来进行攻击和观察;②均匀随机攻击策略[10](uniformly Random Attack algorithm, RA),该攻击策略中每个信道被攻击的概率相等。数值仿真中,攻击者的性能指标除了优化目标即主用户仿冒累积攻击 A_T 以外还包括次用户累积接入 C_T。次用户累积接入 C_T 是所有次用户在总时隙数 T 内累积成功接入的次数,其值等于 $\sum_{t=1}^{T}\sum_{j=1}^{N}d_t(q_{t,j})\cdot 1[q_{t,j}\neq I_t]$。仿真中次用户采取文献[10]中基于对抗多臂赌博机的被动防御策略,为了更好验证攻击者的性能,假定次用户在时隙结束时获得所有信道的主用户使用状态。所有的仿真都在 MATLAB 中进行,每个结果都是 10 000 次独立随机实验的均值。

图 2 比较了在稳态信道场景下攻击策略的性能,具体仿真参数为:$K=10$, $N=1$, $\lambda=1$, $T=2000$, $\gamma=0.99$。可以看出:在线攻击策略的主用户仿冒累积攻击为 376,比 PROLA 和 RA 分别高了 35% 和 119%;相应的次用户累积接入为 1105,比 PROLA、RA 和无攻击者分别少了 8%、28% 和 36%。性能提升的原因有两方面:①考虑并有效处理了次用户行为的非平稳性带来的挑战;② 提出的算法能够更有效而快速地利用观察信息改进攻击策略。值得注意的是,主用户仿冒累积攻击的增加量和次用户累积接入的减少量是密切相关但又不完全相同的。以在线攻击策略为例,相较于无攻击的情况,主用户仿冒累积攻击增加了 376,而次用户累积接入减少了 609,这是因为每一次成功的主用户仿冒攻击对次用户的影响包括直接的阻止次用户接入和间接的破坏次用户对主用户行为规律的学习。

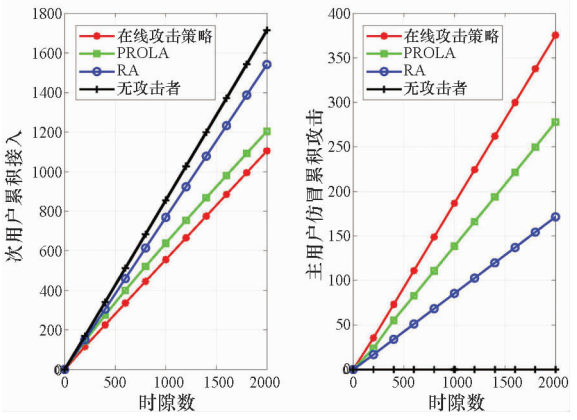

图 2　稳态信道场景下攻击策略性能

Fig. 2　Performance of tested strategies in stationary channel scenario

图 3 比较了在非稳态信道场景下攻击策略的性能，具体仿真参数为：$K=10$，$N=1$，$\lambda=1$，$T=2000$，$\gamma=0.99$，$\Delta T=100$，$P_\Delta\in[0,1]$。随着 P_Δ 的增加，主用户使用状态的非平稳性不断增加，极大影响了次用户和攻击者的性能。随着 P_Δ 从 0 到 1，三种攻击策略的主用户仿冒累积攻击都有着不同程度的下降。其中 PROLA 的主用户仿冒累积攻击下降了 40%，而在线攻击策略的主用户仿冒累积攻击只下降了 8%，与此同时两者之间的差距也从 35% 增加到了 98%。这是因为引入的遗忘因子 γ 可以很好地处理认知无线网络中的非平稳性。值得注意的是，次用户累积接入并没有随着 P_Δ 的增加而产生一致的变化，这是因为次用户在无攻击情况下的累积接入随着 P_Δ 的增加而减少。在线攻击策略的次用户累积接入随着 P_Δ 的增加甚至有小幅增加的原因在于：主用户仿冒攻击的间接影响本质上是为次用户学习的环境添加非平稳性，其影响随着环境本身非平稳性的增加而减弱。

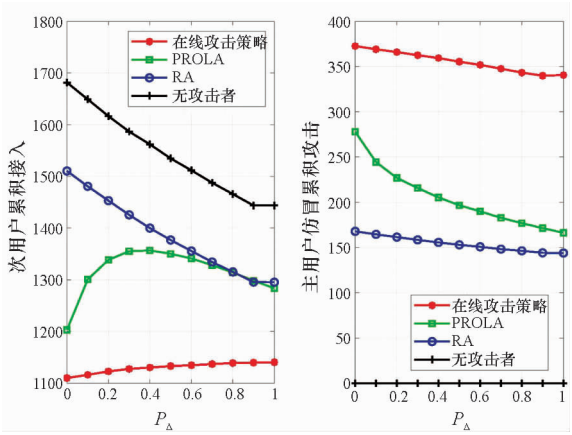

图 3 非稳态信道场景下攻击策略性能

Fig. 3 Performance of tested strategies in non-stationary channel scenario

图 4 比较了不同 λ 下的在线攻击策略性能，具体仿真参数为：$K=40$，$N=4$，$\lambda\in[1,37]$，$T=2000$，$\gamma=0.99$，$\Delta T=100$，$P_\Delta=0.5$。随着 λ 从 1 到 37，在线攻击策略的主用户仿冒累计攻击在稳态和非稳态信道场景下分别提升了 30% 和 90%。这是因为攻击者在每个时隙观察到的信道状态信息随着 λ 的增加而增加，也就是说攻击者的学习能力随着 λ 的增加而增加。相较于稳态信道场景，λ 的增加对非稳态信道场景的影响更大。这是因为非稳态信道场景更为复杂，攻击者学习起来也更困难。值得注意的是，λ 的增加会提高对攻击者计算能力的要求，因此需要在性能和代价间权衡。在刚开始的阶段，通过少量增加 λ（如从 1 到 4），主用户仿冒累积攻击显著提高；当 λ 足够大时（如从 10 到 13），主用户仿冒累积攻击的增加量就几乎可忽略不计了。本次仿真中，当 λ 为 10 时，攻击者的表现已经接近最优。

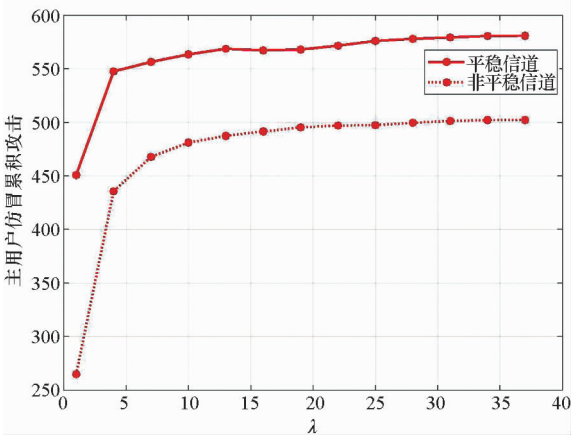

图 4 不同 λ 下的在线攻击策略性能

Fig. 4 Performance of online attacking strategy under different λ

4 结论

以认知无线网络中的主用户仿冒攻击策略问题为例，通过在线学习将攻击者的观察和策略更新相结合以实现更有效地攻击，这为无线网络中干扰机如何通过与环境交互实现干扰决策优化提供了一种思路。具体来说，将主用户仿冒攻击策略问题建模为在线学习问题，并提出了基于汤普森采样的在线攻击策略。该攻击策略可以有效利用观察信息实现在利用高奖赏信道和探索奖赏未知信道之间的权衡。仿真结果表明，与现有攻击策略相比，在线攻击策略在稳态信道和非稳态信道场景下都表现优越，能够有效适应非平稳的认知无线网络，并在少量的观察信道下就达到接近最优的性能。

参考文献(References)

[1] Wang S W, Ge M Y, Wang C G. Efficient resource allocation for cognitive radio networks with cooperative relays[J]. IEEE Journal on Selected Areas in Communications, 2013, 31(11): 2432 - 2441.

[2] Dai J Y, Wang S W. Clustering-based spectrum sharing strategy for cognitive radio networks[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(1): 228 - 237.

[3] Zhou M, Wang T Y, Wang S W. Spectrum sensing across multiple service providers: a discounted Thompson sampling method[J]. IEEE Communications Letters, 2019, 23(12): 2402 - 2406.

[4] 杨鹏，樊昀，黄知涛，等. 基于 Sub-Nyquist 采样的单通道频谱感知技术[J]. 国防科技大学学报，2013，35(4)：

121 – 127.
YANG Peng, FAN Yun, HUANG Zhitao, et al. Single-channel spectrum sensing technique based on Sub-Nyquist sampling[J]. Journal of National University of Defense Technology, 2013, 35(4): 121 – 127. (in Chinese)

[5] 孙伟朝, 王丰华, 黄知涛, 等. 改进多重信号分类算法的宽带频谱快速感知方法[J]. 国防科技大学学报, 2015, 37(5): 155 – 160.
SUN Weichao, WANG Fenghua, HUANG Zhitao, et al. Wideband spectrum fast sensing method based on improved multiple signal classification [J]. Journal of National University of Defense Technology, 2015, 37(5): 155 – 160. (in Chinese)

[6] 裴庆祺, 李红宁, 赵弘洋, 等. 认知无线电网络安全综述[J]. 通信学报, 2013, 34(1): 144 – 158.
PEI Qingqi, LI Hongning, ZHAO Hongyang, et al. Security in cognitive radio networks[J]. Journal on Communications, 2013, 34(1): 144 – 158. (in Chinese)

[7] Wang S W, Zhou Z H, Ge M Y, et al. Resource allocation for heterogeneous multiuser OFDM-based cognitive radio networks with imperfect spectrum sensing[C]// Proceedings of IEEE INFOCOM, 2012.

[8] Chen R L, Park J M, Reed J H. Defense against primary user emulation attacks in cognitive radio networks[J]. IEEE Journal on Selected Areas in Communications, 2008, 26(1): 25 – 37.

[9] Li H S, Han Z. Dogfight in spectrum: combating primary user emulation attacks in cognitive radio systems, part I: known channel statistics [J]. IEEE Transactions on Wireless Communications, 2010, 9(11): 3566 – 3577.

[10] Li H S, Han Z. Dogfight in spectrum: combating primary user emulation attacks in cognitive radio systems—part II: unknown channel statistics [J]. IEEE Transactions on Wireless Communications, 2010, 10(1): 274 – 283.

[11] Dabaghchian M, Alipour-Fanid A, Zeng K, et al. Online learning with randomized feedback graphs for optimal PUE attacks in cognitive radio networks [J]. IEEE/ACM Transactions on Networking, 2018, 26(5): 2268 – 2281.

[12] Chen W, Wang Y J, Yuan Y. Combinatorial multi-armed bandit: general framework, results and applications[C]// Proceedings of the 30th International Conference on Machine Learning, 2013.

[13] Duchi J, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization[J]. Journal of Machine Learning Research, 2011, 12: 2121 – 2159.

[14] Thompson W R. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples[J]. Biometrika, 1933, 25(3/4): 285 – 294.

[15] Agrawal S, Goyal N. Analysis of Thompson sampling for the multi-armed bandit problem [C]// Proceedings of 25th Annual Conference on Learning Theory, 2012.

[16] Zhao Q, Tong L, Swami A, et al. Decentralized cognitive MAC for opportunistic spectrum access in ad hoc networks: a POMDP framework[J]. IEEE Journal on Selected Areas in Communications, 2007, 25(3): 589 – 600.