Computer Science and Application 计算机科学与应用, 2022, 12(2), 304-314
Published Online February 2022 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2022.122030

Hans汉斯

一种基于增强学习的飞行自组网地理路由协议

杨　斌[1]，王辛果[2]

[1]成都信息工程大学计算机学院，四川　成都
[2]中国航空工业无线电电子研究所，上海

收稿日期：2022年1月12日；录用日期：2022年2月9日；发布日期：2022年2月17日

摘　要

飞行自组网(FANETs，Flying Ad-Hoc Networks的缩写)是航空平台组建的自组网网络，不依赖固定通信设施，具备部署快，健壮性高等优势，可广泛应用于应急通信和军事等场景。然而，由于航空平台的移动速率更高，飞行自组网的网络拓扑动态性更高，现有的移动自组网路由协议无法直接适用。在现有的FANETs的路由协议中，基于地理位置的路由协议相较于其他路由协议具有很大的优势，它仅依靠于节点的地理坐标，不建立和维护端到端连接。但是，现有的基于地理位置的路由协议也存在最优转发节点选择困难和端到端延迟较高等问题。为此，本文提出了一种基于增强学习的地理路由协议，称为QEgr。该协议基于Q-Learning算法综合考虑了链路稳定性和延迟，并使用了ns3-gym模拟器与其他路由协议进行了比较。实验表明，与GPSR、Q-Grid、Q-Geo等经典地理路由协议相比，QEgr具有更低的端到端延迟和更高的数据包发送成功率。

关键词

FANETs，Q-Learning，地理路由协议，ns3-Gym

A Geographic Routing Protocol for FANETs Based on Reinforcement Learning

Bin Yang[1], Xinguo Wang[2]

[1]School of Computer Science, Chengdu University of Information Technology, Chengdu Sichuan
[2]China Aviation Industry Institute of Radio Electronics, Shanghai

Received: Jan. 12th, 2022; accepted: Feb. 9th, 2022; published: Feb. 17th, 2022

Abstract

FANETs (the abbreviation of Flying Ad-Hoc Networks) is an Ad-Hoc network formed by aviation platforms. It does not rely on fixed communication facilities. It has the advantages of fast deploy-

文章引用: 杨斌, 王辛果. 一种基于增强学习的飞行自组网地理路由协议[J]. 计算机科学与应用, 2022, 12(2): 304-314.
DOI: 10.12677/csa.2022.122030

ment and high robustness. It can be widely used in emergency communications and military scenarios. However, due to the higher mobile speed of the aviation platform and the higher dynamics of the network topology of the FANETs, the existing mobile Ad-Hoc network routing protocol cannot be directly applied. Among the existing FANETs routing protocols, the geographical location-based routing protocol has great advantages over other routing protocols. It only relies on the geographical coordinates of the nodes and does not establish and maintain end-to-end connections. However, the existing geographic location-based routing protocols also have problems such as difficulty in selecting the optimal forwarding node and high end-to-end delay. To this end, this paper proposes a geographic routing protocol based on enhanced learning called QEgr. This protocol is based on the Q-Learning algorithm and considers the link stability and delay, and uses the ns3-gym simulator to compare with other routing protocols. Experiments show that, compared with classic geographic routing protocols such as GPSR, Q-Grid, and Q-Geo, QEgr has a lower end-to-end delay and a higher packet transmission success rate.

Keywords

FANETs, Q-Learning, Geographic Routing Protocol, ns3-Gym

1. 引言

飞行自组织网络 FANETs (Flying Ad-Hoc Networks)因其多功能性、易于部署、高移动性和低运营成本而越来越受欢迎[1]。FANETs 通常由无人驾驶飞行器(UAV)组成，可以自主飞行或远程控制。自开始用于监视和救援以来，无人机已被全球各地的军队使用[2]。如今，随着技术的进步，无人机已广泛应用于各个领域的敏感任务，例如交通监控、灾害监控、其他自组织网络的中继、遥感和野火监测等等，在各个领域都发挥着越来越重要的作用。由于无人机节点的高移动性而导致的拓扑频繁变化，使得 FANETs 中的路由设计变得具有挑战性。

在现有的许多 FANETs 的路由协议中，主动路由在转发数据包之前创建路由表，但是维护路由表信息会带来更大的控制开销。反应式路由会在转发数据包时创建路由，但由于存在发现交付路径这个过程，它带来了更大的延迟。混合路由在主动路由和反应式路由之间进行了权衡，它结合了主动路由的低延迟和反应路由的低网络控制开销的优点，但是它主要适用于网络拓扑稳定的网络。基于地理位置的路由协议仅利用邻居的位置信息，尽管开销减少了，但是由于其无法感知整个网络拓扑的变化，会造成路由空洞，从而影响路由协议的性能。

针对这些由于拓扑结构频繁变化而引起的问题，我们考虑采用自适应和自治的路由协议来解决，这意味着 FANETs 中的路由协议应该能够通过检测环境的变化来发现一个稳定可靠的邻居来发送数据。在这种情况下，我们提出了一种基于增强学习的飞行自组网地理路由协议 QEgr。Q-Learning 是一种以环境反馈为输入的自适应机器学习技术，这有助于自适应的路由设计。在 Q-Learning 中，智能体可以根据环境反馈的奖励不断调整自己的行动策略，以更好地适应动态和不可持续的拓扑结构[3]。为了克服网络拓扑频繁变化带来的高延迟和高丢包率的问题，在提出的 QEgr 路由协议中，我们综合考虑了链路稳定性以及延迟，以此来优化拓扑频繁变化带来的限制。此外，针对现有的基于 Q-Learning 的路由协议如 Q-Grid [4]、Q-Geo [5]存在参数(学习率和折扣因子)固定等不足，在提出的 QEgr 中，我们分别使用链路稳定性和

延迟的变化来确定学习率和折扣值，以适应 FANETs 的高度动态性。

2. 相关工作

2.1. 基于地理位置的路由协议

基于移动性预测的地理路由(MPGR) [6]是对贪婪周边无状态路由(GPSR)协议的增强。在 MPGR 中，为了最小化无人机的高移动性对路由性能的影响，将移动性预测方案与地理路由机制合并。通过利用高斯概率密度函数来采用移动性估计技术，以在特定时间间隔预测网络中节点的位置分布。仿真结果显示，与 AODV 和 GPSR 相比，MPGR 在 PDR 和传输延迟方面显示出更好的结果，而开销却更少。但是，其并没有考虑发生路由无效的情况。

基于地理位置的路由(GBR) [7]，使用无人机的地理位置和速度的链路预测技术与贪婪的转发方案相结合用以选择下一跳节点。仿真结果表明，该路由协议能降低路由开销，但是在稀疏网络中，丢包和延迟的可能性比较高。

FANETs 中无人机的主要应用之一是在灾后行动中，例如搜索和救援。[8]中的作者提出了一种称为 LADTR 的延迟容忍路由方案，该方案利用了位置辅助转发与存储转发方案相结合的方法。基于 Guess-Markov 模型，基于无人机位置和速度信息的位置预测方案，从机载 GPS 装置中获取用于估计无人机在网络中的未来位置。另外，在 FANETs 中引入了用于运送的 UAV2，以改善具有要发送的数据分组的 UAV 与 GCS 之间的路径连通性，从而减少了端到端延迟并提高了数据包传送率。仿真结果表明，与其他传统的路由方案相比，LADTR 具有更好的控制开销，但是对 LADTR 在 2D 空间(或高度恒定)中的 UAV 机动性的假设不能完全表征 UAV (3D)机动性的特征，可能会使 LADTR 在实际实施中效率低下。

2.2. 基于 Q-Learning 的路由协议

基于 Q-Learning 方法的路由协议有望处理 FANETs 中的动态变化。Q-Grid 是为 VANETs 设计的协议，它利用宏观(通过查询离线学习的 Q 值表获得的最优下一跳网格)和微观方面(最优下一跳网格中的特定车辆)来执行路由决策，将区域划分为不同的网格。通过这种方法，Q-Grid 使用 Q-Learning 算法计算特定目的地的相邻网格之间各种移动的 Q 值。执行的模拟表明，与其他现有的基于位置的路由协议相比，Q-Grid 具有优势。

基于 Q-Learning 的地理路由 Q-Geo 提出了一种系统，可以在高移动性场景中最小化网络开销。作者将 Q-Geo 的性能与使用 ns3 模拟器的其他方法进行了比较，结果表明 Q-Geo 与其它解决方案相比具有更高的数据包传输率和更低的网络开销。

在 MANETs 的背景下，[9]提出了一种基于 Q-Learning 的 CSMA/MAS 协议。在这种方法中，网络中的每个节点都能够同步，然后以循环方式参与，而不必处理争用冲突。在网络层，该方法对 Q-Geo 和 Q-Grid 进行了多次修改。结果表明，与现有传输协议相比，这种传输协议方法提供了更高的数据包到达率和更低的端到端延迟。

QNGPSR [10]是一种路由协议，其灵感来自于自组织网络的 GPSR 协议。它旨在通过使用强化学习来执行下一跳选择来减少网络延迟。结果表明，与 GPSR 的性能相比，QNGPSR 提供了更高的数据包传输率和更低的端到端延迟。

综上所述，基于地理位置的路由协议不能自适应和自主地发现可靠的通信链路，基于 Q-Learning 的路由协议也存在一些不足，一方面很少有路由协议基于 Q-Learning 的基础之上综合考虑链路稳定性和网络延迟，另一方面，Q-Learning 的参数，学习率和折扣因子不能根据网络状况进行自适应调整。针对这些问题，我们提出了 QEgr 协议，在 Q-Learning 的基础上同时考虑链路稳定性和网络延迟用于优化下一

跳节点的选择，并且让 Q-Learning 的参数自适应调整，以适应 FANETs 的高动态性。

3. 网络场景

在本文中，我们的无人机网络是由多架无人机和一个地面站组成，如图 1 所示。地面站作为接收数据的目的节点，一架无人机被视为发送数据的源节点，其余无人机作为中继节点转发数据。核心问题是如何找到一条最优路径，使沿路径传输的数据能够以较低的延迟和较高的数据投递率到达目的地。

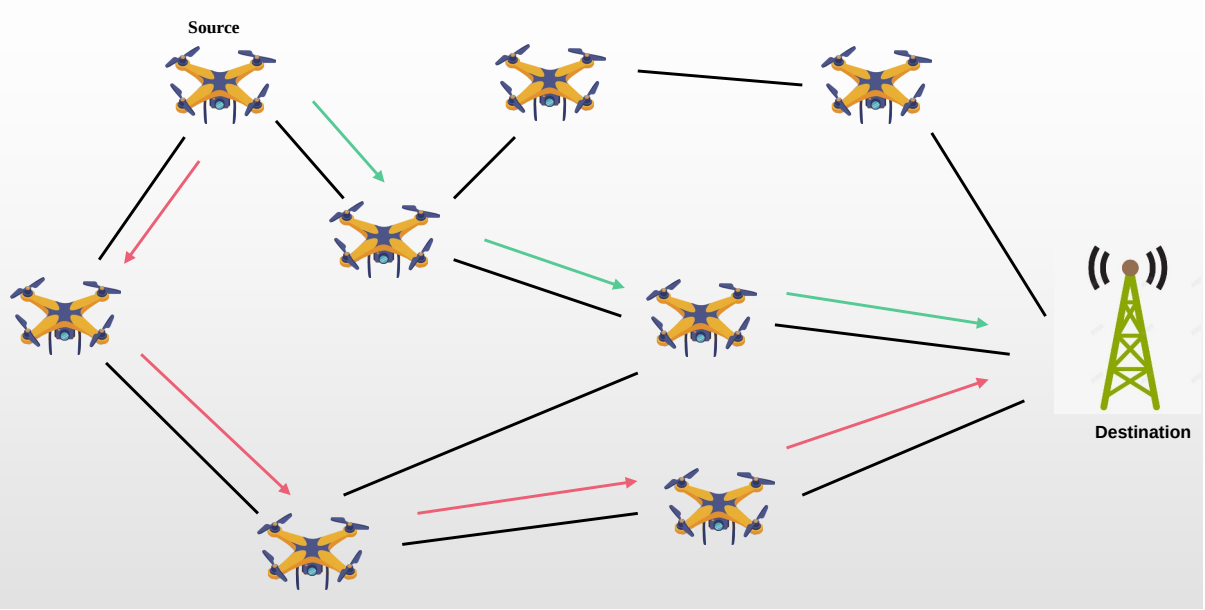

Figure 1. Multi-hop routing in FANETs
图 1. FANETs 中的多跳路由

4. QEgr 协议设计

在本节中，将介绍 QEgr 路由协议，这是一种基于 Q-Learning 的地理路由协议。它由路由邻居发现、Q-Learning 和路由决策三部分组成，其流程图如图 2 所示。

4.1. 路由邻居发现模块

在 QEgr 路由协议中，每个节点都会创建一个邻居表，通过定期发送 HELLO 包来更新邻居节点的信息。每个 HELLO 消息包括邻居的地理位置、移动模式(即移动速度和方向)、折扣值和延迟。邻居表不仅存储了 HELLO 包中的信息，还存储了学习率、HELLO 包达到时间、链路稳定性和 Q 值。每个节点利用其邻居表的信息来感知局部的网络状况。如果某个邻居的信息在指定的时间后没有刷新，则会从邻居表中删除。

4.2. 链路稳定性度量和延迟度量

为了选择一个链路稳定以及延迟较低的节点作为下一跳，在本节中，我们先对链路稳定性以及延迟进行度量。

4.2.1. 链路稳定性度量

在本文中，我们设定链路稳定性度量由两部分组成，包括链路质量和链路持续因子。

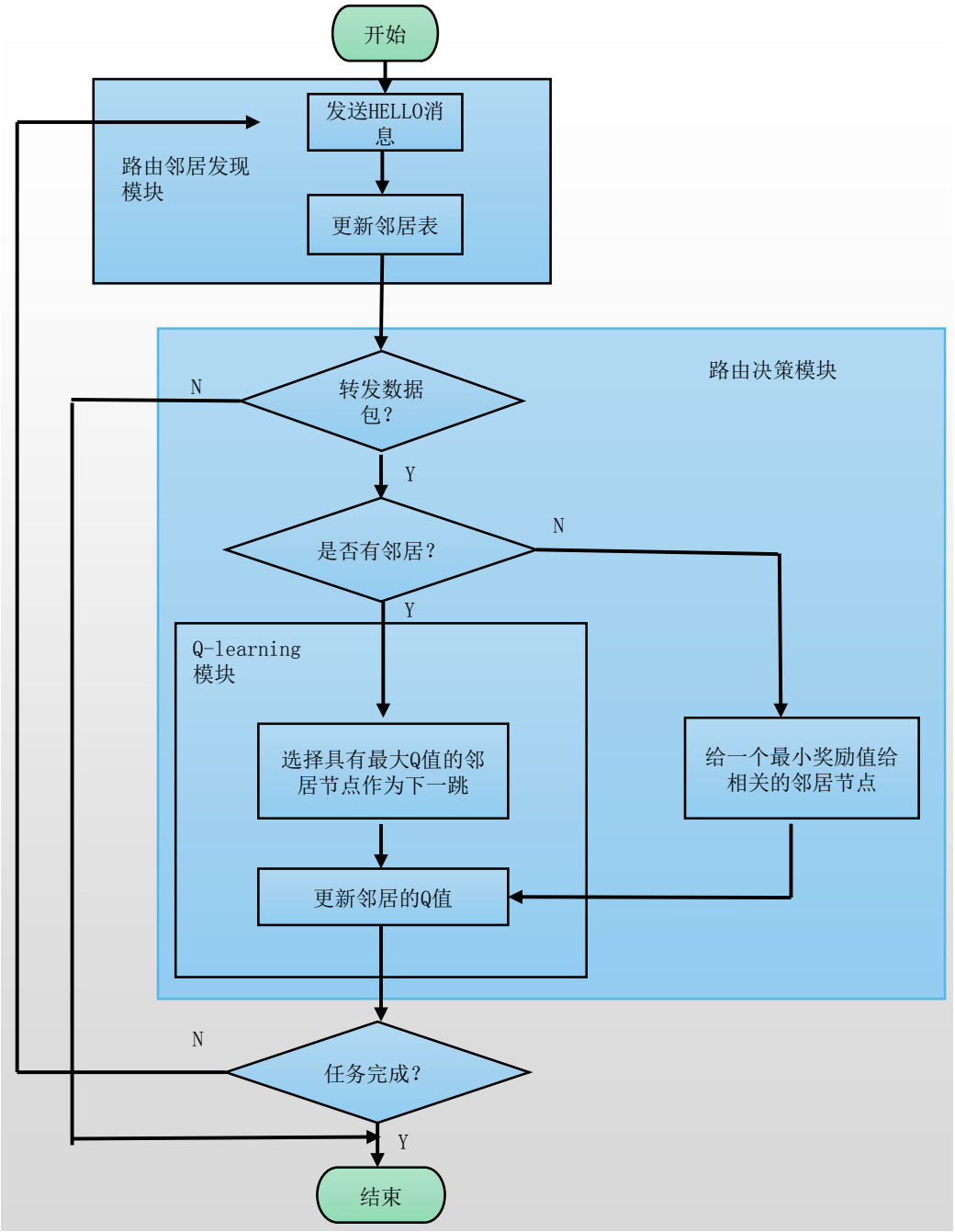

Figure 2. QEgr flow chart
图 2. QEgr 流程图

假设节点 i 和 j 在彼此的通信范围内，用 $STB_{i,j}$ 表示为 i 和 j 之间的链路稳定性，$LQ_{i,j}$ 表示节点 i 和 j 之间的链路质量，$PF_{i,j}$ 表示节点 i 和 j 之间的链路持续因子，则 $STB_{i,j}$ 可以表示如下：

$$STB_{i,j} = \lambda_1 LQ_{i,j} + \lambda_2 PF_{i,j} \tag{1}$$

其中，λ_1 和 λ_2 是给定的权重系数。$STB_{i,j}$ 的值越大，表明该链路的稳定性更好。

$$\lambda_1 + \lambda_2 = 1 \tag{2}$$

需要指出的是，$STB_{i,j}$的值被限制在 0 到 1 的范围内。接下来，我们将详细介绍提到的指标。我们利用前向和反向传递率来衡量链路质量，可以定义为：

$$LQ_{i,j} = \gamma^f * \gamma^r \tag{3}$$

其中，γ^f为前向传递率，表示节点 j 成功接收到节点 i 发送的数据包的概率，γ^r 为反向传递率，表示发送节点 i 成功收到 ACK 数据包的概率。如果可以成功接收到来自前向和反向的数据包，则 $LQ_{i,j}$ 等于 1。在路由协议中可以使用 HELLO 消息估计 γ^f 和 γ^r，γ^f 和 γ^r 可以被如下表示：

$$\begin{cases} \gamma_n = \partial * \gamma_{n-1} + (1-\partial) * h_n \\ \gamma_0 = 0 \end{cases} \tag{4}$$

$$h_n = \begin{cases} 1, \text{如果收到第} n \text{个 HELLO 消息} \\ 0, \text{其他} \end{cases} \tag{5}$$

参数 α 的取值范围是[0,1)。当 α 等于 0 时，传递率仅取决于当前的链路质量，随着 α 增大，传递率的估计会更平均和更稳定。

大多数无人机(UAV)都配备了全球定位系统(GPS)用来获取位置信息。假设每个节点都能获取自己的位置，节点 i 的位置坐标为(x_i, y_i, z_i),链路持续因子 $PF_{i,j}$ 表示当前节点 i 和 j 在距离上的接近程度，表示如下：

$$PF_{i,j} = \frac{R - d_t}{R} \tag{6}$$

其中 R 是无人机的通信半径，d_t 是两个节点的欧几里德距离。

$$d_t = \sqrt{\left[x_j - x_i\right]^2 + \left[y_j - y_i\right]^2 + \left[z_j - z_i\right]^2} \tag{7}$$

$PF_{i,j}$ 越大，两个节点越靠近，它们之间链接断开就需要更长的时间。

4.2.2. 延迟度量

假设节点 i 选择节点 j 作为下一跳转发数据包，由于数据包在无线媒体中以光速传播，因此在数百米量级的通信范围内，传播延迟可以忽略不计。节点 i 到节点 j 的单跳延迟 $D\_hop_{i,j}$ 由介质访问延迟(MAC 延迟)、排队延迟和传输延迟组成。节点 i 到节点 j 的单跳延迟 $D\_hop_{i,j}$ 表示如下：

$$D\_hop_{i,j} = D\_mac_{i,j} + D\_que_{i,j} + D\_tr_{i,j} \tag{8}$$

其中 $D\_mac_{i,j}$ 表示 MAC 延迟，它是媒体访问协议成功传送数据帧或者在重复传输失败的情况下丢弃数据包所需要的时间。$D\_que_{i,j}$ 表示排队延迟，即数据包进入传输队列到数据包从队列头部开始传输的时间。$D\_tr_{i,j}$ 表示传输延迟，即数据包从第一个比特开始传输到最后一个比特传输完毕所用的时间。

MAC 延迟的计算公式为：

$$D\_mac_{i,j} = T_{ack} - T_{send} \tag{9}$$

T_{ack} 是节点 i 从邻居节点 j 收到 ACK 包的时刻，T_{send} 是节点 i 向邻居节点发送数据包的时刻。$D\_mac_{i,j}$ 延迟由 MAC 层提供。

4.3. Q-Learning 模块

4.3.1. Q-Learning 模型

1) Q-Learning 简介：Q-Learning 是强化学习的主要算法之一[11]，Q-Learning 的基本思想是从与环境

的交互中学习，它使用奖励函数从环境中获取反馈。通过 Q 值，智能体可以评估当前状态下的动作有多好，并在下一步做出更好的动作。智能体的目标是最大化长期累积奖励的期望。Q 值的迭代公式如下：

$$Q(s_t,a_t)\leftarrow Q(s_t,a_t)+\partial\left[R(s_t,a_t)+\gamma\max_a Q(s_{t+1},a)-Q(s_t,a_t)\right] \tag{10}$$

其中，$Q(s_t, a_t)$是在时间 t 选择动作 a_t 时当前状态的 Q 值，$R(s_t, a_t)$表示智能体在状态 s_t 时采取动作 a_t 的直接奖励值，$\max Q(s_{t+1}, a)$表示智能体在下一个状态 s_{t+1} 中选择可能的动作 a 时的最大 Q 值。

2) 在我们的 QEgr 路由协议中，考虑一个来自源 UAV 的数据包通过多跳传输到达地面站，整个网络被视为一个环境，网络中的每个数据包代表一个智能体，持有该数据包的节点被认为是智能体的一个状态，因此，网络中所有节点的集合就是状态空间。一个节点只能选择一跳邻居节点作为下一跳，因此智能体动作的集合就是邻居的集合。例如，当数据包位于节点 i 时，与该数据包关联的当前状态就是 s_i，动作 $a_{i,j}$ 表示要从节点 i 转发到邻居节点 j 的数据包(智能体)的决定，通过这个动作，智能体的状态从 s_i 移动到 s_j 并且智能体收到该动作的奖励值。我们使用的利用与开发策略为 ε-greedy，其中 $\varepsilon = 0.9$。

4.3.2. 奖励函数

在我们的基于 Q-Learning 的路由协议中，奖励函数可以定义为：

$$R(s_t,a_t)=\begin{cases}R_{\max}, \text{当} s_{t+1} \text{是目的地}\\ R_{\min}, \text{当} s_t \text{是局部最大值}\\ \dfrac{\Delta d_{i,j} * STB_{i,j}}{D\_hop_{i,j} * R_{pts}}, \text{其他}\end{cases} \tag{11}$$

如果下一个状态是目的地，那么就是等式中的第一项，最大奖励值 $R_{\max}$，第二项是为了避免路由空洞的最小奖励值。在地理路由中，节点在不知道整个网络拓扑的情况下可能会遭遇路由空洞问题，这个过程会导致网络成本的增加，为了解决这个问题，我们设置了一个最小奖励 $R_{\min}$。最后一项则是其他情况的时候，其中$\Delta d_{i,j}$是节点 i 和 j 到目的地的距离差，$STB_{i,j}$是节点 i 与节点 j 之间的链路稳定性，$D_hop_{i,j}$是节点 i 与节点 j 之间的一跳延迟，R_{pts}为一个固定值，是为了防止奖励值过度增加。

4.3.3. 自适应 Q-Learning 参数

在 Q-Learning 中，学习率决定了新获取的信息覆盖旧信息的程度。如果学习率较高，则 Q 值更新得就更快。折扣因子 γ 代表未来 Q 值期望的稳定性，折扣因子值高表示未来 Q 值预期稳定，而低值表示 Q 值预期脆弱。大多数现有的基于 Q-Learning 的路由协议都是固定的学习率和折扣值，然而，在 FANETs 中，节点之间的链接不稳定，固定的学习率和折扣值不能适应拓扑的频繁变化。因此，我们分别使用链路稳定性和延迟的变化来确定学习率和折扣值。学习率可以表示为：

$$\partial_{i,j}=\begin{cases}1-e^{-STB_{i,j}} & \sigma_{i,j}\neq 0\\ 0.3 & \sigma_{i,j}=0\end{cases} \tag{12}$$

其中，$STB_{i,j}$是节点 i 到节点 j 之间的链路稳定性的度量，$\sigma_{i,j}$表示链路稳定性的方差。折扣值可以表示为：

$$\gamma_i=\frac{\sigma_{d,i}^2}{\max\sigma_d^2}*(\gamma_{\max}-\gamma_{\min})+\gamma_{\min} \tag{13}$$

其中 $\sigma_{d,i}^2$ 是节点 i 的延迟的方差，$\max\sigma_d^2$ 表示所有节点延迟方差的最大值，$\gamma_{\max}$ 和 $\gamma_{\min}$ 是常数值。

5. 性能评估

我们使用 ns3-gym 仿真器对 QEgr 路由协议进行了模拟，并将其与 GPSR，Q-Grid，Q-Geo 进行了比

较。ns3-gym 即结合 ns3 和 OpenAl Gym 的系统框架，用于基于强化学习的网络研究[12]。我们在表 1 中总结了有关我们的实验参数的详细信息。模拟场景由 30 个节点组成，部署在 500 m × 500 m × 500 m 的区域中。我们的 MAC 协议采用 IEEE 802.11a，无线传播的范围是 180 m。每个节点都安装了 RWP 移动模型，每个节点的速度为 0~15 m/s。每 100 ms 生成一个 HELLO 消息，每个邻居条目的生命周期为 300 ms。我们随机选择一个节点作为源节点向目的节点传输数据，除了目的节点外的其余节点为中继节点。源节点周期性发出数据包，其数据间隔设为不同的值以进行比较。

Table 1. Main simulation experiment parameters
表 1. 主要仿真实验参数

参数名	设置值
区域大小	500 m × 500 m × 500 m
节点数量	30
节点速度	0~15 m/s
无线传播范围	180 m
HELLO 间隔	100 ms
HELLO 条目过期时间	300 ms
MAC	IEEE 802.11a
数据包大小	128 Bytes
路径损耗模型	Free-space
移动模型	Random Waypoint Model

为了评估 QEgr，我们选择了 PDR、平均重传次数、平均端到端延迟和网络开销作为指标。在仿真实验中，源节点以不同的时间间隔发送 1000 个数据包，对于每个间隔，我们重复模拟 100 次，其仿真结果分析如下。

1) PDR 分析

数据包到达率 $P_{success}$ 计算公式如(14)所示。

$$P_{success} = \frac{Num_{rcv}}{Num_{send}} \tag{14}$$

在公式(14)中，Num_{send} 表示源节点发送的数据包的个数，Num_{rcv} 表示目的节点收到的数据包的个数。

图 3 表明，QEgr 协议的数据包到达率较 Q-Geo 有提高。Q-Grid 和 Q-Geo 相较于 GPSR 采用了强化学习技术，在高度动态的拓扑结构中能够选择相对稳定的邻居作为中继转发点，所以 PDR 有明显的提升。此外，Q-Geo 相较于 Q-Grid 考虑了链路错误和节点位置错误，选择的中继节点更加可靠稳定，故较 Q-Grid 数据包到达率有所提高。在 QEgr 中，我们考虑了链路的质量以及链路的持续时间，这对于数据包的成功传输很有利，故较 Q-Geo 的数据包到达率有提高。

2) 平均重传次数分析

图 4 表明，Q-Grid 在没有考虑链路错误和位置错误的情况下会增加重传次数。QEgr 在考虑了链路稳定性的情况下，选择的中继节点更加可靠稳定，它提供了一个较可靠的网络，重传次数就有所下降。

3) 平均端到端时延分析

平均端到端时延 D_{aver_ete} 计算公式如(15)所示

$$D_{aver\_ete} = \frac{\sum T_{rcv}}{num_{rcv}} \tag{15}$$

在式(15)中，ΣT_{rcv} 表示目的节点收到的所有数据包的传输时间，num_{rcv} 表示目的节点收到的数据包数目。

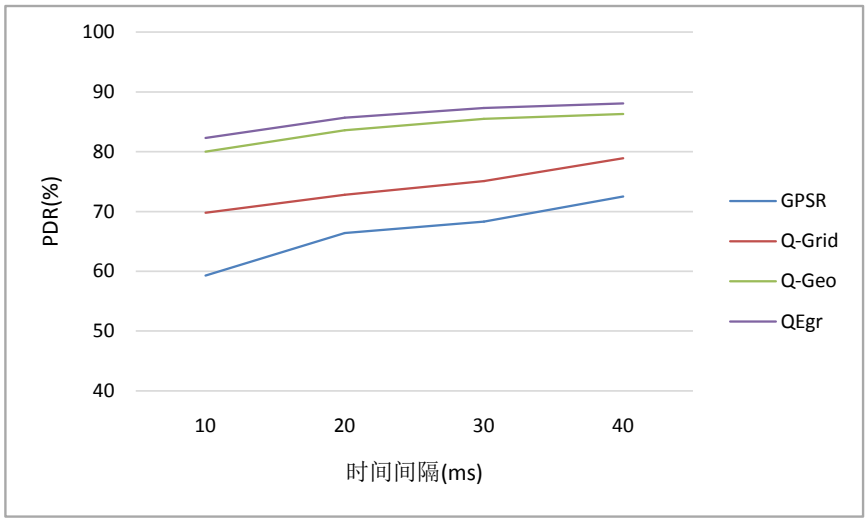

Figure 3. Data packet transmission success rate
图 3. 数据包发送成功率

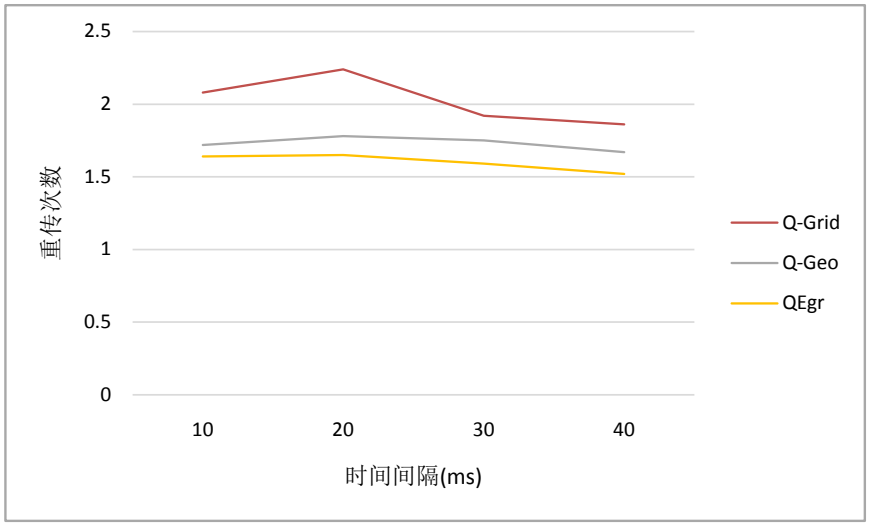

Figure 4. Average number of retransmissions
图 4. 平均重传次数

图 5 说明了端到端延迟比较结果。GPSR 的延迟最大，这是因为由于实验场景的不断移动，模拟拓扑中会产生空隙区域，为了避免这个空白区域，GPSR 经常使用周边转发而不是贪婪转发，这会显著增加延迟。与 GPSR 不同，在 Q-Grid 和 Q-Geo 中，当源节点无法选择转发候选节点时，它会一直持有这个数据包，直到候选节点出现在它的通信范围内。在 QEgr 中，每个节点在选择下一跳节点时，总是会选择 Q 值最大的那个作为下一跳，将每个数据包的延迟纳入到了 Q 值的计算中，根据公式(11)第三个等式可知，延迟越小获得 Q 值就会越大，所以，QEgr 中端到端的时延有所下降。

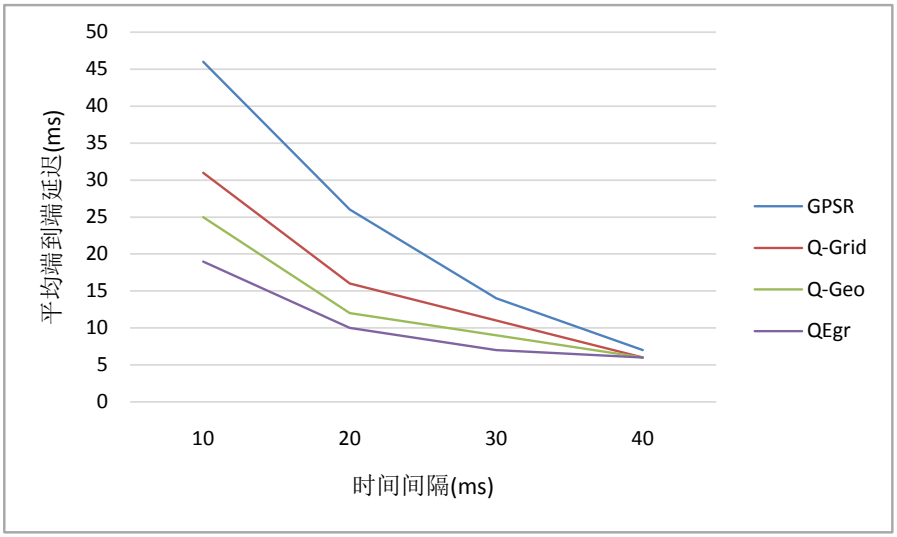

Figure 5. Average end-to-end delay
图 5. 平均端到端时延

4) 平均控制开销率分析

如图 6 所示，QEgr 的平均控制开销率最高，这是因为 QEgr 协议在 HELLO 消息中添加了每个邻居的位置、移动模式等信息，造成了 HELLO 消息开销高于其余三种协议。故 QEgr 协议实际上是通过牺牲一些控制开销来换取数据包到达率和端到端延迟性能的提升。

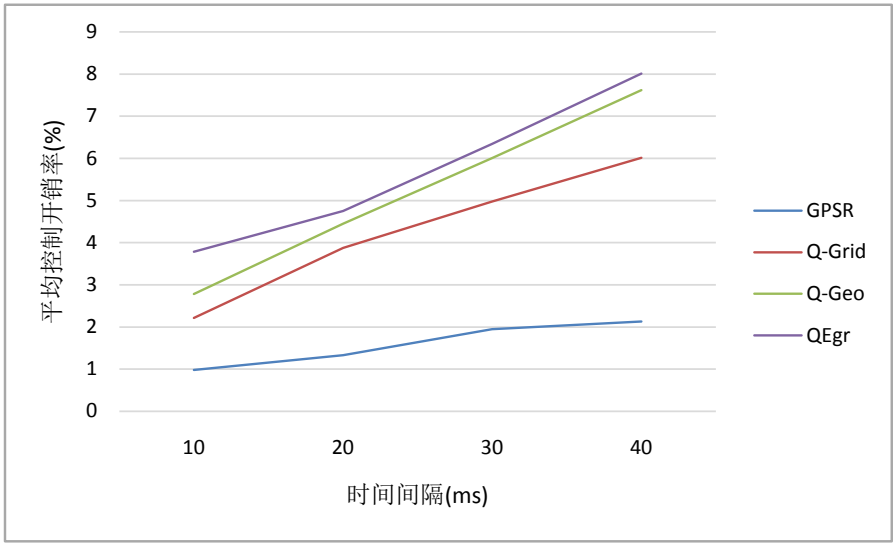

Figure 6. Average control overhead rate
图 6. 平均控制开销率

6. 结论

本文提出了 QEgr，一种基于增强学习的地理路由协议。所提出的方法将使用强化学习把链路稳定性和延迟结合在一起，目标是提出一种更好地适应 FANETs 高度动态变化的协议，从而提高网络可靠性和其它性能。使用 ns3-gym 模拟器评估表明，QEgr 具有更低的端到端延迟和更高的数据包达到率。

基金项目

本论文受四川省科技计划项目 2020YFG0422 支持。

参考文献

[1] Wang, J.J., Jiang, C.X., Han, Z., Ren, Y., Maunder, R.G. and Hanzo, L. (2017) Taking Drones to the Next Level: Cooperative Distributed Unmanned-Aerial-Vehicular Networks for Small and Mini Drones. *IEEE Vehicular Technology Magazine*, **12**, 73-82. https://doi.org/10.1109/MVT.2016.2645481

[2] Alshbatat, A.I. and Dong, L. (2010) Cross Layer Design for Mobile Ad-Hoc Unmanned Aerial Vehicle Communication Networks. 2010 *International Conference on Networking, Sensing and Control* (*ICNSC*), Chicago, 10-12 April 2010, 331-336. https://doi.org/10.1109/ICNSC.2010.5461502

[3] Watkins, C.J. and Dayan, P. (1992) Q-Learning. *Machine Learning*, **8**, 279-292. https://doi.org/10.1007/BF00992698

[4] Li, R.L., Li, F., Li, X. and Wang, Y. (2014) Qgrid: Q-Learning Based Routing Protocol for Vehicular Ad Hoc Networks. 2014 *IEEE* 33*rd International Performance Computing and Communications Conference* (*IPCCC*), Austin, 5-7 December 2014, 1-8. https://doi.org/10.1109/PCCC.2014.7017079

[5] Jung, W.-S., Yim, J. and Ko, Y.-B. (2017) Qgeo: Q-Learning-Based Geographic *Ad Hoc* Routing Protocol for Unmanned Robotic Networks. *IEEE Communications Letters*, **21**, 2258-2261. https://doi.org/10.1109/LCOMM.2017.2656879

[6] Lin, L., Sun, Q., Wang, S. and Yang, F. (2012) A Geographic Mobility Prediction Routing Protocol for Ad Hoc UAV Network. 2012 *IEEE Globecom Workshops*, Anaheim, 3-7 December 2012 1597-1602. https://doi.org/10.1109/GLOCOMW.2012.6477824

[7] Choi, S.-C., Hussen, H.R., Park, J.-H. and Kim, J. (2018) Geolocation-Based Routing Protocol for Flying Ad Hoc Networks (FANETs). 2018 *Tenth International Conference on Ubiquitous and Future Networks* (*ICUFN*), Prague, 3-6 July 2018, 50-52. https://doi.org/10.1109/ICUFN.2018.8436724

[8] Arafat, M.Y. and Moh, S. (2018) Location-Aided Delay Tolerant Routing Protocol in UAV Networks for Post-Disaster Operation. *IEEE Access*, **6**, 59891-59906. https://doi.org/10.1109/ACCESS.2018.2875739

[9] He, C.T., Wang, Q., Xu, Y.J., Liu, J.M. and Xu, Y.D. (2019) A Q-Learning Based Cross-Layer Transmission Protocol for MANETs. 2019 *IEEE International Conferences on Ubiquitous Computing & Communications* (*IUCC*) *and Data Science and Computational Intelligence* (*DSCI*) *and Smart Computing, Networking and Services* (*SmartCNS*), Shenyang, 21-23 October 2019, 580-585. https://doi.org/10.1109/IUCC/DSCI/SmartCNS.2019.00122

[10] Lyu, N.Q., Song, G.H., Yang, B.W. and Cheng, Y.N. (2018) Qngpsr: A Q-Network Enhanced Geographic Ad-Hoc Routing Protocol Based on GPSR. 2018 *IEEE* 88*th Vehicular Technology Conference* (*VTC-Fall*), Chicago, 27-30 August 2018, 1-6. https://doi.org/10.1109/VTCFall.2018.8690651

[11] Sutton, R.S. (1988) Learning to Predict by the Methods of Temporal Differences. *Machine Learning*, **3**, 9-44. https://doi.org/10.1007/BF00115009

[12] Gawlowicz, P. & Zubow, A. (2019). ns-3 Meets OpenAI Gym: The Playground for Machine Learning in Networking Research. *Proceedings of the* 22*nd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, Miami Beach, 25-29 November 2019, 113-120. https://doi.org/10.1145/3345768.3355908