文章编号: 1001-0920(2018)07-1199-08 DOI: 10.13195/j.kzyjc.2017.0365

基于强化学习的无线自组网络多节点干扰策略

颛孙少帅†, 杨俊安, 刘 辉, 黄科举

(1. 国防科技大学 电子对抗学院，合肥 230037；2. 安徽省电子制约技术重点实验室，合肥 230037)

摘 要: 为了实现无线自组网络通信拒止的干扰需求,构建无线自组网络模型,并针对该模型提出一种基于强化学习理论的未知拓扑网络多节点干扰策略选择算法,以实时交互的方式进行在线学习. 该算法无需获悉网络拓扑等先验知识,仅以网络流数目作为反馈信息,以多节点联合干扰的方式逐步学习最佳干扰节点. 在不同参数的无线自组网中的仿真结果表明,所提算法在累积阻断网络流方面优于现有算法,且在新的奖赏标准下,所提算法仍具有优异的干扰性能.

关键词: 无线自组网络；强化学习；拓扑网络；干扰决策；奖赏标准

中图分类号: TP972 **文献标志码:** A

Multi-nodes jamming strategy in wireless Ad hoc network based on reinforcement learning

ZHUANSUN Shao-shuai†, YANG Jun-an, LIU Hui, HUANG Ke-ju

(1. College of Electronic Countermeasure, National University of Defense Technology, Hefei 230037, China; 2. Key Laboratory of Electronic Restriction, Hefei 230037, China)

Abstract: An unknown topology network interdiction strategy based on reinforcement learning is proposed. When the model of the wireless Ad hoc network is established, the proposed interdiction strategy could fulfill the needs of interdicting information transmission by jamming multi nodes and enable the jammer to interdict communication in the underlying network in real time manner. By jamming multi nodes as an operation style and counting stopped network flows as action feedback, the proposed strategy could learn the better nodes to jam without a priori knowledge of the network topology. Simulation results on the established wireless Ad hoc network with various parameters show that, the proposed interdiction strategy has a better performance in accumulate stopped flows than existing algorithms, and still has excellent jamming capability under the proposed new reward standard.

Keywords: wireless Ad hoc；reinforcement learning；topology network；decision making；reward standard

0 引 言

伴随着无线网络去中心化的发展趋势,无线自组网[1-2]引起人们的广泛关注. 自组网技术来源于军事通信协同作战需求,随着世界各国军队网络中心战的转型,自组网技术被军队日渐重视,并应用于军事通信的许多领域,如战术互联网(Tactical internet, TI)、士兵电台波形(Soldier radio waveform, SRW)、宽带组网波形(Wideband networking waveform, WNW)、卫星自组网、传感器自组网络等. 自组网技术使战场通信突破了地域的局限性,提高了网络连通的成功率、抗毁性和抗干扰能力. 然而,无线网络媒介的开放广播特性仍使得对网络进行干扰具有一定的可行性,特别是针对敌方战场无线自组网络进行干扰,阻断战场上通信设备之间互连互通进而实现通信拒止,为我方掌握战场信息的主导权提供技术支撑.

当前针对无线自组网络干扰的研究主要集中在两方面: 1) 利用先验知识对网络进行干扰; 2) 利用强化学习理论学习最优干扰策略. 一方面,当干扰方通过某些途径获取一些先验知识时,可利用这些先验知识有针对性地采取干扰策略. 例如,在获悉网络协议的前提下, Amuru等[3]以试错的方式干扰数据包的不同帧结构,并通过仿真得出结论,干扰CTS帧为最

收稿日期: 2017-03-30；**修回日期:** 2017-05-27.

基金项目: 安徽省自然科学基金项目(1308085QF99, 1408085MKL46).

责任编委: 侯忠生.

作者简介: 颛孙少帅(1990−), 博士生, 从事认知干扰和强化学习的研究；杨俊安(1965−), 教授, 从事信号处理和智能计算等研究.

†通讯作者. E-mail: zhuansunss@sina.com

佳干扰策略; Moon等[4]和Yi等[5]针对无线自组网络分别提出了两种不同的洪泛攻击方法,该方法能够瘫痪网络的大部分节点,进而使得网络丧失通信功能; Proano等[6]对采用TCP协议的无线自组网络提出了一种假冒攻击方法,根据掌握的信息从网络协议不同层进行攻击. 然而,基于协议的干扰策略需要已知网络协议为前提,且算法往往针对协议的漏洞进行攻击,当协议日趋完善时,某些攻击方式将不再奏效. 在已知网络拓扑结构的前提下, Sefair等[7]提出了干扰最短路径,使得通信路径最大化,进而阻断通信的方法; Altner等[8]对最大流网络干扰问题进行了理论分析和仿真验证,该方法同样实现了对无线网络的有效干扰,但是无线网络的拓扑结构在战场环境下难以获得,并且拓扑结构是动态调整的,这进一步增加了算法操作的困难性. 另一方面,强化学习作为一种在线的、无需先验知识的机器学习方法,在一些先验知识匮乏或人为参与受限的领域广受关注,如机器人控制[9]、自动驾驶、推荐系统、游戏设计[10]、深度学习[11]等. 利用强化学习实时交互的特点, Auer等[12]提出了一种多臂老虎机算法,并就该算法在有限时间内的收敛性能和学习能力进行了理论论证; Amuru等[13]提出了一种针对未知拓扑网络单个节点进行干扰的强化学习方法,在BA无标度网络、星形网络、ER网络以及PPP (Poisson point process)网络中均取得了优异的干扰效果; Yi等[14]提出了一种针对网络优化问题的解决方法,将目标问题建模为多臂老虎机模型,所提方法对模型中的各种可行动作联合操作,但要求明确知道每个动作对应的奖赏信息. 尽管基于强化学习的算法无需网络协议、拓扑结构等先验知识,但算法要求有明确的奖赏信息,恰当合适的奖赏依据是算法收敛的必要保证.

本文研究无线自组网络多节点干扰问题,提出一种基于强化学习的未知拓扑网络干扰策略选择算法以及一种新的奖赏标准. 算法利用强化学习环境反馈的特点指导干扰节点的选择,对选择的节点以联合干扰的方式作用于环境,通过与环境反复交互的方式逐渐逼近最佳干扰节点. 为了克服当前奖赏依据需要获悉网络协议的缺点,提出一种以网络节点活跃性变化为依据的新的奖赏标准,该标准无需干扰方获悉网络协议,使得现有算法在干扰未知协议的网络时同样适用. 针对无线自组网络的组网特点构造PPP网络,并在该网络上进行算法性能的验证实验,实验结果表明:对于不同参数下的PPP网络进行干扰,本文提出的算法与当前算法相比具有更大的累积奖赏,即干扰性能更优,且提出的新的奖赏标准经实验验证是可行的,能够作为环境对干扰行为的反馈信息.

1 无线自组网络

1.1 无线自组网模型

无线自组网定义为一种特殊的自组织、对等式、多跳、无线移动网络,包括移动自组网络(Mobile Ad hoc network, MANET)和传感器自组网络(Sensor Ad hoc network)两种. 其中,移动自组网络是由移动的节点通过分布式协议自组织起来的一种无线网络,可用于构建战术互联网,或用于已有军用网络,以提高网络的可靠性和生存性. 目前美军典型的战术自组网产品有: JNN-TE波形、无人机自组网、卫星自组网等. 无线自组网络具有多种组网方式,平面网络结构是最基本的网络结构,本文将其作为研究的重点.

本文根据无线自组网络平面网络结构的特点构造PPP网络. PPP用以描述一定区域内节点个数服从泊松分布的网络,确定节点个数后,所有节点以均匀分布的形式分布在该区域,当该区域足够大以至于两个通信设备之间受制于功率、协议等因素无法直接通信时,节点可根据分布式协议与周围最近邻的若干个节点相连接.

利用图论[15]的方法对无线自组网络进行表示,以拓扑图中的节点表征网络中的通信设备,以连接节点的边表示设备之间存在通信,用“连接度(connection degree)”的概念来衡量一个节点与周围节点的连接数,用两节点间的最短路径——流(flow)来模拟信息传递的过程. 在对网络进行干扰时,可通过干扰节点的方法阻断网络中传输的数据流,而节点的选择问题属于网络节点性能分析范畴,网络节点重要性分析的指标有:点度中心性(Degree centrality)、中介中心性 (Betweenness centrality)、网络流中心性(Flow centrality)、接近中心性(Closeness centrality)、特征向量中心性(Eigenvector centrality)等. 在计算节点的上述指标时,需要明确知道网络的拓扑结构,而度量指标中介中心性被广泛用于网络攻击研究和节点重要性度量. 根据Brandes等[16]的定义,中介中心性是指经过节点v的所有最短路径的数量在网络全部最短路径总量中占的比例. 换言之,该指标用于衡量一个节点在其他任意两节点最短路径上扮演的桥梁角色,如果许多节点对的最短路径经过该节点,则认为该节点的中介中心性较大,反之较小. 节点中介中心性的计算方式如下所示:

$$C_B(v)=\frac{1}{(|V|-1)(|V|-2)}\sum_{s\neq v\neq d\in V}\frac{\sigma_{sd}(v)}{\sigma_{sd}}. \quad (1)$$

其中: $\sigma_{sd}(v)$ 表示节点 s 与 d 之间经过节点 v 的最短路径数, σ_{sd} 表示节点 s 与 d 之间的最短路径数.

1.2　无线自组网干扰分析

对网络中介中心性最大的节点进行干扰的策略称之为Betweenness centrality attacker. 此外, 如果一个全能干扰方在获悉网络拓扑结构的基础上, 还明确知道网络流流经的路径, 则其可以有针对性地对某些重要性节点施加干扰, 以实现最佳网络干扰, 这种干扰策略称之为Omniscient attacker. 该策略同Betweenness centrality attacker均利用网络的先验知识进行攻击, 两者可作为已知先验知识下的基准策略用于衡量其他干扰策略的性能.

当未知网络拓扑信息时, 强化学习理论为干扰方提供了一种在线、交互的干扰方法, 主流算法包括Upper confidence bound learning (UCB learning)算法、Slotted exploit explore learning (SEE learning)算法、Contextual bandit learning算法. 考虑到前两者干扰性能更为优异, 下面主要就UCB learning算法和SEE learning算法进行简要介绍, 并以此作为未知先验知识下本文所提出算法的比较算法.

UCB learning算法[12]: 该算法在解决多臂老虎机问题时被广泛使用, 当用于网络干扰时, 算法每次干扰一个节点, 同时记录该节点获得的奖赏以及被干扰的次数, 将两者的比值(即平均奖赏)作为该节点的真实奖赏, 然后利用相关公式对节点的重要性进行度量, 进而选择出下一次进行干扰的节点.

SEE learning算法[13]: 顾名思义, 该算法按一定规则交替进行“探索”和“利用”两个阶段, “探索”阶段逐个选择网络节点进行干扰, 记录其奖赏信息和干扰次数, “利用”阶段计算各节点的平均奖赏信息, 选择平均奖赏值最高的节点进行干扰. 随着“探索”、“利用”阶段的进行, 算法逐渐学习到最佳干扰节点.

为衡量各种干扰算法的优劣, 可从以下两个方面进行比较: 1) 累积阻断网络流数. 每次选择节点进行干扰时, 必然会对网络中数据流的传输造成影响, 随着干扰的进行, 阻断的网络流数目越多, 算法的干扰性能越好. 2) 算法的鲁棒性. 无线自组网络的规模是上下浮动的, 并且其拓扑结构、网络流数也是动态变化的, 鲁棒性强意味着对不同参数下的无线自组网络均具有优异的干扰性能.

2　未知拓扑网络多节点干扰策略选择算法

2.1　无线自组网络认知干扰

对未知网络协议和拓扑结构的无线自组网络进行干扰, 可利用强化学习无需先验知识的优势逐步学习最佳干扰策略. 强化学习作为一种自学习和在线学习方法, 以试错的方式与动态环境进行持续交互, 交互过程中以环境反馈作为对前一个干扰动作的奖赏, 并根据该奖赏信息指导后一个干扰行为的选择. 图1给出了无线自组网络认知干扰系统的主要构成部分.

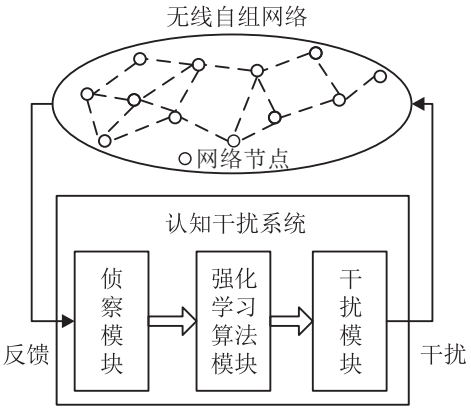


图 1　无线自组网络认知干扰系统

图1中: 强化学习算法模块根据奖赏信息对网络节点进行评估, 并挑选出待干扰的节点由干扰模块采取压制干扰等方式进行攻击; 侦察模块用于对干扰目标进行侦察, 从环境反馈信息中根据奖赏标准确定干扰行为的奖赏信息, 并将该信息传递给算法模块, 用于该模块计算挑选出下一步需要干扰的节点. 明确各个模块的功能后, 可以发现算法模块所采用的学习算法以及侦察模块获取的奖赏信息共同决定着哪些节点需要干扰, 两者属于认知干扰系统的核心部分, 选择恰当的奖赏标准以及构造性能优异的强化学习算法对系统的干扰性能至关重要.

2.2　奖赏标准的选择

强化学习算法的学习动力来源于外界环境的奖赏, 准确、恰当的奖赏信息便于算法迅速收敛到最佳策略. 本文对未知拓扑无线自组网络中的多个节点施加干扰时, 以阻断的网络流数作为干扰动作的奖赏依据, 因此需要干扰方明确知道网络中流的总数, 以及干扰后网络中流的数量, 该信息可通过对网络侦察获得. 例如, 美军战术互联网以TCP/IP协议作为网络通信协议, 目的节点需要发送ACK/NACK帧信息表明是否收到信息, 而该信息为非加密状态, 通过对协议分析后很容易获得. 然而, 当干扰对象为采用其他秘密协议的军用无线自组网络时, 由于无法获悉网络采用的协议信息, 也就无法统计干扰行为阻断的网络流数. 因此, 在未知网络协议和拓扑结构的情况下, 选择能够侦察到的信息作为奖赏依据对执行算法至关重要.

本文提出一种新的奖赏标准,以干扰过后活跃节点的减少量作为反馈信息. 原因在于,当网络流

$$\text{flow} = \{\text{node3} \to \text{node24} \to \text{node17} \to \text{node42}\}$$

中的一个节点node24被干扰时,该节点必须要求上一个节点node3重发信息,同时不再向下一个节点node17发送信息,这样做会导致该节点之后的所有节点node17、node42不再活跃,即活跃节点数由4变为2,网络中被阻断的流越多,由活跃转为静默的节点也相应增多,以此作为对干扰动作的奖赏具有一定的合理性. 此外,该奖赏标准无需获悉目标网络的通信协议,仅需要干扰方通过一定的侦察行为判断网络中节点的活跃性即可.

2.3 Improved CUCB算法

Yi等[14]提出了一种CUCB (Combinatorial upper confidence bound)算法,该算法适用于解决某些优化问题. 在算法执行阶段,学习系统能够对各种不同组合搭配进行尝试,但要求明确知道每一个被选中元素对应的奖赏. 选择网络中的若干个节点进行干扰,达到最佳干扰便属于组合优化问题. 然而,在未知网络拓扑结构的情况下,即便知道干扰过后网络中流数目的减少量,由于被干扰的节点之间并非相互独立,仍无法将该减少量与被选中节点所做的贡献相联系,该原因使得CUCB算法不适用于未知拓扑网络的干扰任务,但其组合搭配的想法值得本文借鉴.

以无线自组网络中恒定流情况为例,若某个节点属于多条流流经的节点,则对该节点进行干扰,能够尽可能地阻断更多的流,即该节点具有更优的干扰性能. 因此,本文提出Improved CUCB算法. 该算法在执行的初始阶段并非每次选择多个节点,而是逐个节点进行干扰,将阻断流的数目作为对该节点干扰的奖赏,并且该奖赏完全属于被干扰的节点;在主循环阶段,对满足下式条件的多个节点联合干扰,将奖赏值直接赋值给被干扰的节点:

$$\text{arm} = \underset{\text{arm}\in F}{\text{argmax}} \sum_{i\in V} \left(\hat{\theta}_i + \sqrt{\frac{(L+1)\ln n}{m_i}}\right). \quad (2)$$

Improved CUCB算法的执行步骤如算法1所示. 其中: NumPoint表示网络中节点总数, Reward表示节点获得的累积奖赏, Playcount表示截止当前时刻每个节点被干扰的次数, arm表示每次被干扰的节点集合, $\hat{\theta}_i$表示i节点获得的平均奖赏, m_i表示截止当前时刻i节点被干扰的次数, F表示所有可能的干扰动作集合.

算法1 Improved CUCB算法.

1) //初始化,以恒定流为例.

$L = \text{NumPoint}$;

$\text{Reward} = [0, 0, \cdots, 0]_{1\times\text{NumPoint}}$;

2) $\text{Playcount} = [0, 0, \cdots, 0]_{1\times\text{NumPoint}}$;

3) **for** $p = 1 : \text{NumPoint}$

4) $\text{arm} = p$;

5) 执行动作arm,并更新Reward和Playcount;

6) **end for**

7) //主循环

8) **while** $n <= \text{TotalTime}$

9) 执行能够使下面公式最大化的动作arm

10) $\text{arm} = \underset{\text{arm}\in F}{\text{argmax}} \sum_{i\in V} \left(\hat{\theta}_i + \sqrt{\frac{(L+1)\ln n}{m_i}}\right)$;

11) 相应更新Reward和Playcount;

12) $n = n + 1$;

13) **end while**

对于随机流的情况,由于网络中每时刻流的数目和流经的路径都在变化,这时只需要对程序的初始化阶段稍加改动,将步骤4)中由一次选择一个节点修改为一次随机选择N个节点,N是干扰方一次能够干扰的最大节点数. 步骤5)在执行arm后需要更新Reward以及Playcount信息,由于无法衡量每个被干扰节点在此次干扰任务中的贡献,奖赏值无法按比例划分,本文将奖赏值直接赋值给每个节点,并相应地更新Playcount信息.

下面就算法的学习性能进行简要分析,假定最佳干扰动作为arm^*,奖赏信息为$r(\text{arm}^*)$,第n次干扰选择的动作为arm_n,对应奖赏信息$r(\text{arm}_n)$. 每次干扰未选择最优干扰的概率为

$$\begin{aligned}
&P[r(\text{arm}_n) > r(\text{arm}^*)] = \\
&1 - P[r(\text{arm}_n) < r(\text{arm}^*)] < \\
&1 - P[r(\text{arm}^*) > \mu(\text{arm}^*) - \Delta/2]\cdot \\
&P[r(\text{arm}_n) < \mu(\text{arm}_n) + \Delta/2] \leqslant \\
&P[r(\text{arm}^*) \leqslant \mu(\text{arm}^*) - \Delta/2] + \\
&P[r(\text{arm}_n) \leqslant \mu(\text{arm}_n) + \Delta/2], \quad (3)
\end{aligned}$$

其中$\Delta = \mu(\text{arm}^*) - \mu(\text{arm})$, $\mu(\text{arm}_n)$表示干扰动作arm_n对应的平均奖赏.

由Chernoff-Hoeffding边界定理可知

$$\begin{cases}
P[r(\text{arm}^*) \leqslant \mu(\text{arm}^*) - \Delta/2] \leqslant \exp(-T_1\Delta_{\min}^2/2), \\
P[r(\text{arm}_n) \leqslant \mu(\text{arm}_n) + \Delta/2] \leqslant \exp(-T_1\Delta_{\min}^2/2).
\end{cases} \quad (4)$$

其中: T_1表示干扰动作arm_n累积被执行的次数, $\Delta_{\min} = \underset{\text{arm}\neq\text{arm}^*}{\min} \Delta$.

结合式(3)和(4)可知,未选择最佳干扰概率的上

边界为

$$P[r(\mathrm{arm}_n) > r(\mathrm{arm}^*)] \leqslant$$
$$2\exp(-T_1\Delta_{\min}^2/2)(C_V^{\mathrm{arm}}-1) <$$
$$2\exp(-T_1\Delta_{\min}^2/2)C_V^{\mathrm{arm}}, \tag{5}$$

C_V^{arm}为非最佳干扰动作的个数. 因此,算法的学习性能Reward的上下边界可以确定为

$$\begin{cases} \mathrm{Reward} > 2\exp(-T_3\Delta_{\min}^2/2)C_V^{\mathrm{arm}}T\Delta_{\min}, \\ \mathrm{Reward} < 2\exp(-T_2\Delta_{\min}^2/2)C_V^{\mathrm{arm}}T\Delta_{\max}. \end{cases} \tag{6}$$

其中

$$\Delta_{\max} = \max_{\mathrm{arm}\neq\mathrm{arm}^*}\Delta,$$
$$T_2 = \min(T_1),$$
$$T_3 = \max(T_2),$$

T表示总的干扰次数.

3　仿真实验

根据无线自组网络的特点,本文将构建的PPP网络作为实验对象,将Joint SEE算法、Omniscient attacker、Betweenness centrality attacker、Independent SEE算法、Independent UCB算法作为比较算法,以累积奖赏和算法鲁棒性作为衡量干扰性能的标准,将200次实验后的数据取均值作为最终结果. 需要注意的是,SEE算法、UCB算法仅适用于干扰一个节点的情况,当需要对N个节点同时进行干扰时,本文假设由N个干扰方分别采取独立(Independent)、联合(Joint)两种方式进行干扰. Omniscient attacker、Betweenness centrality attacker以及Improved CUCB算法适用于单个干扰方一次干扰多个节点的情况,不存在联合的情况. 下面分别从初始化阶段对算法性能的影响,对静态拓扑和动态拓扑PPP网络干扰,干扰不同参数下的PPP网络以及采用新的奖赏标准4个方面对上述算法的性能进行比较.

3.1　初始化阶段对算法性能的影响

在无线自组网络中,就单个节点而言,均具有相似的配置和功能,但从网络的角度来看,每个节点因其地理位置和周围节点分布的不同,在网络中起到的作用有着显著差距. 当已知网络拓扑结构和网络流流经路径时,通过分析能够确定重要性节点,对其干扰能够阻断最大数目的网络流. 若无法获悉网络的拓扑结构信息,便无法判断网络中节点的重要性,一个可行的方法是:利用算法的初始化阶段,逐个干扰网络中的节点,目的在于对每个节点有一个基本度量,便于主循环阶段进一步度量节点重要性. 图2给出了有初始化阶段和无初始化阶段的Improved CUCB算法在累积奖赏方面的比较.

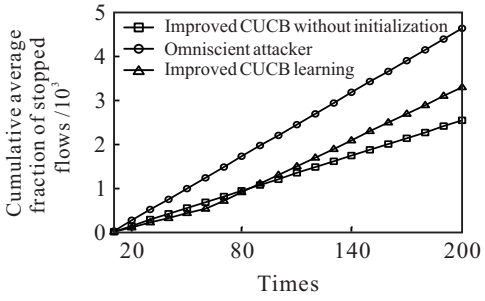


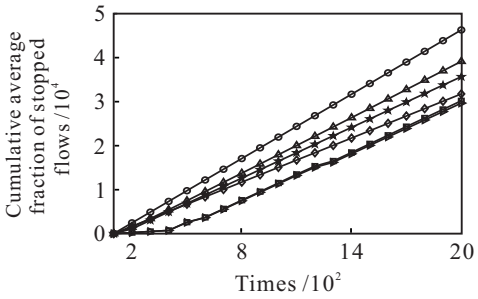
图2　初始化阶段对干扰性能的影响

从图2可以看出,包含初始化阶段的Improved CUCB算法在累积奖赏方面要优于无初始化阶段的情况,即前者累积阻断更多的网络流,干扰效果更优. 同最佳干扰相比,Improved CUCB算法累积奖赏曲线的斜率较小,这意味着该算法并未学习到最优干扰策略. 主要原因是:在文献[14]中,元素间相互独立且明确知道每个元素对应的奖赏,只需选择奖赏值最大的若干元素便能够学习到最优策略. 然而,在无线自组网络多节点干扰任务中,节点是相关的且无法知道每个节点对应的奖赏,原有的算法不再适用,改进后的算法适用于新任务,但代价是无法学习到最佳干扰策略.

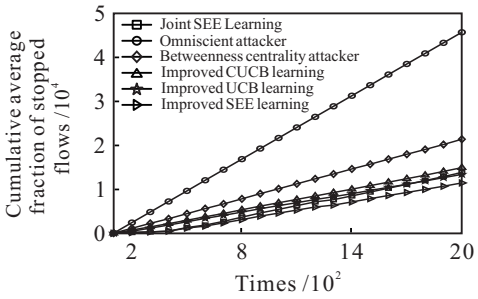
3.2　对静态拓扑和动态拓扑PPP网络干扰

静态拓扑恒定流和动态拓扑随机流属于复杂网络流情况中的两个特例,前者是指网络的拓扑结构未发生变化,且网络中的节点坚持按照既定路线传递信息,不因干扰的存在而改变传输路线. 后者是指网络的拓扑结构和网络流的路径在每次干扰后都发生变化,体现了网络动态变化的特点. 在构造静态拓扑恒定流PPP网络时,将网络流流数设置为30条,每条流的源节点和目的节点随机选择后固定不变,干扰节点数为5个节点. 构造的PPP网络参数为:总节点数为50个,每个节点与最近邻的8个节点相连接;在干扰动态拓扑随机流PPP网络时,假定每次干扰时间段内拓扑结构随机变化,网络流数介于[10, 30],每条流的源节点和目的节点随机选取,单个干扰方能同时干扰节点数为5个,网络总节点数依然为50个,每个节点与最近邻的8个节点相连接. 不同算法的干扰性能如图3所示.

图3(a)给出了干扰静态拓扑恒定流PPP网络不同算法的累积奖赏曲线. 由于Omniscient attacker每次干扰都选择最佳节点进行干扰,其累积奖赏为一条直线,且较其他算法具有最大的累积奖赏值. Betweenness centrality attacker在初始阶段具有较高的累积奖赏值,但随着其他算法不断学习,致使



(a) 静态拓扑恒定流



(b) 动态拓扑随机流

图 3 对PPP网络不同状态网络流进行干扰

其累积奖赏在后期阶段落后于其他算法,表明仅仅对网络的节点进行重要性分析是不够的,需要将网络流信息综合考虑;Improved CUCB算法具有比Independent UCB算法更高的累积奖赏值,仅次于Omniscient attacker. 此外,该算法累积奖赏曲线的斜率仅次于最佳干扰,表明算法学习收敛的结果并非最优结果,性能仍有一定的提升空间. Joint SEE算法和Independent SEE算法拥有相似的累积奖赏值曲线,只不过由于联合学习的原因,使得前者的累积奖赏值更高,在累积奖赏曲线中,两者曲线每隔一定干扰次数会出现增长放缓的现象,原因在于SEE算法交替处于“探索”、“利用”阶段,而“探索”阶段干扰性能有所下降致使增长缓慢. 此外,尽管Joint SEE算法也属于联合干扰范畴,但其本质仍是选取平均奖赏最大的节点,忽略了节点之间的相关性,因此其累积奖赏比Improved CUCB算法要小.

图3(b)给出了动态拓扑随机流PPP网络不同算法干扰性能之间的比较. Omniscient attacker依然具有最大的累积奖赏值,由于Betweenness centrality attacker选择中介中心性更大的节点进行攻击,且网络拓扑和网络流是随机的,使得该攻击策略较Independent UCB算法、Joint SEE算法、Independent SEE算法和Improved CUCB算法具有更大的累积奖赏,即在完全随机的情况下利用网络的拓扑结构信息进行干扰,效果表现优异. 将几种学习型干扰方法相比较可以发现,Improved CUCB算法的累积奖赏曲线是最高的,表明该算法学习性能出众. 此外,Improved CUCB算法与Joint SEE算法的“利用”阶段具有类似的累积奖赏增长速率,只不过从整体累积奖赏的角度来看,前者在开始干扰后的较长一段时间内具有更大的累积奖赏值,即Improved CUCB算法的整体干扰效果更好.

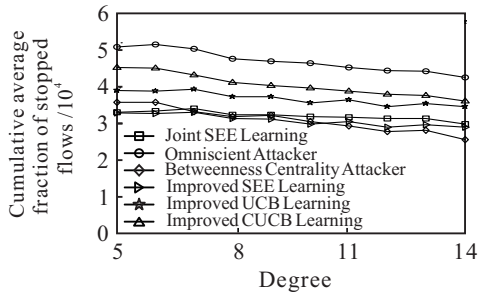
对比图3(a)和图3(b)可以发现:静态拓扑恒定流和动态拓扑随机流两种情况下算法的学习能力有着显著差别,干扰静态拓扑恒定流网络时,网络中节点的重要性程度是不变的,上述几种学习算法能够随着干扰次数的增多逐渐学习到更加重要的干扰节点;干扰动态拓扑随机流网络时,重要性程度较大的节点时刻在变化,而学习算法具有一定的延迟性,无法短时间内学习,因此上述算法的累积奖赏与最佳干扰的累积奖赏相差较大. 需要补充说明的是,由于动态拓扑网络中的流数预置为[10, 30],其最优干扰的累积奖赏比静态拓扑网络中最优干扰的累积奖赏要少.

3.3 对不同参数下的PPP网络干扰

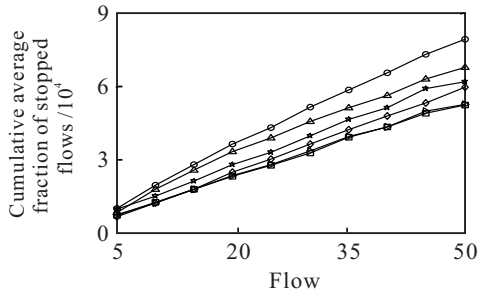
不同的节点总数、节点连接度构成不同拓扑结构的PPP网络,不同的网络流条数和干扰节点数与算法的干扰性能息息相关. 为了进一步验证Improved CUCB算法对PPP网络干扰的鲁棒性,从上述4个方面构造不同干扰环境下的PPP网络,其中网络流设置为恒定情况,网络拓扑设定为静态,不同参数下的PPP网络干扰结果如图4所示.

图4(a)给出了不同节点连接度条件下,不同算法2 000次干扰后的累积奖赏曲线. 可以看出,Improved CUCB算法的累积奖赏仅次于最佳干扰,始终高于其他几种对比算法. 另外,之所以对节点连接度介于[5, 14]的情况进行讨论,主要原因在于当连接度过低时,该网络往往会形成几个独立的子网络,使得某些节点之间的通信完全被阻断,仅当节点连接度较大时,任意两节点能够互联互通,并且随着连接度的增大,两个节点可以通过更少的“跳”数进行通信,更不易因节点被干扰而阻断通信,这也恰好解释了为什么随着连接度的增大,各算法的累积奖赏曲线呈下降趋势.

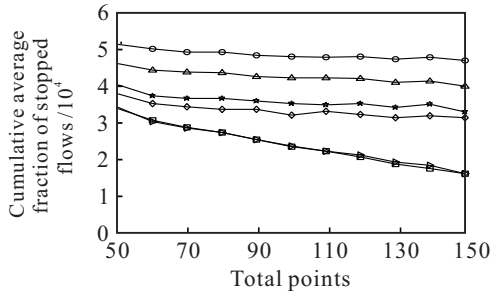
当PPP网络拓扑结构保持不变时,图4(b)给出了几种算法干扰不同数目网络流的性能比较. 可以看出,随着网络中流数的增加,每种算法对应的累积奖赏也在增加,且Improved CUCB算法依然保持着较大的累积奖赏,即算法具有更优的干扰性能. 随着网络流数的不断增多,几种算法与最佳干扰在累积奖赏方面的差值越来越大,说明了当网络中流数过多时,算法的学习能力有所下降.



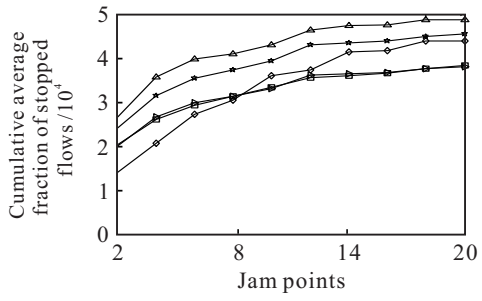
(a) 不同节点连接度



(b) 不同网络流数目



(c) 不同节点总数



(d) 不同干扰节点数

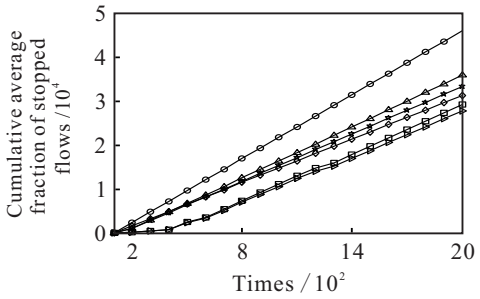
图 4　干扰不同参数的PPP网络

可以看出,图4(c)给出了变化的节点总数对算法性能的影响,随着数目的增加,几种算法的累积奖赏值均有所下降,特别是Joint SEE算法和Independent SEE算法,下降速度较其他算法更快. 累积奖赏曲线整体呈下降趋势的原因在于,当网络中节点总数增加而网络流数目保持不变时,网络流会更加分散而不易交汇,对于同样是5节点干扰的情况而言,所能阻断的网络流数将有所下降,表现为累积奖赏值呈降低趋势.

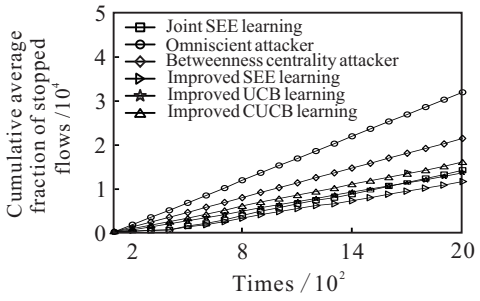
当干扰对象稳定,即PPP网络的节点总数、节点连接度和网络流数目保持不变时,图4(d)给出了干扰不同节点数时各个算法的累积奖赏曲线. 可以看出,随着干扰节点数的增加,网络中更多的流能够被阻断,使得各个算法的累积奖赏曲线整体呈上升趋势. 当干扰节点数超过14个时,所有算法的累积奖赏曲线不再有明显增长,主要原因在于网络流数预设为30条,而2 000次干扰最多能够阻断6×10^4条网络流,干扰14个节点已经能够阻断大部分的网络流. 如果想要进一步提升累积奖赏,需要干扰更多的节点.

3.4　新的奖赏标准下干扰性能比较

同样以3.2节中构造的PPP网络作为干扰对象,从静态拓扑恒定流和动态拓扑随机流两种网络状态对算法的干扰性能进行讨论,所有需要反馈信息参与的算法均采用新的奖赏标准,即以干扰过后活跃节点的减少量衡量干扰效果的优劣, 实验条件(如网络参数、干扰参数等)均与3.2节设置一致,这里不再赘述,各算法的干扰性能如图5所示.



(a) 静态拓扑恒定流



(b) 动态拓扑随机流

图 5　新的奖赏标准下算法的干扰性能

从图5可以看出,Improved CUCB算法对静态拓扑恒定流网络和动态拓扑随机流网络的干扰性能均优于对比算法. 若从累积奖赏曲线的斜率来看,Improved CUCB算法与最佳干扰之间仍具有较大的差距,甚至比图5(a)中的差距还要大,原因在于采用了新的奖赏标准后,对干扰动作的奖赏不够精确具体,导致算法的学习结果弱于原有奖赏标准. 尽管如此,考虑到新的奖赏标准更加容易获得,采用新的奖赏标准后能够对未知通信协议的无线自组网络进行干扰,即在适用领域方面要优于原有奖赏标准.

从以上仿真结果可以看出,以不同参数下的PPP网络作为干扰模型,将静态拓扑恒定流和动态拓扑随

机流作为实验条件,把累积奖赏和算法鲁棒性作为比较标准,本文所提的Improved CUCB算法较当前无线网络多节点干扰算法具有更优的干扰性能. 此外,分析了在不同节点总数、不同节点连接度、不同网络流数、不同干扰节点数等参数下各个算法的累积奖赏曲线,比较了采用新的奖赏标准后算法的干扰性能,进一步验证了所提Improved CUCB算法在不同参数下的PPP网络中均具有更强的鲁棒性.

4 结 论

本文以干扰无线自组网络中的多个节点作为研究对象,在现有算法基础上提出了一种改进的CUCB算法——Improved CUCB算法. 该算法以联合的方式对节点进行干扰,克服了单节点干扰性能差的缺点,同时比联合算法(如Joint SEE算法)具有更优的干扰性能. 文中根据无线自组网络的特点构造PPP网络,并从静态拓扑恒定网络流和动态拓扑随机网络流,以及不同网络参数下的干扰效果验证所提算法的干扰性能和鲁棒性能. 仿真结果表明,本文所提算法在累积奖赏方面均优于现有算法,且算法具有较强的鲁棒性,对不同状态的PPP网络均具有优异的干扰效果. 所提出的新的奖赏标准尽管在学习能力方面弱于原有标准,但因无需获悉网络协议,适用领域更加广泛.

对无线自组网络中的多个节点进行干扰有利于阻断网络通信,实现更佳的干扰效果,有助于达成既定的干扰任务. 今后的工作将主要围绕如何使多节点干扰学习的学习效果收敛至最优干扰,以及如何利用干扰方获得的先验知识进一步加快算法的学习速度,使得算法更加快速高效,向实用性进一步靠拢.

参考文献(References)

[1] 刘颖, 唐艺玮, 邵小桃, 等. 基于无线多跳自组网的混合路由研究[J]. 兵工学报, 2017, 38(1): 184-189.
(Liu Y, Tang Y W, Shao X T, et al. Hybrid routing protocol for wireless multi-hop Ad-hoc networks[J]. Acta Armamentarii, 2017, 38(1): 184-189.)

[2] Venkatesan K G S, Khanaa V. Reliable communication in MANET to communicate in Ad-hoc network[J]. Int J of Pharmacy & Technology, 2016, 8(3): 17770-17774.

[3] Amuru S, Buehrer R M. Optimal jamming using delayed learning[C]. Proc of the IEEE Military Communications Conf. Baltimore: IEEE, 2014: 1528-1533.

[4] Moon A H, Iqbal U, Bashir A, et al. Simulating and analyzing RREQ flooding attack in wireless sensor networks[C]. 2016 Int Conf on Electrical, Electronics, and Optimization Techniques(ICEEOT). Xiamen: DEStech, 2016: 3374-3377.

[5] Yi Ping, Zou Futai, Zou Yan. Performance analysis of mobile ad hoc networks under flooding attacks[J]. J of Systems Engineering and Electronics, 2011, 22(2): 334-449.

[6] Proano A, Lazos L. Selective jamming attacks in wireless network[C]. 2010 IEEE Int Conf on Communication. Cape Town: IEEE, 2010: 1-6.

[7] Sefair J A, Smith J C. Dynamic shortest-path interdiction[J]. Networks, 2016, 68(4): 315-330.

[8] Altner D S, Ergun O, Uhan N A. The maximum flow network interdiction problem: Valid inequalities, integrality gaps, and approximability[J]. Operations Research Letters, 2010, 38(1): 33-38.

[9] 张国良, 杜柏阳, 孙一杰, 等. 基于预测控制的时滞多机器人编队脉冲控制[J]. 控制与决策, 2016, 31(8): 1453-1460.
(Zhang G L, Du B Y, Sun Y J, et al. Impulsive control for multi-robot formation with communication delay based on predictive control[J]. Contol and Decision, 2016, 31(8): 1453-1460.)

[10] 陈兴国, 俞扬. 强化学习及其在电脑围棋中的应用[J]. 自动化学报, 2016, 42(5): 685-695.
(Chen X G, Yu Y. Reinforcement learning and its application to the game of Go[J]. Acta Automatica Sinica, 2016, 42(5): 685-695.)

[11] 赵冬斌, 邵坤, 朱圆恒, 等. 深度强化学习综述: 兼论计算机围棋的发展[J]. 控制理论与应用, 2016, 33(6): 701-717.
(Zhao D B, Shao K, Zhu Y H, et al. Review of deep reinforcement learning and discussion on the development of computer Go[J]. Control Theory & Applications, 2016, 33(6): 701-717.)

[12] Auer P, Cesa-Bianchi N, Fischer P. Finite-time analysis of the multiarmed bandit problem[J]. Machine Learning, 2002, 47: 2-13.

[13] Amuru S, Buehrer R M, Schaar M V D. Blind network interdiction strategies — A learning approach[J]. IEEE Trans on Cognitive Communications and Networking, 2016, 1(4): 435-449.

[14] Yi Gai, Bhaskar Krishnamachari, Rahul Jain. Combinatorial network optimization with unknown variables: Multi-armed bandits with linear reward[J]. IEEE/ACM Trans on Networking, 2012, 20(5): 1466-1478.

[15] 李超, 彭力, 赵龙. 基于图论的无线传感器网络自组织性研究[J]. 计算机工程与科学, 2010, 31(1): 25-28.
(Li C, Peng L, Zhao L. Research on the self organization performance of wireless sensor networks based on graph theory[J]. Computer Engineering & Science, 2010, 31(1): 25-28.)

[16] Brandes U, Borgatti S P, Freeman L C. Maintain the duality of closeness and betweenness centrality[J]. Social Networks, 2016, 44: 153-159.

(责任编辑: 齐 霁)