

专题：低空智联网

基于综合信任约束的Q-learning无人机自组网安全路由协议

柳亚男[1,2]，树赟[1]，张正[3]，邱硕[1]，沈计鹏[1]，王鹏飞[1]，仓基云[1]
（1. 金陵科技学院网络安全学院，江苏 南京 211169；
2. 河南通达电缆股份有限公司，河南 洛阳 471999；
3. 东南大学网络空间安全学院，江苏 南京 211189）

摘　要：针对无人机自组网中自私节点易进入转发路径，导致传输可靠性下降的问题，提出了基于综合信任约束的Q-learning安全路由协议。该协议面向多跳数据汇聚场景，构建包含直接信任、间接信任、能量信任和负载信任的综合信任模型，并采用CRITIC权重法实现动态赋权。在此基础上，节点根据综合信任值生成可信候选邻居集合，并结合Q-learning实现逐跳路由决策。仿真结果表明，该协议能够有效抑制自私节点对路由过程的干扰，提高分组投递率，降低端到端时延，验证了其在高动态无人机自组网中支持安全稳定传输的有效性。

关键词：无人机自组网；自私节点；综合信任；Q-learning；安全路由
中图分类号：TP393
文献标志码：A
doi: 10.11959/j.issn.1000−0801.DXKX260235

Q-learning-based secure routing protocol for UAV ad hoc networks constrained by comprehensive trust

Liu Ya’nan[1,2], Shu Yun[1], Zhang Zheng[3], Qiu Shuo[1], Shen Jipeng[1], Wang Pengfei[1], Cang Jiyun[1]
1. School of Network Security, Jinling Institute of Technology, Nanjing 211169, China
2. Henan Tong-Da Cable Co., Ltd., Luoyang 471999, China
3. School of Cyber Science and Engineering, Southeast University, Nanjing 211189, China

Abstract: To address the problem that selfish nodes in unmanned aerial vehicle (UAV) ad hoc networks could easily enter forwarding paths, thereby degrading transmission reliability, a comprehensive trust-constrained Q-learning secure routing protocol was proposed. This protocol was designed for multi-hop data aggregation scenarios, in which a comprehensive trust model was constructed by integrating direct trust, indirect trust, energy trust, and load trust, and dynamic weighting was achieved by employing the CRITIC weighting method. On this basis, a trusted candidate

收稿日期：2026−04−16；修回日期：2026−05−31
通信作者：树赟，3512846630@qq.com；张正，zhangzheng@jit.edu.cn
基金项目：国家自然科学基金资助项目（No.61902163）；江苏高校“青蓝工程”项目；河南省博士后科研项目
Foundation Items: The National Natural Science Foundation of China (No.61902163), The QingLan Project of Jiangsu Higher Education Institutions, Henan Province Postdoctoral Research Project

neighbor sets were formed according to the comprehensive trust value, and hop-by-hop routing decisions were made in combination with Q-learning. Simulation results showed that the interference of selfish nodes in the routing process was effectively suppressed, the packet delivery ratio was improved, and the end-to-end delay was reduced, thereby validating the effectiveness of the proposed protocol in supporting secure and reliable transmission in highly dynamic UAV ad hoc networks.

Key words: UAV ad hoc network, selfish node, comprehensive trust, Q-learning, secure routing

0　引言

随着无人机技术的发展，无人机自组网已经在协同侦察、灾后搜救、物流配送和环境监测等场景中得到广泛应用[1-3]。与传统地面自组网相比，无人机自组网具有机动性高、拓扑变化快和易受内部攻击等特点。因此，在无人机自组网中，路由选择不仅要保证连通性，还需判断中继节点是否可信、是否仍具备稳定转发能力。

传统的加密认证与密钥管理方法虽能在一定程度上提升网络安全性，但受限于无人机平台的计算资源和续航能力，这类依赖复杂认证过程或静态基础设施的安全机制，难以直接适应高动态组网环境[4]。相比之下，基于行为观测和信任评估的安全路由方法，能够以较低开销判断中继节点可靠性，因此逐渐成为相关研究的重要方向。文献[5-9]主要从多因子信任建模、推荐机制改进和动态赋权等方面开展研究。通过引入直接信任、间接信任、剩余能量，以及距离等因素，提高了异常节点识别能力和安全路径选择能力。其中，文献[9]采用熵权法对多项信任因子进行动态赋权，降低了人工赋权带来的主观性。在多指标客观赋权方面，已有研究将 CRITIC 方法用于三维网络路由权重计算和无人机集群网络优化，通过指标对比强度和相关关系调整权重，提高了多指标评价的客观性和自适应性[10-11]。然而，现有方法大多仍基于当前或历史观测结果进行评价，对高动态环境下节点实际可靠性的反映仍显不足。同时，动态赋权过程多服务于路由代价或簇首选择，尚未充分实现对多维信任因子与逐跳路由决策的统一结合。

除了安全评估外，如何在快速变化的拓扑条件下实现自适应路由决策，也是无人机自组网研究的重要内容。文献[12-14]分别从预测式安全路由、能量感知路由和强化学习路由等角度提升网络传输性能与安全性。文献[15-18]进一步将 Q-learning 引入路由选择过程，并结合链路状态、时延、位置信息以及奖励机制，提高了高动态环境下的路由自适应能力。然而，现有方法大多仍将安全约束与路由决策分开处理，由综合信任评估直接驱动逐跳路由的统一机制仍然不足。

针对上述问题，本文提出一种基于综合信任约束的 Q-learning 无人机自组网安全路由协议（comprehensive trust-constrained Q-learning secure routing protocol，CT-QLSR）。该协议从直接信任、间接信任、能量信任和负载信任四个维度评估邻居节点的状态，并采用 CRITIC 权重法，根据各信任因子的区分能力和相关关系进行动态赋权。在此基础上，根据综合信任门限筛选可信候选邻居，再结合 Q-learning 完成逐跳路由决策，使信任评估结果直接参与下一跳选择，从而提升在自私节点攻击和高动态拓扑条件下的路径稳定性与传输可靠性。

1　动态信任模型

1.1　网络模型与基本假设

本文考虑一种多跳无人机自组网监测场景，其示意图如图 1 所示。假设待监测区域为 $X\times Y\times Z$ 的三维空间，网络中部署多个无人机节点，其中包含 1 个汇聚节点，其余节点为普通节点。各

普通节点在执行任务过程中周期性采集监测数据，并将数据发送至汇聚节点；当源节点与汇聚节点之间无法直接通信时，数据包需通过中继节点逐跳转发完成传输。

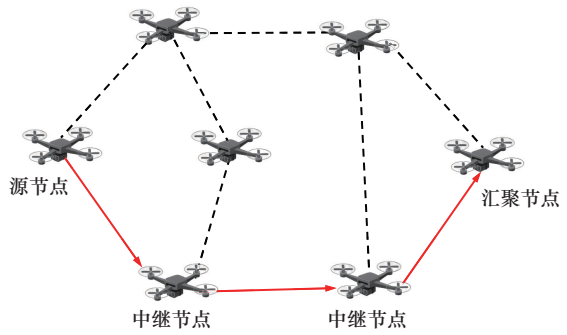


图1　多跳传输场景示意图

使用该模型时，需要做出如下假设：（1）网络不依赖固定控制中心，除汇聚节点承担数据汇聚功能，其余无人机节点地位平等；（2）所有无人机节点均能量有限且无法补充；（3）节点共享无线信道，具备混杂模式，在通信范围内可监听并统计邻居节点的数据包转发情况，为直接信任计算提供数据支撑；（4）网络中存在一定比例的自私节点，这些节点在被选为中继节点后可能为了节省自身资源而拒绝转发、减少转发或丢弃部分数据包。

1.2　直接信任

在无人机自组网中，自私节点通常表现为选择性转发或拒绝转发数据包，该行为会直接体现在邻居节点的转发成功与失败记录中。直接信任用于描述这类本地转发观测结果，反映邻居节点当前的转发合作程度。假设在一个观测周期内，节点i监听到邻居节点j的成功转发次数和失败转发次数分别为$S_{ij}(t)$和$F_{ij}(t)$。考虑无线信道随机丢包可能导致失败观测偏大，本文对失败次数进行修正。设信道随机丢包率为p_c，则修正后的失败转发次数$\tilde{F}_{ij}(t)$可表示为：

$$\tilde{F}_{ij}(t)=\max\left\{F_{ij}(t)-p_c[S_{ij}(t)+F_{ij}(t)],\ 0\right\} \quad (1)$$

其中，当p_c与真实丢包率存在一定偏差时，仅改变失败次数的修正强度，却不会直接决定节点的最终可信状态。

在此基础上，节点i对邻居节点j的直接信任可表示为：

$$D_{ij}(t)=\frac{S_{ij}(t)+1}{S_{ij}(t)+\tilde{F}_{ij}(t)+2} \quad (2)$$

其中，$D_{ij}(t)\in(0,1)$。当节点j在当前观测周期内具有更高的成功转发率时，$D_{ij}(t)$取值更大；反之，则取值减小。

在修正信道随机扰动影响后，$D_{ij}(t)$可用于反映邻居节点当前的转发合作程度。

1.3　间接信任

由于无人机节点高速移动，邻居关系会随拓扑变化频繁更新，因此，单个观测周期内获得的直接交互样本可能不足。此时仅依赖直接信任容易使信任判断受偶然观测影响，因此需要利用可信公共邻居的推荐信息对目标节点进行补充评价。

可信公共邻居示意图如图2所示。假设节点i与邻居节点j的可信公共邻居集合为$\boldsymbol{M}_{ij}(t)$，集合中的任一节点记为c。若节点i对公共邻居节点c的直接信任低于推荐阈值T_r，则认为该节点提供的推荐信息不可靠，不参与间接信任计算。考虑不同公共邻居对目标节点的观测充分程度存在差异，则节点c在间接信任计算中的推荐权重$\omega_{ic}(t)$可表示为：

$$\omega_{ic}(t)=\frac{D_{ic}(t)\,\phi_{cj}(t)}{\sum_{m\in\boldsymbol{M}_{ij}(t)}D_{im}(t)\phi_{mj}(t)} \quad (3)$$

其中，$D_{ic}(t)$表示节点i对公共邻居节点c的直接信任；$\phi_{cj}(t)$表示节点c对邻居节点j的观测充分程度；$\boldsymbol{M}_{ij}(t)$表示时刻t节点i与目标节点j的可信公共邻居集合；m表示该集合中的任一公共邻居

节点；$D_{im}(t)$表示节点i对节点m的直接信任；$\phi_{mj}(t)$表示节点m对目标节点i的观测充分程度。节点越可信、对目标节点的观测越充分，其推荐权重越大。

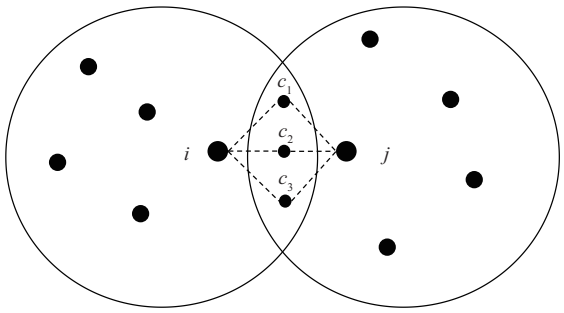


图2　可信公共邻居示意图

在此基础上，节点i对节点j的间接信任$I_{ij}(t)$可表示为：

$$I_{ij}(t)=\sum_{c\in \boldsymbol{M}_{ij}(t)}\omega_{ic}(t)\,D_{cj}(t) \tag{4}$$

其中，$D_{cj}(t)$表示公共邻居节点c对目标节点j的直接信任。

当可信公共邻居集合$\boldsymbol{M}_{ij}(t)$为空时，不引入推荐信息，可令间接信任退化为直接信任值。在局部观测不足时，利用该指标可提高对邻居节点可靠性的评价稳定性。

1.4　能量信任

无人机节点依赖有限电池供能，剩余能量不仅影响节点自身运行状态，也会影响其继续承担中继转发任务的能力。对于剩余能量较低的节点，即使其当前转发行为正常，也可能在后续路由过程中因能量不足而引发路径中断。设邻居节点j的初始能量和当前剩余能量分别为$e_{j,\mathrm{ini}}$和$e_j(t)$，维持基本转发所需的最低能量阈值为$e_{\min}$，则邻居节点j的能量信任可表示为：

$$E_j(t)=\begin{cases}0, & e_j(t)\leqslant e_{\min}\\ \dfrac{e_j(t)-e_{\min}}{e_{j,\mathrm{ini}}-e_{\min}}, & e_j(t)>e_{\min}\end{cases} \tag{5}$$

其中，$E_j(t)\in[0,1]$。

当节点剩余能量低于阈值时，认为其已不具备持续转发能力；当剩余能量高于阈值时，能量信任随剩余能量增加而增大。该指标用于反映邻居节点在当前时刻的持续转发可用性。

1.5　负载信任

在自私节点攻击场景下，节点转发失败既可能源于主动拒绝转发，也可能是受到缓存拥塞和转发压力的影响。若仅依据失败次数更新信任值，容易将拥塞造成的丢包误认为节点自私行为。为提高异常转发行为识别的准确性，将节点负载状态纳入信任评价，并将异常丢包抑制因子$\psi_j(t)$表示为：

$$\psi_j(t)=\exp\left[-\beta\max(l_j(t)-l_{j,\mathrm{th}}(t),0)\right] \tag{6}$$

其中，$l_j(t)$表示邻居节点j在观测窗口内的实际丢包率，$l_{j,\mathrm{th}}(t)$表示由拥塞状态估计得到的丢包阈值，β为调节系数。

当实际丢包率明显高于拥塞阈值时，$\psi_j(t)$将快速减小。在此基础上，设邻居节点j的当前缓冲区占用率为$b_j(t)$，则邻居节点j的负载信任$L_j(t)$可表示为：

$$L_j(t)=\omega_b[1-b_j(t)]+(1-\omega_b)\psi_j(t) \tag{7}$$

其中，ω_b表示权重系数，且$0\leqslant\omega_b\leqslant1$。

当节点缓冲区占用率较低且实际丢包率未明显超过拥塞阈值时，负载信任较高；反之，则说明该节点已处于较高业务压力或异常丢包状态，负载信任降低。与主要反映节点转发合作行为的直接信任不同，负载信任侧重反映节点在当前缓存压力和拥塞状态下的转发可用性。

1.6　CRITIC权重法

不同网络状态下，各信任因子的区分效果存在显著差异。为避免固定权重带来的综合信任评价结果受人为经验参数干扰的问题，引入CRITIC权重法，对直接信任、间接信任、能量信任和负载信任四类信任因子进行动态赋权。其流程示意图如图3所示。

节点根据一跳邻居观测结果构建信任评估矩

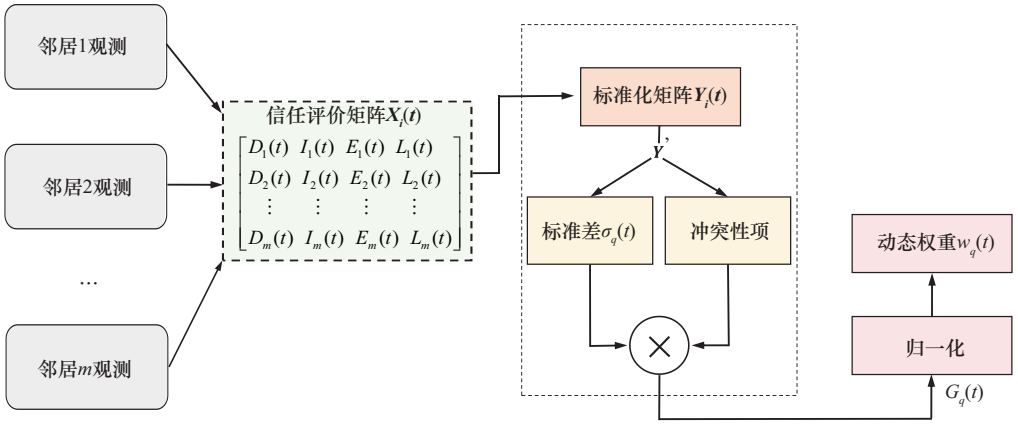


图3　CRITIC动态赋权流程示意图

阵，并据此计算各信任因子的动态权重。设节点i在时刻t的一跳邻居集合为$\mathcal{N}_i(t)=\{j_1,j_2,\cdots,j_n\}$，则由四类信任因子构成的评价矩阵$\boldsymbol{X}_i(t)$可表示为：

$$\boldsymbol{X}_i(t)=\begin{bmatrix} D_{ij_1}(t) & I_{ij_1}(t) & E_{j_1}(t) & L_{j_1}(t) \\ D_{ij_2}(t) & I_{ij_2}(t) & E_{j_2}(t) & L_{j_2}(t) \\ \vdots & \vdots & \vdots & \vdots \\ D_{ij_n}(t) & I_{ij_n}(t) & E_{j_n}(t) & L_{j_n}(t) \end{bmatrix} \tag{8}$$

对评价矩阵进行标准化处理后，结合各因子的离散程度和相关性计算其信息量。设第q个信任因子的标准差为$\sigma_q(t)$，其与第p个信任因子的相关系数为$\rho_{qp}(t)$，则第q个因子的信息量$G_q(t)$可表示为：

$$G_q(t)=\sigma_q(t)\sum_{p=1}^{4}[1-\rho_{qp}(t)] \tag{9}$$

其中，$q\in\{D,I,E,L\}$分别对应直接信任、间接信任、能量信任和负载信任。$G_q(t)$越大，表明该因子的区分能力越强、与其他因子的冗余越小。

在此基础上，第q个信任因子的动态权重$w_q(t)$可表示为：

$$w_q(t)=\frac{G_q(t)}{\sum_{p=1}^{4}G_p(t)} \tag{10}$$

其中，$w_q(t)\in[0,1]$，且满足$\sum_{q\in\{D,I,E,L\}}w_q(t)=1$。

CRITIC权重法同时考虑指标的对比强度与冲突性，能够根据各信任因子的变化特征与关联关系实现权重动态调整。熵权法则依据指标离散程度进行客观赋权，虽然能够减少人工设定权重带来的主观偏差，但未量化考虑不同指标之间的相关性。本文选取的直接信任、间接信任、能量信任和负载信任并非相互独立：直接信任和间接信任均与节点转发行为有关，能量信任和负载信任也会受到中继转发过程与业务负载波动的共同作用。若仅采用熵权法，单纯依据指标差异程度赋权，可能使相关性较强的信任因子被重复放大。相比之下，CRITIC法能够在突出高区分度指标的同时削弱冗余信息影响，因此更适配本文无人机自组网场景下的综合信任评价。此外，当网络处于能量普遍较低或负载普遍较高等极端状态时，相关指标在邻居集合中的差异程度会降低，其对应的信息量和权重变化趋于平缓。这一现象表明该指标在当前状态下的区分能力有限，而非权重机制失效。此时，综合信任评价仍由多类信任因子共同决定，能够避免单一指标在极端状态下对路由决策产生过度影响。

1.7　综合信任值

将直接信任、间接信任、能量信任和负载信任加权融合，可得节点i对邻居节点j的综合信任值。设四类信任因子的动态权重分别为$w_D(t)$、$w_I(t)$、$w_E(t)$和$w_L(t)$，则综合信任值$C_{ij}(t)$可表示为：

$$C_{ij}(t)=w_D(t)D_{ij}(t)+w_I(t)I_{ij}(t)+w_E(t)E_j(t)+w_L(t)L_j(t) \tag{11}$$

其中，$w_D(t)+w_I(t)+w_E(t)+w_L(t)=1$。

综合信任值越大，节点越适合作为候选下一跳；反之，则该节点在行为或资源状态上存在潜在风险。每个观测周期结束后，节点根据综合信任值更新邻居信息表，并形成可信候选邻居集合。

2　基于综合信任的 Q-learning 路由决策方法

2.1　路由决策框架

仅依据当前观测结果选择下一跳，难以适应高动态无人机自组网中的连续转发过程。随着节点运动和链路状态变化，在当前时刻表现可靠的邻居节点，在后续时刻未必依然适用。因此，路由决策不仅需要考虑当前评价结果，还应结合实际转发反馈，持续修正各候选邻居的动作价值。Q-learning 路由决策框架如图 4 所示。

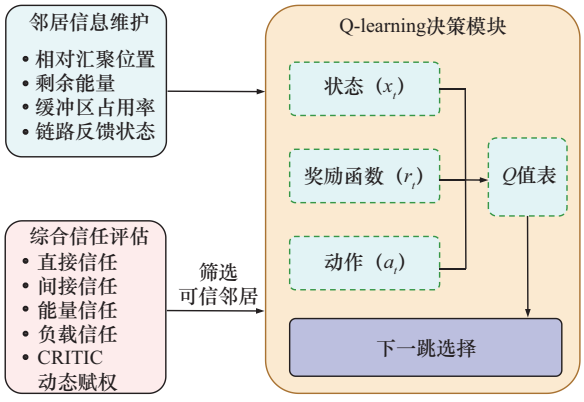


图 4　Q-learning 路由决策框架

节点首先基于局部观测维护邻居信息，并计算综合信任值；随后将可信候选邻居输入 Q-learning 决策模块，依次完成状态构造、动作选择、奖励计算和 Q 值更新；最后，输出下一跳选择结果。

2.2　状态描述

由于无人机节点难以获取全局拓扑信息，逐跳路由决策只能依据一跳邻居的局部观测结果进行。设时刻 t 的当前持包节点为 i，汇聚节点为 S，节点 i 的一跳邻居集合为 $\boldsymbol{N}_i(t)$。节点 i 通过控制报文交互和本地监听维护邻居节点的综合信任值、相对汇聚位置、剩余能量、缓冲区占用率及链路反馈状态，并据此构造局部状态。节点 i 在时刻 t 的状态 x_t 可表示为：

$$x_t=(\tau_t,\rho_t,\xi_t,\lambda_t) \tag{12}$$

其中，τ_t 表示综合信任等级，ρ_t 表示路由推进等级，ξ_t 表示资源状态，λ_t 表示链路稳定性状态。

为控制状态空间规模，对各状态分量进行离散化处理。设节点 i 对候选邻居 j 的综合信任值为 $C_{ij}(t)$，则综合信任等级可表示为：

$$\tau_t=\begin{cases}2, & C_{ij}(t)\geqslant C_{\mathrm{H}}\\ 1, & C_L\leqslant C_{ij}(t)<C_{\mathrm{H}}\\ 0, & C_{ij}(t)<C_{\mathrm{L}}\end{cases} \tag{13}$$

其中，C_{H} 和 C_{L} 分别表示高、低信任划分阈值，且满足 $C_{\mathrm{H}}>C_{\mathrm{L}}$。

设当前节点 i 与候选邻居 j 到汇聚节点 S 的距离分别为 $d_i(t)$ 和 $d_j(t)$，则节点 j 相对节点 i 的路由推进量 $\Delta p_{ij}(t)=d_i(t)-d_j(t)$。据此，路由推进等级可表示为：

$$\rho_t=\begin{cases}2, & \Delta p_{ij}(t)\geqslant P_{\mathrm{H}}\\ 1, & P_L\leqslant\Delta p_{ij}(t)<P_{\mathrm{H}}\\ 0, & \Delta p_{ij}(t)<P_{\mathrm{L}}\end{cases} \tag{14}$$

其中，P_{H}、P_{L} 表示高、低路由推进等级划分阈值，且满足 $P_{\mathrm{H}}>P_{\mathrm{L}}$。

资源状态综合考虑剩余能量占比和缓冲区占用率。设节点 j 的当前剩余能量为 $e_j(t)$，初始能量为 $e_{j,\mathrm{ini}}$，缓冲区占用率为 $b_j(t)$，则资源状态可表示为：

$$\xi_t=\begin{cases}1, & \dfrac{e_j(t)}{e_{j,\mathrm{ini}}}\geqslant e_{\mathrm{th}}\wedge b_j(t)\leqslant b_{\mathrm{th}}\\ 0, & \text{其他}\end{cases} \tag{15}$$

其中，e_{th} 和 b_{th} 分别表示能量门限和缓冲区占用门限。

链路稳定性状态由最近观测窗口内的单跳交互结果确定。设节点i对链路(i,j)的单跳成功率为$p_{ij}(t)$，则链路稳定性状态可表示为：

$$\lambda_t=\begin{cases}1, & p_{ij}(t)\geqslant p_{\text{th}}\\ 0, & p_{ij}(t)<p_{\text{th}}\end{cases} \tag{16}$$

其中，p_{th}为链路稳定性判别阈值。

由式（13）～式（16）可知，τ_t、ρ_t、ξ_t、λ_t分别取3、3、2、2个离散状态，则局部状态空间规模为$3\times3\times2\times2=36$。该状态空间在保留主要决策信息的同时，有效控制了$Q$表的规模。

2.3 候选邻居筛选与动作选择

在状态描述的基础上，当前节点将下一跳选择建模为有限状态下的Q-learning决策过程。为避免不可靠节点进入动作集合，本文先依据综合信任值、资源状态和路由推进等级对一跳邻居进行筛选，再在候选集合内完成动作选择，则可信候选邻居集合表示为：

$$\mathcal{H}_i(t)=\left\{j\in\mathcal{N}_i(t)\middle|\, C_{ij}(t)\geqslant C_{\text{th}},\ \frac{e_j(t)}{e_{j,\text{ini}}}\geqslant e_{\text{th}},\ b_j(t)\leqslant b_{\text{th}},\ d_j(t)<d_i(t)\right\} \tag{17}$$

其中，C_{th}表示综合信任门限。

在状态x_t下，当前节点的动作是从候选邻居集合$\mathcal{H}_i(t)$中选择一个邻居节点作为下一跳，因此，动作集合可记为$\mathcal{A}_i(t)=\mathcal{H}_i(t)$，各节点仅维护本地$Q$表$Q_i(x_t,a)$，并利用局部观测结果进行在线更新。当$\mathcal{A}_i(t)\neq\varnothing$时，节点采用$\varepsilon-$greedy策略选择动作。

$$a_t=\begin{cases}\arg\max\limits_{a\in\boldsymbol{A}_i(t)} Q_i(x_t,a), & \zeta_t>\varepsilon\\ \mathcal{R}\left(\boldsymbol{A}_i(t)\right), & \zeta_t\leqslant\varepsilon\end{cases} \tag{18}$$

其中，a_i表示节点i在时刻t选择的下一跳节点；a表示动作集合$\boldsymbol{A}_i(t)$中的任一候选动作，每个候选动作对应一个可信候选邻居节点；ε表示探索概率；$\zeta_t\sim U(0,1)$为区间$(0,1)$上的均匀随机数；$\mathcal{R}(\mathcal{A}_i(t))$表示从动作集合中随机选择的一个动作。

当节点选定下一跳后，立即将数据包转发给对应邻居节点，并等待单跳反馈结果。若$\mathcal{A}_i(t)=\varnothing$，则说明当前不存在可用下一跳，此时进入路由维护过程。

2.4 奖励函数与Q值更新

当前节点完成一次单跳选择后，需要依据交互结果对所选动作进行评价。为兼顾转发可靠性、路由推进效果和转发代价，将即时奖励设计为反馈项、推进项、时延代价项和代价项的组合。设节点i在时刻t选择邻居节点j作为下一跳后获得的即时奖励为r_t，则有：

$$r_t=\omega_1 f_t+\omega_2\frac{\Delta p_{ij}(t)}{\Delta p_{\max}(t)}-\omega_3\frac{\delta_{ij}(t)}{\delta_{\max}}-\omega_4 b_j(t) \tag{19}$$

其中，f_t表示单跳反馈结果，当本次单跳交互成功时，$f_t=1$；否则，$f_t=-1$。$\Delta p_{ij}(t)$表示邻居节点j相对当前节点i的路由推进量。$\Delta p_{\max}(t)$表示当前候选集合中的最大推进量。$\delta_{ij}(t)$表示链路(i,j)的单跳反馈时延。$\delta_{\max}$表示反馈时延归一化上界。$b_j(t)$表示邻居节点j的缓冲区占用率。ω_1、ω_2、ω_3和ω_4表示权重系数。

奖励函数的权重设置以提高转发可靠性为主要目标，同时兼顾路由推进效果、反馈时延和节点负载。其中，反馈项用于表征单跳转发结果对路径质量评价的影响，推进项用于引导数据包向汇聚节点方向转发，时延代价项和负载代价项用于抑制高时延或高负载节点进入下一跳选择。单跳反馈结果由转发监听和确认反馈联合判定：若当前节点在监听窗口内检测到下一跳节点对该数据包进行了继续转发，则记为成功；若监听条件受限，但在限定时间内收到对应的确认反馈，也记为成功；否则，记为失败。在获得即时奖励后，节点按照标准Q-learning规则更新状态–动作值：

$$Q_i(x_t,a_t)\leftarrow(1-\alpha)Q_i(x_t,a_t)+\eta\left[r_t+\gamma\max_{a\in\mathcal{A}_i(t+1)}Q_i(x_{t+1},a)\right] \tag{20}$$

其中，α表示学习率；γ表示折扣因子；x_{t+1}表示节点完成本次单跳交互后进入的下一状态；$\mathcal{A}_i(t+1)$表示下一状态下的可选动作集合，当$\mathcal{A}_i(t+1)=\varnothing$时，$\max_{a\in\mathcal{A}_i(t+1)}Q_i(x_{t+1},a)=0$。

随着数据包转发过程的持续进行，节点通过累积不同局部状态下的动作收益，使Q值表逐步反映各候选邻居在长期转发过程中的优劣，从而提高逐跳路由决策的自适应性与稳定性。

2.5　路由组建

本文采用基于综合信任约束的逐跳方式完成路由组建。路由组建示意图如图5所示。当前节点依据一跳邻居的局部状态信息选择下一跳，随后由新选中的转发节点继续执行相同过程，直至数据包到达汇聚节点。

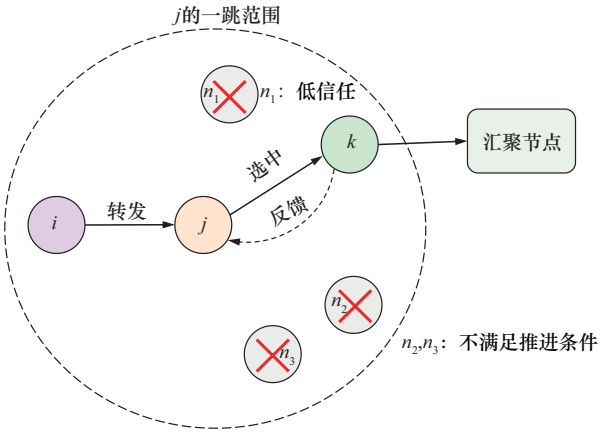


图5　路由组建示意图

路由组建过程可描述如下。

步骤1　当源节点产生待上传数据包时，首先查询本地邻居信息表。若汇聚节点位于当前节点i的一跳通信范围内，则将数据包直接发送至汇聚节点；否则，节点i进入下一跳选择过程。

步骤2　当前节点i根据本地维护的邻居状态信息，对一跳邻居进行筛选，形成满足综合信任、资源可用性和路由推进条件的可信候选邻居集合。若候选集合为空集，则触发局部维护过程。

步骤3　若可信候选邻居集合非空集，则节点i根据当前局部状态和Q-learning决策规则，从候选集合中选择邻居节点j作为下一跳，并将携带分组序号的数据包转发给节点j。

步骤4　节点i根据监听结果和确认反馈判定本次单跳交互是否成功，并据此更新链路状态、邻居信息表以及对应的Q值。若在限定时间内未获得有效交互结果，则将该次交互记为失败，并执行相应更新。

步骤5　若节点j不是汇聚节点，则由节点j作为新的当前转发节点，重复执行步骤2至步骤4；若数据包已成功到达汇聚节点，则本次路由组建结束。

2.6　路由维护

在数据传输过程中，若当前下一跳节点的综合信任值低于门限、剩余能量不足，或链路连续反馈失败，则判定当前转发路径失效。此时，当前节点依据最新邻居状态重新筛选可信候选邻居，并按照Q-learning决策规则完成局部下一跳重选；若局部修复失败，则向上游节点返回路由错误信息，由上游节点继续执行重选过程；若仍不存在可用转发节点，则由源节点重新发起路由组建。该机制能够在节点可靠性下降或链路状态恶化时及时修正转发路径，从而增强路由的连续性与稳定性。

3　仿真实验与性能比较

3.1　仿真环境及参数

本文在OMNeT++6.2平台上进行仿真实验，并选取HQTR[17]和AODV作为对比协议。网络场景设置为1 000 m×1 000 m×600 m的三维空间，默认节点数为60，移动模型采用3D-GM。自私

节点在被选为中继后可能拒绝转发或选择性丢弃部分数据包。综合信任门限用于约束候选邻居集合，其取值需要兼顾自私节点过滤效果与网络连通性保持；Q-learning 相关参数则结合高动态拓扑下的路由学习稳定性和路径探索需求进行设置。基于协议机制和仿真场景设置实验相关参数，其余关键仿真参数设置见表 1。

表 1　仿真参数设置

参数	值
仿真时间/s	300
节点飞行速度/（m·s^{-1}）	[5,30]
自私节点比例	[10%,35%]
发送速率/pps	[2,12]
无线通信半径/m	350
MAC 层协议	IEEE 802.11b
综合信任门限	0.4
折扣因子	0.85
学习率	0.2
探索率	0.1
信道随机丢包率	0.05

注：pps 表示包每秒。

3.2　协议性能分析

实验主要选取分组投递率（packet delivery ratio，PDR）、平均端到端时延和路由开销 3 项指标对 CT-QLSR 协议的性能进行仿真验证。分组投递率是指目的节点成功接收数据包数量与源节点发送数据包数量的比值。平均端到端时延是指数据包从源节点传输到目的节点所经历的平均时延。路由开销为数据传输过程中产生的路由控制消息数量与成功交付数据包数量的比值。

3.2.1　节点速度不同时协议性能的比较

3 种协议在不同节点速度下的分组投递率和平均端到端时延如图 6 所示。随着节点移动速度不断加快，网络拓扑变化加剧，链路保持时间缩短，3 种协议的分组投递率均出现不同程度回落，平均端到端时延则同步增大，说明高速运动条件下的链路波动和路径重建已成为影响路由性能的重要因素。

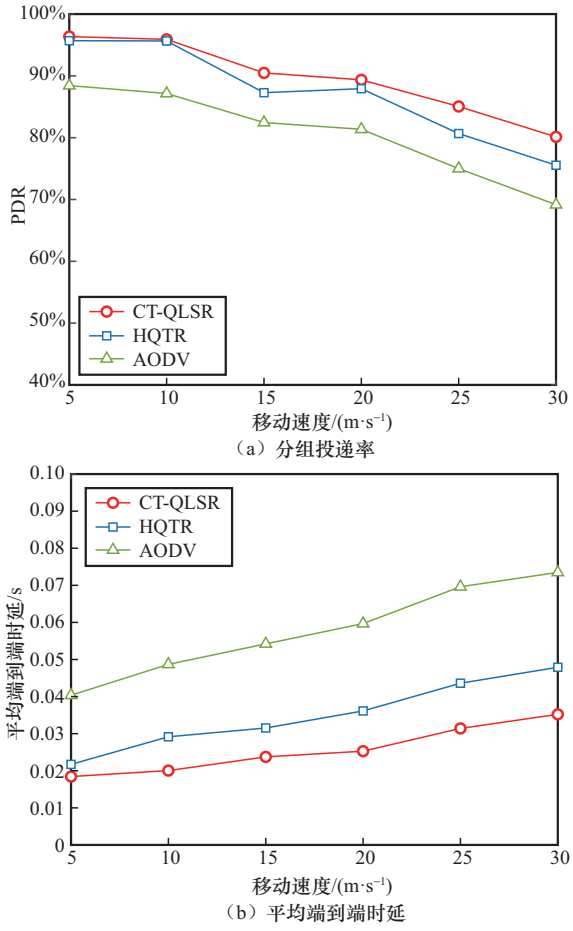


图 6　3 种协议在不同节点速度下的分组投递率和平均端到端时延

就分组投递率而言，CT-QLSR、HQTR 和 AODV 的降幅分别约为 16.8%、21.1% 和 21.8%，其中 CT-QLSR 的下降幅度最小；在最高移动速度 30 m/s 条件下，CT-QLSR 的分组投递率较 HQTR 提高约 6.1%，较 AODV 提高约 15.9%。就平均端到端时延而言，在最高移动速度 30 m/s 条件下，CT-QLSR 较 HQTR 和 AODV 分别降低约 26.5% 和 52.1%，优势较为明显。

综合来看，AODV 在高动态场景下更容易受到链路快速失效的影响；HQTR 虽然通过信任与启发式学习在一定程度上改善了路径选择质量，但面对持续增强的拓扑波动，其性能仍会出现较明显退化。相比之下，CT-QLSR 在路由选择过程中不仅关注节点当前的可信状态，而且兼顾链路可用性与转发条件，能够在高速运动条件下更早

规避不稳定中继和易失效链路，从而减缓拓扑频繁演化所引起的性能衰减，体现出更好的动态适应能力和路径保持能力。

3.2.2　发送速率不同时协议性能的比较

3 种协议在不同发送速率下的分组投递率和平均端到端时延如图 7 所示。随着发送速率不断提高，网络中的业务注入强度、信道竞争程度以及节点缓存压力同步上升，3 种协议的分组投递率均明显走低，平均端到端时延则进一步升高，表明高负载条件下的拥塞扩散与无效转发会对整体传输性能造成持续影响。

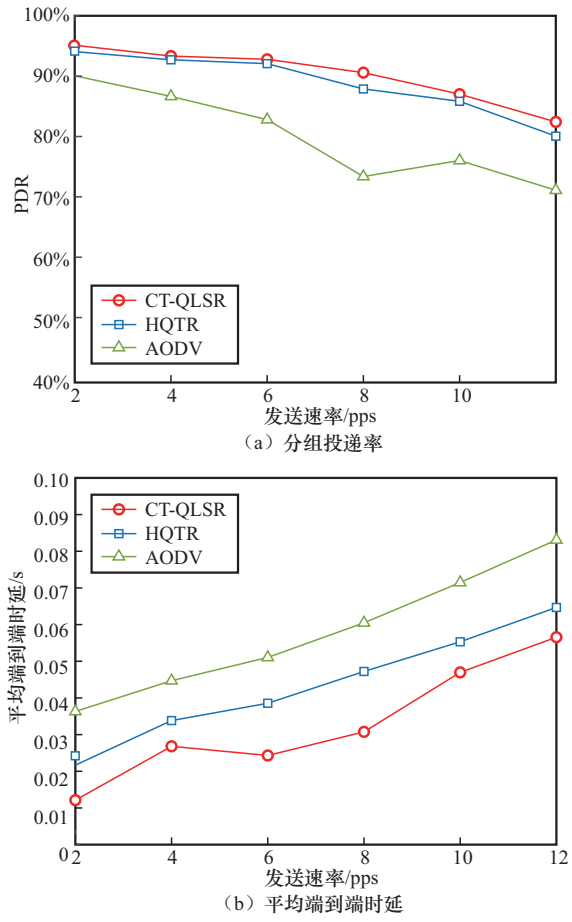


图 7　3 种协议在不同发送速率下的分组投递率和平均端到端时延

就分组投递率而言，CT-QLSR、HQTR 和 AODV 的降幅分别约为 13.2%、14.7% 和 20.8%，其中 CT-QLSR 的回落最为平缓；在最高发送速率（发送速率为 12 pps）条件下，CT-QLSR 的分组投递率较 HQTR 提高约 2.9%，较 AODV 提高约 15.6%。就平均端到端时延而言，在高发送速率条件下，CT-QLSR 较 HQTR 和 AODV 分别降低约 12.6% 和 31.9%，仍保持最优水平。上述结果表明，随着网络负载持续加大，AODV 更容易因局部拥塞、队列积压和链路退化而引起路径质量下降；HQTR 虽具备一定的调节能力，但在高业务压力下，局部观测与 Q 值更新的有效性仍会受到影响，因此性能同样会出现明显下滑。相比之下，CT-QLSR 能够综合节点当前可信状态和转发承载条件，对高负载节点及潜在低效中继进行更及时地规避，使路由选择更符合当前网络状态，因此其优势并不主要体现在低负载条件下的绝对值领先，更体现在高负载条件下退化速度更慢、性能保持更稳。

3.2.3　网络规模不同时协议性能的比较

3 种协议在不同网络规模下的分组投递率和平均端到端时延如图 8 所示。随着网络节点数量增加，网络中可选中继节点增多，但同时也会带来更复杂的邻居关系、信道竞争和路由维护压力。因此，3 种协议的分组投递率整体呈下降趋势，平均端到端时延则逐渐增大。由于节点随机分布和移动过程会影响局部连通性，小规模网络下不同协议性能存在一定波动，但随着节点规模增大，各协议之间的差异逐渐明显。

就分组投递率而言，当节点数量由 40 增加至 90 时，CT-QLSR、HQTR 和 AODV 的分组投递率降幅分别约为 13.8%、19.1% 和 19.6%，其中 CT-QLSR 下降幅度最小。在节点数量达到 90 个时，CT-QLSR 的分组投递率较 HQTR 提高约 5.1%，较 AODV 提高约 11.9%。就平均端到端时延而言，在最高节点规模（节点数量为 90 个）下，CT-QLSR 较 HQTR 和 AODV 分别降低约 10.8% 和 34.0%。综合来看，随着网络规模增大，AODV 缺少对节点可信状态和转发条件的综合判断，容

易受路由竞争和异常转发节点影响；HQTR 虽然引入了信任和学习机制，但在较大规模网络下仍存在一定性能退化。相比之下，CT-QLSR 能够结合综合信任评价和 Q-learning 决策选择更可靠的中继节点，从而在网络规模增大时保持较高的分组投递率和较低的平均端到端时延，体现出较好的规模适应性。

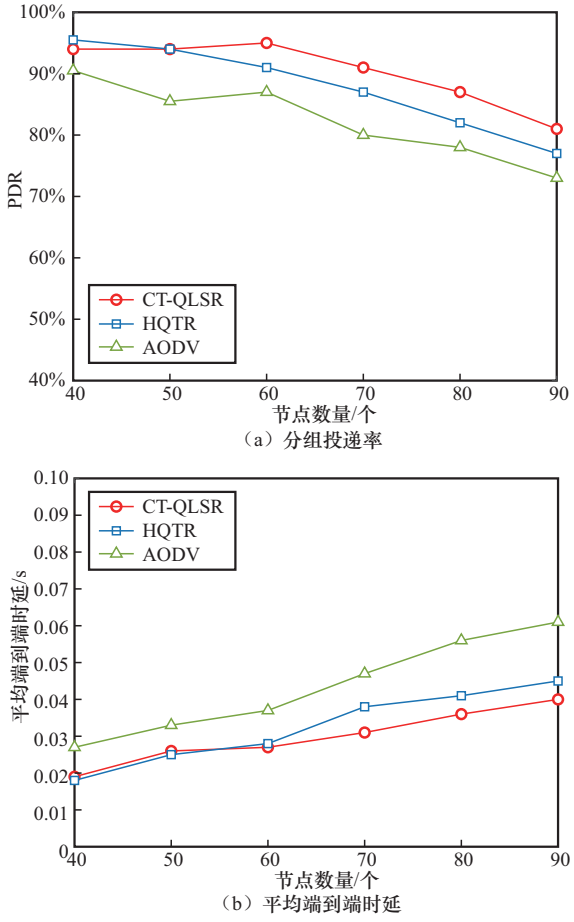


图 8　3 种协议在不同网络规模下的分组投递率和平均端到端时延

3.2.4　开销分析

3 种协议在不同场景下的路由开销比较如图 9 所示。随着发送速率和网络节点数量的增加，3 种协议的路由开销整体呈上升趋势。发送速率提高会使网络中的业务负载增大，节点队列竞争和路径维护压力随之增强；节点数量增加则会带来更复杂的邻居关系和更多候选路径，使协议需要维护的路由状态和控制消息数量增加，因此路由开销均出现不同程度增长。

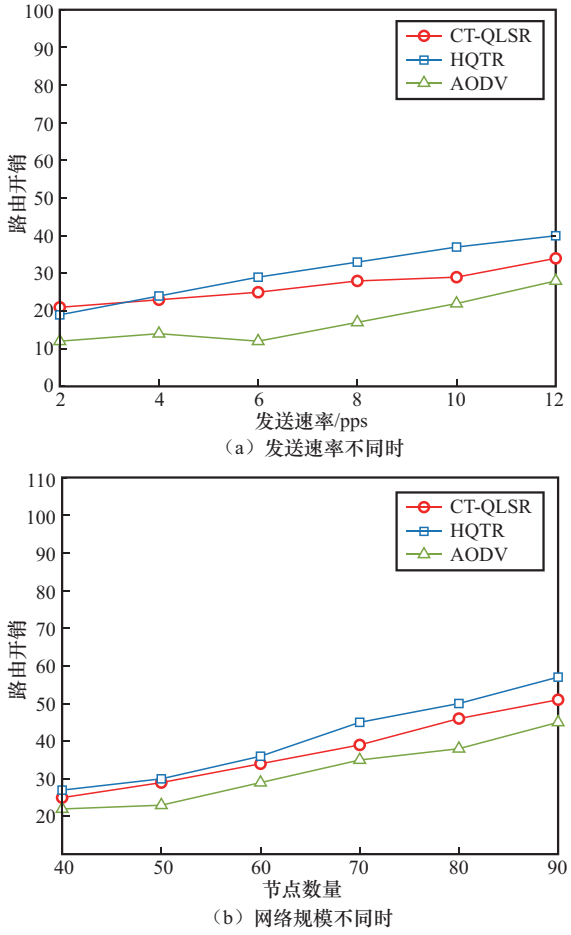


图 9　3 种协议在不同场景下的路由开销比较

从图 9（a）可以看出，在发送速率由 2 pps 增加至 12 pps 时，AODV 由于不进行信任维护和学习决策，其路由开销整体最低；HQTR 在较高发送速率下开销增长较快；CT-QLSR 虽然引入综合信任评价和 Q-learning 决策会产生一定额外控制开销，但其开销整体低于 HQTR。在最高发送速率条件下，CT-QLSR 的路由开销较 HQTR 降低约 15.0%。从图 9（b）可以看出，随着节点数量增加，CT-QLSR 的路由开销始终保持在 AODV 和 HQTR 之间，在节点数量为 90 时，较 HQTR 降低约 12.8%。综合来看，AODV 开销较低主要来源于其机制简单，但其分组投递率和平均端到端

时延表现相对较弱；CT-QLSR 以适度增加的控制开销换取了更高的传输可靠性和更低的时延，体现出较好的性能与开销折中。

3.3　鲁棒性分析

3 种协议在不同自私节点比例下的分组投递率和平均端到端时延如图 10 所示。随着自私节点比例不断升高，网络中的不可靠中继数量随之增加，正常数据转发受到更强干扰，3 种协议的分组投递率均明显下滑，平均端到端时延也进一步拉大，这说明内部攻击强度的增加会直接削弱路径可用性并增加重传与重路由开销。

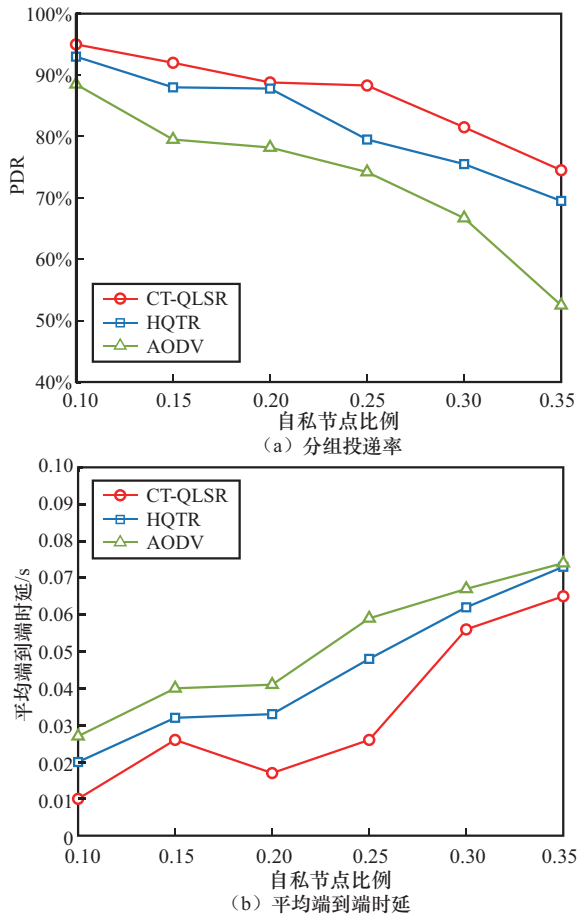


图 10　3 种协议在不同自私节点比例下的分组投递率和平均端到端时延

就分组投递率而言，CT-QLSR、HQTR 和 AODV 的降幅分别约为 21.3%、25.3% 和 40.5%，其中 AODV 的性能衰减最为明显，而 CT-QLSR 始终保持相对平缓的下降幅度；在最高自私节点比例（0.35）条件下，CT-QLSR 的分组投递率较 HQTR 提高约 7.3%，较 AODV 提高约 42.0%。就平均端到端时延而言，在高自私节点比例（0.35）条件下，CT-QLSR 较 HQTR 和 AODV 分别降低约 11.3% 和 12.6%，依然保持最低水平。这说明随着内部威胁不断增强，AODV 由于缺乏对自私节点的识别与规避能力，关键路径更易被破坏，因而性能下降最为迅速；HQTR 通过信任评估和启发式学习能够在一定程度上抑制自私节点行为，但当自私节点持续增多时，其局部观测和动态调整仍会受到明显干扰。相比之下，CT-QLSR 更强调对高风险节点的前置规避，使自私节点更难进入关键转发路径，并且能够减少由错误中继选择引起的分组丢失和时延扩张，因此在强对抗条件下仍能维持较高的传输可靠性和较低的时延开销，表现出更强的鲁棒性。

4　结束语

仅根据当前观测结果判断邻居节点是否可靠，在高动态无人机自组网中往往不够。尤其是当自私节点进入转发过程后，路径质量和传输效率均容易受到显著影响。因此，本文提出了一种基于综合信任约束的 Q-learning 安全路由协议（CT-QLSR）。节点先从直接信任、间接信任、能量信任和负载信任 4 个维度评估邻居节点状态，再将综合信任值用于下一跳选择。仿真结果表明，CT-QLSR 能够降低自私节点进入关键路径的概率，并在分组投递率和平均端到端时延方面取得较好性能。由此可见，将信任评估结果进一步转化为路径选择性约束，是提升安全路由稳定性和有效性的重要途径。

参考文献：

[1]　Ogunbunmi S, Chen Y, Blasch E, et al. A survey on reputation systems for UAV networks[J]. Drones, 2024, 8(6): 253.

[2] Gong Y Y, Chen Y X, Li K, et al. DSCR: a dynamic secure clustering routing scheme for UANETs based on reputation mechanism[J]. IEEE Internet of Things Journal, 2025, 12(10): 14109-14123.

[3] Luo H X, Wu Y F, Sun G G, et al. ESCM: an efficient and secure communication mechanism for UAV networks[J]. IEEE Transactions on Network and Service Management, 2024, 21(3): 3124-3139.

[4] Kharjana M, Sahana S C, Saha G. Securing autonomous UAV cluster with blockchain-based threshold key management system utilizing crypto-asset and multisignature[J]. IEEE Transactions on Mobile Computing, 2025, 24(7): 5765-5778.

[5] Kundu J, Alam S, Koner C, et al. Trust-based dynamic leader selection mechanism for enhanced performance in flying ad-hoc networks (FANETs) [J]. IEEE Transactions on Intelligent Transportation Systems, 2024, 25(12): 20616-20627.

[6] Hosseinzadeh M, Mohammed A H, Alenizi F A, et al. A novel fuzzy trust-based secure routing scheme in flying ad hoc networks[J]. Vehicular Communications, 2023, 44: 100665.

[7] 韩优佳, 胡黄水, 姚美琴. 基于信任感知的无线传感器网络安全路由协议[J]. 计算机工程, 2021, 47(9): 145-152.
Han Y J, Hu H S, Yao M Q. Trust-aware secure routing protocol for wireless sensor networks[J]. Computer Engineering, 2021, 47(9): 145-152.

[8] 孙子文, 吴平. 基于信任评估模型的IWSN安全路由研究[J]. 传感技术学报, 2019, 32(6): 858-865.
Sun Z W, Wu P. Research on IWSN secure routing based on trust evaluation model[J]. Chinese Journal of Sensors and Actuators, 2019, 32(6): 858-865.

[9] 郑浩泽, 潘巨龙. 一种基于多因子信任机制的无线传感网络安全路由协议[J]. 传感技术学报, 2024, 37(8): 1395-1403.
Zheng H Z, Pan J L. A new secure routing protocol of wireless sensor networks based on multi-factor trust[J]. Chinese Journal of Sensors and Actuators, 2024, 37(8): 1395-1403.

[10] 崔颖, 李巧珏, 高山, 等. 基于改进最小生成树的三维路由算法[J]. 应用科技, 2023, 50(6): 76-81.
Cui Y, Li Q J, Gao S, et al. Research on 3D routing algorithm based on improved minimum spanning tree[J]. Applied Science and Technology, 2023, 50(6): 76-81.

[11] 党元章, 华翔, 张金金, 等. 自适应ANP-CRITIC的无人机集群网络优化算法[J]. 西安工业大学学报, 2023, 43(6): 568-577.
Dang Y Z, Hua X, Zhang J J, et al. Optimized algorithm for UAV cluster networks based on adaptive ANP-CRITIC[J]. Journal of Xi'an Technological University, 2023, 43(6): 568-577.

[12] Hutchins C. Using contextual reinforcement learning to design FANET defence protocols to combat grey hole attacks[C]//Proceedings of the NOMS 2024-2024 IEEE Network Operations and Management Symposium. Piscataway: IEEE Press, 2024: 1-4.

[13] Zhou Z H, Tang J, Yang Z T, et al. A new predictive based secure geographic routing strategy for UAV network under location spoofing attack[J]. Chinese Journal of Aeronautics, 2025, 38(10): 103498.

[14] Hussain A, Ahmad W. Delay and energy aware routing (DEAR) protocol for UAV networks[J]. ICCK Transactions on Intelligent Unmanned Systems, 2025, 2(1): 1-14.

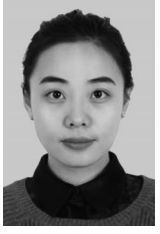
[15] 吴麒, 左琳立, 丁建, 等. 基于Q-learning的自适应链路状态路由协议[J]. 重庆邮电大学学报(自然科学版), 2024, 36(5): 945-953.
Wu Q, Zuo L L, Ding J, et al. Q-learning based adaptive link state routing protocol[J]. Journal of Chongqing University of Posts and Telecommunications (Natural Science Edition), 2024, 36(5): 945-953.

[16] 谭周正, 范琅, 李宇峰, 等. 基于强化学习的无人机网络自适应QoS路由算法[J]. 计算机应用研究, 2025, 42(4): 1177-1184.
Tan Z Z, Fan L, Li Y F, et al. Adaptive QoS routing algorithm of UAV network based on reinforcement learning[J]. Application Research of Computers, 2025, 42(4): 1177-1184.

[17] 赵蓓英, 姬伟峰, 翁江, 等. 基于启发式Q学习的FANET可信路由算法[J]. 计算机工程, 2022, 48(5): 162-169.
Zhao B Y, Ji W F, Weng J, et al. Trusted routing algorithm based on heuristic Q-learning for FANET[J]. Computer Engineering, 2022, 48(5): 162-169.

[18] 刘星宇, 强楠楠, 付银娟. 面向无人机多状态模糊推理的Q学习路由机制研究[J]. 信息技术, 2025(6): 17-22, 29.
Liu X Y, Qiang N N, Fu Y J. Research on Q-learning routing mechanism for UAV based on multi-state fuzzy reasoning[J]. Information Technology, 2025(6): 17-22, 29.

[作者简介]



柳亚男（1984–），女，博士，金陵科技学院网络安全学院副教授，主要研究方向为密码协议、物联网安全。



树赟（2001– ），男，金陵科技学院网络安全学院硕士生，主要研究方向为物联网、网络安全。



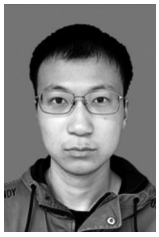
张正（1973– ），男，东南大学网络空间安全学院博士生，主要研究方向为网络安全、通信安全。



邱硕（1989– ），女，博士，金陵科技学院网络安全学院副教授，主要研究方向为云计算安全、隐私保护。



沈计鹏（2003– ），男，金陵科技学院网络安全学院硕士生，主要研究方向为物联网、网络安全。



王鹏飞（2002– ），男，金陵科技学院网络安全学院硕士生，主要研究方向为物联网、网络安全。



仓基云（1993– ），女，金陵科技学院网络安全学院讲师，主要研究方向为计算机网络、局域网安全。