

空中传感器网络中负载均衡的地理路由协议

黄鑫权 刘爱军 梁小虎 王 桁

陆军工程大学通信工程学院天基信息系统教研室 南京 210007

(huangxinquan1993@sina.com)



摘 要 针对多跳空中传感器网络(Aerial Sensor Network, ASN)中的负载不均衡问题,提出了强化学习(Reinforcement Learning, RL)理论辅助的队列高效地理路由(Reinforcement-Learning Based Queue-Efficient Geographic Routing, RLQE-GR)协议。RLQE-GR 协议首先将 ASN 路由问题抽象为强化学习(RL)任务,其中每个无人机抽象为一个 RL 状态,而数据包的每跳成功转发则抽象为一个 RL 动作。其次,RLQE-GR 协议中引入了新的奖赏函数来评估每次动作,该奖赏函数的值不仅与无人机节点地理位置和每跳链路质量相关,而且与无人机节点的可用路由队列长度密切相关。然后,根据所设计的奖赏函数,RLQE-GR 协议利用 Q 函数分布式地更新每个动作的长期累积奖赏值(Q 值),并使得每个节点根据本地 Q 值的大小采用贪婪策略转发数据包。最后,为了使全网的 Q 值快速收敛且最小化收敛过程中造成的路由性能损失,RLQE-GR 采用周期性信标机制对 Q 值进行迭代更新。当 Q 值收敛时,RLQE-GR 协议能够实现可靠有效的多跳数据传输性能。与现有地理路由协议相比,所提协议在转发数据包的同时考虑了节点之间的相对距离、每跳链路质量和中间节点路由队列利用率。这使得 RLQE-GR 协议能够在保证路由跳数以及数据包重传次数的限制下,实现 ASN 的负载均衡。此外,利用强化学习理论,所提协议可以实现近乎最优的路由性能。

关键词: 空中传感器网络;地理路由协议;强化学习;奖赏函数;信标机制

中图法分类号 TN927

Load-balanced Geographic Routing Protocol in Aerial Sensor Network

HUANG Xin-quan, LIU Ai-jun, LIANG Xiao-hu and WANG Heng

College of Communication Engineering, Army Engineering University, Nanjing 210007, China

Abstract The unbalanced burden on the nodes nearing the ground station pose challenges on the multi-hop data transmission in aerial sensor networks(ASNs). In order to achieve reliable and efficient multi-hop data transmission in ASNs, a reinforcement-learning based queue-efficient geographic routing(RLQE-GR) protocol is proposed. The RLQE-GR protocol maps routing problem into the general reinforcement learning(RL) framework, where each UAV is treated as one state and each successful packet forwarding is treated as one action. Based on the framework, the RLQE-GR protocol designs a reward function related to geographical location, link quality and available transmission queue length. Then, the Q-function is employed to converge all the state-action values(Q-values), and each packet is forwarded based on potential state-action values. To converge all Q values and minimize performance deterioration during the convergence process, a beacon mechanism is employed in RLQE-GR protocol. In contrast to existing geographic routing protocols, the RLQE-GR protocol simultaneously takes the queue utilization, link quality and relative distance into consideration for forwarding packets. This makes the RLQE-GR protocol achieve load balancing, meanwhile not introducing strict performance deteriorations on routing hop and link quality. Moreover, due to the near-optimization character of RL theory, the RLQE-GR protocol can achieve routing performance optimization on packet delivery ratio and end-to-end delay.

Keywords Aerial sensor network, Geographic routing protocol, Reinforcement learning, Reward function, Beacon mechanism

到稿日期:2020-10-26 返修日期:2021-03-15 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金(61671476, 61901516);江苏省自然科学基金(BK20180578);中国博士后科学基金(2019M651648)

This work was supported by the National Natural Science Foundation of China(61671476, 61901516), Natural Science Foundation of Jiangsu Province of China(BK20180578) and China Postdoctoral Science Foundation(2019M651648).

通信作者:刘爱军(liuaj.cn@163.com)

1 引言

空中传感器网络^[1-3]是一类由大量装载传感器单元的小型无人机(Unmanned Aerial Vehicle, UAV)^[4]构成的无线自组网。空中传感器网络是 UAV 网络的一个分支,我们将用于战术侦察、灾区搜救等场景下的 UAV 网络称为 ASN 网络。在这些应用场景中,每个无人机搜集关于地面目标的信息并生成相应数据包,通过 ASN 将数据包交付至地面数据收集中心(简称地面站)。然而,相比大多数 UAV 网络,ASN 网络的单个汇聚节点(地面站)的特点,使得 ASN 网络容易发生负载均衡现象,导致多跳数据的传输性能降低。

由于 ASN 网络拓扑动态变化和网络规模大的特点,使得地理路由(Geographic Routing, GR)协议^[5-12]在 ASN 中的应用引起了广泛关注。相比先应式路由协议^[13-14]和反应式路由协议^[15],分布式和可扩展性的优点使得 GR 协议在大规模 ASN 网络环境中能够实现更加可靠和有效的多跳数据传输^[16-17]。然而,大部分地理路由协议并没有加入负载均衡机制。因此,在地理路由协议中加入负载均衡机制对于在 ASN 中实现更好的多跳数据包传输具有重大意义。

为了实现上述目标,我们可以借鉴现有无线自组网中负载均衡路由协议的设计思路。近年来,无线自组网负载均衡路由协议逐渐受到研究者的关注。然而,大部分负载均衡路由协议基于拓扑路由协议^[18-21],仅有少数研究者在地理路由协议的基础上提出了负载均衡路由协议^[9]。这些负载均衡路由协议可以划分为基于信道忙率的路由协议^[18-19]和基于可用队列长度的路由协议^[9, 20-21]。然而,信道忙率会错误判断节点对网络是否出现负载均衡的现象。这是由于除了节点队列出现高负载现象之外,节点密度增加以及物理信道恶化(误码率增加或者信号冲突程度加剧)带来的数据包频繁重传,均会增大信道忙率。

对于基于可用队列长度的路由协议,其核心思想是使每个节点根据本地路由队列的使用情况来调整路由决策,以避免或者缓解负载均衡现象。然而,现有基于可用队列长度的路由协议仍存在一些不足。首先,文献^[20-21]中的负载均衡路由协议均基于无线自组网按需多径距离向量(Ad Hoc On-demand Multipath Distance Vector, AOMDV)路由协议。这些协议在路由发现过程中,通过路由请求包(Route Request, RREQ)来选择未发生拥塞的路由路径。然而,在两次路由发现过程的时间间隔中,这些被选择的路由路径中可能会出现新的拥塞节点。而这类协议^[20-21]只有在下一次路由发现过程中才能感知到这种变化,即它们无法保证这段时间间隔内网络业务的传输性能。其次,文献^[9]中的基于贪婪算法的队列感知路由协议(Greedy Traffic Light and Queue Aware Routing Protocol, GTLQR)利用贪婪算法选择距离减少比率与路由队列使用率乘积较低的邻节点作为下一跳节点。GTLQR 协议在做路由决策时,为距离减少比率和路由队列使用率这两个度量值分配了相同的权重系数。对于距离汇聚节点较远的网络节点而言,由于距离减少比率的分母增大,路由队列使用率将对下一跳节点的选择占据更大的比重,

进而贪婪算法的局部队列资源最优化特点有可能会导导致远端节点数据包的丢失。这是因为,总是选择可用路由队列长度较大的下一跳节点不一定能保证数据包朝着目的节点的方向进行转发。

此外,在有效缓解网络负载不均衡问题的同时,应尽可能避免牺牲路由跳数和路径链路质量方面的性能来进一步保证多跳数据业务的服务质量(Quality of Service, QoS)需求。然而,现有的负载均衡路由协议^[9, 18-21]缺乏对链路质量的考虑。综上所述,为了进一步提高地理路由协议在 ASN 中应用时的多跳传输性能,需要提出一种新型地理路由协议。该协议应能有效缓解负载均衡问题,同时应尽可能减少路由路径上每跳的丢包率。针对上述背景和问题,本文提出了一种基于强化学习的队列高效地理路由协议。该协议兼顾了地理路由协议的分布式和可扩展性优点的同时,还实现了上述关于负载均衡和无线链路质量的目标。首先,RLQE-GR 协议通过将 ASN 中的路由问题映射到一般的强化学习任务框架中,引入基于强化学习的 ASN 路由(RL Based Routing, RLR)框架。在 RLR 框架中,ASN 中的每个无人机被映射为一个 RL 状态,而每跳数据包的成功转发则被映射为一个 RL 动作。基于 RLR 框架,RLQE-GR 协议设计了新的奖赏函数,该奖赏函数的值与邻接节点相对距离、每跳无线链路质量以及邻节点可用路由队列长度密切相关。随后,RLQE-GR 协议利用 Q 函数所设计的奖赏函数,分布式地更新所有状态-动作对的长期奖赏值,即 Q 值。基于本地不同 Q 值的大小,所有节点采用贪婪策略转发数据包。在 RLQE-GR 协议中,为了使所用 Q 值快速收敛且最小化收敛过程带来的路由性能损失,RLQE-GR 协议采用信标机制对 Q 值进行迭代更新。当 Q 值收敛时,RLQE-GR 协议能够实现近乎最优的路由性能。仿真结果表明,与现有的地理路由协议相比,RLQE-GR 协议能够实现更高的数据包投递成功率和更低的端到端传输时延。

本文第 2 节对系统模型进行了详细描述,并确定了需要解决的问题;第 3 节对 RLQE-GR 协议进行了详细介绍;第 4 节对该协议进行了性能评估;最后总结全文。

2 系统模型和问题描述

假设 ASN 为平面同构网络,该假设的合理性来源于以下几个方面:1)本文提出的路由协议是一种分布式的路由协议,不依赖于平面图或者其他网络拓扑,因此本文提出的路由协议可以适用于二维或者三维 ASN 网络场景;2)ASN 平面网络可以采用一致性控制算法来实现^[22],通过设置一致性状态为飞行高度,即可在有限时间内使所有无人机的飞行高度趋于一致;3)本文重点研究单个汇聚节点对地理路由协议在 ASN 中应用时带来的挑战,而非 ASN 的三维欧氏分布对协议的影响。

假设 ASN 中包含 N 个无人机节点和一个地面站,其中每个无人机的 ID 为 $i=1, \dots, N$ 。ASN 的所有无人机节点在动力、功能特性和配置等方面都不存在差异。在 ASN 中,每个无人机的无线通信设备均使用半双工全向天线,且所有

无人机具有同样的最大通信范围 r 。对于 ASN 中的某个节点 u_i , 我们称 u_j 为 u_i 的一跳邻节点当且仅当 $d(u_i, u_j) < r$, 其中二元函数 $d(u_i, u_j)$ 的值表示节点 u_i 和 u_j 之间的欧氏距离。此外, 假设地面站的无线电信号辐射范围足以覆盖 ASN 中的所有无人机。

2.1 ASN 应用场景和相关假设

本文考虑战术侦察、灾区搜救等 ASN 应用场景^[4,7]。在这些应用场景中, 每个无人机装载一个或多个微小型传感器

单元(如摄像机、激光雷达或声纳^[23])。在任务执行初期, 所有无人机从基地起飞前往目标环境。到达目标环境后, 无人机集群采用先进的任务分配算法^[24-25]使每个无人机节点前往该环境中的不同区域。无人机在各自的区域通过机载传感器搜集地面目标的信息, 并将该信息封装成数据包, 通过 ASN 网络传输给地面站。在 ASN 中, 并非所有的无人机都能与地面站进行一跳通信, 因此无人机以多跳传输的形式将数据包交付至地面站, 如图 1 所示。

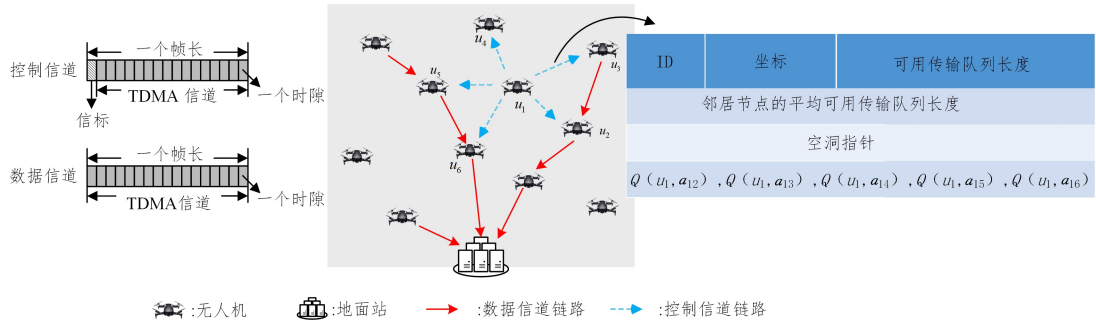


图 1 ASN 系统模型、信道划分以及信标帧结构

Fig. 1 ASN system model, channel division and beacon overhead

此外, ASN 网络环境满足以下几点假设:

- (1) 每个无人机节点可以获得自身的地理位置信息; 每个无人机可通过先进的定位感知技术获取自身的地理位置信息。例如, 每个无人机可装载全球定位系统 (Global Positioning System, GPS) 和惯性测量单元 (Inertial Measurement Units, IMU) 来获取自身的地理位置。
- (2) ASN 网络规模大且网络具有连通性: 在本文所考虑的场景中, 我们假设 ASN 中的无人机节点数量大, 且任务分配算法使 ASN 网络具有连通性。
- (3) 单个地面站: 本文只考虑一个地面站, 并假设该地面站是静止的或者移动非常缓慢。
- (4) 本文假设所有节点具有相同的最大可用路由队列长度。
- (5) 本文假设 ASN 网络拓扑每隔固定周期动态变化, 且在该周期内可以认为 ASN 的拓扑为静态或者缓慢变化。

2.2 ASN 信道接入

本文为 ASN 分配两个在频段上相互正交的物理信道: 一个控制信道 (Control Channel, CCH) 和一个数据信道 (Service Channel, SCH)。如图 1 所示, 控制信道在时间上被划分为多个等长的时间帧。每一帧的第一个时隙为信标时隙, 用于地面站向其无线电信号辐射范围内的所有无人机周期性地广播信标帧。该信标帧中标识了地面站 (即数据包目的节点) 的位置信息, 因此收到信标帧的无人机节点可以知道地面站的地理位置。信标时隙之后的时间段是一个时分多址 (Time Division Multiple Access, TDMA) 信道, 这部分信道被预先划分成 s 个等长的控制时隙, 且每个控制时隙只能允许一个无人机向其一跳邻节点广播一条控制帧。所有无人机生成的控制帧的大小固定且相同, 控制帧包含的内容将在后文进行详细描述。

数据信道 SCH 用于传输 ASN 中的数据包。与控制信道类似, 数据信道也被划分为多个等长的时间帧, 而且每一帧被预先划分为 s 个数据时隙。每个数据时隙只允许一个无人机节点转发一个数据包。所有无人机传输的数据包的大小固定且相同。

在控制信道 CCH 和数据信道 SCH 上, ASN 中的每个无人机节点均采用 VeMAC 协议^[26]在每一帧中预留一个控制时隙或者数据时隙。当无人机在数据信道 (控制信道) 上获得一个时隙时, 该无人机将在之后的每一帧占用该时隙, 直到该无人机节点不能使用这个时隙为止, 即两跳通信范围内的另一个无人机也接入了该时隙。上文中, s 是 ASN 所有无人机无冲突接入信道的最小数据 (控制) 时隙数, 即当一帧中有 s 个时隙数时, VeMAC 协议可以使每个无人机能够预留与其两跳通信范围内任一节点不同的时隙。

本文路由协议的研究主要针对数据信道上的数据业务传输。上文介绍的控制帧的信道访问机制并不是本文的重点, 它仅仅是为了说明路由协议中邻节点实现信息交互的物理基础。需要指出的是, 数据包的大小远大于控制帧的大小, 控制帧的传输也是本文路由协议的重要组成部分。关于控制帧中所包含的信息及它在该路由协议中所起到的作用将在 4.3.2 节进行详细介绍。

3 问题描述

由于仅有单个地面站节点, ASN 的网络负载具有汇聚特性, 因此路由过程中的某些中间节点尤其是靠近地面站区域的某些节点将被过度使用。过度使用这些中间节点将会使这些节点出现负载不均衡的现象, 成为网络的瓶颈, 进而导致多跳数据传输性能降低。遵循先入先出 (First Input First Output, FIFO) 的路由队列管理原则, 一个数据包必须等待路由

队列前面的所有数据包完成转发后才可以进行转发操作。当一个中间节点发生负载不均衡时,其可用队列长度较小,因此该节点上的数据包需要等待较长的排队时间,增加了路由过程的端到端传输时延。此外,当中间节点的路由队列饱和(发生拥塞)时,新到达的数据包将无法被正确接收,相应的发送节点将重传该数据包。当数据包重传次数达到上限时,发送节点将丢弃该数据包,这会降低路由过程的数据包投递成功率。

此外,ASN 中无线链路具有不可靠性,且每跳无线链路所能实现的链路质量不同。对于每跳无线链路而言,我们用该链路所能实现的丢包率来量化它的链路质量。丢包指数据包无法被接收节点正确解析,相邻无线链路的干扰、误码率等都会造成丢包现象。对于丢包率很高的无线链路,在该链路上转发的大量数据包需要被多次重传,这会给路由过程引入额外的端到端传输时延。此外,当数据包达到重传上限时,该数据包将被转发节点丢弃,降低了路由过程的数据包投递成功率。

针对上述背景,本文考虑如下问题:在保证数据包在有限跳数内可靠到达目的节点,且数据包在有限重传次数内成功转发的前提下,尽可能实现网络负载均衡。

为了更清晰地理解本文的研究目标,我们对上述问题进行了数学建模。首先给出以下集合和矩阵的概念:

(1)集合 $U = \{u_i\}$ 表示 ASN 中所有无人机节点构成的集合,其中 $i = 1, 2, \dots, n$ 表示每个无人机的 ID 值。

(2) u_{N+1} 表示地面站,它是所有数据包的目的节点。

(3)矩阵 $\mathbf{B} = [b_{ij}]_{(n+1) \times (n+1)}$ 称为邻接矩阵,它是一个 $N+1$ 阶方阵,其中 $\forall i, j = 1, 2, \dots, N+1$, 其第 i 行第 j 列元素 b_{ij} 定义为:

$$b_{ij} = \begin{cases} 1, & d(u_i, u_j) \leq r \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

此外,对于任意 $i = 1, 2, \dots, n+1$, 我们规定 $b_{ii} = 0$ 。

(4)对角矩阵 $\mathbf{D} = [d_i]_{(n+1) \times (n+1)}$ 称为度矩阵,其中 $\forall i = 1, 2, \dots, n+1$, 其第 i 个对角元素 $d_i = \sum_{j=1}^{n+1} b_{ij}$ 。

(5) $n+1$ 阶方阵 $\mathbf{W} = [\omega_{ij}]_{(n+1) \times (n+1)}$ 称为链路质量矩阵。 $\forall i, j = 1, 2, \dots, n+1$, 其第 i 行第 j 列元素 ω_{ij} 的值表示从节点 u_i 到节点 u_j 的单向链路质量,该元素的值等于该链路上的丢包率。此外,规定当 $b_{ij} = 0$ 时, $\omega_{ij} = 0$ 。

(6)方阵 $\mathbf{Q} = \mathbf{IE}$ 表示矩阵 $\mathbf{I}$ 和 $\mathbf{E}$ 的叉乘。其中, $\mathbf{I}$ 是一个元素全为 1 的 $(n+1) \times 1$ 列矩阵,而 $\mathbf{E} = [e_i]_{1 \times (n+1)}$ 是一个 $1 \times (n+1)$ 行矩阵。 $\mathbf{E}$ 中第 $i = 1, 2, \dots, n$ 个元素 e_i 的值等于无人机 u_i 的可用路由队列长度。规定 $e_{n+1} = 1$ 。

(7) $n+1$ 阶方阵 $\mathbf{O} = [o_{ij}]_{(n+1) \times (n+1)}$ 称为路由策略矩阵。其中,矩阵 $\mathbf{O}$ 中的第 i 行第 j 列元素定义为:

$$o_{ij} = \begin{cases} \delta, & \text{如果 } u_i \text{ 选择 } u_j \text{ 作为下一跳转发节点} \\ 1, & \text{如果 } i = j = n+1 \\ 0, & \text{其他} \end{cases} \quad (2)$$

其中, δ 是一个常数且 $\delta > 1$ 。本节考虑的地理路由协议中,

每个节点每跳仅向一个邻节点转发数据包。因此,在矩阵 $\mathbf{O}$ 的每一行中,只有一个元素是非零值。

基于上述定义,本文拟解决的问题可以建模为如下约束方程:

$$\min_{\mathbf{O}} \text{sumf}[-(\mathbf{Q} * \mathbf{O})^{h_m}] \quad (3)$$

$$\text{s. t. } \text{eig2}(\mathbf{D} - \mathbf{B}) > 0 \quad (4)$$

$$z_{i(n+1)}(\mathbf{O}^{h_m}) > 0, \forall i = 1, 2, \dots, n+1 \quad (5)$$

$$\text{sumf}[(\mathbf{W} * \mathbf{O})^{h_m}] > \phi \quad (6)$$

其中,对于任意方阵 $\mathbf{X} = [x_{ij}]_{(n+1) \times (n+1)}$, 函数 $\text{sumf}(\mathbf{X})$ 定义为:

$$\text{sumf}(\mathbf{X}) = \sum_{i=1}^{n+1} x_{i(n+1)} \quad (7)$$

其中, $\mathbf{X}^{h_m} = \underbrace{\mathbf{X} \mathbf{X} \mathbf{X} \cdots \mathbf{X}}_{h_m}$ 表示方阵 $\mathbf{X}$ 的 h_m 次连续矩阵叉乘,其中

正整数 h_m 表示 ASN 网络允许的最大跳数。而 $\mathbf{X} * \mathbf{Y}$ 形式表示两个任意方阵 $\mathbf{X}$ 和 $\mathbf{Y}$ 的点乘。此外,对于任意方阵 $\mathbf{X} = [x_{ij}]_{(n+1) \times (n+1)}$, 函数 $\text{eig2}(\mathbf{X})$ 返回 $\mathbf{X}$ 的第二个最小特征值。

约束条件利用了图论知识^[27], 限制了 ASN 的网络连通性。在约束条件中, $z_{i(n+1)}(\mathbf{O}^{h_m})$ 函数返回矩阵 $\mathbf{O}^{h_m}$ 的第 i 行第 $n+1$ 列元素。对于任意整数 $h < h_m$, 如果矩阵 $\mathbf{O}^h$ 的第 i 行第 j 列元素为非零,则表示无人机 u_i 生成的数据包在策略 $\mathbf{O}$ 下经过 h 跳后最终到达无人机 u_j 。因此,约束条件保证了数据包以不超过 h_m 的跳数成功投递到地面站。

根据矩阵 $\mathbf{O}$ 中元素的定义可知, $\mathbf{W} * \mathbf{O}$ 矩阵中每行只有一个元素为非零值,该值为相应路由策略所对应的链路质量。如果 $(\mathbf{W} * \mathbf{O})^h$ ($h < h_m$) 的第 i 行第 $n+1$ 列元素 (其中 $i \neq n+1$) 为非零,则表示在策略 $\mathbf{O}$ 下,节点 u_i (其中 $i = 1, 2, \dots, n$) 生成的数据包经历 h 跳后被交付至目的节点。此外,对于任意正整数 h_1 , 矩阵 $\mathbf{O}^h \mathbf{O}^{h_1}$ 相同位置元素的值将保持不变。因此,在策略 $\mathbf{O}$ 下,如果节点 u_i (其中 $i = 1, 2, \dots, n$) 生成的数据包经历 h 跳后被交付至目的节点,则 $\mathbf{O}^{h_m}$ 的第 i 行第 $n+1$ 列元素的值除以 δ^h ($h = 1, 2, \dots, h_m$) 即为该路由路径每跳链路质量的乘积。而且 $\text{sumf}[(\mathbf{W} * \mathbf{O})^{h_m}]$ 的值越大,路由策略 $\mathbf{O}$ 在 ASN 中路由路径的链路质量就越好。因此,使 $\text{sumf}[(\mathbf{W} * \mathbf{O})^{h_m}]$ 大于我们预定的门限值 ϕ , 即式(6)限制了路由过程中数据包由于发生丢包而重传的次数。上述对 $(\mathbf{W} * \mathbf{O})^{h_m}$ 的相关阐述也可用于分析 $-(\mathbf{Q} * \mathbf{O})^{h_m}$ 。式(3)表达了本文的可用路由队列长度的优化目标,即希望每跳可用路由队列长度之和更大,以此来均衡网络中的负载。

4 基于强化学习的 RLQE-GR 协议

为了求解式(3),需要 ASN 的全局信息,这给大规模的 ASN 网络环境带来了巨大的复杂度。另外,由于 ASN 的应用场景一般是未知的环境(如战术环境),这给提前求解式(3)带来了巨大的挑战。因此,本文提出了 RLQE-GR 协议,通过利用强化学习技术使网络自动学习近似最优的路由策略 $\mathbf{O}$, 从而满足上述对地理路由协议所提出的需求。

本节将对本文提出的 RLQE-GR 协议进行详细描述。

首先介绍了 RLQE-GR 协议,如何将 ASN 的路由问题抽象为一个强化学习任务。在此基础上,对 RLQE-GR 协议中提出的奖赏函数进行详细描述。最后,对 RLQE-GR 协议中的长期累积奖赏值收敛策略进行了描述。

4.1 基于强化学习的路由框架

强化学习任务通常用马尔可夫决策过程(Markov Decision Process, MDP)来描述;智能体处在环境 E 中,状态空间为 S , S 中的每个状态都是智能体对环境的描述,智能体采取的动作构成了动作空间 A ,若 A 中的某个动作作用在当前状态上,则潜在的转移矩阵 P 将使环境从当前状态按某种概率转移到另一个状态。在转移到另一个状态的同时,环境会根据潜在的奖赏函数 R 反馈给智能体一个奖赏值。关于强化学习的更多介绍,可参考文献[28]。

RLQE-GR 协议将 ASN 中的路由问题抽象为一个强化学习任务,进而构建了基于强化学习的路由框架(RLR 框架)。在 RLR 框架中,每个数据包被视作一个智能体,每个无人机节点被映射为强化学习任务中的一个状态。显而易见,RLR 框架中的状态空间即为 ASN 中的所有节点。此外,在 RLR 框架中,一次成功的数据包转发被映射为强化学习任务中的一个动作。对于一个数据包(智能体)而言,只有当它成功被某个邻节点接收(收到该邻节点的 ACK 确认包),我们才称该数据包采取了一个完整的动作。因此,数据包采取的一个动作可能包含多次数据包重传。

在 RLR 框架中,如果一个数据包处于无人机节点 u_i 的路由队列中,则称该数据包处于 s_i 状态,并且当该数据包被 u_i 成功转发到其邻节点 u_j 时,称该数据包采取了一个动作,并将该动作记为 a_{ij} 。如图 2 所示,无人机节点 u_1 的传输队列中有个数据包等待被转发,那么该数据包的当前状态是 s_1 ;如果无人机节点 u_2 成功接收到该数据包(u_2 发送的 ACK 确认包被无人机节点 u_1 成功接收),我们称该数据包采取了一个动作 a_{12} 并从状态 s_1 转移到状态 s_2 。对于转移概率矩阵 P 中的第 i 行第 j 列元素 p_{ij} ,它表示从状态 s_i 转移到状态 s_j 的概率。基于上述对状态和动作的描述, p_{ij} 的值为:

$$p_{ij} = \begin{cases} 1 - \omega_{ij}^{h_r}, & d(u_i, u_j) < r \\ 0, & d(u_i, u_j) > r \end{cases} \quad (8)$$

其中, h_r 表示重传上限值, $1 - \omega_{ij}^{h_r}$ 表示在重传次数达到上限之前成功转发数据包的概率。

如图 2 所示,在 RLR 框架中,当无人机节点 u_i 收到来自无人机节点 u_j 的 ACK 确认包的同时(意味着状态从 s_i 转移到 s_j),无人机节点 u_i 将会收到一个直接奖赏值 $R(s_i, a_{ij})$ 来评价动作 a_{ij} 的好坏。而且,当转移的下一个状态为地面站时,无人机节点 u_i 收到的直接奖赏值也称最终奖赏值。每一个无人机在本地维护一个状态-动作值列表,其中列表里的每个状态-动作值都表示无人机采取各个动作的长期累积奖赏值。在无人机收到直接奖赏值的同时,它通过 Q 函数^[12]更新相应的状态-动作值。经过多次迭代更新后,状态-动作值最终会收敛到一个固定值。基于 Q 函数得到的状态-动作值也

称为 Q 值。考虑无人机 u_i (状态 s_i) 及动作 a_{ij} , 定义第 t 次更新之后(即数据包第 t 次从无人机 u_i 转发至无人机 u_j) 该状态-动作对的 Q 值为 $Q_t(s_i, a_{ij})$, 那么第 $t+1$ 次更新后,该状态-动作对的 Q 值 $Q_{t+1}(s_i, a_{ij})$ 可以按照式(9)计算得出:

$$Q_{t+1}(s_i, a_{ij}) = (1 - \alpha)Q_t(s_i, a_{ij}) + \alpha[R(s_i, a_{ij}) + \gamma \max_{a'} Q_t(s_j, a')] \quad (9)$$

其中, α 和 γ 是 0 到 1 之间的参数,分别表示学习速率和折扣因子,学习速率规定了 Q 值更新的速度。此外, a' 表示状态 s_j 采取的一个动作。

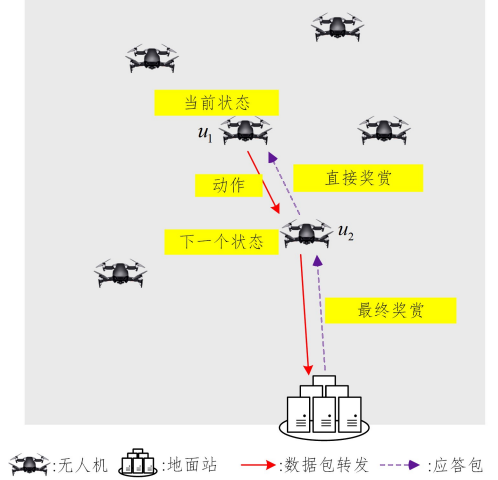


图 2 基于强化学习的路由框架

Fig. 2 Reinforcement learning based routing framework

4.2 奖赏函数设计

奖赏函数是 RLQE-GR 协议中的重要环节,它决定了每次执行动作后获得的直接奖赏值的大小。在 RLQE-GR 协议中,所提奖赏函数与邻接节点相对距离、每跳无线链路质量和邻节点可用路由队列长度密切相关。下文将以状态-动作对 (s_i, a_{ij}) 为例来介绍所设计的奖赏函数。

在 RLQE-GR 协议中,若数据包被无人机节点 u_i 成功转发至其一跳邻节点 u_j ,则无人机节点 u_i (状态 s_i) 因采取动作 a_{ij} 而得到的直接奖赏值 $R(s_i, a_{ij})$ 可由式(10)计算得出:

$$R(s_i, a_{ij}) = \begin{cases} R_{\max}, & \text{if } u_j = u_{n+1} \\ -R_{\max}, & \text{if } d(u_j, u_{n+1}) < d(u_k, u_{n+1}), \\ & \text{for } u_k \in N_j \\ R^{\text{link}}(s_i, a_{ij})[p_1 R^{\text{dis}}(s_i, a_{ij}) + R^{\text{rel}}(s_i, a_{ij})], & \\ \text{otherwise} & \end{cases} \quad (10)$$

其中, p_1 为 0 到 1 之间的固定参数,而二元函数 $d(u_j, u_{n+1})$ 为返回邻节点 u_j 和地面站之间的距离。此外, N_j 表示无人机 u_j 的邻节点集合,即 $N_j = \{u_k | d(u_j, u_k) < r\}$ 。从式(10)中的第一个等式可以看出,当无人机节点 u_i 成功地将数据包转发到目的节点时,它将会收到一个非常大的最终奖赏值 $R_{\max}$ 。根据式(10)中的第二个等式可以看出,当下一跳节点 u_j 为空洞节点时,即 u_j 比它的所有邻节点在距离上都更靠近目的

地,则无人机节点 u_i 将获得一个非常低的直接奖赏值 $R_{\max}$ 。

而函数 $R^{dis}(s_i, a_{ij})$, $R^{link}(s_i, a_{ij})$ 和 $R^{ql}(s_i, a_{ij})$ 的返回值分别与邻接节点相对距离、每跳无线链路质量和邻节点可用路由队列长度相关。我们将在下文对这 3 个奖赏函数形式进行描述,并详细分析每个奖赏函数形式会如何影响 RLQE-GR 协议的路由性能。

4.2.1 基于相对距离的奖赏函数

奖赏函数 $R^{dis}(s_i, a_{ij})$ 的形式如式(11)所示:

$$R^{dis}(s_i, a_{ij}) = \frac{d(u_i, u_{n+1}) - d(u_j, u_{n+1})}{r} \quad (11)$$

从式(11)可以看出,无人机节点 u_j 距离地面站越近,无人机节点 u_i 选择 u_j 作为下一跳节点就能获得越高的直接奖赏值。在采用 Q 函数对状态-动作函数值进行多次迭代更新后,式(11)的函数形式将会使跳数较少的路径中所有节点相应的状态-动作对呈现更高的 Q 值。此外,跳数较少也能实现更小的端到端传输时延。

4.2.2 基于链路质量的奖赏函数

定义 $N_{\text{total}}(u_i, a_{ij})$ 为无人机节点 u_i 在数据信道 SCH 上向其一跳邻节点 u_j 转发数据包的总次数(包括被 u_j 成功接收的次数和未被 u_j 成功接收的次数)。此外,将 $N_{\text{succ}}(u_i, a_{ij})$ 定义为 $N_{\text{total}}(u_i, a_{ij})$ 中被 u_j 成功接收的次数。我们用 $N_{\text{succ}}(u_i, a_{ij})$ 与 $N_{\text{total}}(u_i, a_{ij})$ 之比表示无人机节点 u_i 观察到的它与无人机节点 u_j 之间链路的链路质量。当 $N_{\text{total}}(u_i, a_{ij})$ 趋于无穷大时,该比值将会正确反映该链路的实际丢包率。基于该链路质量的定义,奖赏函数 $R^{link}(s_i, a_{ij})$ 的形式如式(12)所示:

$$R^{link}(s_i, a_{ij}) = \begin{cases} \frac{N_{\text{succ}}(u_i, a_{ij})}{N_{\text{total}}(u_i, a_{ij})}, & \text{if } N_{\text{total}}(u_i, a_{ij}) \neq 0 \\ 1, & \text{if } N_{\text{total}}(u_i, a_{ij}) = 0 \end{cases} \quad (12)$$

由式(12)可得,发送节点 u_i 与接收节点 u_j 之间的链路质量越高,无人机 u_i 选择 u_j 作为下一跳节点获得的直接奖赏值就越高。对于一条完整的源到地面站的多跳路由路径,如果该路径上的链路质量总和越高,则根据式(9)的 Q 函数进行多次迭代更新后,式(12)的函数形式将使该路径上所有节点相应的状态-动作对收敛到更高的 Q 值。

4.2.3 基于可用路由队列长度的奖赏函数

奖赏函数 $R^{ql}(s_i, a_{ij})$ 基于邻节点的可用路由队列长度,其形式如式(13)所示:

$$R^{ql}(s_i, a_{ij}) = -p_2 f_1(Q_{re}^i, Q_{\text{total}}) - p_3 f_2(Q_{re}^i, Q_{\text{total}}, \{Q_{re}^{ni}\}_j) \quad (13)$$

其中, p_2 和 p_3 是介于 0 和 1 之间的固定参数。 Q_{re}^i 是无人机 u_i 存储的关于邻节点 u_j 的可用路由队列长度信息,而 Q_{total} 是每个无人机的最大可用路由队列长度。集合 $\{Q_{re}^{ni}\}_j$ 表示无人机节点 u_i 的所有邻节点的可用路由队列长度,即 $\{Q_{re}^{ni}\}_j = \{Q_{re}^k | u_k \in N_j\}$ 。

函数 $f_1(Q_{re}^i, Q_{\text{total}})$ 的形式如式(14)所示:

$$f_1(Q_{re}^i, Q_{\text{total}}) = \begin{cases} R_{\max}, & \text{if } Q_{re}^i = 0 \\ (Q_{\text{total}} - Q_{re}^i) / Q_{\text{total}}, & \text{otherwise} \end{cases} \quad (14)$$

由式(13)和式(14)可以看出,邻节点 u_j 的可用路由队列长度越小,无人机节点 u_i 选择 u_j 作为下一跳节点所获得的直接奖赏值就越低。对于一条源到目的节点的完整多跳路径,如果该路径上节点的可用路由队列长度总和越大,则在采用式(9)的 Q 函数进行多次迭代更新后,奖赏函数的形式将使该路径上所有节点相应的状态-动作对获得更高的 Q 值。

考虑节点 u_i 的另一个邻节点 u_m ,并定义 u_m 的邻节点集合为 N_m 。假设节点 u_m 和 u_j 的可用路由队列长度几乎相同,但是 u_j 的可用路由队列长度略大于 u_m 。由式(14)可知,无人机 u_i 采取动作 a_{ij} 比采取动作 a_{im} 能获得更高的直接奖赏值。然而, N_m 中节点的平均可用路由队列长度可能远高于 N_j 。由于 Q 值反映了长期的累计奖赏值,因此根据式(9)进行长期迭代后,状态-动作对 (s_i, a_{im}) 的 Q 值将高于状态-动作对 (s_i, a_{ij}) 的 Q 值。同时为了加快 Q 值的收敛,我们设计了函数 $f_2(Q_{re}^i, Q_{\text{total}}, \{Q_{re}^{ni}\}_j)$,该函数基于两跳邻节点的可用路由队列长度信息,其具体形式如式(15)所示:

$$f_2(Q_{re}^i, Q_{\text{total}}, \{Q_{re}^{ni}\}_j) = \frac{Q_{\text{total}} - \text{ave}([Q_{re}^i, \{Q_{re}^{ni}\}_j])}{Q_{\text{total}}} \quad (15)$$

其中,函数 $\text{ave}(X)$ 返回一维数组 X 中所有元素的平均值。定义无人机集合为 M_j ,该集合包含了无人机 u_j 和 N_j 中的无人机。由式(13)和式(15)可以看出, M_j 中无人机可用路由传输队列长度越小,无人机 u_i 因采取动作 a_{ij} 而获得的直接奖励就越低。

4.3 RLQE-GR 协议实现

4.3.1 基于信标机制的 Q 值收敛过程

为了使 Q 值快速收敛,RLQE-GR 协议设计了信标机制来实现对 Q 值的迭代更新,如算法 1 所示。信标机制规定每个无人机节点在 CCH 信道上周期性地向其一跳邻节点广播信标帧。当任一无人机收到某个邻节点的信标帧时,该无人机利用信标帧中标识的信息和式(10)计算的直接奖赏值,并根据式(9)进一步更新相应状态-动作对的 Q 值。基于信标机制的收敛策略可实现频繁遍历所有状态-动作对。

算法 1 基于信标机制的 Q 值更新策略

输入:无人机节点 u_i ;无人机节点 u_{j+1} ;第 u_{i+1} 次被节点 u_i 接收的来自 u_j 的信标帧

输出: $Q_{i+1}(s_i, a_{ij})$

1. If 无人机节点 u_i 成功接收无人机节点 u_j 的信标帧;
2. 节点 u_i 获得节点 u_j 的位置坐标;
3. 节点 u_i 获得节点 u_j 的可用路由队列长度;
4. 节点 u_i 获得节点 u_j 的邻节点平均可用路由队列长度;
5. 节点 u_i 获得节点 u_j 的空洞指针;
6. End
7. 节点 u_i 根据式(10)计算直接奖赏值;
8. 节点 u_i 查询当前的 Q 值;
9. 根据式(9)计算 $Q_{i+1}(s_i, a_{ij})$ 。

4.3.2 信息交换和列表维护

为了计算直接奖赏值和更新 Q 值,每个无人机节点广播的信标帧应包括以下内容:无人机节点 ID、无人机节点地理位置坐标、无人机当前可用路由队列长度、无人机邻节点平均

可用路由队列长度、空洞指针、无人机所有可能的 Q 值,如图 1 所示。其中,无人机邻节点平均可用路由队列长度可以根据一跳邻节点列表计算得出,关于一跳邻节点列表的建立与维护将在下文进行详细介绍。而无人机的空洞指针用于表征该无人机是否为空洞节点,空洞指针的值可以是 1 或 0,其中 1 表示该无人机节点是空洞节点,0 表示该无人机节点不是空洞节点。一个无人机是空洞节点则说明它比所有邻节点距离目的节点更近。判断自身是否为空洞节点,可以通过比较自身的位置坐标和本地列表中存储的邻节点的位置坐标来判断。信标帧中的其他信息可通过直接访问无人机的相应模块获得。

当一个无人机在 CCH 信道上的时隙到来时,该无人机从本地模块中获取位置坐标、可用路由队列长度和 Q 值,并根据一跳邻节点列表中的信息计算出邻节点平均可用路由队列长度和空洞指针的值。然后,这些信息被封装到信标帧中,并在 CCH 信道上广播该控制帧。

此外,每个无人机节点在本地维护一个一跳邻节点列表,该列表中包含多个条目,每个条目应对应一个邻节点。每个条目包括以下内容:邻节点 ID、邻节点可用路由队列长度、邻节点地理位置坐标和时间戳。当无人机节点接收到来自邻节点的一条控制帧时,它将检查自身一跳邻节点列表是否已为该邻节点创建了一个条目。如果已有条目,则该接收节点用该控制帧中的信息覆盖之前的信息,否则该接收节点为邻节点创建一个新条目。如果在指定的时间戳内没有接收到来自同一个邻节点的控制帧,则从列表中删除该邻节点的信息。

4.3.3 路由策略

RLQE-GR 协议利用贪婪 Q 学习策略来采取动作(转发数据包),如算法 2 所示。对于无人机节点 u_i ,如果 $Q(s_i, a_{ij})$ 是 u_i 所有 Q 值中最大的状态-动作值,则 u_i 选择 u_j 作为下一跳转发节点。随后, u_i 尝试转发数据包给节点 u_j ,直到达到重传次数上限或者收到 u_j 关于该数据包的 ACK 确认包。数据包的转发是在 SCH 信道上实现的,与上文在 CCH 信道中采用信标机制更新 Q 值是两个独立的过程。对于大规模的空中传感器网络,在 CCH 信道上采用信标机制需要一定的时间才能使 Q 值收敛。因此,在所有 Q 值收敛之前,贪婪 Q 学习策略可能会给 SCH 信道上的数据转发带来不理想的数据包投递性能和端到端传输时延。然而,当 Q 值收敛时,该策略将获得接近最优的路由性能。

算法 2 贪婪 Q 学习策略

输入:无人机节点 u_i ; K 个邻节点的 Q 值

输出:下一跳转发节点 u_j

- 1. 初始化: $Q_MAX=0$;
- 2. if $K \neq 0$;
- 3. $Q_MAX=\max_K\{Q(s_i, a_{ik})\}$
- 4. end;
- 5. 从 K 个 Q 值中找出值为 Q_MAX 的邻节点集合,并随机选出一个节点 u_j 。

5 性能评估

本节将对本文提出的 RLQE-GR 协议进行仿真,并与

CAGR 协议^[7]、QGEO 协议^[12]、ECC-AOMDV 协议^[21] 以及 GTLQR 协议^[9] 的仿真结果进行比较,进而评估 RLQE-GR 协议的性能。CAGR 地理路由协议^[7] 仅仅基于节点的地理位置坐标进行路由决策;而 QGEO 地理路由协议^[12] 基于节点的地理位置坐标和每条无线链路质量进行路由决策;ECC-AOMDV 协议^[21] 以及 GTLQR 地理路由协议^[9] 均利用路由队列长度进行路由决策。

5.1 仿真场景

在仿真中,我们首先生成 100 个无人机节点,并使这些无人机节点随机分布在 $1000\text{ m} \times 1000\text{ m}$ 的正方形区域内。这些无人机在整个仿真过程中以选择的位置为圆心且圆半径为 10 m 盘旋飞行,来模拟无人机对地面目标进行搜索时的运动模型^[29]。所有无人机节点在仿真过程中不生成数据包。盘旋的目的是使 ASN 拓扑在仿真过程中缓慢变化,以对应 2.1 节中的假设(5)。仿真中,我们将横坐标位于 $[1000/3\text{ m}, 2000/3\text{ m}]$ 范围内的无人机称为参考无人机(Reference UAV, RU)节点。在每次仿真的初始时刻,我们使 RU 节点的路由队列存储一定数量的数据包,而其余无人机的路由队列为空。在每次仿真运行中,每个 RU 节点初始时刻路由队列中存储的数据包个数是相同的。设置 RU 节点的实际意义可以理解为,ASN 所处未知环境中仅 RU 节点可检测到地面目标并采集了相关数据。

此外,在仿真区域内,坐标为(500 m, 1000 m)处有一个源无人机节点,并保持盘旋飞行,而地面站坐标为(500 m, 0 m)。在仿真过程中,源无人机以 1 packet/s 的速率生成数据包。地面站是所有数据包(包括参考无人机节点在仿真初始时刻路由队列中的数据包和源 UAV 在仿真运行时生成的数据包)的目的节点。图 3 给出了 ASN 在初始时刻的节点分布情况,其中每个五角星都代表一个参考无人机节点,而圆圈分别代表源无人机以及地面站。RLQE-GR 协议以及其他 4 种协议的仿真参数如表 1 所列。对于 RLQE-GR 协议,我们设置 $R_{\max}$ 的值为 10, α 和 γ 的值均为 0.6,而 p_1, p_2, p_3 的值分别为 1, 0.5 和 0.05。在仿真中,如果一个数据包因未被邻节点正确接收而在 MAC 层重传,则该数据包跳数增加 1。在每次仿真中,源无人机总共产生 100 个数据包,当 ASN 中没有数据包时(被地面站接收或者由于跳数限制而被丢弃),仿真终止。对于每个协议,我们进行 100 次相同的仿真以获得路由性能统计值。在每次仿真运行中, RU 中的数据包和源节点生成的数据包均采用相同的协议进行传输。

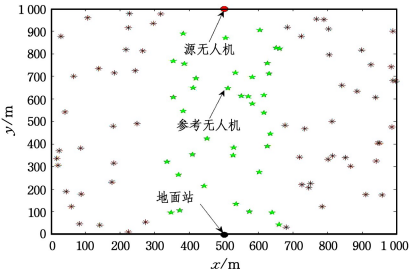


图 3 初始时刻 ASN 网络拓扑(2D 场景)

Fig. 3 ASN topology at initial time instant(2D scenario)

表 1 RLQE-GR 协议及其他 4 种协议仿真参数
Table 1 Simulation parameters of different protocols

参数名称	参数值
SCH/CCH 时隙长度/ms	10
SCH/CCH 信道上每一帧包含的时隙数	60
源无人机的数据包生成速率	1
每条链路上的丢包率	0~0.2
每个无人机的通信半径/m	200
数据包大小/Byte	900(根据文献[30]设置的)
信标帧/Byte	64
传输速率/(Mbit/s)	1
每个数据包允许的最大跳数	10
每个无人机的最大队列长度/packets	32

5.2 性能指标

在仿真中,我们仅对源节点生成的数据包的传输性能进行统计,并以此来评估每个路由协议的性能。本节考虑了以下性能指标:平均路由队列长度(Queue Length of Each Route,QLER)、平均拥塞重传次数(Retransmission Times Caused by Congestion of Each Route,RTCER)、平均链路重传次数(Retransmission Times Caused by Link Error of Each Route,RTLEER)、数据包投递成功率(Packet Delivery Ratio)和端到端传输时延(End-to-End Delay)。

在仿真中,源无人机节点生成的每个数据包的包头中都包含一个“路由队列长度”的字段。当源无人机节点生成数据包时,该字段的值设为 0。当数据包在路由过程中被转发至一个中间节点时,该字段中的值通过添加中间节点的已用路由队列长度来更新。当该数据包被交付至地面站时,地面站记录该数据包中“路由队列长度”字段所标识的值。在每次仿真结束时,地面站对所有“路由队列长度”字段求平均值,即可得到 QLER 的值。

此外,源 UAV 生成的每个数据包的包头中还包含一个“拥塞重传次数”字段和一个“丢包重传次数”字段。当源无人机节点生成数据包时,这两个字段的值均设为 0。在路由过程中,如果中间节点发生拥塞(中间节点路由队列饱和)而引起数据包重传,则“拥塞重传次数”字段的值递增 1。类似地,在路由过程某一跳转发中,如果丢包发生而引起数据包重传时,“丢包重传次数”字段的值递增 1。在每次仿真结束时,地面站对所有“拥塞重传次数”字段和“丢包重传次数”字段求平均值,即可得到 RTCER 和 RTLEER 的值。

数据包投递成功率是源节点生成的数据包被地面站成功接收的比率。端到端传输时延描述了从源节点的数据包被放入 SCH 信道时刻,到该数据包被地面站成功接收时刻的平均时间间隔。

5.3 仿真结果分析

本文用 RU 节点在仿真初始时刻路由队列中存储的数据包个数来表示 RU 节点的队列负载程度。如果初始时刻存储的数据包个数越多,则认为 RU 节点的队列负载程度越高。仿真中,我们使初始时刻 RU 节点的已存储数据包个数在 10,15,20,25,30(单位为 packets)之间变化来表示不同的队列负载程度。本节仿真了不同 RU 节点的队列负载程度下,不同路由协议的传输性能。

在图 4 和图 5 中,我们仿真了不同路由协议的 RTCER 值和 QLER 值。

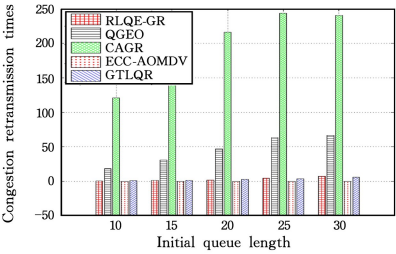


图 4 不同协议的平均拥塞重传次数
Fig. 4 RTCER of different protocols

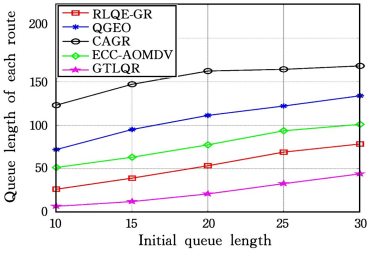


图 5 不同协议的平均路由传输队列长度
Fig. 5 QLER of different protocols

由图 4 和图 5 可以看出,CAGR 协议、QGEO 协议的 RTCER 值和 QLER 值明显高于其他 3 个协议。这是由于 CAGR 协议和 QGEO 协议没有加入负载均衡机制,RU 节点相比非 RU 节点距离目的节点更近。因此,在这两种协议下,ASN 中的数据包更倾向于在 RU 节点之间进行多跳转发,导致 RU 节点被过度使用。随着队列负载程度加剧,RU 节点会更快出现负载不均衡现象且更易发生拥塞。即使 RU 节点出现负载不均衡现象或者发生拥塞,这两种协议仍会使源节点的数据包在 RU 节点之间进行转发。因此,CAGR 协议、QGEO 协议的 RTCER 值和 QLER 值相对较高。此外,QGEO 协议的 RTCER 值和 QLER 值始终小于 CAGR 协议。这是因为 QGEO 协议会选择链路质量较高的多跳路径,所以它由于丢包而引起的重传次数比 CAGR 协议更少,如图 6 所示。因此,QGEO 协议的网络负载比 CAGR 协议更小,使得 QGEO 协议下 RU 节点负载不均衡程度小于 CAGR 协议。

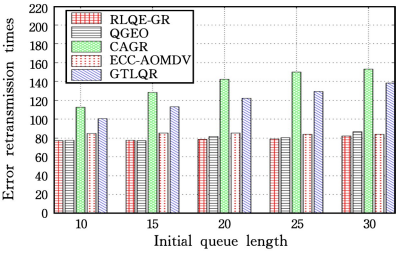


图 6 不同协议的平均路由链路重传次数
Fig. 6 RTLEER of different protocols

RLQE-GR 协议、GTLQR 协议以及 ECC-AOMDV 协议均加入了负载均衡机制。当 RU 节点的队列负载程度较高时,这 3 种协议更倾向于使网络中的数据包(RU 节点的数据包和源节点的数据包)尽可能地选择非 RU 节点作为中继

节点。这样便有效避免了 RU 节点被过度使用,缓解了负载不均衡现象。因此,这 3 种协议的 RTCER 值和 QLER 值相对较低。

从图 4 中可以看出,GTLQR 协议的 RTCER 值略高于 RLQE-GR 协议,这是贪婪算法无法实现全局最优化导致的。例如,对于节点 A 而言,其一跳邻节点 B 的可用队列长度大于一跳邻节点 C。因此,在 GTLQR 协议下,即使 B 的下一跳节点发生拥塞,节点 A 还是始终转发数据包至节点 B。此外,GTLQR 利用贪婪算法选择距离减少比率与路由队列使用率乘积较低的一跳邻节点作为下一跳节点。GTLQR 协议在做路由决策时,为距离减少比率和路由队列使用率这两个度量值分配了相同的权重系数。即对于距离汇聚节点较远的源节点而言,由于距离减少比率的分子增大,路由队列使用率将对下一跳节点的选择占据主导作用。因此,图 5 中 GTLQR 协议的 QLER 值反而小于 RLQE-GR 协议。

此外,从图 4 中还可以看出,ECC-AOMDV 协议的 RTCER 值几乎为 0。这是因为当节点发生拥塞时,ECC-AOMDV 协议会放弃选择与该节点相关的所有路由路径。然而,由图 5 可以看出,ECC-AOMDV 协议的 QLER 值高于 RLQE-GR 协议和 GTLQR 协议。这是因为 ECC-AOMDV 协议是一种拓扑路由协议。在两次路由表更新的时间间隔中,某些中间节点虽然没有发生拥塞,但是出现了严重的负载不均衡现象。而 ECC-AOMDV 协议并不能及时发现这种现象,只能在下次路由表更新时重新选择更优路径。

图 6 给出了不同协议的 RTLEER 性能。由图 6 可以看出,RLQE-GR 协议和 QGEO 协议的 RTLEER 值远远小于 CAGR 协议和 GTLQR 协议。这是因为 RLQE-GR 协议和 QGEO 协议在路由过程中考虑了链路质量,使数据包更倾向于在链路质量更高的路径上进行多跳转发。另外,从图中还可以看出,RLQE-GR 协议和 QGEO 协议的 RTLEER 值几乎相同。图 4—图 6 表明本文所设计的奖励函数形式能够在不牺牲链路质量性能的同时满足负载均衡的要求。此外,ECC-AOMDV 协议的 RTLEER 值和 RLQE-GR 协议的 RTLEER 值几乎相同。这是因为 ECC-AOMDV 协议基于 AOMDV 协议,其中链路质量较好的路由路径将会先返回应答包。因此,ECC-AOMDV 协议在某种程度上也能实现对链路质量的优化。

图 7 和图 8 给出了不同协议的数据包投递成功率性能和端到端传输时延。从图 7 可以看出,RLQE-GR 协议和 ECC-AOMDV 协议实现了更高的数据包投递成功率。然而,从图 8 可以看出,RLQE-GR 协议的端到端传输时延比 ECC-AOMDV 协议缩短了近 1/2。此外,相比其他协议,RLQE-GR 协议和 GTLQR 协议实现了更低的端到端传输时延。这主要是由于 RLQE-GR 协议和 GTLQR 协议在路由过程中考虑了节点的路由队列资源,因此减少了由中间节点发生拥塞造成的数据包重传的次数(见图 4),同时也减少了数据包在中间节点路由队列排队的时间。然而,从图 7 可以看出,GTLQR 协议的数据包投递成功率随着 RU 节点的队列负载程度增加而大幅下降,甚至当 RU 节点初始已用队列长度

等于 30 时,GTLQR 协议实现了比 CAGR 协议更低的数据包投递成功率。这是因为贪婪算法的局部队列资源最优化特点有可能造成远端数据包的丢失。换句话说,总是选择可用路由由队列长度较大的下一跳节点不一定能保证数据包朝着目的节点的方向进行转发。

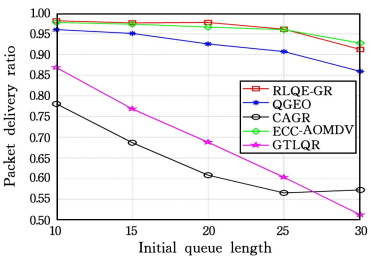


图 7 2D 场景下不同协议的数据包投递成功率
Fig. 7 Packet delivery ratio of different protocols in 2D scenario

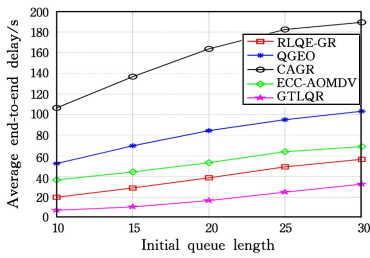


图 8 2D 场景下不同协议的数据包交付时延
Fig. 8 End-to-end delay of different protocols in 2D scenario

此外,我们还仿真了 3D 场景下不同协议的路由性能,而初始时刻 ASN 的网络拓扑如图 9 所示。其中源无人机节点的位置坐标为(500 m,1 000 m,50 m),其他设置与 2D 场景相同。由于篇幅限制,本文只给出了数据包投递成功率和端到端传输时延的性能曲线,分别如图 10 和图 11 所示。

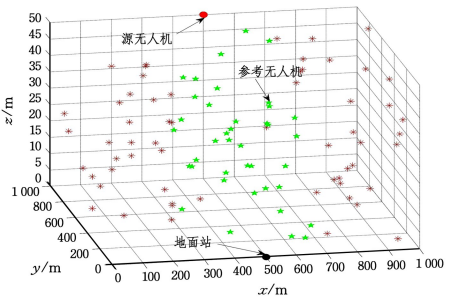


图 9 初始时刻 ASN 网络拓扑(3D 场景)
Fig. 9 ASN topology at initial time instant(3D scenario)

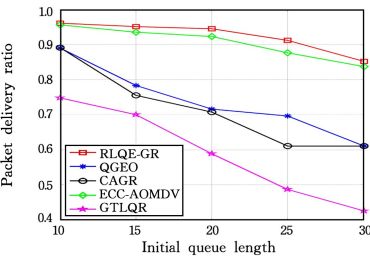


图 10 3D 场景下不同协议的数据包投递成功率
Fig. 10 Packet delivery ratio of different protocols in 3D scenario

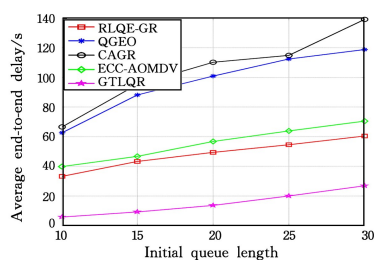


图 11 3D 场景下不同协议的数据包交付时延

Fig. 11 End-to-end delay of different protocols in 3D scenario

可以看出,在 3D 场景下,所提路由协议仍能实现更优的路由性能,说明所提路由协议在二维或者三维 ASN 网络场景中都适用。

结束语 针对地理路由协议在 FANET 中传输多跳数据业务时出现的负载不均衡现象,本文基于强化学习理论提出了 RLQE-GR 路由协议。RLQE-GR 协议将 FANET 路由问题抽象为强化学习(RL)任务。在该映射中,将 FANET 中的每个无人机抽象为一个 RL 状态,将无人机一次成功的数据包转发抽象为一个 RL 动作。此外,RLQE-GR 协议中引入了新的奖赏函数,该奖赏函数的值不仅与无人机节点地理位置和每跳链路质量相关,而且与无人机节点的可用路由队列长度密切相关。基于所设计的奖赏函数,RLQE-GR 协议利用 Q 函数分布式地更新所有状态-动作对的 Q 值。基于 Q 值,所有节点均采用贪婪策略转发数据包。此外,为了使所用的 Q 值能快速收敛,且最小化收敛过程带来的路由性能损失,RLQE-GR 协议设计了合理的信标机制对 Q 值进行更新迭代。当 Q 值收敛时,RLQE-GR 协议能够实现近乎最优的路由性能。与现有地理路由协议和负载均衡路由协议相比,本文协议能够有效缓解负载不均衡现象,同时优化路由路径的总链路质量。

这种协议仍然具有一定的局限性。本文在 2.1 节中假设 ASN 网络拓扑在一段足够长的时间内保持不变或者缓慢变化。因此,在这段时间内,强化学习模型中的状态转移概率可视为保持不变。进而通过在控制信道设置合适的信标帧广播周期,可使 Q 值在有限时间内收敛。然而,当网络动态性非常大时,即 ASN 拓扑可视为静态的时间间隔非常小时,需要设置极大的信标帧生成速率才可使 Q 值收敛。因此,在网络动态性非常大的场景下,RLQE-GR 路由协议实现 Q 值收敛将带来巨大的通信开销。在未来的工作中,可以将网络分成多个网格,并采用分级强化学习算法来实现多跳数据通信。

参 考 文 献

[1] ASADPOUR M, HUMMEL K A, GIUSTINIANO D, et al. Route or carry: Motion-driven packet forwarding in micro aerial vehicle networks[J]. IEEE Transactions on Mobile Computing, 2017, 16(3): 843-856.

[2] ZHANG C. Progress in Time Synchronization Technology for Wireless Sensor Networks[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2019, 36(6): 88-94.

[3] YANG J, GU Y H, XU Q, et al. A Node Sleeping Routing Algorithm for Underwater Wireless Sensor Networks[J]. Journal of Chongqing University of Technology (Natural Science), 2020, 34(1): 226-234.

[4] ASADPOUR M, VAN DEN BERGH B, GIUSTINIANO D, et al. Micro aerial vehicle networks: An experimental analysis of challenges and opportunities[J]. IEEE Communications Magazine, 2014, 52(7): 141-149.

[5] BIOMO J D M M, KUNZ T, ST-HILAIRE M. Routing in unmanned aerial ad hoc networks: A recovery strategy for greedy geographic forwarding failure[C]// 2014 IEEE Wireless Communications and Networking Conference (WCNC). 2014: 2236-2241.

[6] MOSTEFAOUI A, MELKEMI M, BOUKERCHE A. Localized routing approach to bypass holes in wireless sensor networks[J]. IEEE Transactions on Computers, 2013, 63(12): 3053-3065.

[7] HUANG H, YIN H, MIN G, et al. Coordinate-assisted routing approach to bypass routing holes in wireless sensor networks[J]. IEEE Communications Magazine, 2017, 55(7): 180-185.

[8] HUANG H, YIN H, MIN G, et al. Energy-aware dual-path geographic routing to bypass routing holes in wireless sensor networks[J]. IEEE Transactions on Mobile Computing, 2017, 17(6): 1339-1352.

[9] XIA Y, QIN X, LIU B, et al. A greedy traffic light and queue aware routing protocol for urban VANETs[J]. China Communications, 2018, 15(7): 77-87.

[10] WANG X, LIU X, WANG M, et al. Energy-Efficient Spatial Query-Centric Geographic Routing Protocol in Wireless Sensor Networks[J]. Sensors, 2019, 19(10): 2363.

[11] SINGH P, CHEN Y C. Energy efficient greedy forwarding based on residual energy for wireless sensor networks[C]// 2018 27th Wireless and Optical Communication Conference (WOCC). 2018: 1-6.

[12] JUNG W S, YIM J, KO Y B. QGeo: Q-learning-based geographic ad hoc routing protocol for unmanned robotic networks[J]. IEEE Communications Letters, 2017, 21(10): 2258-2261.

[13] ZHANG K, ZHANG W, LI W, et al. Research of applicability for UAV Ad Hoc networks proactive routing protocols[J]. Computer Engineering and Applications, 2010, 46(2): 4-6, 18.

[14] TABBANA F. Performance Comparison and Analysis of Proactive, Reactive and Hybrid Routing Protocols for Wireless Sensor Networks[J]. International Journal of Wireless & Mobile Networks, 2020, 12(4): 20.

[15] ANSHORI H A, ABDUROHMAN M. Comparison of Reactive Routing Protocol Dynamic Manet on Demand and Ad Hoc on Demand Distance Vector for Improving Vehicular Ad hoc Network Performance[J]. Advanced Science Letters, 2015, 21(1): 20-23.

[16] JIANG J, HAN G. Routing protocols for unmanned aerial vehicles[J]. IEEE Communications Magazine, 2018, 56(1): 58-63.

[17] LEMMON C, LUI S M, LEE I. Geographic Forwarding and Routing for Ad-hoc Wireless Network: A Survey[C]// Proceedings

dings of 2009 Fifth International Joint Conference on INC, IMS and IDC. Seoul, South Korea, 2009; 188-195.

[18] AOUIZ A A, HACENE S B, LORENZ P. Channel Busyness Based Multipath Load Balancing Routing Protocol for Ad hoc Networks[J]. IEEE Network, 2019, 33(5): 118-125.

[19] DSOUZA M B, MANJIAH D H. Congestion Free And Bandwidth Aware Multipath Protocol for MANET[C]//2019 1st International Conference on Advances in Information Technology (ICAIT). 2019; 267-270.

[20] POURBEMANY J, MIRJALILY G, ABOUEI J, et al. Load Balanced Ad-Hoc On-Demand Routing Based on Weighted Mean Queue Length Metric[C]//Iranian Conference on Electrical Engineering (ICEE). 2018; 470-475.

[21] SINGH G, SHARMA A K, BAWA O S, et al. Effective Congestion Control In MANET[C]//2020 International Conference on Intelligent Engineering and Management (ICIEM). 2020; 86-90.

[22] REN W, BEARD R W, ATKINS E M. Information consensus in multivehicle cooperative control [J]. IEEE Control Systems Magazine, 2007, 27(2): 71-82.

[23] SON J, CHOI S, CHA J. A brief survey of sensors for detect, sense, and avoid operations of small unmanned aerial vehicles [C]//Proceedings of 17th International Conference on Control, Automation and Systems. Jeju, South Korea, 2017; 279-282.

[24] LUO W. An efficient sensor-mission assignment algorithm based on dynamic alliance and quantum genetic algorithm in wireless sensor networks[C]//Proceedings of 2010 International Conference on Intelligent Computing and Integrated Systems. Guilin, China, 2010; 854-857.

[25] TALGINI A, SHAKARAMI V, SHEIKHOLESLAM F, et al. Aerial node placement in wireless sensor networks using Fuzzy K-means clustering[C]//Proceedings of 8th International Conference on e-Commerce in Developing Countries; With Focus on e-Trust. Mashhad, Iran, 2014; 1-7.

[26] OMAR H A, ZHUANG W, LI L. VeMAC: A TDMA-based MAC protocol for reliable broadcast in VANETs [J]. IEEE Transactions on Mobile Computing, 2013, 12(9): 1724-1736.

[27] SENGOKU M, TAMURA H, MASE K, et al. A routing problem on ad-hoc networks and graph theory[C]//Proceedings of International Conference on Communication Technology Proceedings. Beijing, China, 2000; 1710-1713.

[28] NOWE A, BRYN T. A Gentle Introduction to Reinforcement Learning[C]//Proceedings of International Conference on Scalable Uncertainty Management. Nice, France, 2016; 18-32.

[29] KUIPER E, NADJM-TEHRANI S. Mobility models for UAV group reconnaissance applications[C]//2006 International Conference on Wireless and Mobile Communications (ICWMC'06). 2006; 33.

[30] WANG W, DONG C, WANG H, et al. Design and implementation of adaptive MAC framework for UAV ad hoc networks [C]//Proceedings of 12th International Conference on Mobile Ad-Hoc and Sensor Networks. Hefei, China, 2016; 195-201.

 **HUANG Xin-quan**, born in 1993, post-graduate. His main research interests include multi-agent systems, and flying ad-hoc network.

 **LIU Ai-jun**, born in 1970, professor. His main research interests include satellite communication system theory, signal processing, channel coding, and information theory.

(责任编辑:喻黎)