

无人机网络中基于分层博弈的干扰对抗频谱接入优化

范超琼, 赵成林, 李斌

(北京邮电大学信息与通信工程学院, 北京 100876)

摘要: 针对无人机通信网络中的干扰对抗问题, 提出了一种基于分层博弈的自适应频谱接入优化机制。考虑到无人机网络节点的动态特性, 将干扰器视为分层博弈领导者, 无人机用户视为分层博弈跟随者, 不同层级间博弈参与者具有不同的效应函数, 采用斯坦伯格均衡分析所构建的博弈并证明了均衡解的存在性和唯一性。在此基础上设计一种分层信道选择学习算法来求解博弈的均衡解, 并分析其收敛性能。仿真表明, 所提算法能使网络节点智能地调整信道选择策略, 从而获得良好的吞吐量性能。

关键词: 无人机网络; 干扰对抗; 频谱接入; 分层博弈; 强化学习

中图分类号: TN92

文献标识码: A

doi: 10.11959/j.issn.1000-436x.2020114

Hierarchical game based spectrum access optimization for anti-jamming in UAV network

FAN Chaoqiong, ZHAO Chenglin, LI Bin

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: For the anti-jamming spectrum access optimization problem in unmanned aerial vehicle (UAV) communication networks, considering the complex and diverse malicious jamming from jammers, a Bayesian Stackelberg game was proposed to formulate the competitive relations between UAV users and jammers. Specifically, jammers acted as the leader, whereas users acted as followers of the proposed game. Based on their different utility functions, the jammer and users independently and selfishly selected their optimal strategies and obtained the optimal channels selection. Due to the NP-hard nature, it was challenging to obtain the Stackelberg Equilibrium of the proposed game. To this end, a hierarchical learning framework was formulated, and a hierarchical channel selection-learning algorithm was proposed. Simulations demonstrate that with the proposed hierarchical learning algorithm, UAV nodes can adjust their channel selection and obtain superior performance.

Key words: unmanned aerial vehicle network, anti-jamming, spectrum access, hierarchical game, reinforcement learning

1 引言

作为一种新型的无线通信网络, 无人机网络近来受到了广泛的关注并成为当下的研究热点^[1]。与传统蜂窝网络相比, 无人机具有成本低廉、易于部署、机动性强、用途广泛等优势, 因此在多个领域展现出了巨大的应用潜力。然而, 由于无线网络的开放特性, 无人机通信容易受到“欺骗”、干扰和

窃听等恶意攻击, 这对无人机通信安全造成严重威胁。如何保证干扰对抗场景中无人机网络可靠稳健的信息传输是一个严峻的挑战。另一方面, 考虑频谱资源的日益紧缺性和现有频谱分配方式的效率低下性, 动态频谱接入成为提高资源利用率的重要技术手段, 在无人机网络中引入稳健高效的频谱接入技术成为大势所趋^[2]。综上所述, 研究无人机网络干扰对抗场景中有效的频谱接入优化机制以有

收稿日期: 2020-03-24; 修回日期: 2020-05-20

基金项目: 国家自然科学基金资助项目 (No.U1805262, No.61971050)

Foundation Item: The National Natural Science Foundation of China (No.U1805262, No.61971050)

效避免外部复杂多样的恶意干扰，进而实现节点间可靠安全的信息传输对无人机通信具有重大意义^[3]。

在干扰对抗场景中，无人机用户的目标是在复杂干扰情况下实现最佳效益（最大信干噪比、最大数据速率、最大吞吐量等）；干扰器的目标是最大程度地损害用户效益，进而破坏其正常通信传输。博弈论^[4]作为一种强大的理论分析工具，适用于建模分析多用户节点竞争优化问题。在多种博弈模型框架中，分层博弈更加契合于所研究的不同节点（无人机用户和干扰器）间目标相悖且自发独立进行决策的优化问题。

为了有效降低或避免开放无线传输环境中多种恶意干扰的影响，研究人员已开展了大量的抗干扰对策相关工作。然而针对无人机通信网络干扰对抗场景中的频谱接入优化问题，当前的研究尚有一些不足之处。首先，大多数研究仅考虑单个干扰器进行恶意干扰攻击的情况，场景设置和优化问题构造相对简单；其次，未对无人机节点的动态移动特性进行建模分析；最后，大多数研究基于对全局信息的理想假设，即无人机用户端和干扰器端分别能够获知对方的信道状态、行动选择、决策收益等信息，并依赖全局信息进行自身的决策分析。然而这一假设与实际应用不相符，一方面，对于能量受限的无人机节点来说，进行全局的信息交互需要额外的信令开销，从而增大能量消耗；另一方面，在实际通信场景中，不同类型节点间的信息为封锁状态，难以实现信息交互。

鉴于对上述问题的挑战和研究不足，本文考虑节点移动性，提出一种基于分层博弈的分布式频谱接入优化算法，以期有效降低或避免无人机网络中干扰器的恶意干扰。本文的主要创新及贡献如下。

1) 针对无人机通信网络，充分考虑网络中节点的动态移动特性，对其时变位置进行建模；在此基础上，无人机用户以总数据速率为目标、干扰器以总破坏效益为目标分别构建在干扰对抗场景下频谱接入优化的数学模型。

2) 基于分层博弈模型，将干扰器视为领导者，无人机用户视为跟随者，建立无人机用户和干扰器间的分层博弈框架；随后采用斯塔伯格均衡表征所构建的博弈，并对均衡解的存在性和唯一性进行严格分析证明。

3) 对于所构建的分层博弈，提出一种基于 Q -学习的分布式信道选择学习算法，并分析其收敛性能。利用所提算法，参与者能够根据历史经验和局部信息对环境进行学习和适应，从而调整自己的信道选择策略，最终收敛到稳定的均衡解。

2 相关工作

目前，在无线通信网络的背景下，抗干扰传输优化已获得了广泛的研究，包括对连续/离散发射功率、用户调度、频谱接入等参数的优化。

对于传输功率维度的优化问题，文献[5]采用离散集合对功率进行建模，提出了基于博弈论的分层学习算法来优化干扰对抗环境中用户发射功率；文献[6]针对协作式无线网络中连续功率控制优化，采用分层博弈构建合法用户和非法干扰器间的对抗；文献[7]重点关注网络中的非理想信息，采用贝叶斯分层博弈建模功率控制问题，并利用对偶优化理论来求解最优功率分配策略。

对于频谱接入维度的优化问题，文献[8]研究了认知无线网络中基于随机博弈的抗干扰频谱共享技术；文献[9]将认知无线网络中次用户与干扰器之间交互干扰博弈建模，研究了以用户为中心的抗干扰优化机制；文献[10]采用分层博弈，将合法用户视为领导者，非法干扰器视为跟随者，分别设计效用函数，通过自适应学习从而获得优化的均衡解。

此外，从用户个数角度来说，立足于典型的强化学习算法，即 Q -学习方法，文献[11]基于马尔可夫决策过程研究了单用户场景中的抗干扰决策问题，随后，Aref等^[12]采用马尔可夫博弈框架将其扩展至多用户场景，并考虑合法用户间的相互耦合干扰对网络性能的影响，Hanawal等^[13]基于马尔可夫博弈研究了抗干扰无线系统中跳频和传输速率的联合自适应优化问题。

综上所述，对于干扰对抗场景中用户传输参数优化以保证其可靠通信的研究已取得了较大的进展，但仍存在场景适应性问题，应用范围具有局限性。本文关注于面向无人机传输通信的干扰对抗频谱接入优化问题，具有一定的研究意义。

3 系统模型与问题构建

3.1 无人机网络模型

考虑恶意干扰对抗场景中无人机网络中的信

道选择优化问题, 网络模型如图 1 所示。假设网络中存在 N 个无人机用户 (发射端-接收端), 其集合表示为 $\{1, 2, \dots, N\}$; M 个恶意干扰器, 其集合表示为 $\{1, 2, \dots, M\}$; F 个可用信道, 其集合表示为 $\{1, 2, \dots, F\}$ 。本文聚焦于网络中恶意干扰的消除, 因此, 假设无人机用户间的信道选择相互正交, 某一信道上至多有一个无人机用户, 即 $F \geq N$ 。每个无人机用户在一个时隙内选择一个信道进行数据传输; 每个干扰器在一个时隙内选择一个信道发射干扰攻击信号。假设无人机用户和干扰器均具有环境感知适应和策略优化更新功能: 对于无人机用户来说, 为了保证其可靠传输, 需优化信道选择以最小化干扰器的恶意干扰; 对于干扰器来说, 动态调整所攻击的信道以最大化其干扰破坏效果。

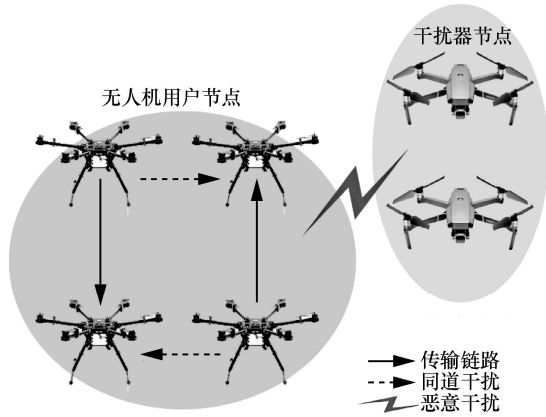


图 1 无人机网络干扰对抗模型

假设网络中无人机用户和恶意干扰器节点均按照预设的航行轨迹运动。将某一节点 i ($i \in (N \cup M)$) 在时隙 t 的空间位置记为 $\mathbf{l}_i(t) = (x_i(t), y_i(t), z_i(t))$, 其飞行速度为 $\mathbf{v}_i(t)$, 则节点 i 在下一时隙 $t+1$ 的位置表示为

$$\mathbf{l}_i(t+1) = \mathbf{l}_i(t) + \mathbf{v}_i(t)\theta_i(t) \quad (1)$$

其中, $\theta_i(t)$ 表示节点 i 在时隙 t 的航行方向。节点 i 与节点 j ($j \in (N \cup M)$) 在时隙 t 间的距离 d_{ij} 如式(2)所示。

$$d_{ij}(t) = \sqrt{(x_i(t) - x_j(t))^2 + (y_i(t) - y_j(t))^2 + (z_i(t) - z_j(t))^2} \quad (2)$$

为了表述简单, 下文描述中省略时隙 t 的表示。

根据文献[14], 本文采用服从瑞利分布的自由空间路径损耗模型来描述链路间的信号传输衰落,

记节点 i 与节点 j 在信道 f ($f \in F$) 上的功率增益为 H_{ij}^f , 则有

$$H_{ij}^f = (d_{ij})^{-\alpha} \beta_f \quad (3)$$

其中, α 表示路径损耗指数, β_f 表示信道 f 上的瞬时随机衰落增益。

3.2 优化问题构建

记无人机用户 n ($n \in N$) 选择接入的信道为 a_n ($a_n \in F$), 干扰器 m ($m \in M$) 选择攻击的信道为 b_m ($b_m \in F$)。所有用户的信道选择形式为 $\mathbf{a} = \{a_1, a_2, \dots, a_N\}$, 所有干扰器的信道攻击形式为 $\mathbf{b} = \{b_1, b_2, \dots, b_M\}$ 。基于以上假设, 无人机用户 n ($n \in N$) 选择信道 a_n 受到的恶意攻击干扰可以表示为

$$I_n = \sum_{m \in M} \delta(a_n, b_m) P_J H_{mn}^{b_m} \quad (4)$$

其中, P_J 表示干扰器的发射功率, $\delta(x, y)$ 是二元指示函数, 即

$$\delta(x, y) = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases} \quad (5)$$

无人机用户 n 的数据速率可表示为

$$R_n = \text{lb} \left(1 + \frac{P_U H_{nn}^{a_n}}{I_n + \sigma^2} \right) \quad (6)$$

其中, P_U 表示无人机用户 n 的发射功率, σ^2 表示信道加性高斯白噪声方差。

所有无人机用户的总数据速率可表示为

$$R_{\text{sum}}(\mathbf{a}) = \sum_{n \in N} R_n \quad (7)$$

为了保证无人机用户间信息的可靠传输, 记其所需的数据速率门限值为 R_0 , 即对于用户 n 来说, 如果 $R_n \geq R_0$, 则表明数据传输成功; 否则, 则受到恶意干扰, 数据传输失败。

基于此, 干扰器 m ($m \in M$) 的效应值 C_m 定义为攻击信道 f 使数据传输遭受破坏的无人机用户个数, 即

$$C_m = \sum_{n \in N} \delta(a_n, b_m) l_{R_n < R_0} \quad (8)$$

其中, $l_{x < y}$ 是二元指示函数, 当 $x < y$ 时, $l_{x < y} = 1$; 否则, $l_{x < y} = 0$ 。

所有干扰器的总效应值 $C_{\text{sum}}(\mathbf{b})$ 表示为

$$C_{\text{sum}}(\mathbf{b}) = \sum_{m \in M} C_m \quad (9)$$

综上所述, 无人机用户的目标为通过感知环境

动态自适应地调整信道选择 \mathbf{a} ，以最大化其总数据速率，即

$$\max_{\mathbf{a}} R_{\text{sum}}(\mathbf{a}) \quad (10)$$

干扰器的目标为通过感知环境动态自适应地调整攻击信道 \mathbf{b} ，以最大化其干扰总效应值，即

$$\max_{\mathbf{b}} C_{\text{sum}}(\mathbf{b}) \quad (11)$$

4 干扰对抗分层博弈优化

本节首先采用分层博弈对以上构建的非多项式复杂度干扰对抗频谱接入优化问题进行建模；然后给出分层博弈的均衡解，并对其存在性和唯一性进行证明；最后提出一种自适应的分布式强化学习算法求解构建博弈的均衡解，并分析其收敛性能。

4.1 分层博弈

干扰对抗分层博弈可数学化地表示为 $G=(N, M, F_U, F_J, u_n, w_m)$ ，其中， $F_U=F^N$ 和 $F_J=F^M$ 分别表示用户和干扰器的策略空间； u_n 和 w_m 分别表示用户 n 和干扰器 m 的效应函数。在分层博弈中干扰器作为领导者，无人机用户作为跟随者，各个博弈参与者独立地进行环境感知和策略更新，以优化其效应函数。下面分别介绍领导者和跟随者的博弈模型。

4.1.1 领导者博弈

对于干扰器节点来说，目标为尽可能多地破坏无人机用户的数据传输，则其效应函数 w_m 可定义为

$$w_m(\mathbf{a}, \mathbf{b}_m) = C_m = \sum_{n \in N} \delta(a_n, b_m) l_{R_n < R_0} \quad (12)$$

因此，领导者博弈可数学化地表示为

$$G_J = (M, F_J, w_m(\mathbf{a}, \mathbf{b}_m)) \quad (13)$$

则领导者博弈问题可描述为寻找最佳的信道攻击选择，以最大化其效应函数，即

$$\mathbf{b}_m^* = \arg \max_{\mathbf{b}_m} w_m(\mathbf{a}, \mathbf{b}_m), \quad m \in M \quad (14)$$

4.1.2 跟随者博弈

与领导者博弈类似，对于无人机用户来说，目标为保证其可靠的数据传输，则其效应函数 u_n 可定义为

$$u_n(a_n, \mathbf{b}) = R_n = \text{lb} \left(1 + \frac{P_U H_{nm}^{a_n}}{I_n + \sigma^2} \right) \quad (15)$$

因此，跟随者博弈可数学化地表示为

$$G_U = (N, F_U, u_n(a_n, \mathbf{b})) \quad (16)$$

则跟随者博弈问题可描述为寻找最佳的信道接入选择，以最大化其效应函数，即

$$\mathbf{a}_n^* = \arg \max_{a_n} u_n(a_n, \mathbf{b}), \quad n \in N \quad (17)$$

4.2 分层博弈均衡解

对于以上构建的干扰对抗分层博弈框架，斯塔伯格均衡可用来分析博弈 G 的性质。斯塔伯格均衡的定义如下所示。

定义 1 如果无人机用户和干扰器的策略选择形式 $(\mathbf{a}^*, \mathbf{b}^*)$ ，满足在分层博弈框架中没有参与者能够通过单方面改变策略选择来提高其效应值，则 $(\mathbf{a}^*, \mathbf{b}^*)$ 构成了分层博弈 G 的斯塔伯格均衡。即对于 $n \in N$ 、 $m \in M$ ，式(18)成立。

$$\begin{cases} u_n(\mathbf{a}_n^*, \mathbf{a}_{-n}^*, \mathbf{b}^*) \geq u_n(a_n, \mathbf{a}_{-n}, \mathbf{b}^*) \\ w_m(\mathbf{a}^*, \mathbf{b}_m^*, \mathbf{b}_{-m}^*) \geq w_m(\mathbf{a}^*, \mathbf{b}_m, \mathbf{b}_{-m}^*) \end{cases} \quad (18)$$

其中，向量 \mathbf{a}_{-n}^* 表示无人机集合 N 中除 n 之外的其他用户的策略选择形式，向量 \mathbf{b}_{-m}^* 表示干扰器集合 M 中除 m 之外的其他干扰器最佳的策略选择形式。

需要指出的是，分层博弈是一种非合作博弈，而在一般的非合作博弈中纳什均衡是最常用的均衡解概念，其构成了非合作博弈的一个稳定点，在该点处没有参与者能够通过单方面策略变化来提高收益。基于此，可以通过求解领导者博弈 G_J 和跟随者博弈 G_U 的纳什均衡解来得到分层博弈 G 的斯塔伯格均衡解^[15]。

4.3 均衡解分析

接下来，分析分层博弈 G 均衡解斯塔伯格均衡的存在性和唯一性。

4.3.1 存在性分析

定理 1 所构建的分层博弈 G 中存在斯塔伯格均衡。

证明 对于无人机用户来说，当干扰器的信道攻击策略确定时，跟随者博弈 G_U 构成一个非合作博弈。在 G_U 中参与者的策略空间 F_U 是欧氏空间上的非空、紧凑、凸子集。此外，每个参与者的效应函数是关于信道选择的连续凹函数。根据文献[16]，则当给定干扰器的信道攻击策略选择形式 \mathbf{b} 时， G_U 中至少存在一个纳什均衡解。同理，当给定无人机用户的信道接入策略选择形式 \mathbf{a} 时， G_J 中至少存在一个纳什均衡解。领导者博弈 G_J 和跟随者 G_U 博弈的稳定策略构成了一个斯塔伯格均衡，即

$$\mathbf{a}^* = \arg \max_{\mathbf{a}} u_n(\mathbf{a}, \mathbf{b}) \quad (19)$$

$$\mathbf{b}^* = \arg \max_{\mathbf{b}} w_m(\mathbf{a}(\mathbf{b}), \mathbf{b}) \quad (20)$$

则 $(\mathbf{a}^*(\mathbf{b}^*), \mathbf{b}^*)$ 构成了分层博弈 G 的一个斯塔伯格均衡解。

证毕。

4.3.2 唯一性分析

定理 2 在干扰对抗频谱接入分层博弈 G 中斯塔伯格均衡解是唯一的。

证明 分析无人机用户 n 效应函数 u_n 的二阶导数, 可得

$$\begin{aligned} \frac{\partial^2 u_n}{\partial (H_{mn}^{a_n})^2} &= \frac{\partial^2 \text{lb} \left(1 + \frac{P_U H_{mn}^{a_n}}{I_n + \sigma^2} \right)}{\partial (H_{mn}^{a_n})^2} = \\ &= \frac{d \left\{ \frac{P_U}{(\ln 2)(I_n + \sigma^2 + P_U H_{mn}^{a_n})} \right\}}{d(H_{mn}^{a_n})} = \\ &= -\frac{P_U}{\ln 2} \frac{1}{(I_n + \sigma^2 + P_U H_{mn}^{a_n})^2} < 0 \end{aligned} \quad (21)$$

由此看出, 当给定干扰器信道攻击策略 \mathbf{b} 时, 无人机用户 n 的效应函数 u_n 是关于其信道增益的凸函数。则基于对偶优化理论^[17], 跟随者的最佳相应 $\mathbf{a}^*(\mathbf{b}^*)$ 是唯一的。同理可得, 干扰器 m 的最佳相应 \mathbf{b}^* 是唯一的。因此, 分层博弈 G 的斯塔伯格均衡解 $(\mathbf{a}^*(\mathbf{b}^*), \mathbf{b}^*)$ 是唯一的。

证毕。

4.4 强化学习算法

4.4.1 算法描述

为了获得所构造分层博弈 G 的斯塔伯格均衡解, 基于 Q -学习机制^[18-19], 提出一种分层信道选择学习算法。在所提算法中, 干扰器和无人机用户以不同的时间尺度更新其策略选择, 即领导者以周期 k 更新一次信道选择, 而跟随者以时隙 t 更新一次信道选择。每个周期 k 包含 T 个时隙 t , 即 $k=Tt$ 。为了便于表述, 将博弈参与者 i 的信道选择扩展为混合策略。具体来说, 对于无人机用户 n 来说, 其在时隙 t 的混合策略可表示为向量 $\mathbf{p}_n(t)=(p_{n1}(t), \dots, p_{nf}(t), \dots, p_{nF}(t))$, 显然, $\sum_{f \in F} p_{nf}(t)=1$, 其中, $p_{nf}(t)$

表示用户 n 在时隙 t 从可选信道集合 F 中选择信道 f 的概率; 对于干扰器 m 来说, 其在周期 k 的混合策略可表示为向量 $\mathbf{s}_m(k)=(s_{m1}(k), \dots, s_{mf}(k), \dots, s_{mF}(k))$, 显

然, $\sum_{f \in F} s_{mf}(k)=1$, 其中, $s_{mf}(k)$ 表示用户 m 在周期 k 从可选信道集合 F 中选择信道 f 的概率。

基于以上混合策略表示, 无人机用户 n 选择信道 a_n 的 Q 值的更新规则为

$$Q_{na_n}^N(t+1)=(1-\lambda_t)Q_{na_n}^N(t)+\lambda_t u_n(a_n, \mathbf{a}_{-n}, \mathbf{b}) \quad (22)$$

其中, $\lambda_t \in [0, 1]$ 表示无人机用户的学习速率, 满足式(23)所示条件。

$$\sum_{t=0}^{\infty} \lambda_t = \infty, \sum_{t=0}^{\infty} (\lambda_t)^2 < \infty \quad (23)$$

则无人机用户 n 的策略选择按式(24)所示的规则进行更新。

$$p_{na_n}(t+1) = \frac{\exp \left[\frac{Q_{na_n}^N(t+1)}{\tau} \right]}{\sum_{f \in F} \exp \left[\frac{Q_{nf}^N(t+1)}{\tau} \right]} \quad (24)$$

其中, τ 用于控制参与者对策略空间“探索”和“扩张”的平衡。

与之相似, 干扰器 m 选择信道 b_m 的 Q 值的更新规则为

$$Q_{mb_m}^M(k+1)=(1-\rho_k)Q_{mb_m}^M(k)+\rho_k w_m(\mathbf{a}, \mathbf{b}_m, \mathbf{b}_{-m}) \quad (25)$$

其中, $\rho_k \in [0, 1]$ 表示干扰器的学习速率, 满足式(26)所示条件。

$$\sum_{k=0}^{\infty} \rho_k = \infty, \sum_{k=0}^{\infty} (\rho_k)^2 < \infty \quad (26)$$

则干扰器 m 的策略选择按式(27)所示规则进行更新。

$$s_{mb_m}(k+1) = \frac{\exp \left[\frac{Q_{mb_m}^M(k+1)}{\tau} \right]}{\sum_{f \in F} \exp \left[\frac{Q_{mf}^M(k+1)}{\tau} \right]} \quad (27)$$

所提分层信道选择学习算法执行过程中, 参与者(领导者和跟随者)的策略更新迭代如图2所示, 详细步骤如算法1所示。首先, 干扰器选择其在当前周期要攻击的信道; 然后, 无人机用户学习环境, 直到其子博弈收敛到纳什均衡解; 重复该过程, 直到达到最大迭代次数或相邻时隙的信道选择策略相同。在多次迭代后, 领导者子博弈和跟随者子博弈均收敛到纳什均衡解, 即为分层博弈的斯塔伯格均衡解。

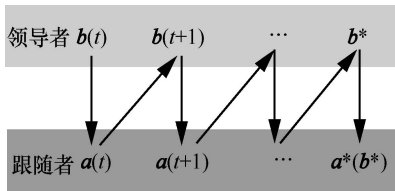


图2 策略更新迭代示意

算法1 分层信道选择学习算法

输入 设置 $k=0, t=0$, 初始化无人机用户和干扰器的混合策略 $p_n(0)$ 、 $s_m(0)$ 及 Q 值。

输出 分层博弈的斯塔伯格均衡解 $(a^*(b^*), b^*)$, 即无人机用户和干扰器优化的信道选择策略

迭代更新

步骤1 在周期 k 中, 干扰器 m 根据其混合策略 $s_m(k)$ 在信道集合 F 中选择攻击信道 b_m 。

步骤2 无人机用户学习过程

① 在时隙 t 无人机用户 n 根据其混合策略 $p_n(t)$ 在信道集合 F 中选择接入信道 a_n 。

② 无人机用户 n 评估效应收益 $u_n(a_n, b)$ 。

③ 无人机用户 n 根据式(22)更新 Q 值。

④ 无人机用户 n 根据式(24)更新其下一时隙信道选择策略 p_n 。

⑤ 令 $t=t+1$, 直到满足结束条件。

步骤3 干扰器 m 评估其效应收益 $w_m(a, b_m)$ 。

步骤4 干扰器 m 根据式(25)更新其 Q 值。

步骤5 干扰器 m 根据式(27)更新其下一周期的信道选择策略 s_m 。

步骤6 返回步骤1, 直到满足结束条件。

4.4.2 收敛性能分析

定理3 所提分层信道选择学习算法收敛。

证明 根据文献[20], 当式(23)和式(26)的条件满足时, 算法的收敛性能能够得到保证。因此, 本文所提学习算法是收敛的。关于强化学习算法收敛证明的详细过程请参考文献[19-20]。

证毕。

5 仿真结果**5.1** 参数设置

仿真中参数的设置具体如下。首先, 对于网络中无人机节点: 用户数 $N=5$, 干扰器个数 $M=6$, 其飞行速度均为 $v=20$ m/s, 无人机用户发射功率 $P_U=30$ dBm, 干扰器发射功率设置为 $P_J=20$ dBm, 为保证无人机用户的可靠传输, 其数据速率门限为 $R_0=2$ bit/s; 其次, 对于网络信道参数: 可用信道数

为 $F=5$, 其路径损耗设置为 $\alpha=2$, 瞬时随机衰落增益 β_f 是区间 $[0.6, 1]$ 上的随机数, 信道中加性高斯白噪声功率设为 $\sigma^2=-114$ dBm; 最后, 对于算法执行中涉及的参数: 策略空间“探索”和“扩张”控制参数 $\tau=1$, 无人机用户和干扰器的学习速率动态变化, 设为 $\lambda_t = \frac{1}{t+1}$, $\rho_k = \frac{1}{k+1}$ 。

5.2 结果分析

下面, 主要从收敛性能和系统性能对算法进行分析, 本文选择的对比算法为遗传优化算法和随机选择算法。同时, 为了消除仿真实验的随机性, 除特别说明的单次运行结果之外, 实验均进行了 500 次, 取 500 次结果的平均值作为最终的结果。

5.2.1 收敛性能

从单次执行结果和多次统计性能 2 个方面对所提算法的收敛性进行分析。首先, 给出在单次执行过程中无人机用户和干扰器的信道选择概率随迭代次数的变化演进, 以无人机用户 1 和干扰器 1 为例, 实验结果分别如图 3 和图 4 所示。从图 3 中可以看出, 在初始时隙时, 无人机用户以相同的信道概率 ($p_{11}=p_{12}=p_{13}=p_{14}=p_{15}=0.2$) 随机选择信道, 在经过 100 次迭代后 $p_{12}=1$, 其他概率变为 0, 说明无人机用户 1 最终选择了信道 2。同理, 对于干扰器来说, 在初始周期时, 以相同的概率随机选择信道, 在大约 40 次迭代达到收敛状态, 干扰器 1 最终选择了信道 5 进行攻击。

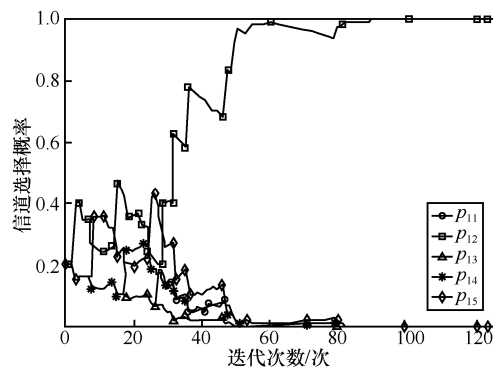


图3 无人机用户1在某一周期中的信道选择概率演进

其次, 从统计角度分析所提算法收敛性能, 无人机用户和干扰器所需迭代次数的累积分布函数如图 5 所示。从图 5 中可以看出, 干扰器收敛所需的平均迭代次数 (30 次) 小于无人机用户收敛所需的平均迭代次数 (100 次), 这与图 3 和图 4 的结果相一致。这是因为无人机用户为跟随者, 其在干扰

器策略的基础上进行决策，因此其获得纳什均衡解所需的迭代次数增多。

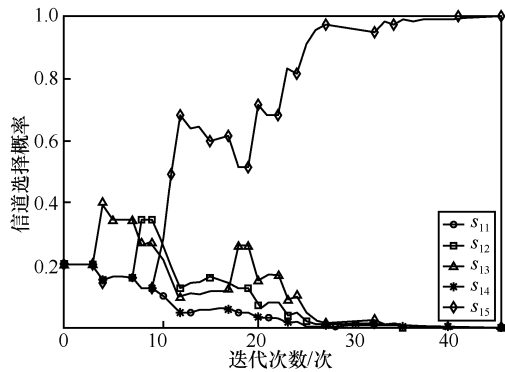


图 4 干扰器 1 在单次执行中的信道选择概率演进

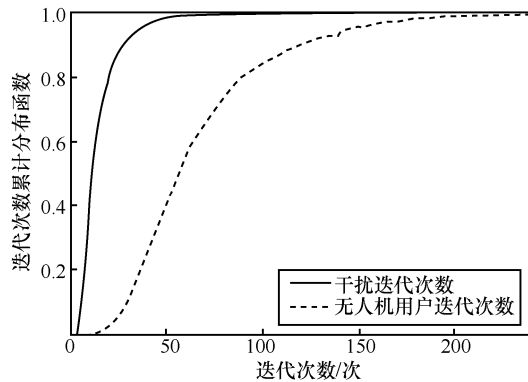


图 5 无人机用户和干扰器所需迭代系数的累积分布函数

5.2.2 系统性能

首先给出不同算法下无人机用户的总数据速率性能对比。采用本文所提算法、遗传优化算法和随机选择算法所获得的无人机用户总数据速率性能结果如图 6 所示。从图 6 中可以看出，所提分层学习算法的性能显著优于 2 种对比算法的性能。由于随机选择算法没有优化机制来进行资源的协调和分配，用户有很大的概率集中于某一信道上，因此其性能最差。遗传优化算法所取得的系统性能介于本文所提算法和随机选择算法之间，具有一定的优化效果。本文所提算法能够实现优化的频谱接入，最大限度地降低同道干扰和恶意干扰，因而能够获得良好的系统性能。

进一步对本文所提算法在不同干扰器个数下所取得的总数据速率性能进行分析，仿真结果如图 7 所示。从图 7 中可以看出，随着干扰器个数的增多，无人机用户的总数据速率性能下降。显然，网络中越多的干扰器对用户产生越大的恶意干扰，这使其有效的数据传输严重被破坏，从而降低网络总数据速率。

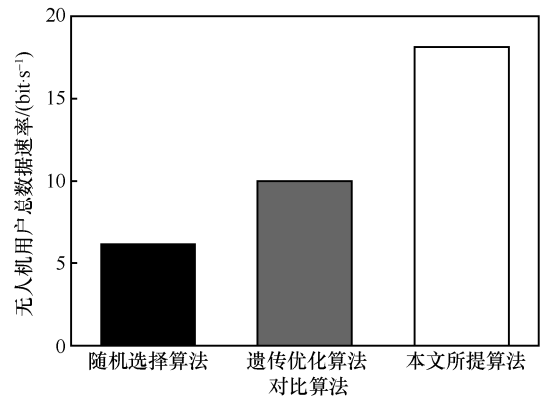


图 6 不同优化算法下的总数据速率性能对比

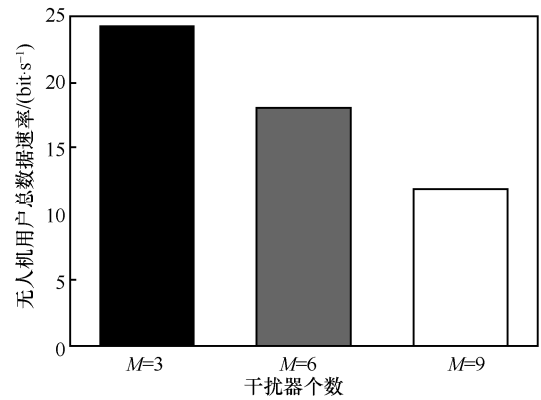


图 7 不同干扰器个数下所提算法的总数据速率对比

6 结束语

干扰对抗问题在无人机通信网络中具有重要研究意义，如何在资源受限的网络中自适应地进行资源优化分配以降低内部干扰并避免外部恶意攻击是一个重大挑战。本文考虑无人机网络中节点动态移动特性，研究干扰对抗场景下的频谱接入优化问题，并提出基于分层博弈的自主信道选择机制。在分层博弈中，干扰器为领导者，其目标为最大化攻击破坏；无人机用户为跟随者，其目标为最大化数据传输速率。对于构建的分层博弈模型，对其斯塔伯格均衡解的存在性和唯一性进行了分析。在此基础上设计一种基于 Q -学习的分层强化学习算法来求解分层博弈的均衡解，并分析其收敛性能。最后，通过实验仿真对本文所提算法和遗传优化算法、随机选择算法的优化性能进行对比分析，从实验结果可以看出，所提分层学习算法优于以上 2 种对比算法，在干扰对抗频谱接入中具有良好的优化性能。

参考文献:

[1] GUPTA L, JAIN R, VASZKUN G. Survey of important issues in UAV

- communication networks[J]. IEEE Communications Surveys & Tutorials, 2015, 18(2): 1123-1152.
- [2] 沈旭. 作为空中基地的无人机覆盖与资源分配研究[D]. 北京: 北京邮电大学, 2018.
SHEN X. Research on UAV coverage and resource allocation as air base station[D]. Beijing: Beijing University of Posts and Telecommunications, 2018.
- [3] ZOU Y, ZHU J, WANG X, et al. A survey on wireless security: technical challenges, recent advances, and future trends[J]. Proceedings of the IEEE, 2016, 104(9): 1727-1765.
- [4] 张建英. 博弈论的发展及其在现实中的应用[J]. 理论探索, 2005, 2(3): 36-37.
ZHANG J Y. Game theory and its application[J]. Theory Exploration, 2005, 2(3): 36-37.
- [5] XIAO L, LIU J, LI Q, et al. User-centric view of jamming games in cognitive radio networks[J]. IEEE Transactions on Information Forensics and Security, 2015, 10(12): 2578-2590.
- [6] JIA L, YAO F, SUN Y, et al. A hierarchical learning solution for anti-jamming Stackelberg game with discrete power strategies[J]. IEEE Wireless Communications Letters, 2017, 6(6): 818-821.
- [7] LI Y, XIAO L, LIU J, et al. Power control stackelberg game in cooperative anti-jamming communications[C]//The 2014 5th International Conference on Game Theory for Networks. Piscataway: IEEE Press, 2014: 1-6.
- [8] WANG B, WU Y, LIU K J R, et al. An anti-jamming stochastic game for cognitive radio networks[J]. IEEE Journal on Selected Areas in Communications, 2011, 29(4): 877-889.
- [9] JIA L, YAO F, SUN Y, et al. Bayesian Stackelberg game for antijamming transmission with incomplete information[J]. IEEE Communications Letters, 2016, 20(10): 1991-1994.
- [10] YANG D, XUE G, ZHANG J, et al. Coping with a smart jammer in wireless networks: a Stackelberg game approach[J]. IEEE Transactions on Wireless Communications, 2013, 12(8): 4038-4047.
- [11] MACHUZAK S, JAYAWEERA S K. Reinforcement learning based anti-jamming with wideband autonomous cognitive radios[C]//2016 IEEE/CIC International Conference on Communications. Piscataway: IEEE Press, 2016: 1-5.
- [12] AREF M A, JAYAWEERA S K, MACHUZAK S. Multi-agent reinforcement learning based cognitive anti-jamming[C]//2017 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2017: 1-6.
- [13] HANAWAL M K, ABDEL-RAHMAN M J, KRUNZ M. Joint adaptation of frequency hopping and transmission rate for anti-jamming wireless systems[J]. IEEE Transactions on Mobile Computing, 2015, 15(9): 2247-2259.
- [14] SAGDUYU Y E, BERRY R A, EPHREMIDES A. Jamming games in wireless networks with incomplete information[J]. IEEE Communications Magazine, 2011, 49(8): 112-118.
- [15] 王学婷, 朱琦. 分层异构网络中基于斯坦克尔伯格博弈的资源分配算法[J]. 信号处理, 2017, 33(2): 168-177.
WANG X T, ZHU Q. Stackelberg game based resource allocation in hierarchical heterogeneous networks[J]. Signal Processing, 2017, 33(2): 168-177.
- [16] ROSE N J B. Existence and uniqueness of equilibrium points for concave n-person games[J]. Journal of the Econometric Society, 1965: 520-534.
- [17] DING G, WU Q, WANG J. Sensing confidence level-based joint spectrum and power allocation in cognitive radio networks[J]. Wireless Personal Communications, 2013, 72(1): 283-298.
- [18] 高阳, 陈世福, 陆鑫. 强化学习研究综述[J]. 自动化学报, 2004, 30(1): 86-100.
GAO Y, CHEN S F, LU X. A survey of reinforcement learning[J]. Journal of Automation, 2004, 30(1): 86-100.
- [19] FAN C, LI B, ZHAO C, et al. Learning-based spectrum sharing and spatial reuse in mm-wave ultra-dense networks[J]. IEEE Transactions on Vehicular Technology, 2017, 67(6): 4954-4968.
- [20] SUN Y, SHAO H, LIU X, et al. Traffic offloading in two-tier multi-mode small cell networks over unlicensed bands: a hierarchical learning framework[J]. Transactions on Internet and Information Systems, 2015, 9(11): 4291-4310.

[作者简介]



范超琼 (1992-), 女, 河北石家庄人, 北京邮电大学博士生, 主要研究方向为无线资源管理、博弈论、强化学习等。



赵成林 (1964-), 男, 河北石家庄人, 博士, 北京邮电大学教授, 主要研究方向为信号处理、无线通信技术等。



李斌 (1985-), 男, 甘肃天水人, 博士, 北京邮电大学副教授, 主要研究方向为统计信号估计与检测、认知无线电、60 GHz 毫米波通信、高性能优化算法设计等。