

DOI: 10.7652/xjtuxb201806014

未知拓扑无线自组网络多节点干扰决策算法

颢孙少帅^{1,2}, 杨俊安^{1,2}, 刘辉^{1,2}, 黄科举^{1,2}

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

摘要: 为满足战场环境下无线自组网络通信拒止的干扰需求,提出了一种未知拓扑无线自组网络多节点干扰决策算法(CUCB)。首先,根据战场无线自组网络结构特点构建泊松点过程(PPP)网络模型,并利用其模拟网络中数据流传输过程;其次,随机对 PPP 网络中多个节点进行干扰,通过监听确认帧信息或侦察节点活跃度判断阻断网络流数,根据干扰结果构造节点相关性矩阵;最后,利用强化学习与环境实时交互的特点,在干扰过程中不断更新节点相关性矩阵并将其用于后续节点选择。所提算法无需获悉目标网络拓扑结构、节点重要性等先验信息,仅以阻断网络流数目或节点活跃性作为奖赏标准,适用网络类型更为广泛。仿真结果表明,对不同参数下的无线自组网络进行干扰,所提算法在鲁棒性方面优于现有算法,在累积阻断网络流数量方面比联合利用探索算法提高了 27.1%。

关键词: 无线自组网;强化学习;拓扑网络;干扰决策;奖赏标准

中图分类号: TN975 **文献标志码:** A **文章编号:** 0253-987X(2018)06-0091-07

An Algorithm of Multi-Nodes Jamming Decision in Blind Wireless Ad-hoc Networks

ZHUANSUN Shaoshuai^{1,2}, YANG Jun'an^{1,2}, LIU Hui^{1,2}, HUANG Keju^{1,2}

(1. School of Electronic Countermeasure, National University of Defense Technology, Hefei 230037, China;

2. Key Laboratory of Electronic Restriction, Anhui Province, Hefei 230037, China)

Abstract: An algorithm focus on multi-nodes jamming decision is proposed to fulfill the need of interdicting information transmission in wireless Ad-hoc networks. Firstly, a model of Poisson point process (PPP) network is constructed in terms of the structure of wireless Ad-hoc networks, and then it is used to simulate the process of data transmission. Secondly, random interference to multiple nodes are performed, and the number of stopped network flows is counted by monitoring ACK information or reconnoitering nodes activities. A correlation matrix is constructed from jamming effects. Finally, the correlation matrix is continuously updated in jamming process by taking the advantage of interaction in time of reinforcement learning and is used for the selection of subsequent nodes. The proposed algorithm does not need to have a priori knowledge of information such as network topology or importance of nodes, and interaction is only needed in learning nodes correlation matrix which would be helpful when choose nodes to jam. Simulation results by jamming wireless Ad-hoc in different circumstances and a comparison with the joint slotted exploit explore learning show that the proposed interdiction algorithm

收稿日期: 2017-12-03。 作者简介: 颢孙少帅(1990—),男,博士生;杨俊安(通信作者),男,教授,博士生导师。 基金项目: 国家自然科学基金资助项目(11375263)。

网络出版时间: 2018-03-15

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1069.T.20180315.2055.014.html>

increases by 27.1% in accumulate stopped flows, and its robustness is superior to existing algorithms.

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

无线自组网技术具有去中心化、动态组网的特点,将该技术应用于战场通信时,能够突破地域的局限性,提高网络连通的成功率、抗毁性和抗干扰能力^[1-2]。在战场环境下,为实现对目标网络的有效干扰,需要采取一定的手段以实现通信拒止。

当前对无线自组网络的干扰研究可归纳为以下两个方面:①利用先验信息有针对性的对网络进行干扰;②利用强化学习不断试错的方法搜索最佳干扰策略。例如,已知网络协议时,Amuru等人从MAC层角度出发,对数据流中不同的帧结构进行干扰,并指出干扰确认帧/非确认帧(Acknowledge/Not Acknowledge)所需的干扰能量最少但干扰效果最优^[3]。Moon等人利用目标网络协议漏洞提出一种洪泛攻击方法^[4]。刘书建等人对基于Zamp洪泛攻击方法的可行性进行了分析研究^[5]。Proano等人对网络中传输的数据包进行分类处理,然后根据分析结果对其中“重要的”数据包进行干扰^[6]。此外,已知目标网络拓扑结构时,Sefair提出一种干扰网络关键节点的方法,该方法迫使通信方采用更多跳进行通信,甚至无法有效通信^[7]。Borrero在文献[8]中指出,通过观察网络中数据流传输路径,能够获取有关网络拓扑结构的部分信息,进而使用该信息指导干扰决策。尽管上述方法能够取得优异的干扰性能,但是在战场环境下,网络协议或拓扑结构等信息是无法获得的,即研究的前提条件无法满足;另一方面,强化学习无需环境信息的优势使得其在机器人控制^[9]、游戏设计^[10]、深度学习^[11]等领域得到充分利用。Amuru在文献[12]中提出了一种未知拓扑无线自组网络单节点干扰决策算法,通过不断试错的方式找到单个最佳干扰节点。Gai等人文献[13]中提出一种多智能体强化学习方法,但要求在学习过程中明确知道每个智能体对应的奖赏信息。然而在对目标网络进行干扰时,需要对多个节点进行联合干扰才能实现一定程度的通信拒止,并且干扰每个节点对应的奖赏信息也是未知的。

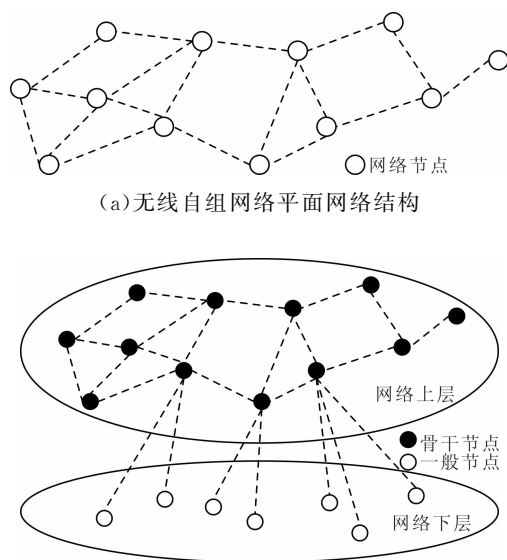
为解决上述问题,本文研究了未知拓扑无线自组网络多节点干扰问题,提出了一种利用节点间相关性的干扰决策算法,算法根据每次干扰后外界反馈信息更新节点相关性矩阵,并利用该矩阵指导后

续干扰节点的选择,通过与环境反复交互的方式不断完善相关性矩阵。此外,将战场无线自组网络建模为泊松点过程(Poisson Point Process, PPP)网络,并在该网络模型上进行算法性能的验证实验,实验结果表明:对目标网络进行多次干扰后,本文提出的算法同现有算法相比能够阻断更多的网络流数,即在干扰性能和鲁棒性能方面更优。

1 无线自组网络

1.1 无线自组网络的概念

无线自组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, SANET)两种。移动自组网技术在美军的主要通信装备中普遍使用,典型的战术自组网产品有:JNN-TE波形、弹药数据链、无人机自组网^[14]、卫星自组网等。无线自组网络具有多种组网方式,如图1所示,按照节点功能及结构层次可分为平面网络结构和分级网络结构。在两种网络结构中,平面网络结构是最基本的网络结构,同时也是分层网络的重要组成部分,在平面网络结构中,所有节点为对等结构,具有完全一致的功能特性,采用自组织协同算法形成网络,故本文将其作为研究的重点。



(b) 无线自组网络分级网络结构

图1 无线自组网络拓扑结构

本文根据该结构的特点构造 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。假设指定区域内节点总数为 50,节点连接度为 5,图 2 为根据该参数构造的 PPP 网络。

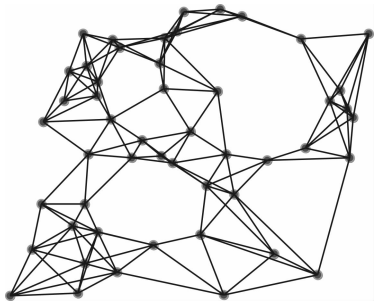


图 2 PPP 网络示例图

1.2 无线自组网络干扰分析

图论理论^[15]常用于无线自组网络分析,在网络拓扑图中,以节点表示通信设备,以边表示设备之间存在通信,用两节点间的最短路径——流(flow)来模拟信息传递的过程。在衡量节点重要性程度时常使用中介中心性进行度量,根据 Freeman 的定义^[16],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,计算方式如下

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

式中: $\sigma_{sd}(v)$ 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。

对网络中介中心性最大的节点进行干扰的策略称之为中介中心性 (Betweenness Centrality Attacker,BCA)算法。此外,将最佳干扰策略简称为 OA(Omniscient Attacker),该策略同中介中心性算法均利用网络的先验信息进行攻击。将强化学习理论用于干扰节点选择,主流算法有置信上界 (Upper Confidence Bound Learning,UCB)算法^[17]、利用探索 (Slotted Exploit Explore Learning,SEE)算法^[12]、上下文情景 (Contextual Bandit Learning)算法,本文将上述算法中性能较优的 UCB 和 SEE 作为比较算法。

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。每次选择若干节点进行干扰时,必然会阻断网络中的若干条网络

流,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 无线自组网络干扰

2.1 节点相关性

在对目标网络中的多个节点进行干扰时,往往能够阻断一定数目的网络流,但由于干扰方不知道该网络拓扑结构,无法对被干扰节点的重要性进行评估,进而在面对信度分配^[18]问题时无从下手。例如,图 3 是一个小型的网络拓扑结构示意图,并假设网络中的数据流如表 1 所示,当同时干扰节点 1 和节点 4 时,共阻断 5 条网络流,两节点分别贡献 1 条和 3 条,并共同贡献 1 条(flow2),只不过对于干扰方而言,在不具备额外信息的情况下,无法衡量节点 1 和节点 4 的重要性,只能将两个节点同等对待,即各自贡献 2.5 条。倘若被干扰的节点是 1 和节点 2,此时共有 3 条网络流被阻断,可认为节点 1 和节点 2 各自贡献 1.5 条。比较上述两种情况可以看出,同时干扰节点 1 和节点 4 能够阻断更多的网络流,即节点 4 与节点 1 之间的相关性更强,在执行网络多节点干扰任务且已经确定干扰节点 1 时,更倾向于选择干扰与该节点相关性更强的节点 4 而非节点 2,定义 1 给出了节点相关性的概念与性质。因此,干扰方如何根据干扰过程中获取的信息,充分挖掘网络节点之间的相关性,对实现目标网络干扰最

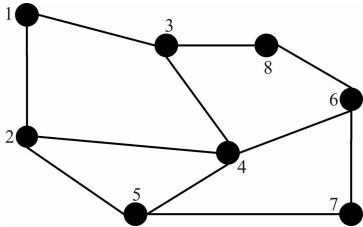


图 3 小型网络拓扑结构示意图

表 1 网络流及流经路径

网络流	流经路径
flow1	{1→3→8→6}
flow2	{1→2→4}
flow3	{2→4→6}
flow4	{3→4→5→7}
flow5	{6→8}
flow6	{3→4→5}

大化至关重要。

定义 1 节点相关性 给定一个无向网络拓扑图 G , 节点 $\text{node } p \in V = \{1, 2, \dots, n\}$, 网络中存在固定数目和路径的网络流, 当同时从 G 中删除任意两个节点 i 和 j 时, $i, j \in V$, 且 $i \neq j$, 网络流数目减少了 d 条, 规定该条件下节点 i 与节点 j 的相关性为 $d/2$ 。

性质 1 节点间的相关性与拓扑图中网络流状态相关, 网络流状态改变时, 节点间的相关性也会相应的变化。

性质 2 多个节点之间也存在相关性, 若同时从 G 中删除节点 i , 节点 j 以及节点 k , $i, j, k \in V$ 且 $i \neq j \neq k$, 网络流数目减少 e 条, 规定此时 3 个节点中任意两个节点间的相关性为 $e/3$ 。

性质 3 当多次从 G 中删除节点 i 和 j 时, 两节点之间的相关性可通过对每次相关性求和取均值获得。

网络节点之间的相关性如下

$$\mathbf{R}(t) = \begin{bmatrix} r_{11}(t) & \cdots & r_{1i}(t) & r_{1j}(t) & \cdots \\ \vdots & & \vdots & \vdots & \\ r_{i1}(t) & \cdots & r_{ii}(t) & r_{ij}(t) & \cdots \\ r_{j1}(t) & \cdots & r_{ji}(t) & r_{jj}(t) & \cdots \\ \vdots & & \vdots & \vdots & \end{bmatrix} \quad (2)$$

在执行干扰任务前, 将 $\mathbf{R}(t)$ 初始化为全 0 矩阵, 元素值随着干扰任务的进行不断更新。假定 $r_{ij}(t)$, $i, j \in V$, $t \in \{1, 2, \dots, T\}$ 表示第 t 次干扰时节点 i 与节点 j 之间的相关性, 对于任意一次干扰, 节点 i 与节点 j 之间的相关性是相同的, 即 $r_{ij}(t) = r_{ji}(t)$, 矩阵内元素的更新方式如下

$$\left. \begin{aligned} r_{ij}[t] &= \frac{N}{N+1} \left[r_{ij}[t-1] + \frac{r_{ij}(t)}{N} \right] \\ t &\in \{2, 3, \dots, T\}, r_{ij}[1] = r_{ij}(1) \end{aligned} \right\} \quad (3)$$

式中: N 表示节点 j 在本次任务以前被选中的次数。

该矩阵的使用方法如下: 已知一个节点 i 被选中, 那么观察该矩阵第 i 行, 将本行中最大元素对应的列的编号(假设为第 j 列)作为下一个被选中的节点, 同时将该列元素值置零, 防止再次被选中; 继续观察该矩阵第 j 行, 将最大元素对应的列的编号作为下一个被选中的节点并清空相应的列, 以此类推, 直至以序列的方式挑选出本次干扰的节点集合, 最后将相关性矩阵恢复至节点选择前状态并用于内部元素的下一次更新。

在节点相关性矩阵中, 每个元素表示两个节点之间的相关性, 如果采取贪婪选择的方式, 直接从相关性矩阵中选取值最大的元素对应的节点作为干扰节点, 无法保证不同节点对之间的相关性最大, 且本

文的创新点在于构造节点相关性矩阵, 并以此作为干扰节点选择的依据。如何从相关性矩阵中选择干扰效果更优的节点, 有待于下一步深入研究。

2.2 奖赏依据

本文对未知拓扑结构的无线自组网络中的多个节点施加干扰时, 需要对选择动作的干扰效果进行评估, 即对于干扰动作给予奖赏。根据掌握信息的不同, 奖赏依据可分为有先验信息情况下的奖赏和无先验信息情况下的奖赏。对于前者, 算法根据侦察获得的 ACK/NACK 信息推测阻断网络流数^[12], 将网络流减少数目直接作为干扰动作对应的奖赏值。对于后者, 算法以干扰过后活跃节点的减少量作为奖赏依据。为了便于同现有算法作比较, 本文以阻断网络流数作为奖赏依据。

2.3 采用节点相关性的无线自组网络干扰决策算法

利用节点间的相关性选择待干扰节点, 需要解决两个最根本的问题: 一是干扰方如何具备目标网络所有节点间的相关性信息; 二是干扰方如何确定待干扰节点序列中第一个节点。对于第一个前提条件, 可通过设置包含若干干扰次数的初始化阶段来满足, 该阶段内每次干扰时随机选择干扰节点, 并根据网络反馈结果对节点相关性矩阵进行更新, 该矩阵可用于后续干扰节点选择。第二个前提条件依然可通过上述初始化阶段来满足, 统计初始化阶段每个节点被选中的次数以及获得的奖赏, 根据 UCB 算法计算每个节点的重要性, 将最重要的节点作为待干扰节点序列中的第一个节点。

结合节点相关性概念以及 UCB 算法, 本文提出了一种相关置信上界(Correlation UCB, CUCB)算法来实现无线自组网络多节点干扰策略选择。下面为算法的详细步骤, 其中步骤 2 为初始化阶段, 步骤 3 为主循环阶段。

步骤 1 初始化操作(以恒定流为例)。将网络中节点总数赋值为 L , 生成全零向量 $\mathbf{F}$ 和 $\mathbf{P}$ 分别用于记录每个节点获得的累积奖赏和被选择次数, 构造全零矩阵 $\mathbf{R}(t)$ 统计节点间的相关性。

步骤 2 for $t=1:L$ 执行以下(1)~(3)步:

(1) 从网络中随机选择指定数目节点作为干扰动作 a ;

(2) 执行干扰动作 a , 根据网络流数目的变化情况更新 $\mathbf{F}$ 和 $\mathbf{P}$;

(3) 根据式(3)对节点相关性矩阵 $\mathbf{R}(t)$ 的元素进行更新。

步骤 3 for $t < T$ 执行以下(1)~(4)步:

(1)选择能够使下面公式最大化的节点

$$\varphi = \arg \max_{\varphi \in V} \sum_{i \in V} \left(\hat{\theta}_i + \sqrt{\frac{(L+1)\ln n}{m_i}} \right) \quad (4)$$

式中: n 表示当前干扰次数; $\hat{\theta}_i = F_i/P_i$ 为节点 i 对应的平均奖赏; m_i 表示截止当前时刻节点 i 被干扰的次数。

(2)以该节点作为起始节点,并根据相关性矩阵 $\mathbf{R}(t)$ 依次选择出待干扰节点,即新的干扰动作 a 。

(3)执行干扰动作 a ,根据网络流数目的变化情况更新 $\mathbf{F}$ 和 $\mathbf{P}$ 。

(4)根据式(3)对节点相关性矩阵 $\mathbf{R}(t)$ 的元素进行更新。

3 实验仿真

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,将联合 SEE 算法(Joint SEE, JSEE)、最佳干扰策略、中介中心性算法、独立 SEE 算法(Independent SEE, ISEE)、独立 UCB 算法(Independent UCB, IUCB)等作为比较算法,分别从累积奖赏和鲁棒性两个方面衡量上述算法性能。

3.1 初始化阶段对算法性能的影响

2.1 节中明确指出,可通过初始化阶段构造节点相关性矩阵,还可以根据该阶段确定后续干扰时应首先干扰的节点,也就是说,初始化阶段是本文提出算法的必要环节,该阶段包含干扰次数的多少直接决定着算法的干扰性能。图 4 给出了包含不同干扰次数的初始化阶段对 CUCB 算法在累积奖赏方面的比较,总干扰次数为 2 000 次。

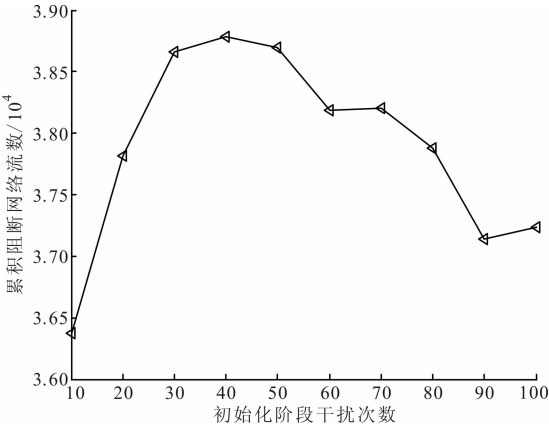


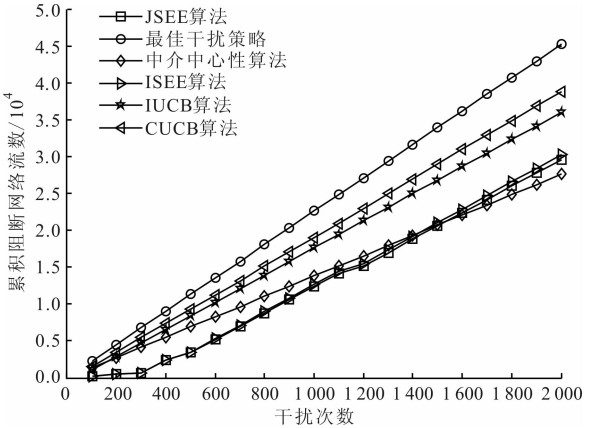
图 4 初始化阶段干扰对本文算法学习结果的影响

从图 4 可以看出,当初始化阶段的干扰次数设置为 40 次时,利用 CUCB 算法能够获得最大的累积奖赏,过小或过大的设置都会在一定程度上导致累积奖赏的降低,原因在于,较少次数的干扰,使得

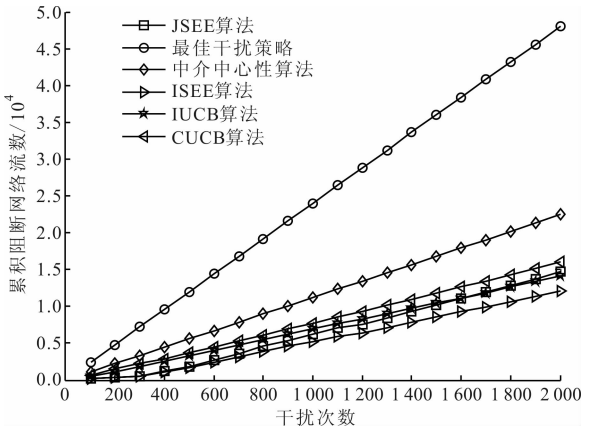
节点相关性矩阵不够完整,未能对节点之间的相关性进行准确的衡量;而过多次数的干扰,尽管学习到了更加准确的相关性矩阵,但由于该阶段随机选择节点的原因,使得干扰效果较差,进而导致累积奖赏偏低。鉴于 SEE 算法和 UCB 算法均包含初始化阶段,且初始化阶段内干扰次数设置与总节点数相同,为了便于同上述两种算法相比较,本文算法初始化阶段的干扰次数同样设置与总节点数相同,即 50 次干扰。

3.2 对静态拓扑和动态拓扑 PPP 网络干扰

假定待干扰的 PPP 网络中共 50 个节点,节点的连接度为 5,对于静态拓扑结构,网络流数为 30 条,源节点和目的节点固定不变,干扰方每次干扰 5 个节点;对于动态拓扑结构,网络流数介于[10, 30]条,源节点和目的节点随机选择,干扰节点数依然为 5 个。图 5 给出了 6 种算法在不同网络流状态下干扰性能比较。



(a)静态拓扑恒定流 PPP 网络 5 节点干扰



(b)动态拓扑随机流 PPP 网络 5 节点干扰

图 5 对 6 种算法在不同网络流状态下的干扰性能比较

从图 5a 可以看出,最佳干扰策略在累积阻断的网络流数目最多,CUCB 算法累积阻断的网络流数

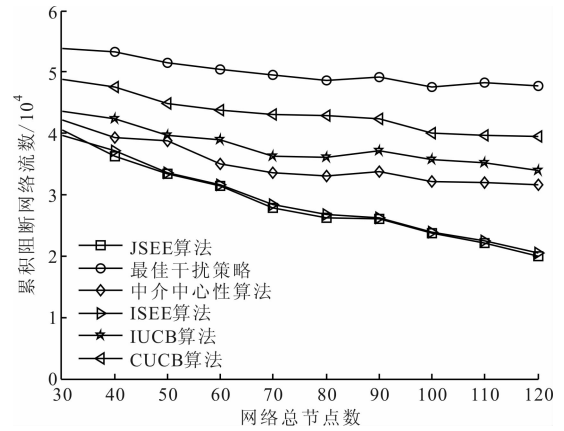
仅次于最佳干扰策略,数目约为其 89.6%,比 JSEE 算法高出 27.1%。此外,所有算法的奖赏曲线在 1 000 次干扰后以不同斜率线性增长,线性增长表明每次干扰阻断的网络流数不再发生变化,即算法收敛到固定节点,不同斜率意味着算法每次干扰阻断的网络流数同样有差异,斜率越高意味着阻断的网络流数越多。

从图 5b 中可以看出,当目标网络的拓扑结构动态变化时,几种干扰算法的干扰性能均不理想,原因在于本文提出的算法与对比算法均为学习型算法,时刻变化的干扰对象不利于算法学习。但是,比较上述几种学习型干扰算法可以发现,CUCB 算法累积阻断的网络流数依然是最高的,表明本文算法干扰性能出众。

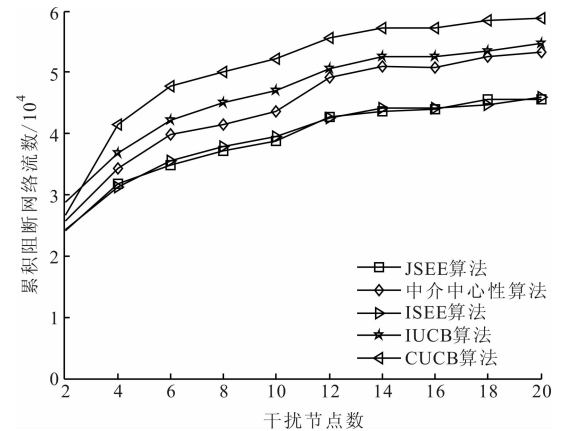
3.3 对不同参数下的 PPP 网络进行干扰

为了进一步验证 CUCB 算法的鲁棒性,图 6 从不同节点连接度、不同网络流数目、不同节点总数、不同干扰节点数等 4 个方面进行对比。

从图 6a~6d 可以看出,同最佳干扰策略相比,CUCB 算法累积阻断的网络流数仍有一定的差距,



(c)不同节点总数条件下



(d)不同干扰节点数条件

图 6 6 种算法对不同参数的 PPP 网络的干扰性能比较

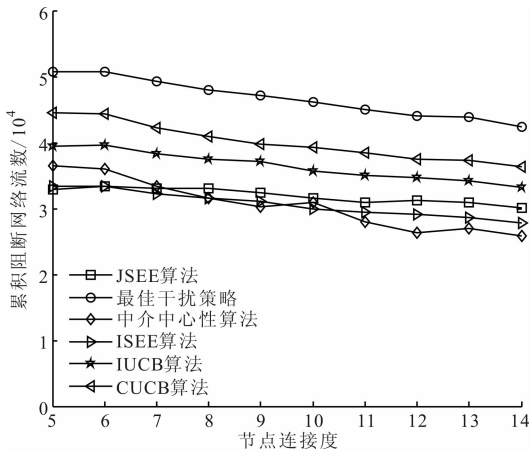
但同对比算法比较而言,阻断的网络流数始终是最多的,反映出本文算法的干扰性能和鲁棒性能优于对比算法。

4 结 论

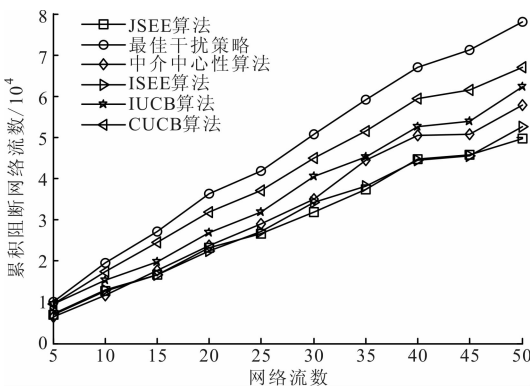
本文以干扰无线自组网络中的多个节点作为研究对象,在充分挖掘节点相关性的基础上结合现有 UCB 算法,提出了一种未知拓扑无线自组网络多节点干扰决策算法——CUCB 算法,该算法将战场无线自组网建模为 PPP 网络,通过对目标网络中多个节点同时进行干扰实现通信拒止。文中从静态拓扑恒定网络流和动态拓扑随机网络流,以及不同网络参数下的干扰效果验证所提算法的干扰性能和鲁棒性能。仿真结果表明,本文所提算法在累积奖赏方面均优于 JSEE 算法,且算法具有较强的鲁棒性,对不同状态的 PPP 网络均具有优异的干扰效果。

参考文献:

[1] 刘蔚,赵宇,陈锐. 无线 Ad-hoc 网络中基于 0-1 优化



(a)不同节点连接度条件下



(b)不同网络流数目条件下

- 的两步骤资源分配算法 [J]. 计算机科学, 2017, 44(1): 103-108.
- LIU Wei, ZHAO Yu, CHEN Rui. Zero-one integer programming based optimization model and two phase resource optimization algorithm for wireless Ad-hoc networks [J]. Computer Sciences, 2017, 44(1): 103-108.
- [2] MUKHERJEE A, KESHARY V, PANDYA K. Flying Ad-hoc networks: a comprehensive survey [C] // 3rd International Conference on Computer and Communication Technologies. Berlin, Germany: Springer, 2016: 1254-1270.
- [3] AMURU S, BUEHRER R M. Optimal jamming using delayed learning [C] // Proceedings of 33rd Annual IEEE Military Communications Conference. Piscataway, NJ, USA: 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] // International Conference on Electrical, Electronics, and Optimization Techniques. Piscataway, NJ, USA: IEEE, 2016: 3374-3377.
- [5] 刘书建, 吴晟. 基于 Zamp 的 Dos 攻击可行性分析与研究 [J]. 化工自动化及仪表, 2016, 43(7): 725-727. LIU Shujian, WU Sheng. Analysis and study of Dos attack feasibility based on Zmap [J]. Control and Instruments in Chemical Industry, 2016, 43(7): 725-727.
- [6] PROANO A, LAZOS L. Selective jamming attacks in wireless network [C] // 2010 IEEE International Conference on Communication. Piscataway, NJ, USA: IEEE, 2010: 1141-1143.
- [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] 段勇, 崔宝侠, 徐心和. 多智能体强化学习及其在足球机器人角色分配中的应用 [J]. 控制理论与应用, 2009, 26(4): 371-376. DUAN Yong, CUI Baoxia, XU Xinhe. Multi-agent reinforcement learning and its application to role assignment of robot soccer [J]. Control Theory & Application, 2009, 26(4): 371-376.
- [10] 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 Dongbin, SHAO Kun, ZHU Yuanheng, et al. Review of deep reinforcement learning and discussion on the development of computer Go [J]. Control Theory & Applications, 2016, 33(6): 701-717.
- [11] 丁乐乐. 基于深度学习和强化学习的车辆定位与识别 [D]. 成都: 电子科技大学, 2016: 42-60.
- [12] AMURU S, MICHAEL B R, VAN DER SCHAAR M. Blind network interdiction strategies: a learning approach [J]. IEEE Transactions on Cognitive Communications and Networking, 2015, 1(4): 435-449.
- [13] GAI Y, KRISHNAMACHARI B, JAIN R. Combinatorial network optimization with unknown variables: multi-armed bandits with linear reward [J]. Networking IEEE/ACM Transaction on Networking, 2012, 20(5): 1466-1478.
- [14] 张国峰. 无人机自组网路由协议研究 [D]. 沈阳: 沈阳工业大学, 2017: 26-45.
- [15] 蒲潇. 战术自组网网络结构及分群算法研究 [D]. 大连: 大连理工大学, 2009: 14-15.
- [16] BRANDES U, BORGATTI S P, FREEMAN L C. Maintain the duality of closeness and betweenness centrality [J]. Social Networks, 2016, 44: 153-159.
- [17] AUER P, CESA B N, FISCHER P. Finite-time analysis of the multi-armed bandit problem [J]. Machine Learning, 2002, 47(2/3): 2-13.
- [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 Yang, CHEN Shifu, LU Xin. Research on reinforcement learning technology: a review [J]. Acta Automatica Sinica, 2004, 30(1): 86-100.

[本刊相关文献链接]

- 颢孙少帅, 杨俊安, 刘辉, 等. 采用双层强化学习的干扰决策算法. 2018, 52(2): 63-69. [doi: 10. 7652/xjtuxb201802010]
- 李清伟, 郭黎利. DS-CDMA 系统多波形优化的多址干扰抑制方法. 2017, 51(10): 94-99. [doi: 10. 7652/xjtuxb201710016]
- 杨少奇, 田波, 周瑞钊. 应用双谱分析和分形维数的雷达欺骗干扰识别. 2016, 50(12): 128-135. [doi: 10. 7652/xjtuxb201612020]
- 刘立, 张衡阳, 毛玉泉, 等. 变换域通信系统抗干扰编码幅度谱成型算法. 2017, 51(2): 91-96. [doi: 10. 7652/xjtuxb201702015]
- 孙黎, 徐洪斌. 协作式终端直通系统中星座旋转辅助的干扰避免策略. 2015, 49(12): 6-11. [doi: 10. 7652/xjtuxb201512002]

(编辑 刘杨)