GP-SR：基于GNN-PPO的低轨卫星网络安全路由优化算法

尹斌[1,2]，刘建[1,2]，唐小妹[1,2]，马春江[1,2]，马鹏程[1,2]
（1. 国防科技大学电子科学学院，湖南 长沙 410000；2. 导航与时空技术国家级重点实验室，湖南 长沙 410000）

摘要：针对低轨卫星网络在拓扑快速变化、状态局部可观测及安全攻击并存条件下的路由优化问题，本文提出一种基于图神经网络与近端策略优化的风险敏感智能路由算法GNN-PPO based Secure Routing（GP-SR）。该方法将分布式路由决策建模为部分可观测马尔可夫决策过程，利用历史增强观测缓解状态不完备，并通过性能收益与安全代价联合优化，引导策略在提升网络性能的同时主动规避高风险链路与异常节点。仿真结果表明，GP-SR在端到端时延、吞吐量和包投递率等指标上均优于对比算法，并在渐进增强攻击场景下表现出更好的鲁棒性与适应能力。

关键词：低轨卫星网络；GP-SR算法；部分可观测马尔可夫决策过程；安全攻击
中图分类号：TP393.0
文献标志码：A
doi：10.11959/j.issn.1000–436x.xing

GP-SR: A GNN-PPO-Based Secure Routing Optimization Algorithm for Low Earth Orbit Satellite Networks

YIN Bin[1,2], LIU Jian[1,2], TANG Xiaomei[1,2], MA Chunjiang[1,2], MA Pengcheng[1,2]
1. College of Electronic Science and Technology, National University of Defense Technology, Changsha 410073, China
2. National Key Laboratory for Positioning, Navigation and Timing Technology, Changsha 410073, China

Abstract: To address the routing optimization problem in low Earth orbit (LEO) satellite networks under conditions of rapidly changing topology, locally observable states, and concurrent security attacks, this paper proposes GNN-PPO-based Secure Routing(GP-SR), a risk-sensitive intelligent routing algorithm based on graph neural networks and proximal policy optimization. The method models distributed routing decision-making as a partially observable Markov decision process, leverages history-enhanced observations to mitigate incomplete state information, and jointly optimizes performance rewards and security costs, guiding the policy to improve network performance while proactively avoiding high-risk links and anomalous nodes. Simulation results show that GP-SR outperforms baseline algorithms in end-to-end delay, throughput, and packet delivery ratio, and demonstrates stronger robustness and adaptability under progressively intensified attack scenarios.

Key words: Low Earth Orbit Satellite Networks,, GP-SR Algorithm, Partially Observable Markov Decision Process,, Security Attacks

0 引言

随着数字经济驱动全域通信需求激增，低轨卫星网络凭借广覆盖与低时延特性，成为弥合地面网络“数字鸿沟”的关键[1]。相比传统卫星通信，低轨卫星网络在实现全球覆盖的同时，能够降低传输时延，因此以低轨卫星网络为核心的通信已成为前

收稿日期：XXXX–XX–XX；修回日期：XXXX–XX–XX
通信作者：刘建，ljabc730@nudt.edu.cn
基金项目：国家自然科学基金资助项目（No.12501031）
Foundation Items：The National Natural Science Foundation of China (No.12501031)

沿热点[2-6]。然而，低轨卫星节点能力均匀、星上资源受限，无法像地面网络那样通过部署高速骨干链路疏导流量。与此同时，暴露的空间信道使星间链路容易受到攻击与干扰，要求路由算法具备在不可靠环境中维持通信的抗毁韧性。因此，在拓扑时变及安全威胁并存的约束下，设计兼顾负载均衡与抗毁生存的高效路由机制极具挑战。

现有低轨卫星网络路由研究大体可分为传统规则驱动方法与机器学习方法两类。传统方法主要围绕负载均衡、故障恢复和协议轻量化展开。例如，Liu等人通过区分轻载与重载区域设计分段负载均衡路由机制，以提升网络吞吐量[7]；Lu等人利用卫星轨道运动的可预测性预计算主备路径，实现动态链路故障下的快速恢复[8]；Wang等人进一步结合链路状态与流量强度因子进行多路径选择，以缓解网络拥塞[9]；He等人则通过轻量化OSPF和跨域聚合路由降低星上控制开销[10]。总体而言，传统路由方法具有机制清晰、实现复杂度较低和可解释性较强等优点，但其多依赖固定规则或局部启发式优化，难以适应低轨卫星网络中复杂时变状态下的全局长期优化需求。

随着人工智能技术的发展，基于机器学习的低轨卫星网络路由方法逐渐受到关注。这类方法通常将路由建模为序贯决策问题，通过与网络环境交互学习长期优化策略。Park等人较早将MDP引入卫星路由，依据轨道几何关系设计分布式角度路由算法[11]；Lyu等人面向集成星地网络提出约束多智能体强化学习方法，在优化端到端时延的同时满足多项QoS约束[12]。进一步地，为刻画卫星网络动态拓扑结构，GNN与DRL的结合成为研究热点。Zhang等人利用GNN提取拓扑特征，并结合Actor-Critic框架实现分布式路由[13]；Xu等人在GNN-DRL框架中引入K最短路径以缩减动作空间、提升训练效率[14]；Li等人进一步结合DRL、GNN与Stackelberg博弈，实现多路径搜索与流量分配的联合优化[15]。总体来看，机器学习方法在动态环境适应性和多目标优化方面具有优势，但仍面临训练开销较大、在线部署复杂和决策可解释性不足等问题。

总结现有研究，大部分算法通常假设网络状态可充分观测且网络运行环境安全可信，较少同时考虑状态部分可观测与主动攻击威胁。然而，现实低轨卫星网络拓扑和链路状态快速变化，节点难以获取全局完备信息；同时，开放信道和星地网络融合使网络攻击与链路干扰成为常态。因此，路由算法需同时兼顾部分可观测性与安全威胁。

为此，本文面向拓扑快速变化、状态局部观测性且易受主动攻击威胁的低轨卫星网络的场景，提出一种融合GNN与近端策略优化（Proximal Policy Optimization，PPO）的智能安全路由算法，动态优化网络数据传输的端到端时延、丢包率和吞吐量。本文具体工作如下：

1) 提出一种面向攻击威胁环境的风险敏感智能路由算法GNN-PPO based Secure Routing（GP-SR）。该方法在保持GNN对动态拓扑关系表征能力和PPO稳定优化能力的基础上，构造风险代价项，并通过性能收益与风险代价的联合优化构建奖励函数，引导路由策略在优化时延、吞吐量与投递率的同时降低高风险链路与异常节点的影响。

2) 构建面向对抗环境的非完全信息路由决策模型。本文将路由过程建模为部分可观测马尔可夫决策过程（Partially Observable Markov Decision Process，POMDP），利用局部观测和历史观测信息进行状态推断，从而更贴合低轨卫星网络中的信息获取约束，并提升动态对抗场景下的决策鲁棒性。

3) 本文基于离散事件技术构建了低轨卫星互联网络仿真平台，可有效模拟卫星星座的动态拓扑切换、链路丢包、节点排队缓存等关键行为。仿真结果表明，GP-SR在端到端时延、吞吐量和包投递率等指标上优于OSPF、DRL和GNN-DRL等基准算法，并在高强度攻击条件下表现出更强的抗毁韧性与鲁棒性。

1 系统模型

1.1 网络模型

本文构建包含空间段与地面段的低轨卫星网络模型。空间段采用基于铱星系统架构的LEO星座，由P个轨道面、每轨S颗卫星组成，形成$P \times S$星座，轨道高度为H，轨道倾角为β；地面段包含N个地面站节点。星间链路中，每颗卫星与相邻四颗卫星建立连接，同轨相邻卫星保持永久连接，异轨卫星根据实时距离与最近邻轨卫星建立链路[16-17]。星地链路采用动态接入机制，地面站与当前最近的K颗卫星连接。无攻击时，各类链路采用固定带

宽和丢包率参数[18]。

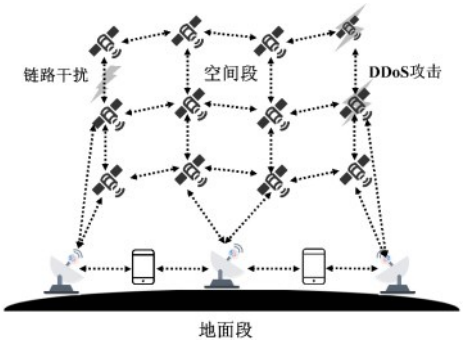


图1　网络与通信模型示意图

在网络节点参数的配置中，为体现卫星与地面节点在资源能力上的固有差异，模型对节点参数选取不同配置策略。卫星节点配置参数为处理速率 μ_{LEO} 和队列容量 Q_{LEO}，地面节点配置参数为处理速率 μ_{GS} 和的队列容量 Q_{GS}，其中 $\mu_{LEO} < \mu_{GS}$，$Q_{LEO} < Q_{GS}$，表示星上相比地面的资源严格受限特性。

针对低轨卫星高速运动引发的拓扑动态性，本文采用离散时间片方法对连续时间进行近似。假设在单个时间片 t 内，整个网络拓扑保持恒定[19]。此时可将网络建模为一个无向图 $\mathcal{G}_t = (\mathcal{V},\mathcal{E}_t)$[20]，其中 $\mathcal{V}$ 为节点集合，$\mathcal{E}_t$ 为时间片 t 内链路集合，整个网络模型由 $\mathcal{T}$ 个连续的时间片的无向图集合 $\{\mathcal{G}_1,\mathcal{G}_2,...,\mathcal{G}_{\mathcal{T}}\}$ 构成。

为模拟真实的网络业务负载，本模型使用五元组 $(src_k,dst_k,\Delta T_k,M_k,L_k)$ 来定义流量 f_k，其中 src_k 和 dst_k 分别表示源和目的节点，ΔT_k 表示源节点发包时间间隔，M_k 表示本条数据流最大发包数量，L_k 表示数据包长度，在此，本文没有采用现有研究将数据包长度设置为定长以简化分析的做法[21-24]，而是考虑实际网络中包长呈双峰分布的特点[25]，将数据包长度按照45%大包、35%小包、20%中间长度。

数据包在链路上的传输过程主要传输时延、传播时延与丢包率影响[26]。在链路 l 上传输时延 T^l_{trans} 由数据包长度 L_j 和链路带宽 B_l 决定，可表示为 $T^l_{trans} = \dfrac{L_j}{B_l}$。传播时延 T^l_{prop} 由链路两端的物理距离 D_l 和光速 c 决定，为 $T^l_{prop} = \dfrac{D_l}{c}$。数据包在通过链路 l 上所需的总时间 T^l 为 $T^l = T^l_{trans} + T^l_{prop}$。此外，为模拟信道干扰，设定链路存在一个固有的丢包率 ε_l，数据包在链路传输过程中会以概率 ε_l 被随机丢弃。

数据包通过链路成功传递到下一网络节点后，节点被建模成一个容量为 Q_{max} 的先进先出（First In First Out，FIFO）队列处理器。若当前队列长度 $Q < Q_{max}$，数据包进入队列等待被处理，否则将被丢弃。若第 i 节点的处理速率为 μ_i，该数据包排队时延为 T^i_{queue}，则该数据长度为 L_j 的数据包在 i 节点内经历的总时延 T^i_{node} 为 $T^i_{node} = T^i_{queue} + \dfrac{L_j}{\mu_i}$。

1.2　路由多智能体模型

为突破传统路由算法的局限，本文引入智能体作为网络中的核心决策单元。将每个智能体部署于网络节点中，具备感知局部网络环境、自主进行路由决策并持续优化策略的能力。智能体的路由决策工作是一个持续的交互-决策-学习循环的过程，图2为智能体路由决策示意图。

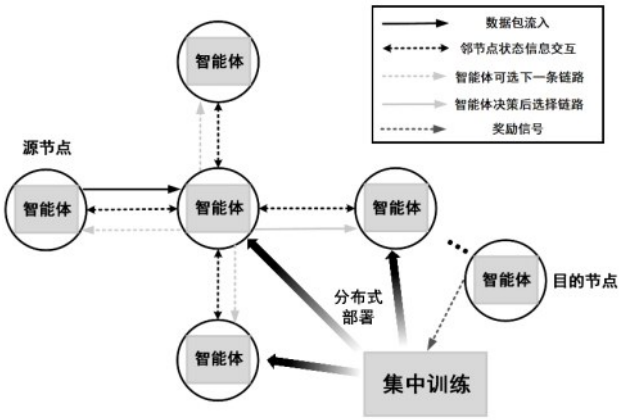


图2　智能体路由决策示意图

训练阶段，智能体周期性获取本节点及邻居节点状态，构建局部观测并输入决策网络，由网络输出各可用下一跳的动作概率分布，进而完成数据包转发；数据包到达目的地后，端到端性能作为奖励反馈，用于更新策略网络。使用阶段，智能体直接基于当前局部观测和已训练策略完成下一跳选择。

随后，鉴于低轨卫星网络同时受到拓扑时变、资源约束和安全攻击的共同影响，本文进一步将路由问题形式化为一个受多重约束的多目标优化模型，以统一描述通信性能与安全风险之间的权衡关系。该模型综合考虑了网络拓扑动态性、资源限制和安全威胁等多重因素，严格遵循节点处理能力、队列容量等系统约束。为避免将安全因素仅作为启

发式惩罚嵌入单一奖励函数，本文将路由优化目标拆分为性能收益项与风险代价项。其中，性能收益项刻画端到端通信质量，风险代价项刻画链路干扰、DDoS 攻击带来的风险累计代价。最终通过联合目标实现性能与安全之间的平衡优化，具体表达式为

$$\begin{aligned}
&\mathcal{J}_{perf}(\pi)=\mathbb{E}_{\pi}|R_{success}-\alpha_1 D_{end-to-end}-\alpha_2 H_{path}|\\
&\mathcal{J}_{risk}(\pi)=\mathbb{E}_{\pi}|C_{risk}|\\
&\max\ \mathcal{J}(\pi)=\mathcal{J}_{perf}(\pi)-\lambda\mathcal{J}_{risk}(\pi)\\
&\text{s.t. } C_1{:}TTL(p)>0,\forall p\in\mathcal{P}_{active}\\
&C_2{:}\sum_{l\in\mathcal{L}_{out(i)}} f_l(t)\leqslant\mu_i,\forall i\in V,\forall t \qquad \#(1)\#\\
&C_3{:}f_l(t)\leqslant B_l(t),\forall l\in E(t),\forall t\\
&C_4{:}\sum_{l\in\mathcal{L}_{in(i)}} f_l(t)+g_i(t)=\sum_{l\in\mathcal{L}_{out(i)}} f_l(t),\forall i\notin\{src,dst\}\\
&C_5{:}N_k\leqslant N_{max}^k,\forall k\in\mathcal{K}
\end{aligned}$$

数学期望符号 $\mathbb{E}$ 强调对全网的、长期的统计平均性能优化，而非针对单次传输结果的局部优化，体现了算法在动态网络环境中对系统整体性能的追求。约束 C1 表示环路抑制与最大跳数约束，TTL（Time To Live）设为最大允许跳数，并在每经过一个转发节点时减 1；当 TTL 减至 0 时，数据包被丢弃。该机制可限制数据包的最大转发次数，避免其因路由环路在网络中持续循环，从而减少资源浪费；约束 C2 要求在任意时刻 t，节点 i 发送数据的总速率，不能超过 μ_i；约束 C3 表示在任何时刻 t，任何一条通信链路 l 上正在传输的数据量，不能超过 $B_l(t)$；约束 C4 表示对于网络中任何一个中间节点，流入它的总数据量 $\sum_{l\in\mathcal{L}_{in(i)}} f_l(t)$ 加上它自己产生数据量 $g_i(t)$，必须等于从它流出的总数据量 $\sum_{l\in\mathcal{L}_{out(i)}} f_l(t)$；约束 C5 表示任何一条数据流所能发送的数据包总数 N_k，不能超过该条数据流设定的最大上限 N_{max}^k。

1.3 攻击威胁模型

低轨卫星网络固有的全球广域覆盖、信道动态开放、物理维护不可达等特性，使其面临的网络与电磁攻击威胁日益严峻，将成为一种新常态。基于此，本文引入安全攻击模型，模拟针对空间段的持续攻击场景，以提升路由算法在对抗性环境下的鲁棒性、自适应性。

低轨卫星网络可能面临欺骗、窃听、DDoS、链路干扰和路由欺骗等多种安全威胁。本文重点关注攻击扰动对路由性能的影响，因此选取 DDoS 攻击和链路干扰攻击作为典型威胁模型[27-30]。DDoS 攻击适用于模拟攻击者利用大量受控终端或恶意接入节点向目标卫星持续发送高频无效业务请求的场景。在该场景下，被攻击卫星的星上处理资源和缓存资源被大量占用，表现为处理速率下降、队列容量受限以及排队丢包增加。星间链路干扰攻击适用于模拟开放空间信道受到恶意电磁干扰、链路压制或局部通信阻断的场景。在该场景下，受干扰链路的传输能力下降，表现为带宽降低、传播时延增加和丢包率升高。上述两类攻击分别从节点侧和链路侧刻画低轨卫星网络面临的典型对抗威胁，能够直接影响端到端时延、吞吐量和数据包成功送达率等路由性能指标，因此与本文安全路由优化问题具有较高一致性。

具体攻击威胁建模形式为：在每个时间片内，随机选取全网卫星节点总数比例 $\mathrm{P}_{\mathrm{node}}^{\mathrm{attack}}$ 作为 DDoS 攻击目标。被攻击节点的处理速率 μ_i 和队列容量 Q_{max} 按衰减因子 α_{node} 动态降低，其中 α_{node} 服从期望为 μ_{node}，标准差为 σ_{node} 的截断正态分布（截断区间 $[a_{\mathrm{node}},b_{\mathrm{node}}]$）$\alpha_{node}\sim TN(\mu_{\mathrm{node}},\sigma_{\mathrm{node}},a_{\mathrm{node}},b_{\mathrm{node}})$，第 i 个卫星节点被攻击后的数据处理速率 μ_i^{attack} 和队列容量 Q_i^{attack} 为

$$\mu_i^{\mathrm{attack}}=\alpha_{\mathrm{node}}\times\mu_i,Q_i^{\mathrm{attack}}=\alpha_{\mathrm{node}}\times Q_{\max}\#(2)$$

类似地，在每个时间片内，随机选取全网链路总数比例 $P_{\mathrm{link}}^{\mathrm{attack}}$ 的链路作为干扰目标。被干扰链路的丢包率、传播时延和带宽分别按恶化因子 α_{link} 动态变化，其中 $\alpha_{\mathrm{link}}\sim TN(\mu_{\mathrm{link}},\sigma_{\mathrm{link}},a_{\mathrm{link}},b_{\mathrm{link}})$，第 j 条链路被干扰之后的丢包率 ε_l^{attack}、传播延迟 $T_{prop}^{l-attack}$ 和带宽 B_l^{attack} 的表达式如下所示

$\varepsilon_l^{\text{attack}} = (1 + \alpha_{\text{link}}) \times \varepsilon_l, T_{prop}^{l\text{ - }attack} = (1 + \alpha_{link}) \times T_{prop}^{l}, B_l^{attack} = (1 - \alpha_{link}) \times B_l \#(3)$

2　GP-SR智能安全路由算法设计

本节提出一种基于GNN与PPO的智能安全路由算法，该算法通过拓扑表征学习与多目标策略优化的深度融合，实现自适应的路由决策。具体而言，首先将路由问题建模为POMDP，利用GNN的拓扑表征能力，以增强智能体对非完全信息的推断能力；随后采用近端策略优化算法，通过训练，使各节点在局部观测基础上协同学习逼近全局最优的多目标路由策略。

2.1　基于POMDP的路由问题建模

本文场景下的路由优化面临两大核心挑战：网络拓扑的动态性及网络易受攻击威胁带来的结构或参数的不稳定性，以及网络状态的部分可观测性。传统的MDP过程无法准确描述这一复杂场景，一方面，实际网络中节点无法获取全网实时状态；另一方面，MDP缺乏对历史时序信息的利用。相比之下，POMDP模型通过引入历史观测序列，能够为智能体积累时空上下文，有效捕捉与风险相关的时序变化模式。为此，本文将分布式路由问题形式化为POMDP，为后续智能安全路由算法提供理论基础。POMDP框架通过七元组$(\mathrm{S,A,P,R,\Omega,O},Y)$对路由决策问题进行建模。

其中，S表示系统的真实状态空间，包含网络中所有节点的队列状态、处理负载、链路质量以及完整的拓扑连接关系。

A表示动作空间，将每个节点作为智能体，A表示其在每个决策时刻可执行的操作集合。具体的，动作$a_t \in A$即是为数据包选择下一跳节点。

$P(s_{t+1}|s_t,a_t)$表示状态转移函数，描述了网络状态（如拓扑结构、负载）的动态演化，定义为在全局状态s_t下，所有节点执行动作a_t后，网络转移到新状态s_{t+1}的概率。

$R(s_t,a_t)$表示奖励函数，用于评价动作a_t在状态s_t下的优劣，它直接对应上一节最后所确立的优化目标。由于s_t不可直接获得，在实际算法中，本文使用基于观测o_t和后续结果（如数据包是否成功投递、端到端时延）计算奖励r_t。其设计原则与之前一致，旨在通过最大化累积奖励的期望，来间接优化端到端时延、传输成功率等目标。

Ω表示观测空间，在本文中，单时刻局部观测记为o_i^t，但实际决策输入并非单一o_i^t，而是由最近L个时刻局部观测构成的历史窗口$h_i^t = \left\{o_i^{t\text{ - }L+1}, o_i^{t\text{ - }L+2}, ..., o_i^t,\right\}$。因此，本文的策略学习采用$\pi(a_i^t|h_i^t)$而非$\pi(a_i^t|o_i^t)$的形式，以体现部分可观测环境下对历史上下文的依赖。

$O(o_{t+1}|s_{t+1},a_t)$表示观测函数，主要通过对局部网络的感知和邻居信息交换获得，定义了在状态s_{t+1}下，节点获得观测o_{t+1}的概率。

$Y \in [0,1]$表示折扣因子，用于平衡即时奖励与长期回报。在本问题中，尽量将Y设定为接近于1的数，以强调路由决策的长期影响，引导智能体不仅关注下一跳的即时成本，更要考虑到网络拥塞或者被安全攻击后网络整个端到端路径的长期性能。

通过上述POMDP建模，本文将低轨卫星网络路由问题明确定义为一个在部分可观测、随机动态环境下（体现在因网络拓扑易变和被威胁攻击后导致的网络参数改变）的序列决策问题，此模型为下一节GP-SR智能安全路由算法定义了清晰的问题框架。

根据POMDP理论，最优策略通常依赖于信念状态$b_t = P(s_t|o_{1:t},a_{1P:t\text{ - }1})$，即由历史观测与历史动作共同决定的对当前隐含全局状态的后验表示。然而，在低轨卫星网络中，网络拓扑高速变化、状态维度高且安全威胁具有突发性，显式维护信念状态的计算代价过高。为此，本文不直接求解精确的信念状态，而采用有限历史窗口进行近似。对于节点i，定义其在时刻t的历史观测窗口为h_i^t。进一步地，通过历史编码函数$\phi(\cdot)$将h_i^t映射为历史增强特征$\tilde{o}_i^t = \phi(h_i^t)$，并将其作为图神经网络的节点输入特征，从而得到潜在状态表示$z_t = f_\theta(G_t,\tilde{O}_t)$。其中$G_t$表示时刻$t$的局部拓扑图，$\tilde{O}_t$表示所有节点的历史增强特征集合。该潜在表示同时刻画当前局部状态及其短期演化趋势，可视作一种面向部分可观测场景的历史增强状态表征，并供后续策略网络与价值网络使用。

2.2　GP-SR智能安全路由算法框架

基于2.1节所定义的POMDP路由问题模型，构建了一个面向低轨卫星网络动态安全路由问题的智能决策框架，GNN通过对网络拓扑的特征提取，使网络节点智能体具备拓扑学习与关系推理能力，能够适应并且通过图结构学习节点邻居连接关系的动态变化与链路参数变化，并采用PPO算法保证

策略学习的稳定性。图3为GP-SR算法框架图，本框架通过GNN特征提取器、Actor-Critic决策网络与PPO优化器三大核心模块协同，实现了基于局部观测到近似全局最优决策的端到端学习。

（1）POMDP重点要素实例化

第2.1节从抽象层面定义了POMDP理论框架，为将这一理论模型转化为可训练、可部署的智能路由算法，本小节首先对POMDP的核心要素进行具体化实例。这一步骤旨在将抽象的数学框架映射为算法中可操作的数据结构与函数，为后续GP-SR算法实现奠定基础。接下来本文将重点实例化三个关键要素：观测空间、动作空间与奖励函数。

观测空间实例化：在POMDP框架下，节点i在时隙t的局部观测状态公式为式6所示。需要说明的是，式6中的各状态特征并非真实全局状态的精确值，而是智能体在部分可观测条件下可获得的带噪观测量或局部统计估计量。util_i表示当前节点队列利用率；$\widetilde{\text{util}}_{\text{neighbors}}$表示邻居节点队列利用率的观测向量；$\text{nei}_i$表示当前节点的邻节点数量；$\text{loss}_{\text{avg}}$是由历史窗口内链路丢包事件统计得到的平均丢包率估计；$\text{delay}_{\text{est}}$表示由静态路由表计算的$d_{\text{src}}$到$d_{\text{dst}}$预估端到端时延；$\text{delay}_{\text{acc}}$表示数据包传输到本节点的累计传输时延；$\widetilde{\text{delay}}_{\text{avg}}$是邻居平均延迟，用来间接反映局部网络拥塞程度。上述观测量可能受到链路测量误差、信息交换延迟或随机扰动的影响，因此本文将其作为带噪声的真实状态观测样本。

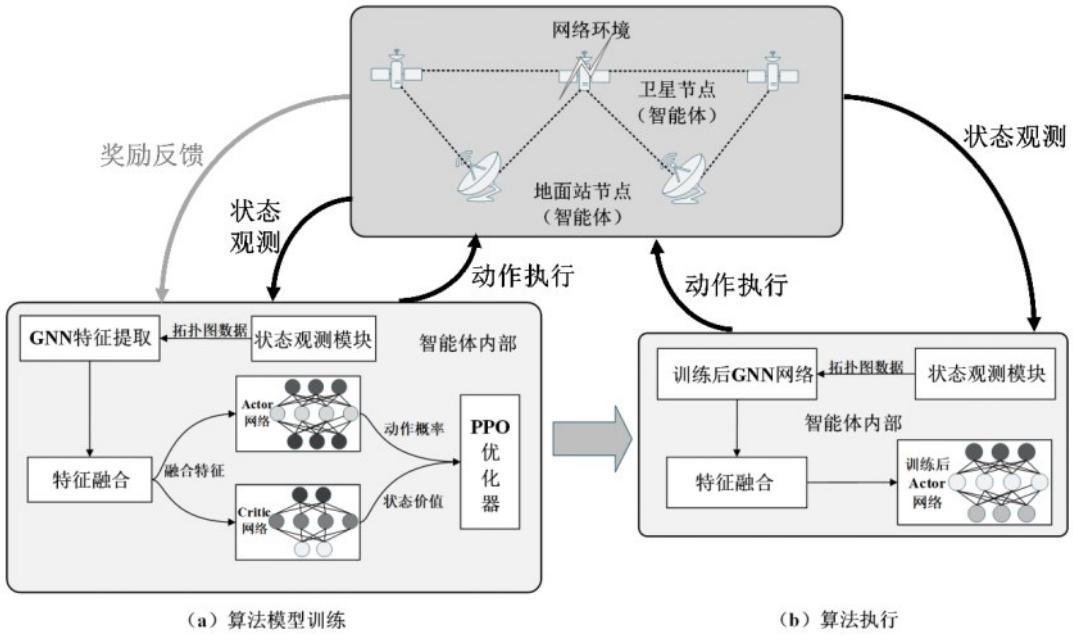


图3　GP-SR路由算法框架图

$$o_t^i = \text{Concat}\left(\widetilde{\text{util}}_i, \widetilde{\text{util}}_{\text{neighbors}}, \text{nei}_i, \text{loss}_{\text{avg}}, \text{delay}_{\text{est}}, d_{\text{src}}, d_{\text{dst}}, \text{delay}_{\text{acc}}, \widetilde{\text{delay}}_{\text{avg}}\right) \#(4)$$

考虑到单时刻局部观测难以充分反映网络状态的短期演化趋势，本文构造长度为L的历史观测窗口h_i^t，并采用时间拼接与统计增强相结合的方式构造历史增强观测$\tilde{o}_i^t = \left[o_i^{t-L+1} \| \cdots \| o_i^t \| \bar{o}_i^t \| \Delta o_i^t\right]$。其中，$\bar{o}_i^t$表示窗口内各维特征的均值统计量，$\Delta o_i^t = o_i^t - o_i^{t-1}$表示最近一次变化量。历史窗口用于平滑瞬时观测噪声，并刻画链路退化、队列拥塞和异常扰动等短期变化趋势，从而为部分可观测环境下的策略学习提供时序上下文。

动作实例化：对于本文的研究的路由算法来说，本质就是让节点将数据包从可以优化整网性能的最优链路上转发，故在此将动作空间定义为当前节点可用输出链路的离散选择集合，其中N为当前节点的邻节点个数，每个动作a_n对应选择第n个邻居作为数据包转发路径下一跳节点。

$$A = \left\{a_n | a_n \in \{0,1,...,N-1\}\right\} \#(5)$$

奖励实例化：反馈信号的设计直接决定了智能体强化学习的优化方向。本文将训练反馈拆分为性

能收益项与安全代价项：前者用于刻画端到端通信质量，后者用于刻画链路干扰、异常波动带来的潜在风险暴露。在性能收益部分，其中α_1到α_4为权重系数，通过调整权重系数可以使得模型更加关注特定的性能指标。其中R_{base}为基础奖励，是固定正向奖励，用来避免策略学习初期因负奖励占主导而导致的梯度消失问题；R_{delay}为时延奖励，$R_{delay} = k_1 \times \tanh(\frac{D_{actual}}{D_{max}})$，本部分设计的优势在于$D_{max}$的引入，$D_{max}$为使用基准路由运行仿真的P90的平均端到端时延（在所有请求中，有90%的请求的响应时间小于等于该值），认为是可接受的最大时延值，采用双曲正切函数使智能体对时延变化更加敏感，特别是当$D_{actual} < D_{max}$时，奖励惩罚随延迟近似线性增长；$R_{success}$为成功传输奖励，设置明显的正向奖励，作为路由任务完成的最终目标信号；P_{hop}为跳数惩罚，对超过最小必要跳数的路径施加线性惩罚。

$$
\begin{aligned}
r_t^{perf} &= \alpha_1 R_{base} - \alpha_2 R_{delay} + \alpha_3 R_{success} - \alpha_4 P_{hop} \\
\hat{c}_t^{risk} &= \hat{c}^{risk}(h_t, a_t) = \beta_1 \hat{c}_{loss} + \beta_2 \hat{c}_{queue} + \beta_3 \hat{c}_{anom} \quad \#(6)\# \\
\hat{r}_t &= r_t^{perf} - \lambda \hat{c}_t^{risk}
\end{aligned}
$$

$$
\begin{aligned}
\hat{c}_{loss} &= \widetilde{\text{lossavg}}_t \\
\hat{c}_{queue} &= \left[\widetilde{\text{util}}_t - u_{th} \right]_+ \qquad \#(7)\# \\
\hat{c}_{anom} &= Y_1 |\Delta \widetilde{\text{lossavg}}_t| + Y_2 |\Delta \widetilde{\text{util}}_t| + Y_3 |\Delta \widetilde{\text{delayavg}}_t|
\end{aligned}
$$

在安全代价方面，本文不再将安全因素简单视为一次性丢包后的强负奖励，而是将其建模为由链路退化风险、局部拥塞风险与异常波动风险共同构成的连续代价项。其核心思想是：当某一转发方向持续表现出较高丢包率、资源紧张或状态突变时，即使尚未立即导致任务失败，该方向也应被视为潜在高风险路径，并在策略优化中提前受到抑制。需要强调的是，在POMDP建模下，智能体无法获得真实全局状态，因此式8中的$\hat{c}_t^{risk}$并非真实状态下的确定性风险代价，而是基于历史观测窗口h_t和动作a_t构造的随机风险估计量，即$\hat{c}_t^{risk} = \hat{c}^{risk}(h_t, a_t)$。

其中，β_1到β_3为权重系数，用于平衡不同风险因素对策略学习的影响。其中$\hat{c}_{loss}$为链路退化风险项，用链路平均丢包率刻画当前转发方向的传输可靠性下降程度；$\hat{c}_{queue}$为局部拥塞风险估计，u_{th}为队列利用率安全阈值，$[x]_+ = \max(x, 0)$，当前节点或候选转发方向的资源使用超过阈值后，超出的部分将被计入安全代价，从而引导策略主动避开局部高拥塞区域；$\hat{c}_{anom}$表示异常波动风险项，用于刻画短时间窗口内网络状态的突变程度，$\Delta\widetilde{\text{lossavg}}_t$、$\Delta\widetilde{\text{util}}_t$、$\Delta\widetilde{\text{delayavg}}_t$分别表示当前时刻相对于上一时刻在链路丢包率、节点利用率和平均时延上的变化幅度。该项越大，说明局部网络状态越不稳定，越可能存在潜在攻击、链路扰动或拥塞恶化趋势。

进一步地，若真实链路退化、节点拥塞和局部时延状态分别对应于不可直接观测的环境状态变量，则智能体实际获得的链路丢包率、队列利用率和平均时延等特征应理解为其带噪观测估计，而非真实状态的精确值。对于链路平均丢包率、队列利用率等取值位于[0,1]区间的特征，本文采用$\tilde{x}_t = \text{clip}(x_t + \varepsilon_t, 0, 1)$表示其观测估计，其中$\varepsilon_t \sim N(0, \sigma^2)$表示观测误差、采样误差和统计窗口波动等因素引入的观测噪声；对于平均时延等具有物理量纲的特征，则采用归一化后加噪声或相对扰动形式建模，以避免不同量纲特征受到不一致的扰动影响。因此，本文中的奖励信号是基于历史观测得到的随机反馈信号，策略优化目标是在观测历史诱导的随机轨迹分布上最大化期望累积回报，而非假定智能体能够获得真实全局状态下的确定性奖励。本文所定义的安全代价不依赖外部入侵检测器或显式攻击标签，而是基于局部观测、历史统计量及其短时变化构造，更符合低轨卫星网络中信息受限、状态部分可观测的实际场景。

（2）状态表征与GNN特征提取

在低轨卫星网络这种拓扑高度动态的环境中，智能体无法直接获取全局网络状态。为使部分可观测环境下的历史信息得到有效利用，本文首先在时间维度上对局部观测进行编码。对于节点i在时隙t的局部观测o_t^i，长度为L的历史观测窗口为h_i^t，进一步地，通过历史编码函数$\phi(\cdot)$将h_i^t映射为$\tilde{o}_i^t$，其中$\phi(\cdot)$可采用特征拼接与统计量提取的方式实现。

在完成时间维度建模后，将智能体通信范围内的局部网络表示为动态图结构$G_t = (V_t, E_t, X_t)$。V_t

表示节点集合，包含自身及其所有一跳邻居节点；E_t表示边集合，对应当前星间或星地链路；$X_t=\{\tilde{o}_1^t,\tilde{o}_2^t,...,\tilde{o}_N^t\}$为节点特征矩阵，$\tilde{o}_i^t$中在包含局部观测信息$o_i^t$的同时，也包含了安全相关状态信息，例如链路的近期丢包率p_{ij}、节点异常变化度α_i，这些特征共同构成了模型感知安全威胁的“原始信号”。

为了精准捕捉多跳范围内的拓扑变化与潜在攻击特征，采用基于注意力机制的消息传递架构（Graph Attention Network, GAT）。其核心在于，消息传递过程中的注意力权重α_{ij}是通过学习动态生成的，它使节点能够自适应地衡量来自不同邻居的信息价值与可信度。在训练中，模型自动学习为那些表现异常的邻居分配较低的注意力权重，从而在特征聚合阶段自然抑制了潜在不可靠或恶意节点的影响，第l层的节点特征更新公式定义如下：

$$\mathbf{h}_i^{(l+1)}=\sigma\Big(\sum_{j\in\mathcal{N}(i)}\alpha_{ij}\mathbf{W}^{(l)}\mathbf{h}_j^{(l)}\Big)\#(8)$$

$\mathbf{h}_i^{(l+1)}$为是节点i在第$l+1$层的输出特征嵌入，由可学习的注意力机制动态计算，使智能体能够自适应地关注不同邻居节点的重要性差异；σ代表非线性激活函数；$\mathbf{W}^{(l)}$是参数权重矩阵。经过多轮消息传递后，通过全局池化操作将整个局部图的节点嵌入聚合为统一的图级表征$\mathbf{h}_{\text{gnn}}$。最后在特征融合曾进行拼接，得到$\mathbf{h}=\text{Concat}(\mathbf{h}_{\text{gnn}},\mathbf{h}_{\text{state}})$。$\mathbf{h}$同时包含了拓扑结构信息和节点特征信息，为后续Actor-Critic网络决策提供了全面而强大的状态表征，图4为基于注意力聚合的GNN特征提取示意图。

（3）Actor-Critic网络架构

本框架采用Actor-Critic架构实现策略学习，其中Actor网络负责生成路由策略，Critic网络负责评估状态价值，两者协同工作以实现稳定的策略优化。Actor网络承担将环境状态映射为动作概率分布的关键职能，其采用深度神经网络结构，主要包含三个部分。输入层：接受GNN提取输出的256维特征向量$\mathbf{h}\in\mathbb{R}^{256}$，输出向量$\mathbf{h}_1$；隐藏层：使用多层全连接网络，使用高斯误差线性单元（Gaussian Error Linear Unit，GeLU）作为激活函数，输出向量$\mathbf{h}_2$；输出层：采用Softmax函数生成动作概率分布。

对于Critic网络来说，它负责状态价值函数估计，为策略梯度提供基线值以降低方差。其整体架构与Actor网络相似，输入层同样是接收GNN所提取的特征向量，具有差异的是Critic网络的输出层，它是单神经元输出，表示状态的长期期望回报。

（4）基于PPO的协同训练机制

GNN特征提取器与Actor-Critic网络共同构成了智能体在部分可观测环境下的决策架构，而PPO算法则是实现其协同训练与稳定优化的核心引擎。针对网络攻击引起的高方差奖励信号，本文采用一种面向对抗环境的on-policy鲁棒更新机制，通过轨迹采样、广义优势估计（Generalized advantage estimation，GAE）与裁剪目标函数，实现策略网络与价值网络的稳定协同优化。

智能体通过执行动作与动态网络环境进行交互。在每个决策时刻t，智能体维护长度为L的历史观测窗口h_i^t，并基于h_i^t生成动作概率分布。因而训练样本被组织为$(h_i^t,a_i^t,r_i^t,h_i^{t+1})$，而非传统MDP的$(o_i^t,a_i^t,r_i^t,o_i^{t+1})$。所有样本仅在当前策略下采集，

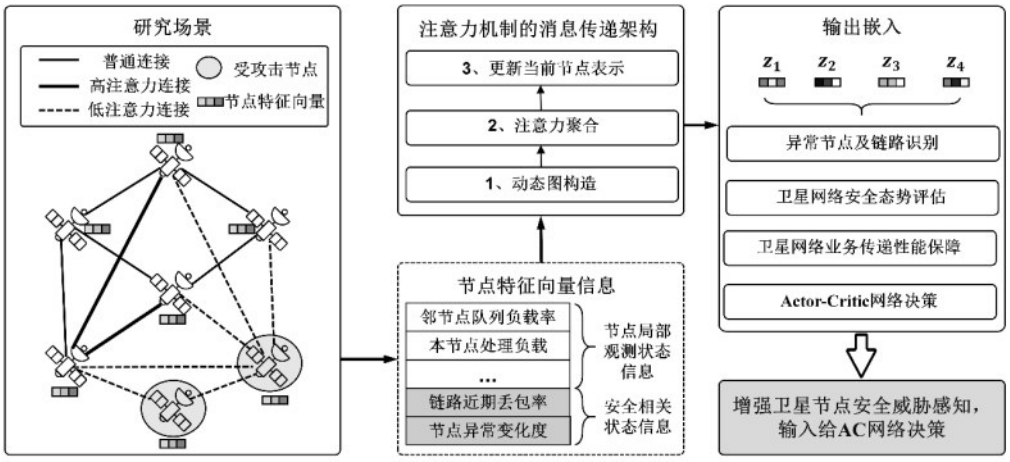


图4　基于注意力聚合的GNN特征提取示意图

并临时存入当前轮轨迹缓存D中；在完成一轮PPO更新后即清空，不跨轮复用，以保证on-policy训练的一致性。

在训练阶段，首先根据历史观测窗口构造历史增强节点特征，并结合当前局部拓扑输入GNN编码器，得到状态嵌入表示z_t。随后，将z_t输入Actor网络和Critic网络，分别输出动作概率分布$\pi_\theta(a_t|z_t)$和状态价值估计$V_\phi(z_t)$。在此基础上，基于风险调整回报$\widetilde{r}_t$与状态价值估计结果，采用GAE计算优势函数$\hat{A}_t$以降低高方差回报对训练稳定性的影响。

在策略更新阶段，采用PPO-Clip目标函数对Actor网络参数进行优化，其表达式如式11所示。概率比$r_t(\theta)=\dfrac{\pi_\theta(a_t|z_t)}{\pi_{old}(a_t|z_t)}$用于衡量新旧策略对同一动作的选择倾向变化。裁剪函数将$r_t(\theta)$限制在$[1-\epsilon,1+\epsilon]$区间内，从而抑制单次更新幅度过大，避免策略在高波动奖励环境中发生剧烈震荡。进一步地，目标函数通过在未裁剪项与裁剪项之间取较小值，实现对策略更新步长的保守控制。当$\hat{A}_t>0$时，目标函数倾向于适度提高对应动作的选择概率；当$\hat{A}_t<0$时，则倾向于降低该动作的选择概率。由此，PPO能够在策略性能提升与训练稳定性之间取得平衡。

$$L^{\mathrm{CLIP}}(\theta)=E_t\left[\min\left(r_t(\theta)\hat{A}_t,\mathrm{clip}\left(r_t(\theta),1-\epsilon,1+\epsilon\right)\hat{A}_t\right)\right]\#(9)$$

综上所述，本框架通过历史轨迹采集、前向传播、优势估计与PPO-Clip更新等一系列步骤，实现了GNN、Actor、Critic三大核心模块的协同优化。完整的训练流程为算法1所示的伪代码。

算法1　GP-SR智能安全路由算法

输入　环境env、折扣因子γ、历史窗口大小L、裁剪系数ϵ，批量大小B，训练轮数M，单轮仿真总时间T，更新轮数K

输出　优化后的路由策略参数θ^*

1) 初始化Actor网络参数θ，Critic网络参数φ，GNN编码器参数θ_g

2) for *episode* = 1 to M do

3) 重置环境，获得初始局部观测o_i^0

4) 对每个节点初始化历史窗口$h_i^0=\{o_i^0,..,o_i^0\}$

5) 初始化当前轮轨迹缓存D

6) for $t=0$ to T-1 do

7) 根据h_i^t构造历史增强特征$\tilde{o}_i^t$

8) 构建当前图结构G_t

9) 将$(G_t,\tilde{O}_t)$输入GNN，得到状态嵌入z_t

10) 根据Actor输出策略$\pi_\theta(a_t|z_t)$

11) 按策略分布采样动作a_t

12) 执行动作，计算性能收益r_t^{perf}和风险代价c_t^{risk}，以及获得新观测o_i^{t+1}

13) 更新历史窗口h_i^{t+1}

14) 将$(h_i^t,a_i^t,r_i^t,h_i^{t+1})$存入轨迹缓存$D$

15) end for

16) 根据D计算回报和GAE优势$\hat{A}_t$

17) for $k=0$ to K do

18) 从D中按小批量采样当前轮轨迹子集

19) 计算PPO-Clip损失与Critic损失

20) 更新GNN、Actor和Critic参数

21) end for

22) 清空轨迹缓存D

23) end for

24) 返回优化后的网络参数

本框架采用集中式训练分布式执行的训练模式：在训练阶段，所有节点的路由决策统一收集经验并进行模型更新；在执行阶段，各节点基于训练好的策略网络独立进行分布式路由决策。

3　仿真结果及分析

为支撑本文算法分析验证，我们构建了离散事件网络仿真平台，具备卫星轨道动力学建模、星地/星间链路建模、节点队列与报文转发行为建模的全流程仿真能力，模块化架构可灵活接入各类路由策略，支持在高动态、大规模星座场景下完成路由算法性能验证与参数优化。平台中，我们基于PyTorch框架实现了GNN与PPO相结合的智能路由模型，模型参数如表2所示。在后续实验中采用包含66颗铱星星座节点与10个地面站节点的低轨卫星网络拓扑。为模拟网络动态性，在卫星运动过程中将总仿真时长划分为$T=10$个时间片拓扑，并在每个时间片内保持拓扑结构静态不变，以离散化地表征网络演化过程，具体其他参数设置如表1所示。仿真实验设备配置为Intel(R) Core(TM) i9-14900HX (2.20 GHz)，NVIDIA GeForce RTX 5070 Laptop GPU和64位Windows 11操作系统。本节

中，我们依次开展算法收敛性、网络性能与稳定性对比、动态威胁下抗毁能力以及复杂性分析等实验。

表1 仿真参数

仿真参数	参数值
轨道面P，每轨卫星数S	6，11
轨道高度为H，轨道倾角为β	780km，86.4°
地面站建链参数K	10
卫星处理速率μ_{LEO}，队列容量Q_{LEO}	2000，6000
μ_{GS}	7000，10240
星间链路参数$B_{\text{LEO - LEO}}$,$\varepsilon_{\text{LEO - LEO}}$	1000，0.5%
星地链路参数$B_{\text{LEO - GS}}$,$\varepsilon_{\text{LEO - GS}}$	6000，2%
攻击威胁参数P_{node}^{attack}、P_{link}^{attack}	5%，10%
DDoS攻击参数μ_{node},σ_{node},a_{node},b_{node}	0.5，0.15，0.1，0.9
链路干扰参数μ_{link},σ_{link},a_{link},b_{link}	0.2，0.08，0.05，0.5

3.1 算法训练阶段收敛性分析评估

本小节对算法训练阶段的收敛性进行分析评估。为模拟真实网络中的高负载运行场景，并提升算法在网络攻击条件下的鲁棒性，训练阶段引入两条持续的大流量背景数据流，同时按照表1中的攻击威胁参数，对随机节点和链路施加DDoS攻击与链路干扰。该设置旨在主动构造网络资源竞争和安全扰动并存的复杂环境，促使GP-SR智能体学习在高负载与随机攻击条件下进行有效的负载均衡与路由决策，避免模型在理想化低负载环境中过拟合，并提升其对动态威胁场景的适应能力。

与此同时，本文选取基于DRL的Actor-Critic（AC）框架作为对比算法，重点比较两种算法在复杂对抗环境下的训练收敛行为，从收敛速度和最终性能两个方面评估GP-SR算法的优势。进一步地，考虑到链路丢包率、队列利用率等风险相关特征均由局部测量、邻居信息交换和历史统计得到，因而不可避免地受到观测噪声影响。为分析局部观测噪声对风险代价估计和策略收敛稳定性的影响，本文进一步开展观测噪声敏感性实验。

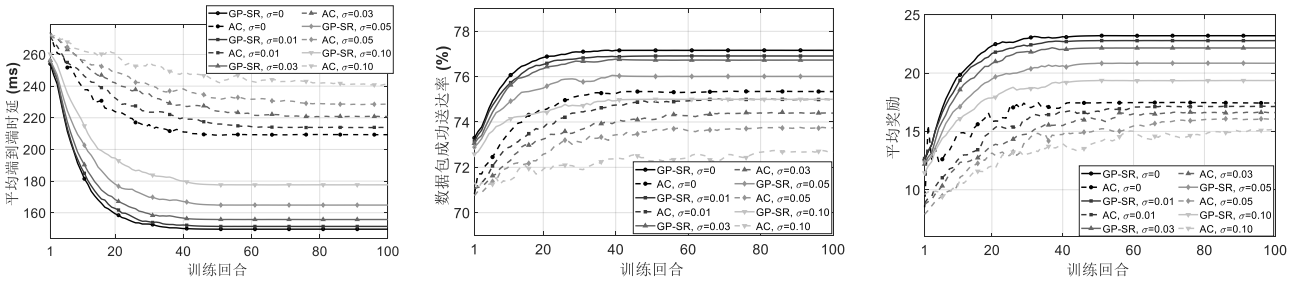
具体而言，在保持网络拓扑、业务流量、攻击模型以及真实环境状态转移过程不变的条件下，仅对智能体可获得的局部观测的归一化特征加入零均值高斯噪声$\varepsilon_t \sim \mathcal{N}(0,\sigma^2)$，$\sigma$表示观测噪声强度，以模拟测量误差、信息交换延迟和统计波动等因素造成的观测不确定性。本文分别设置σ=0，0.01，0.03，0.05，0.10，用于模拟从无噪声到强观测扰动的不同条件。

数据集的数据流组$F_1=\{f_1,f_2\}$，其具体参数为：$f_1=(src_1,dst_1,0.01,20000,L_1\sim\mathrm{U}(1200,1500))$，$f_2=(src_2,dst_2,0.015,20000,L_1\sim\mathrm{U}(1200,1500))$。训练轮数为50轮，每回合保存一次模型，当本回合训练之前首先继承上一回合的模型参数，训练结果如图5所示。

由图5可见，随着观测噪声强度σ增大，GP-SR的收敛轮数由25轮增加至36轮，平均端到端时延由149.62ms上升至177.63ms，数据包成功送达率由77.16%下降至75.01%，平均奖励标准差由0.41增大至0.92，说明观测噪声会增加风险代价估计的不确定性，并进一步引起奖励波动和收敛速度下降。相比之下，在相同观测噪声条件下，AC基线受噪声影响更为明显，其收敛轮数由40—44轮增加至78—86轮，平均端到端时延由209.33 ms上升至240.83 ms，数据包成功送达率由75.34%下降至72.69%。综合来看，在不同观测噪声条件下，GP-SR均较AC表现出更少的收敛轮数、更优的网络性能和更高平均奖励，说明其历史观测增强、风险代价估计与GNN-PPO协同优化机制能够有效提升部分可观测带噪环境下的策略收敛性与路由鲁

表2 深度强化学习参数

模型参数	参数值
Actor，Critic网络损失函数	PPO-Clip函数，MSE
优化器，激活函数	Adam，（GeLU，Softmax）
GNN网络	3层GAT卷积（128，128，64）、全局池化和2层全连接
Actor网络，Critic网络	2个512单元的隐藏层
折扣因子Y	0.99
批量大小B	256



（a） 平均端到端时延对比图　　（b） 数据包成功送达率对比图　　（c） 平均奖励对比图

图5　训练结果图

棒性。

参考带噪状态观测强化学习研究中常用的随机观测噪声设置，本文在归一化局部观测特征上加入零均值高斯噪声，并设置σ=0，0.01，0.03，0.05，0.10进行敏感性分析。其中，σ=0.03属于中低强度观测扰动，既能够模拟局部测量误差和邻居信息交换延迟带来的观测不确定性，又不会使噪声成为影响算法性能的主导因素，因此后续实验默认采用σ=0.03。

3.2　网络性能与稳定性分析评估

本小节围绕网络性能与稳定性分析评估展开，重点从历史窗口敏感性、大规模星座扩展性、复杂业务负载下的路由性能以及多随机种子条件下的结果稳定性四个方面，全面验证GP-SR算法在不同网络规模、业务负载和随机扰动条件下的有效性与鲁棒性。

在模型训练完成之后，为进一步测试算法的泛化能力，对比实验采用了更为复杂的测试集。该测试集包含以大业务量为主的7种不同类型的数据流$F_2=\{f_1',f_2',f_3,f_4,f_5,f_6,f_7\}$（如稳定高吞吐量流、突发性流、平稳背景流等），以模拟真实的复合业务模式。同时，引入了网络安全攻击模拟，在每个时间片内，随机选择网络节点或链路施加攻击威胁，以评估算法在攻击威胁下的韧性。由于本文方法在POMDP框架下采用长度为L的历史观测窗口对部分可观测状态进行近似表征，因此有必要进一步分析窗口长度选择对模型性能的影响。为验证历史增强机制的有效性，并避免后续基线对比结论依赖于经验性参数设定，本文在总体性能对比实验之前，首先开展历史窗口大小敏感性分析实验。分别设置$L\in\{1,2,4,6,8\}$，对应的网络性能结果如表3所示。

由表3可见，随着历史窗口长度L增大，算法性能整体呈先提升后趋于饱和的趋势。当L=6时，整体网络性能基本处于最佳。当L继续增大到8时，虽然成功送达率略有提升，但时延回升、吞吐量下降，且过长的历史窗口会引入冗余信息。综合考虑上述影响之后，本文最终选取L=6作为历史观测窗口长度。

为进一步验证所提GP-SR算法在较大规模低轨卫星星座中的适用性，本文进一步开展大规模星座扩展性实验。除前文采用的66颗铱星星座外，本文参考OneWeb低轨星座构型[33]，构建由720颗卫星组成的OneWeb大规模低轨卫星网络。该星座包含18个轨道面，每个轨道面40颗卫星，并设置10个地面站节点。与66颗铱星星座相比，OneWeb星座的卫星节点规模扩大约10.9倍，能够更充分反映大规模低轨卫星网络中节点数量增加、候选转发路径增多以及动态拓扑变化加剧对路由算法性能的影响。

为保证实验对比的公平性，OneWeb星座实验沿用前文相同的历史观测窗口长度、业务流类型、攻击扰动方式和性能评价指标，仅改变星座规模与

表3　　**不同历史观测窗口网络性能结果表**

	1	2	4	6	8
平均端到端时延（ms）	184	176	175	170	175
数据包成功送达率（%）	79.2	79.9	80.2	80.8	80.9
吞吐量（包/s）	63.9	66.3	67.3	67.9	67.4

星间链路拓扑结构。不同星座规模下GP-SR算法的网络性能结果如表4所示。

表4 不同星座类型网络性能结果表

星座类型	卫星数（轨道数×每轨卫星数）	平均端到端时延（ms）	成功送达率（%）
铱星星座	66（6×11）	170.2	80.8
OneWeb	720（18×40）	163.3	78.5

由表中结果可见，在OneWeb星座规模为720时，平均端到端时延为163.3ms，数据包成功送达率为78.5%。相较于铱星星座，OneWeb星座的卫星数量更多，星座分布更加密集，卫星之间的平均距离相对缩短。因此，在数据流源节点和目的节点位置保持不变的情况下，数据包在相邻卫星间传播所需的距离有所减小，使得传播时延略有降低。同时，由于源节点到目的节点之间的最短路径跳数基本没有发生明显变化，因此平均端到端时延仅表现为小幅下降，而不是出现大幅改善。

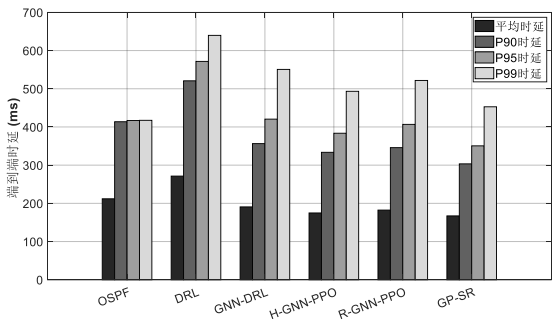
统计结果表明，GP-SR在铱星星座下的平均丢包率为9.6%，而在OneWeb星座下上升至12.4%。在单链路固有丢包率ε固定为0.5%的设定下，该差异主要与端到端路径上的多跳累积丢包效应有关。具体而言，随着星座规模扩大，可选转发节点和链路数量显著增加，路由决策空间更加复杂。GP-SR在路径选择时会综合考虑链路状态、节点负载和安全风险代价，因此在部分情况下会选择规避高风险区域的绕行路径，从而增加平均转发跳数。由于整网平均链路丢包率理论上可近似表示$P_{link_loss} \approx 1-(1-\varepsilon)^h$，平均转发跳数$h$增加会放大链路丢包的累积效应，进而导致OneWeb场景下整网；链路丢包率上升、数据包成功送达率下降。

综上，补充的大规模低轨星座仿真结果表明，所提GP-SR算法在卫星节点数量显著增加的情况下仍能保持较低端到端时延、较高数据包送达率和稳定收敛性能，验证了其在大规模低轨卫星星座场景中的可扩展性与适用性。

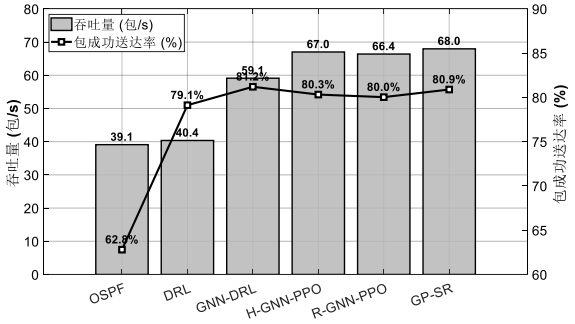
随后，本文选取五种不同的路由算法进行对比实验，来全面评估GP-SR算法的网络性能，分别是：（1）OSPF路由算法[31]：是一种基于链路状态通告的分布式路由协议，它采用洪泛机制在全网范围内同步拓扑信息，使每个路由器能够独立构建以自身为根的最短路径树，在本文的实现中，OSPF算法以最小传播时延作为路径成本度量标准，确保数据包沿网络延迟最低的路径传输；（2）基于DRL的AC框架算法：采用端到端的学习范式，使智能体通过与环境交互试错，直接学习从局部观测到路由决策的映射关系，其以时延、丢包多目标指标的累积奖励为优化目标。（3）基于GNN-DRL路由算法[13]：该论文算法代表了当前基于GNN和DRL的主流智能路由方案。它利用GNN提取当前时刻的网络拓扑特征，并结合DRL进行端到端的路由决策。基线模型仅基于当前时间片的瞬时观测进行状态表征，未引入历史观测序列进行状态增强，且其奖励函数仅为常规QoS指标，缺乏针对网络攻击的主动防御机制与安全风险惩罚项。（4）历史增强GNN-PPO算法（H-GNN-PPO）：该算法在GNN-PPO基础上引入与GP-SR相同的历史观测窗口，将最近 个时刻的局部观测进行拼接与统计增强，以刻画网络状态的短期变化趋势。但其奖励函数仍仅包含时延、送达率和跳数惩罚等常规QoS指标，不引入风险代价项。（5）风险敏感GNN-PPO算法（R-GNN-PPO）：该算法在GNN-PPO基础上引入与GP-SR相同的风险代价项，用于惩罚高丢包链路、高队列利用率节点和短时异常波动状态，但其状态输入仅采用当前时刻局部观测，不使用历史观测窗口。

为评估不同智能路由算法在随机训练条件下的性能稳定性，本文设置5个不同随机种子开展重复实验。需要说明的是，OSPF属于确定性的传统规则驱动路由方法，不涉及随机初始化与训练过程，其性能不受随机种子影响。因此，在多随机种子稳定性分析中，本文主要选取DRL、GNN-DRL和GP-SR三种学习型算法进行对比。

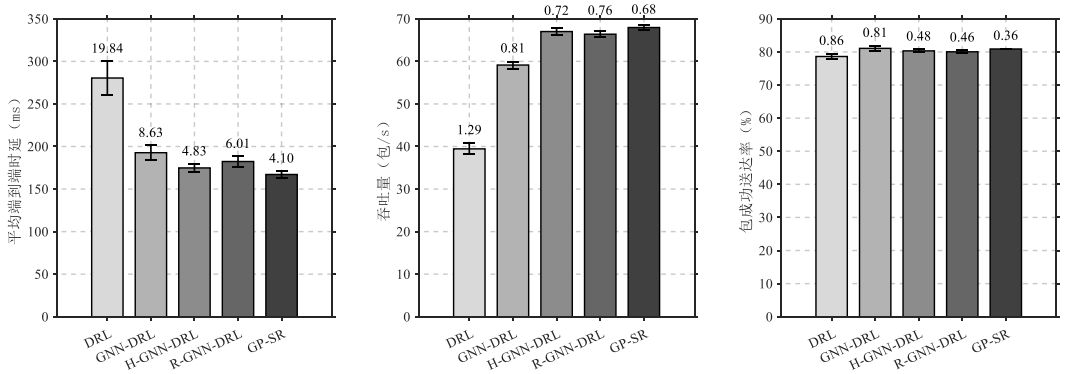
图6(a)给出了六种算法在平均时延以及P90、P95、P99尾部时延上的对比结果。可以看出，GP-SR整体时延性能最优，其平均端到端时延为167.12ms，较OSPF、DRL和GNN-DRL分别降低21.1%、38.4%和12.4%，较H-GNN-DRL和R-GNN-DRL也分别降低4.4%和8.4%。这表明，单独引入历史观测机制或风险代价项均能够改善GNN-DRL的路由性能，而GP-SR同时融合二者后，可以更充分地刻画局部网络状态的短期变化趋势，并提前规避潜在高风险路径，因此获得了更低的端到端时延。



（a） 不同路由算法的平均端到端时延性能对比图



（b） 不同路由算法的吞吐量与包成功送达率对比图



（c） 多随机种子条件下不同算法性能稳定性对比

图6　不同路由算法网络性能对比图

需要特别解释的是，OSPF 在 P99 时延上低于部分学习型算法，并不意味着其整体服务质量更优。本文的时延百分位统计基于成功送达的数据包，因此该指标反映的是成功样本的条件时延分布。结合丢包原因分析可知，OSPF 的队满丢包率达到 30.1%，明显高于其他学习型算法约 9.5%—10.9% 的队满丢包率。这表明 OSPF 由于采用最短路径转发，容易在高负载条件下将流量集中到少数固定路径和关键节点上，导致队列拥塞加剧，大量数据包在队列溢出时被直接丢弃。由于这些被丢弃的数据包并未进入成功送达时延统计，部分原本可能产生极大排队时延的样本被排除在 P99 计算之外，从而使 OSPF 的极端尾部时延表现出一定的成功样本筛选效应。相比之下，学习型算法通过动态路由和负载均衡降低了队满丢包率，能够成功传输更多复杂拥塞场景下的数据包，但这部分数据包可能经历绕行或排队等待，因此会使成功样本中的 P99 时延略有升高。

图 6(b)展示了不同算法的吞吐量和包成功送达率。GP-SR 的吞吐量达到 67.96，明显高于 OSPF、DRL 和 GNN-DRL，较 GNN-DRL 提升约 15.0%，同时也略优于 H-GNN-DRL 和 R-GNN-DRL。在包成功送达率方面，GP-SR 达到 80.89%，与 GNN-DRL 基本持平，并高于 DRL 和 OSPF。该结果说明，GP-SR 在降低时延和提升吞吐量方面具有一定优势，同时仍能保持与 GNN-DRL 相近的包成功送达率，表明其在传输效率和路由可靠性之间取得了较好的平衡。

图 6(c)进一步比较了五种学习型算法在不同随机种子条件下的性能稳定性。结果显示，GP-SR 在平均端到端时延、吞吐量和包成功送达率上的标准差分别为 4.10、0.68 和 0.36，均小于 DRL、GNN-DRL、H-GNN-DRL 和 R-GNN-DRL，说明 GP-SR 在不同训练随机性下具有更稳定的性能表现。这主要得益于历史观测窗口对部分可观测状态的平滑表征，以及风险敏感奖励对策略优化方向的约束，从而有效降低了训练过程中的随机波动。

进一步比较 H-GNN-DRL 与 R-GNN-DRL 可以发现，H-GNN-DRL 在平均时延、尾部时延、吞吐量和包成功送达率方面整体略优于 R-GNN-DRL，说明在本文的部分可观测动态网络环境中，历史观测窗口对路由性能提升的贡献更加直接。其原因在

于，历史窗口能够弥补单时刻局部观测信息不足的问题，帮助智能体捕捉链路状态、队列拥塞和流量负载的短期演化趋势。相比之下，风险奖励主要通过惩罚高丢包链路、高队列利用率节点和异常波动状态来增强策略的安全性与鲁棒性，其作用更偏向于约束策略选择并降低高风险路径被选中的概率。因此，从单独消融结果来看，历史观测机制对整体路由性能提升更为关键，而风险敏感奖励对攻击扰动下的鲁棒性提升具有重要补充作用。GP-SR 同时融合二者，因此取得了最优的综合性能。

3.3　动态威胁抗毁能力分析评估

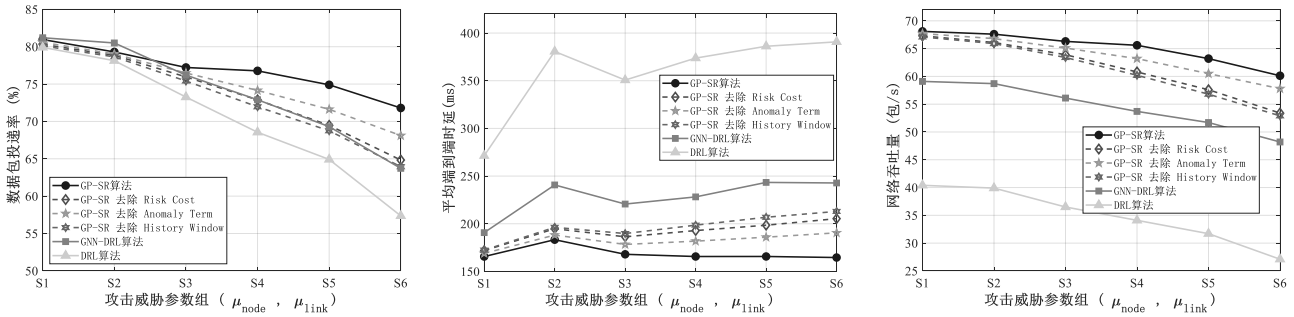
本小节中，我们重点围绕动态威胁抗毁能力分析评估，通过逐步提升 DDoS 攻击与链路干扰强度，系统考察 GP-SR 及其对比算法在高安全威胁场景下的性能退化规律与防御能力。

为了全面评估 GP-SR 算法在动态高安全威胁环境中的生存能力与防御效能，本节设计了一组渐进式压力测试。该实验旨在通过逐步提升 DDoS 攻击强度与链路干扰水平，模拟从基础威胁到极限阻断的全谱系对抗场景，定量考察各路由算法在不同威胁等级下的性能衰减趋势。DDoS 攻击参数 μ_{node} 由 0.5 逐步增加至 1.0，链路干扰参数 μ_{link} 由 0.2 逐步增加至 0.95，设置六组安全威胁参数，第一组：μ_{node}=0.5，μ_{link}=0.2；第二组：μ_{node}=0.6，μ_{link}=0.35，以此类推。本节除比较 GP-SR 与学习型基线算法外，还进一步引入消融版本，用于分析历史增强观测和风险代价建模等关键模块在高威胁场景中的作用。需要说明的是，本实验未纳入传统 OSPF 协议，这是因为其基于静态链路开销和周期性状态更新，难以与动态攻击环境下的学习型在线决策方法形成有效对比。

如图 7(a) 所示，在低威胁场景下，三种算法性能接近。然而，在 S4 至 S6 强干扰场景下，GNN-DRL 与 DRL 算法因缺乏显式安全惩罚，包成功送达率分别迅速下滑至 63.72% 和 57.36%。相比之下，GP-SR 在 S6 场景下仍维持了 71.8% 的投递率。这验证了奖励函数中风险惩罚项的有效性，引导智能体规避了受攻击的高丢包区域。

图 7(b) 显示，三种算法在高攻击威胁场景中，GP-SR 算法受攻击威胁参数影响最小，有一个特殊的地方是：在 S2 情况下，三种算法的平均端到端时延出现激增，但在后来逐渐平缓。为解释这一现象，本文进一步统计了 S1 – S3 阶段各算法的缓冲区溢出丢包率。结合攻击模型可知，DDoS 攻击主要降低被攻击节点的数据处理速率和队列容量，星间链路干扰攻击主要降低链路带宽、增加传播时延并提高丢包率。在 S2 阶段，GP-SR、GNN-DRL 和 DRL 的队满丢包率分别为 8.9%、9.0% 和 9.5%，相较 S1 阶段并未明显升高甚至稍有下降；三个消融模型 w/o Risk Cost、w/o Anomaly Term 和 w/o History Window 的队满丢包率也分别为 8.9%、8.9% 和 9.0%，同样未出现明显上升。说明链路干扰导致的带宽下降和传播时延增加较为明显，且网络尚未进入大规模缓冲区溢出状态，较多数据包仍会在缓冲区中等待转发，这两者共同导致成功送达样本的平均端到端时延升高。

当攻击强度进一步增加至 S3 时，DDoS 攻击对节点处理速率和队列容量的压制作用增强，缓冲区溢出开始加剧。GP-SR、GNN-DRL 和 DRL 的队满丢包率分别上升至 10.6%、11.7% 和 13.3%；三个消融模型的队满丢包率也分别上升至 10.9%、11.1% 和 11.2%。这说明部分原本可能产生较长排



（a）　数据包成功送达率对比　　（b）　平均端到端时延对比　　（c）　整网吞吐量对比

图 7　不同攻击威胁等级下的算法性能对比

队时延的数据包因缓冲区溢出被提前丢弃，未进入成功送达时延统计，因此平均时延出现回落。

图7(c) 所示，GP-SR的吞吐量在全过程中仅损失约 11.7%，而基础DRL算法损失达 32.9%，GNN-DRL 损失18.4%。这表明在部分链路被攻击流量挤占的情况下，GP-SR 凭借GNN的拓扑表征能力，能更精准地感知剩余可用链路，实现了对抗环境下有效吞吐量的最大化。

消融实验表明，GP-SR的性能提升来源于多个关键模块的协同作用，而非单一结构带来的偶然收益。首先，去除历史窗口后，模型在三项指标上均出现最明显退化，尤其在高强度场景下，成功投递率由71.80%降至63.94%，时延由164.58ms上升至212.94ms，吞吐量由60.12下降至52.90，说明历史信息对于部分可观测环境下的状态补全与趋势判断至关重要。其次，去除风险代价项后，模型在中后段场景中退化同样显著，末组场景PDR仅为64.83%，吞吐量下降至53.40，表明显式风险约束有效提升了策略对脆弱路径和潜在受损节点的规避能力。相比之下，去除异常项后的性能下降相对缓和，但仍持续劣于完整模型，说明异常状态表征能够为策略提供额外的判别依据，增强其对局部异常和链路退化的敏感性。

实验证明，GP-SR的性能优势源于其“感知-防御”协同机制：GNN负责全局态势感知，POMDP历史序列负责状态推断，而安全奖励函数则赋予了策略主动防御的本能。综上，GP-SR在复杂对抗场景下展现出了极强的鲁棒性与内生安全潜力。

3.4　算法复杂度分析评估

在最后，我们来分析GP-SR路由算法模型的时间复杂度，本算法模型由GNN模块和深度强化学习决策模块构成。GNN模块由三层GAT卷积层、全局池化层和两层全连接层组成，对于GAT卷积层的第l层，节点需要进行特征线性映射，边需要计算注意力权重并聚合，单层复杂度为$\mathcal{O}(N_L\cdot d_l\cdot d_{l+1}\cdot H_l + E_L\cdot d_{l+1}\cdot H_l)$，$N_L$是局部子图的节点数，$d_l$是第$l$层输入维度，$H_l$是第$l$层注意力头数；全局池化复杂度和全连接层复杂度分别为$\mathcal{O}(N_L\cdot d_{out})$和$\mathcal{O}(d_{out}\cdot d_h + d_h\cdot d_{out})$。深度强化学习模块包括状态处理网络和AC网络，其本质都是多层感知机（Multi-Layer Perceptron，MLP），总复杂度为$\mathcal{O}(S\cdot D_{mlp} + D_{mlp}^2 + (D_{mlp} + d_{out})\cdot D_{mlp} + D_{mlp}\cdot A)$，$S$为强化学习状态向量维度，$D_{mlp}$为隐藏层维度，$A$为动作空间维度。在模型使用阶段，时间复杂度公式如下所示

$$\begin{aligned} C_{GNN} &= \mathcal{O}\Big(\sum_{l=1}^{L_G}(N_L\cdot d_l\cdot d_{l+1}\cdot H_l + E_L\cdot d_{l+1}\cdot H_l) + d_{out}\cdot d_h\Big) \\ C_{MLP} &= \mathcal{O}(S\cdot D_{mlp} + D_{mlp}^2 + (D_{mlp} + d_{out})\cdot D_{mlp} + D_{mlp}\cdot A) \\ T_{inference} &= \mathcal{O}(C_{GNN} + C_{MLP}) \approx \mathcal{O}(N_L d_h^2 H + E_L d_h H + D_{mlp}^2) \end{aligned} \quad \#(10)$$

基于表2设定的网络参数进行量化评估，本模型单次智能路由决策的计算开销约为1.1 MFLOPs。若将其部署于具备256 GFLOPS峰值算力的抗辐射星载计算平台上，理论推理耗时仅为0.0043 ms。相较于毫秒级的星际链路传播延迟，该计算开销完全可以忽略不计[32]，充分验证了本算法在实际工程中的部署可行性。并且从式（11）来看，本路由算法的时间复杂度与星座规模无关，取决于局部子图节点数N_L和边数E_L，完全独立于全网总结点数。因此，若将此算法推广至上万颗低轨卫星的巨型星座在理论上成为可能。

模型训练阶段，PPO更新需基于当前轮轨迹缓存计算GAE优势函数，并在多轮策略迭代中执行前向传播和反向传播，其计算量显著高于推理阶段。因此，训练阶段时间复杂度可表示为：

$$\begin{aligned} T_{train} &= \mathcal{O}(B\cdot T_{inference}^{'} + K\cdot B\cdot (T_{inference}^{'} + T_{backward})) \\ T_{inference}^{'} &= T_{inference} + T_{Critic} \end{aligned} \quad \#(11)$$

本算法中批次大小为B为256，循环K为10，依据深度学习算力评估的通用准则，单次反向传播的浮点运算量约为前向传播的两倍，即$T_{backward} \approx 2T_{inference}^{'}$，单次计算开销约为13.33GFLOPs，在256 GFLOPS峰值算力的抗辐射星载计算平台上的理论耗时为52.07ms，大于星间链路时延（5-10ms）[30]，在星上训练不可行。需要采用“地面离线训练、星上在线推理”的部署范式，即依托高保真网络仿真系统构建大规模交互环境，并融合真实星间链路测量数据与业务流统计特征，对模型进

行联合训练；待策略收敛后，仅将训练完成的轻量化路由模型部署至星上执行在线决策。。

尽管 GP-SR 在关键性能指标上表现优异，但其训练过程仍存在较大的计算开销。较高的训练时间成本与算力需求，使其在资源受限场景下面临一定部署挑战，来将围绕模型轻量化与训练效率提升进一步开展研究。

4 结束语

本文聚焦于低轨卫星网络在拓扑动态时变与安全威胁环境下的路由决策难题，提出了一种 GNN-PPO 与 POMDP 的智能安全路由算法 GP-SR。首先，构建了融合时空特征的深度强化学习架构，利用 GNN 聚合全网拓扑的空间结构信息，并通过 POMDP 引入历史观测序列来推断网络流量与攻击的演化趋势，有效克服了传统 DRL 算法在非完全信息下的决策盲区。其次，建立了包含时延、丢包率及安全风险惩罚项的多目标优化模型，利用 PPO 策略梯度实现端到端的路径寻优。仿真实验与渐进式压力测试结果表明，GP-SR 算法不仅在常规场景下降低了传输时延与丢包率，提升了约 15.2% 的网络吞吐量；更在高强度 DDoS 攻击与链路干扰的极限条件下，凭借“感知-防御”协同机制展现出良好的鲁棒性。

参考文献：

[1] ZHU Xiangming and JIANG Chunxiao. Integrated satellite- terrestrial networks toward 6G: Architectures, applications, and challenges[J]. IEEE Internet of Things Journal, 2022, 9(1): 437–461.

[2] 李学华,廖海龙,张贤,等.面向低轨卫星通信网络的联邦深度强化学习智能路由方法[J].电子与信息学报,2025,47(08):2652-2664.

[3] Tang H , Zhang Q , Li Y ,et al.Regional Resilient Routing Algorithm for LEO Satellite Network[J].2024 22nd International Conference on Optical Communications and Networks (ICOCN), 2024:1-3.

[4] Tu J , Dang J , Wu K ,et al.Segment Routing Algorithm for Mega LEO Constellations[J].2024 4th International Conference on Electronic Information Engineering and Computer (EIECT), 2024:892-896.

[5] Hsu Y H, Lee J I, Xu F M. A deep reinforcement learning based routing scheme for LEO satellite networks in 6G[C]//2023 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2023: 1-6.

[6] Niu Z, Xie Y, Li Z, et al. LEO Satellite Routing Method based on Dynamic A-Star Algorithm[C]//2024 International Conference on Artificial Intelligence and Power Systems (AIPS). IEEE, 2024: 233-237.

[7] Liu W, Tao Y, Liu L. Load-balancing routing algorithm based on segment routing for traffic return in LEO satellite networks[J]. IEEE Access, 2019, 7: 112044-112053.

[8] Lu Z, Zhi R, Ma W. Quick routing response to link failure in low-earth orbit satellite networks[C]//2022 IEEE 8th International Conference on Computer and Communications (ICCC). IEEE, 2022: 690-695.

[9] Wang K, Miao X, Liu P, et al. Traffic-Load-Aware Multipath Routing in LEO Satellite Networks[C]//2024 4th International Conference on Intelligent Communications and Computing (ICICC). IEEE, 2024: 135-138.

[10] He Y, Chen H, Ma C. A cross-domain aggregation routing based on lightweight OSPF protocol for the GEO satellite-ground integrated network[C]//2022 2nd International Conference on Computer Science, Electronic Information Engineering and Intelligent Control Technology (CEI). IEEE, 2022: 459-463.

[11] Park S, Kim G S, Jung S, et al. Markov decision policies for distributed angular routing in leo mobile satellite constellation networks[J]. IEEE Internet of Things Journal, 2024.

[12] Lyu Y, Hu H, Fan R, et al. Dynamic routing for integrated satellite-terrestrial networks: A constrained multi-agent reinforcement learning approach[J]. IEEE Journal on Selected Areas in Communications, 2024, 42(5): 1204-1218.

[13] Zhang S, Liu A, Han C, et al. GRLR: Routing with graph neural network and reinforcement learning for mega LEO satellite constellations [J]. IEEE Transactions on Vehicular Technology, 2024, 74(2): 3225-3237.

[14] Xu P, Feng M, Zhou J, et al. Inter-satellite routing for LEO satellite networks: A GNN and DRL integrated approach[C]//2024 IEEE/CIC International Conference on Communications in China (ICCC). IEEE, 2024: 1346-1351.

[15] Li S, Wu Q, Wang R, et al. Efficient Multipath Differential Routing and Traffic Scheduling in Ultra-Dense LEO Satellite Networks: A DRL With Stackelberg Game Approach[J]. IEEE Transactions on Mobile Computing, 2025.

[16] Wang L, Xu Z, Zhi R, et al. Adaptive Load Balancing Routing Algorithm for Low Earth Orbit Satellite Cluster Networks[C]//2024 9th International Conference on Computer and Communication Systems (ICCCS). IEEE, 2024: 666-671.

[17] Yan H, Zhang Q, Sun Y. A novel routing scheme for LEO satellite networks based on link state routing[C]//2014 IEEE 17th International Conference on Computational Science and Engineering. IEEE, 2014: 876-880.

[18] 王慧姊,孙雷,王健全,等.基于网络演算的LEO卫星网络时延上界分析研究[J].通信学报,2025,46(04):80-90.

[19] 杨世松.低轨卫星网络路由算法研究[D].天津理工大学,2025.

[20] Tao J, Na Z, Zhang N. Time-varying graph model for LEO satellite network routing[C]//2022 9th International Conference on Dependable Systems and Their Applications (DSA). IEEE, 2022: 486-491.

[21] 杨明川,薛冠昌,李清毅.基于邻居卫星负载状态的低轨卫星分布式路由算法[J].通信学报,2021,42(08):43-51.

[22] 汪昊,冉泳屹,赵雷,等.基于深度图强化学习的低轨卫星网络动态路由算法[J].重庆邮电大学学报(自然科学版),2023,35(04):596-605.

[23] Zuo P, Wang C, Yao Z, et al. An intelligent routing algorithm for LEO satellites based on deep reinforcement learning[C]//2021 IEEE 94th Vehicular Technology Conference (VTC2021-Fall). IEEE, 2021: 1-5.

[24] 史琰,尹斌,白卫岗,等.卫星网络数字域仿真平台构建半实物系统方法与实现[J/OL].计算机工程与应用,1-18[2026-04-08].

[25] Paxson V. End-to-end Internet packet dynamics[C]//Proceedings of the ACM SIGCOMM'97 conference on Applications, technologies, architectures, and protocols for computer communication. 1997: 139-152.

[26] Liming H, Shaoli K, Shaohui S, et al. A load balancing routing method based on real time traffic in LEO satellite constellation space networks [C]//2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring). IEEE, 2022: 1-5.

[27] Li Q, Zhou Q, Sun H, et al. Research on DDoS attack technology for satellite communication network[C]//2025 IEEE 2nd International Conference on Electronics, Communications and Intelligent Science (ECIS). IEEE, 2025: 1-5.

[28] Lu T, Ding X, Shang J, et al. DoSat: a DDoS attack on the vulnerable time-varying topology of LEO satellite networks[C]//International Conference on Applied Cryptography and Network Security. Cham: Springer Nature Switzerland, 2024: 265-282.

[29] Giuliari G, Ciussani T, Perrig A, et al. {ICARUS}: Attacking low earth orbit satellite networks[C]//2021 USENIX Annual Technical Conference (USENIX ATC 21). 2021: 317-331.

[30] Du X K, Shu N N, Liu C S, et al. Overview of security issues and defense technologies for Low Earth Orbit satellite network[J]. 电子与信息学报, 2025, 47(6): 1609-1622.

[31] Lv Y, Xing C, Xu N, et al. Research of adaptive routing scheme for LEO network[C]//2019 IEEE 5th International Conference on Computer and Communications (ICCC). IEEE, 2019: 987-992.

[32] Xiang J, He X, Zhao Y, et al. Distributed Dynamic Routing for LEO Satellite Networks with Temporal Graph Convolutions and Imitation Acceleration[J]. IEEE Communications Letters, 2025.

[33] Naz A, Verma K, Sikka G. GNN-TASR: Graph Neural Network based Trust Aware Secure Routing for LEO satellite network[J]. Ad Hoc Networks, 2026: 104223.

[作者简介]

尹斌（2000– ），男，湖南株洲人，国防科技大学电子科学学院、密码研究中心博士生，主要研究方向为空地网络安全、网络人工智能等。

刘建（1986– ），男，山东泰安人，博士，国防科技大学电子科学学院、密码研究中心副教授，主要研究方向为天基信息网络安全、密码安全智能化。

唐小妹（1982– ），女，江苏海安人，博士，国防科技大学电子科学学院、密码研究中心研究员，主要研究方向为安全信号设计、导通融合体系设计。

马春江（1991– ），男，湖南邵阳人，博士，国防科技大学电子科学学院、密码研究中心讲师，主要研究方向为卫星互联网通导融合、空天网络信号安全。

马鹏程（1992– ），男，山东日照人，博士，国防科技大学电子科学学院助理研究员，主要研究方向为卫星导航欺骗检测、侧信道分析等。