

doi: 10.12052/gdutxb.200113

基于强化学习的无人机安全通信轨迹在线优化策略

郑思远, 崔 苗, 张广驰

(广东工业大学 信息工程学院, 广东 广州 510006)

摘要: 物理层安全是无线通信实现安全通信的一种有效手段, 无人机基站在存在地面窃听者的情况下向地面合法用户传输机密信息时, 已有研究一般采用离线方法对无人机基站的飞行轨迹进行优化, 得到的轨迹是固定的, 保密传输缺乏应对通信环境变化的能力。针对该问题, 本文研究无人机飞行轨迹的在线优化策略, 使机密信息被安全传输的同时实现通信平均保密率最大化。文中采用了更符合实际的视距/非视距混合信道模型对空地通信链路进行建模, 提出一种基于Q-learning算法的保障无人机通信系统安全的飞行轨迹在线优化方法。仿真结果表明, 与两种基准算法相比, 本文提出的算法能够有效提高无人机通信的平均保密率。

关键词: 物理层安全; 无人机基站; 在线优化; 强化学习

中图分类号: TN929.5 文献标志码: A 文章编号: 1007-7162(2021)04-0059-06

Reinforcement Learning-Based Online Trajectory Optimization for Secure UAV Communications

Zheng Si-yuan, Cui Miao, Zhang Guang-chi

(School of Information Engineering, Guangdong University of Technology, Guangzhou 510006, China)

Abstract: Physical-layer security is an effective technique to achieve secure communication in wireless communication. Physical-layer security in an unmanned aerial vehicle (UAV) communication system is investigated, where a UAV base station transmits confidential information to a legitimate receiver in the presence of a potential eavesdropper. The existing works usually assume that the flight environment of the UAV is stable, and the corresponding trajectory optimization methods are offline, which are not able to adjust the trajectory according to changing environment. To solve this issue, online trajectory optimization for the UAV base station is studied to ensure the information transmission security and maximize the communication secrecy rate. Different from the existing works, a more accurate hybrid light-of-sight/non-light-of-sight channel model is adopted for the air-to-ground communication links and an online trajectory optimization method proposed based on the Q-learning algorithm to ensure the physical-layer security of UAV communication. Numerical results show that the proposed algorithm significantly improves the secrecy rate of the UAV communication system, as compared with two benchmark schemes.

Key words: physical-layer security; UAV base station; online optimization; reinforcement learning

随着信息时代的快速发展, 无人机凭借自身独特的优势被广泛地应用到诸多领域, 比如空中侦察、航空影像以及灾后紧急救援等。在无线通信领域中, 无人机因其高度灵活性、造价低廉等特点备受欢迎, 其主要应用在3个方向, 移动基站、移动中继和数据搜集^[1-3]。在无人机协作的无线通信系统中, 无人机在

空中的高度优势使得其与地面通信时的链路极大概率是视距信道(Light-of-Sight, LoS), 很大程度上提高了通信质量。然而无人机的参与也带来了新的安全挑战, 由于无线信道具备广播的开放特性, 无人机与地面通信在提高通信质量的同时也提高了潜在窃听者接收信息的强度, 增加了信息传输时的安全隐患。

收稿日期: 2020-09-07

基金项目: 广东省科技计划项目(2017B090909006, 2018A050506015, 2019B010119001, 2020A050515010, 2020A0505100012, 2021A0505030015); 广东特支计划项目(2019TQ05X409)

作者简介: 郑思远(1994-), 男, 硕士研究生, 主要研究方向为强化学习和无线安全通信等

通信作者: 崔苗(1978-), 女, 讲师, 博士, 主要研究方向为电子信息与无线通信技术, E-mail: cuimiao@gdut.edu.cn

如何确保机密信息在传输过程中不被窃听是一个很重要的应用技术。物理层安全作为一个有效的窃听技术被广泛研究^[4],本文通过研究物理层安全提高无人机无线通信的安全性能。

根据窃听信道模型可知,当合法信道的通信质量优于窃听信道时,可以实现安全通信性能的提升^[5]。保密率是物理层安全中一个极为重要的概念,当合法链路的通信质量比窃听链路的通信质量高时,可以得到一个非零的保密率。因此,物理层安全研究的关键就是考虑如何在提高合法链路通信质量的同时降低窃听链路的通信质量。在已有的物理层安全相关文献中,根据是否利用到无人机的高度机动性可以分为两种类型,即通信节点是静态(或准静态)的无线通信和无人机协作的无线通信。当通信节点固定时,物理层安全的研究主要集中在信道、功率控制以及信号检测等技术手段^[6-7]。例如文献[8-9]提出利用人工噪声在避开合法接收者的情况下降低窃听链路的通信质量。在多输入多输出系统中,可以利用波束成形的技术方法改善合法链路的通信质量,降低窃听链路的通信质量^[10-11]。相对于将通信节点布置在地面的传统方法,无人机协作的物理层安全研究充分利用了无人机的机动灵活性,使其在三维空间内部署,可以明显提高合法链路的通信质量,提高保密率。文献[12-13]提出联合优化无人机轨迹和通信资源分配以提高物理层安全。文献[14]将无人机作为空中基站,通过有效分配带宽、功率等资源提高无线通信系统的通信质量。文献[15-16]提出通过可移动的无人机发出噪声干扰窃听者,从而提高通信安全的方法。此外,考虑到单个无人机的活动范围较小,对系统通信质量的提升有限,文献[17-18]对多无人机进行部署以提高系统的通信质量和吞吐量。值得注意的是,以上所述文献的无线通信系统是基于视距信道模型的,将通信环境视为完全已知。若无人机和地面用户的坐标确定,可以在无人机出发前完成整个飞行路径的设计,轨迹优化策略也是离线的。若通信环境发生变化,离线策略得到的无人机轨迹不能做出相应变动,其弊端在于忽略了由于存在障碍物位置的随机性造成无人机与地面节点间信道衰落的随机性,真实的信道应是视距信道和非视距信道(Non-Line-of-Sight, NLoS)的混合,故信道增益也是随机的。

针对这个问题,本文研究无人机轨迹的在线优化策略,提出一种基于强化学习Q-learning算法的优化方法^[19]。动态规划(Dynamic Programming, DP),蒙特卡洛方法(Monte Carlo Methods, MC)和时序差分学习(Temporal-Difference Learning, TD)是强化学习

中常用的方法,而Q-learning算法是TD方法中一个重要突破。相比于DP方法,TD方法的最大优势在于其可以直接从与环境互动的经验中学习,不需要构建环境模型。相比于MC方法,TD方法不需要等待与环境交互的最终结果,而是基于已得到的其他状态的估值以在线、完全增量的方式更新当前状态的动作价值函数,可用于连续任务。

相比于已有的无人机轨迹设计算法,本文提出的算法不需要环境完全已知,借助无人机在飞行过程中每个位置得到的反馈信号作为收益值可以不断训练、更新无人机的轨迹,最终实现路径优化。若是环境发生变化,无人机也可以通过与环境的自主交互实现轨迹重新优化。算法的目标是通过优化无人机的飞行轨迹使得平均保密率达到最大。此外,本文的空地信道模型采用更符合实际的视距/非视距混合信道模型,由文献[20]可知,空地通信链路是视距链路或者是非视距链路的概率大小与无人机到地面节点的仰角以及环境因素有关。由仿真结果可知,当无线通信系统的通信环境处于动态变化时,本文提出的算法可以在线优化无人机飞行路径,并取得优于基准策略的平均保密率性能。

1 系统模型

本文考虑一个无人机基站通信系统,该系统由一个无人机基站和一个地面合法接收节点组成,同时地面存在一个窃听节点,系统模型如图1所示。地面窃听者的位置可借助雷达扫描定位等技术手段探得^[12,21],所以本文设定合法接收节点和窃听节点的位置已知,两者之间的距离为 Z ,其位置坐标分别是 $(0,0,0)$ 和 $(Z,0,0)$ 。为了方便计算,设无人机的起飞、降落过程所需时间忽略不计,假设无人机基站一次完整的飞行时间为 T ,在时刻 t 无人机的坐标是 $\Omega(t) = (x(t), y(t), H)$, $0 \leq t \leq T$,其中 H 是无人机固定的飞行高度。大多数情况下无人机的最小飞行高度需要比地面上的建筑高而且与地面接收者的通信链路主要由视距链路构成。为了维持良好的通信性能,无人机会固定在最小飞行高度上飞行,因此本文考虑无人机的飞行高度固定。无人机的起始位置和终点位置分别是 $\Omega_0 = (x_0, y_0, H)$ 和 $\Omega_F = (x_F, y_F, H)$ 。在飞行过程中,无人机飞行时速度设为定值,即 $\|v(t)\| = v$,无人机悬停时速度为0。无人机在时刻 t 与终点的距离是

$$D(t) = \|\Omega(t) - \Omega_F\|, \forall t \quad (1)$$

本文采用文献[22]中的视距/非视距混合信道模型,对无人机基站到地面两个节点的链路进行建模。其信道增益包含大尺度的路径损耗和小尺度的瑞利

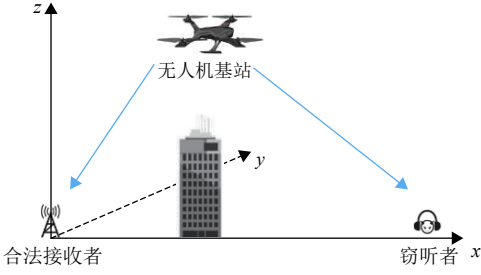


图1 一个基于无人机的无线通信系统

Fig.1 A UAV-enabled wireless communication

衰落。用 h_k 表示无人机到地面节点的信道增益,其中 k 取值为1和2,分别表示合法接收节点与窃听节点,由此可以得到

$$h_k(t) = \frac{\rho_k(t)}{\sqrt{\text{PL}_k(t)}} \quad (2)$$

其中, $\rho_k(t)$ 表示小尺度衰落, $\text{PL}_k(t)$ 表示通信链路在时刻 t 的平均路径损耗。由于信道是视距或非视距的概率与无人机到地面节点的仰角大小有关,为了明确视距信道和非视距信道对平均路径损耗的影响,需要根据无人机到地面节点的仰角计算出其分别发生的概率。具体而言,无人机与地面节点 k 在时刻 t 为视距信道的概率为

$$\text{pr}_k^{\text{LoS}}(t) = \frac{1}{1 + \omega_1 \exp(-\omega_2(\theta_k(t) - \omega_1))} \quad (3)$$

其中, $\theta_k(t)$ 是无人机与地面节点 k 在 t 时刻的仰角, ω_1 和 ω_2 是根据环境(如郊区、城市等)得到的常量参数。仰角的计算表达式为 $\theta_k(t) = \arctan\left(\frac{H}{L(t)}\right)$, H 和 $L(t)$ 分别是无人机的飞行高度和无人机到地面节点 k 的水平距离。由式(3)可以得到无人机到节点 k 在时刻 t 的平均路径损耗

$$\text{PL}_k(t) = \text{pr}_k^{\text{LoS}}(t)\text{PL}^{\text{LoS}}(t) + (1 - \text{pr}_k^{\text{LoS}}(t))\text{PL}^{\text{NLoS}}(t) \quad (4)$$

其中, $\text{PL}^{\text{LoS}}(t)$ 和 $\text{PL}^{\text{NLoS}}(t)$ 分别是视距和非视距两种信道的自由空间路径损耗, $\text{PL}^{\text{LoS}}(t)$ 的表达式是

$$\text{PL}^{\text{LoS}}(t) = \left(\frac{4\pi f d_k(t)}{c}\right)^2 \times \eta^{\text{LoS}} \quad (5)$$

其中,无人机在 t 时刻到节点 k 的距离表示为 $d_k(t)$, π 是圆周率, f 是载波频率, c 是光速, η^{LoS} 是视距信道因环境因素造成的额外损耗。同样地,非视距信道的自由空间路径损耗表示为

$$\text{PL}^{\text{NLoS}}(t) = \left(\frac{4\pi f d_k(t)}{c}\right)^2 \times \eta^{\text{NLoS}} \quad (6)$$

其中, η^{NLoS} 是非视距信道因环境因素造成的额外损耗。不失一般性地,文中把小尺度衰落 $\rho_k(t)$, $\forall t$ 看作 $\mathbb{E}\{\rho_k^2\} = 1$ 的独立同分布。根据模型,无人机到地面

节点 k 链路的信道增益是 $g_k = |h_k(t)|^2$,无人机的传输功率被标记为 P ,设为常量。根据香农公式,从无人机到地面节点 k 的可达率(bps/Hz)为

$$\Gamma_k(t) = \log_2\left(1 + \frac{Pg_k(t)}{\sigma^2}\right) \quad (7)$$

其中, σ^2 是地面节点接收机密信息时产生的热噪声功率,无人机在时刻 t 的保密率为^[9,12]

$$\Gamma_{\text{sec}}(t) = [\Gamma_1(t) - \Gamma_2(t)]^+ \quad (8)$$

无人机完成一次飞行的时间为 T ,本文把 T 分成 N 个相等的时间间隔,即 $T = Nd_t$, d_t 表示时间间隔的长度,将其设置为足够小使得每个时间间隔内的无人机可以视为处于静止状态。式(7)则可以改写为

$$\Gamma_k[n] = \log_2\left(1 + \frac{Pg_k[n]}{\sigma^2}\right) \quad (9)$$

其中, $n=1,2,\dots,N$,由式(8)可知,在无人机从出发点到终点飞行的 N 个时段中,无人机基站的平均保密率是

$$\Gamma_{\text{ave-sec}} = \frac{1}{N} \sum_{n=1}^N [\Gamma_1[n] - \Gamma_2[n]]^+ \quad (10)$$

其中, $[x]^+ = \max(x, 0)$ 。

本文的目标是通过通过对无人机轨迹优化使得式(10)中的平均保密率 $\Gamma_{\text{ave-sec}}$ 最大化,已有文献中传统方案对此类问题的解决方式是利用连续凸近似和块坐标下降法的方法,把非凸问题转化为凸问题。但是此方法是离线的,没有考虑无人机到地面节点间信道的随机性,并且不适用于更符合实际的视距/非视距混合信道模型。考虑到信道增益的随机性,本文提出采用在线优化方法对无人机的轨迹进行优化。由于强化学习方法不需要环境完全已知,无人机可以直接从环境中学习,对环境进行探索和利用,因此本文提出基于强化学习的方法解决无人机轨迹在线优化的问题。

2 强化学习的实现机理

强化学习明确考虑了目标导向的智能体与未知环境之间的交互问题。马尔可夫决策过程(Markov Decision Process, MDP)是一种通过交互式学习来实现目标的理论框架,也是强化学习问题在数学上的理想形式。智能体是进行环境学习和实施决策的个体,与智能体相互作用的事物被称为环境,也可以将环境理解为智能体要经历的状态集合。智能体做出决策,即选择一个合适的动作 A_t ,环境对动作做出响应,智能体所处的状态会发生改变,从 S_t 变为 S_{t+1} ,同时得到采取动作 A_t 后的收益 R_{t+1} ,也是智能体在动作选择过程中想要最大化的数值。图2展示了智能体与

环境的交互过程,过程完成后智能体继续选择下一个合适的动作,不断重复此过程直到达到结束条件,这一系列状态和动作构成智能体的策略 Π 。

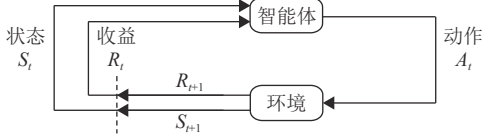


图2 马尔可夫决策过程中智能体与环境的交互

Fig.2 The agent-environment interaction in a Markov decision process

使用收益来形式化目标是强化学习最显著的特征之一,智能体的目标不是最大化当前收益,而是最大化长期收益 $G_t \triangleq \sum_{i=0}^{T-1} \gamma^i R_{t+i+1}$,其中 γ 是折扣率,取值 $0 \leq \gamma \leq 1$, T 是最终时刻。动作价值函数是强化学习中一个重要概念,定义为在策略 Π 下从状态 s 开始,执行动作 a 后所有可能的决策序列的期望收益,记为 $Q_{\Pi}(s, a) = \mathbb{E}_{\Pi}[G_t | S_t = s, A_t = a]$ 。强化学习的最终目标是找到最优策略实现最大化长期收益,相应的最优动作价值函数是 $Q_*(s, a) = \max_{\Pi} Q_{\Pi}(s, a)$ 。

3 Q-learning算法优化无人机轨迹

Q-learning算法是一种使用时序差分求解强化学习控制问题的方法,是不基于环境状态转化模型的求解方法。该算法以迭代的方式计算最优动作价值函数,即通过动作价值函数更新策略,由策略产生新的状态和收益,进而更新动作价值函数,直到动作价值函数和策略都收敛。动作价值函数的更新表示为

$$Q(S_t, A_t) = Q(S_t, A_t) + \alpha [R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)] \quad (11)$$

其中,步长 $\alpha \in (0, 1]$ 。Q-learning算法最大优势是待学习的动作价值函数 Q 直接近似于最优动作价值函数 $Q_*(s, a)$,而与生成智能体决策序列的策略无关。

根据提出的模型,本文将无人机飞行过程分为两个阶段。设定时间 T_{mid} ,当无人机的飞行时间 $t \leq T_{\text{mid}}$ 时是第一阶段,无人机从起始位置 Ω_0 飞到保密率最大的位置,并在此位置保持悬停; $t > T_{\text{mid}}$ 时是第二阶段,无人机从保密率最大的位置飞到终点位置 Ω_F ,两阶段的收益值不同。考虑到无人机通信环境的随机性,对于 T_{mid} 的设定分为两部分,第一部分是无人机从起始位置到保密率最大位置的飞行时间,其值在确定保密率最大位置时可以计算得到;第二部分是无人机在保密率最大位置上的悬停时间,此值根据平均保密率的增量设定。具体来讲,无人机在保密率最大位置悬停的时间越久,整体的平均保密率值

越大,当平均保密率的增长速率低于阈值时,则以满足此条件的悬停时间进行计算。由Q-learning算法的框架,对智能体、状态、动作以及收益作出如下定义:

智能体:无人机 U 。

状态 S_t :无人机在规定区域中 t 时刻的位置。一般来说无人机飞行的位置是连续的,为了限制状态的数量,本文把飞行区域离散化,分割成一定数量的格子,无人机沿着格子的坐标进行移动。

动作 A_t :无人机在某状态的动作有5个,即:往东、往西、往南、往北以及悬停。

收益 R_t :由式(1)和式(8)可知 $D(t)$ 和 $\Gamma_{\text{sec}}(t)$ 分别是无人机到终点的距离和无人机在时刻 t 的保密率,收益值被定义为:(1)当无人机的飞行时间 $t \leq T_{\text{mid}}$ 时,收益 $R_t = \beta_1 \Gamma_{\text{sec}}(t)$;(2)当无人机的飞行时间 $t > T_{\text{mid}}$ 时,收益 $R_t = \beta_2 \Gamma_{\text{sec}}(t) + \beta_3 D(t)$;其中 β_1, β_2 和 β_3 是系数。

算法如下:

参数:幕数 N_{ep} ,状态集 S ,动作集 $A(s)$,步长 α ,探索率 ε ,衰减因子 γ 。

(1) 初始化两个 Q 表 $Q_1(s, a)$ 和 $Q_2(s, a)$,且 $\forall s \in S, a \in A(s)$;

(2) 设定2个飞行阶段的终点位置的状态分别是 $Q_1(\text{Terminalstate}, \bullet)$ 和 $Q_2(\text{Terminalstate}, \bullet)$,其值均为0;

(3) 开始迭代以下步骤,直到幕值为 N_{ep} ;

a) 初始化无人机起飞位置;

b) 用 ε -贪婪算法根据当前状态 S_t 选择动作 A_t ;

c) 执行动作 A_t ,得到新的状态 S_{t+1} 和收益 R_{t+1} ;

d) 按照下式更新动作价值函数 Q ,当无人机的飞行时间 $t \leq T_{\text{mid}}$ 时,更新 $Q_1(s, a)$;当无人机的飞行时间 $t > T_{\text{mid}}$ 时,更新 $Q_2(s, a): Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \alpha [R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)]$;

e) $S_t \leftarrow S_{t+1}$;

f) 如果 S_{t+1} 是终止状态,当前迭代结束,否则转到步骤b)。

4 仿真结果

此节将通过仿真结果对文中提出的无人机轨迹优化算法的优越性进行验证,并且与两种基准算法进行比较。

(1) 直线飞行轨迹(标记为Line)。此算法考虑视距/非视距混合信道,无人机从出发位置以恒速飞向合法接收节点的位置,到达后并在该节点悬停,最后在飞行时间结束前飞向终点位置。

(2) 凸近似方法优化轨迹(标记为Convex)。此算法只考虑视距信道,即忽略了信道衰落的随机性,利用连续凸近似法和块坐标下降法对无人机的离线轨

迹进行优化^[12]。

此文中合法接收节点和窃听节点之间的距离设为 $Z = 200$ m,无人机的飞行高度设为 $H = 100$ m,无人机的初始位置和终点位置分别设置 $(x_0, y_0) = (100, 200)$ m和 $(x_F, y_F) = (100, -200)$ m。本文考虑无人机的飞行区域是 250 m $\times$ 400 m,为了方便无人机飞行过程中状态数目的计算,把飞行区域划分成 $25\,000$ 个状态。飞行过程中无人机的功率设定为 $P = -5$ dBm。其他的参数为载波频率 $f = 5$ GHz,热噪声功率 $\sigma^2 = -90$ dBm,环境参数 $(\omega_1, \omega_2) = (5, 0.5)$, $\eta^{\text{LoS}} = 1$, $\eta^{\text{NLoS}} = 20$,收益值参数 $(\beta_1, \beta_2, \beta_3) = (10, 10, -0.3)$, $v = 2$ m/s, $\alpha = 0.5$, $\gamma = 1$ 。

图3显示的是在利用Q-learning算法训练过程中,每幕的平均保密率值随着幕数的增加而收敛。训练开始的时候,无人机因为对未知环境探索,平均保密率有很大的波动。随着迭代次数增加,动作价值函数和策略逐渐收敛,无人机飞行的优化路径逐渐明晰,相应的平均保密率也开始收敛。由于利用 ϵ -贪婪算法选取飞行方向使得无人机小概率处于探索状态,所以平均保密率即便是趋于收敛也存在偶尔扰动,但随着迭代进行,无人机轨迹最终实现优化,使得平均保密率达到最大值。本文采用的是视距/非视距混合信道模型,因为存在障碍物等因素使得通信环境的信道增益是随机的,相应的无人机优化轨迹也是不唯一的。

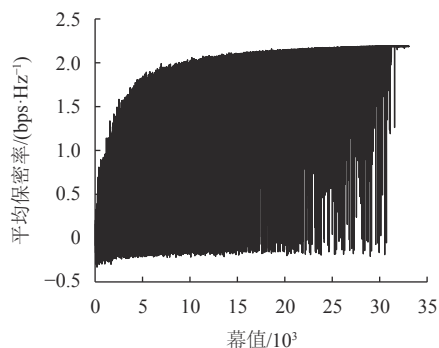


图3 每幕平均保密率的收敛图

Fig.3 Convergence of average secrecy rate versus episode

图4和图5分别展示了文中提出的算法和基准算法在因信道增益随机性而造成的2种不同通信环境下的无人机飞行轨迹。其中合法接收节点、窃听节点的位置以及无人机初始、终点位置分别用 $\circ$, Δ , $\times$, $+$, 这4种符号标记。为了限制无人机的状态数目,文中将其飞行区域离散化并分割成一定数目的格子,因此图中无人机的飞行轨迹是曲折的。由于图4和图5中显示的2种基准算法是离线策略,所以得到的无人机路径是固定不变的。而Q-learning算法是在线策略,其轨迹图(标记为Q-learning)是随着不同通信情况而

变化的,此算法使得无人机飞行时会避开窃听节点的位置,先飞到保密率最大的位置保持悬停状态,最后在飞行时间结束前飞向终点位置。保密率最大的位置在图4和图5中分别用五角星 $\star$ 标记。

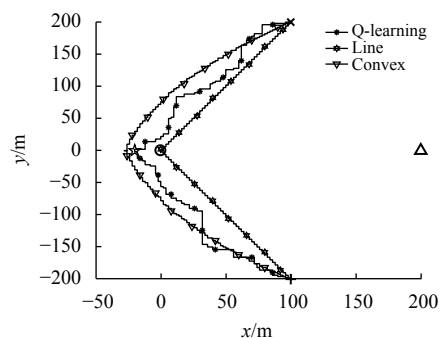


图4 3种算法的轨迹对比图(环境I)

Fig.4 Trajectories of different algorithms (Environment I)

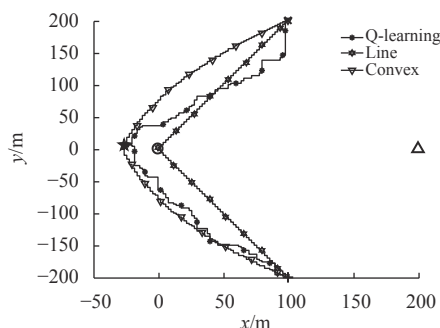


图5 3种算法的轨迹对比图(环境II)

Fig.5 Trajectories of different algorithms (Environment II)

随着无人机在保密率最大位置停留的时间增加,平均保密率也是增加的。图6是在飞行环境I的情况下无人机的平均保密率随着悬停时间增加的变化曲线。当平均保密率的增长率低于设定的阈值时(阈值 $=0.001\,9$ bps/Hz),选择此时间作为无人机的一次完整的飞行时间 T 。

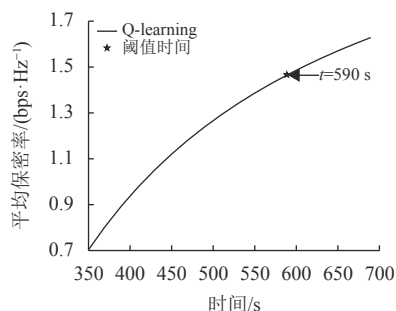


图6 平均保密率随时间变化图

Fig.6 Average secrecy rate versus flight time

图7展示的是在 $T = 590$ s为飞行时间的条件下,3种不同算法所对应的平均保密率随飞行时间的变化。由图可知3种算法的平均保密率都是随着时间加长而出现增长。需要注意的是,利用凸近似方法得到

的无人机轨迹在通信环境只考虑视距传输的情况下是最优的,而在图7中Q-learning算法得到的平均保密率的值高于该算法。原因是本文采用的信道模型是视距/非视距混合模型,其信道衰落存在随机性,而凸近似方法得到的轨迹完全根据视距模型设计的,不能完全反映真实的信道情况,所以该算法的性能不如文中所提的在线优化算法。直线飞行轨迹悬停的位置是在合法接收节点的正上方,由仿真结果可知该位置并不是保密率最大的位置。综上,相对于2种基准算法,本文提出的算法的平均保密率是最大的。

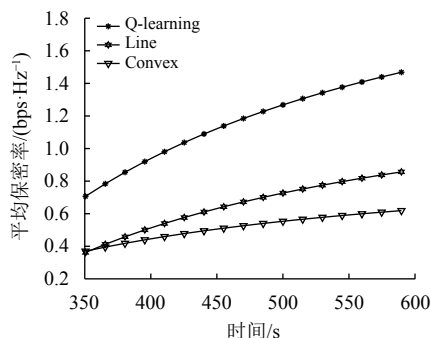


图7 不同算法的平均保密率变化图

Fig.7 Average secrecy rates of different algorithms

5 结论

本文基于强化学习中的Q-learning算法提出了一种无人机与环境自主交互、可在线优化无人机飞行轨迹的保障物理层通信安全的方法,目的是在存在窃听者的情况下使得无人机基站发送的机密信息被安全地传输,最大化平均通信保密率。由仿真结果可知,在视距/非视距混合信道下,利用Q-learning算法能够对无人机进行有效训练使得其飞行轨迹实现在线优化,所提算法的平均保密率性能优于已有的离线基准算法。

参考文献:

- [1] ZENG Y, ZHANG R, LIM T J. Wireless communications with unmanned aerial vehicles — opportunities and challenges [J]. *IEEE Communications Magazine*, 2016, 54(5): 36-42.
- [2] LYU J, ZENG Y, ZHANG R. Cyclical multiple access in UAV-aided communications — a throughput-delay tradeoff [J]. *IEEE Wireless Communications Letters*, 2016, 5(6): 600-603.
- [3] SOTHEARA S, ASO K, AOMI N, *et al*. Effective data gathering and energy efficient communication protocol in wireless sensor networks employing UAV[C]//Wireless Communications & Networking Conference. Istanbul: IEEE, 2014: 2342-2347.
- [4] ZHENG G, KRIKIDIS I, LI J, *et al*. Improving physical layer secrecy using full-duplex jamming receivers [J]. *IEEE Transactions on Signal Processing*, 2013, 61(20): 4962-

- 4974.
- [5] WYNER A D. The wire-tap channel [J]. *Bell System Technical Journal*, 1975, 54(8): 1355-1387.
- [6] SHIU Y S, CHANG S Y, WU H C, *et al*. Physical layer security in wireless networks — a tutorial [J]. *IEEE Wireless Communications*, 2011, 18(2): 66-74.
- [7] 童辉志, 张广驰, 周绪龙, 等. 具有能量获取基站的相邻多蜂窝小区的能量与频谱分配研究[J]. *广东工业大学学报*, 2018, 35(4): 72-78.
- TONG H Z, ZHANG G C, ZHOU X L, *et al*. Joint energy and spectrum allocation in multiple adjacent cells with energy harvesting base stations [J]. *Journal of Guangdong University of Technology*, 2018, 35(4): 72-78.
- [8] GOEL S, NEGI R. Guaranteeing secrecy using artificial noise [J]. *IEEE Transactions on Wireless Communications*, 2008, 7(6): 2180-2189.
- [9] GOPALA P K, LAI L, EL GAMAL H. On the secrecy capacity of fading channels [J]. *IEEE Transactions on Information Theory*, 2008, 54(10): 4687-4698.
- [10] WANG H, YIN Q, XIA X. Distributed beamforming for physical layer security of two-way relay networks [J]. *IEEE Transactions on Signal Processing*, 2012, 60(7): 3532-3545.
- [11] ZHANG G, LI X, CUI M, *et al*. Signal and artificial noise beamforming for secure simultaneous wireless information and power transfer multiple-input-multiple-output relaying systems [J]. *IET Communications*, 2016, 10(7): 796-804.
- [12] ZHANG G, WU Q, CUI M, *et al*. Securing UAV communications via joint trajectory and power control [J]. *IEEE Transactions on Wireless Communications*, 2019, 18(2): 1376-1389.
- [13] ZHANG J, ZENG Y, ZHANG R. Spectrum and energy efficiency maximization in UAV-enabled mobile relaying[C]//International Conference on Communications. Paris: IEEE, 2017: 1-6.
- [14] HUANG Y, CUI M, ZHANG G, *et al*. Bandwidth, power and trajectory optimization for UAV base station networks with back haul and user QoS constraints [J]. *IEEE Access*, 2020(8): 67625-67634.
- [15] ZHOU Y, YEOH P L, CHEN H, *et al*. Secrecy outage probability and jamming coverage of UAV-enabled friendly jammer[C]//2017 11th International Conference on Signal Processing and Communication Systems. Gold Coast: IEEE, 2017: 1-6.
- [16] LIU J, LIU Z, ZENG Y, *et al*. Cooperative jammer placement for physical layer security enhancement [J]. *IEEE Network the Magazine of Global Internetworking*, 2016, 30(6): 56-61.
- [17] FAN J, CUI M, ZHANG G, *et al*. Throughput improvement for multi-hop UAV relaying [J]. *IEEE Access*, 2019(7): 147732-147742.
- [18] ZHANG G, YAN H, ZENG Y, *et al*. Trajectory optimization and power allocation for multi-hop UAV relaying communications [J]. *IEEE Access*, 2018(6): 48566-48576.
- [19] SUTTON R, BARTO A. Reinforcement learning — an introduction[M]. 2nd ed. Cambridge: MIT Press, 2018.
- [20] FENG Q, TAMEH E, NIX A R, *et al*. Modelling the likelihood of line-of-sight for air-to-ground radio propagation in urban environments[C]//GlobeCom2006. San Francisco: IEEE, 2007: 1-5.
- [21] CUI M, ZHANG G, WU Q, *et al*. Robust trajectory and transmit power design for secure UAV communications [J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(9): 9042-9046.
- [22] AL HOURANI A, KANDEEPAN S, LARDNER S. Optimal lap altitude for maximum coverage [J]. *Wireless Communications Letters IEEE*, 2014, 3(6): 569-572.

(责任编辑:王威娜)