

基于环境可靠性预测的低轨卫星网络弹性智能路由技术

陈 啸¹, 纪 哲¹, 吴 胜¹, 姬思敬², 盛 敏²,

(1.北京邮电大学泛网无线通信教育部重点实验室, 北京 100876;

2.西安电子科技大学空天地一体化综合业务网全国重点实验室, 陕西 西安 710071)

摘要: 低轨卫星网络在复杂空间环境中面临严峻的可靠性挑战, 而传统路由协议多采用被动式响应机制, 难以应对空间环境效应导致的链路不稳定风险。为此, 提出一种基于环境可靠性预测的低轨卫星网络弹性智能路由方法。首先设计融合长期预测性辐射风险、瞬态空间天气事件与实时链路通信质量的空间环境可靠性指数, 并给出端到端路由问题模型; 其次构建基于消息传递机制的星间邻域状态信息聚合模型, 提出环境状态预测下基于多智能体协同的分布式路由决策算法, 实现低轨卫星网络路由决策从“被动响应”向“主动规避”的转变。仿真结果显示, 相较于基准算法, 该方法能主动规避高风险区域, 有效提升动态空间环境下的低轨卫星网络端到端路由可靠性。

关键词: 低轨卫星网络; 空间环境效应; 可靠性预测; 端到端路由; 多智能体强化学习

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.2096-8930.2026005

Resilient Intelligent Routing Technology for LEO Satellite Networks Based on Environmental Reliability Prediction

Chen Xiao¹, Ji Zhe¹, Wu Sheng¹, Ji Sijing², Sheng Min²

1. The Key Laboratory of Universal Wireless Communications Ministry of Education, Beijing University of Posts and Telecommunications, Beijing 100876, China

2. The State Key Laboratory of Integrated Service Networks, Xidian University, Xi'an 710071, China

Abstract: Low earth orbit satellite networks face severe reliability challenges in the complex space environment, where conventional routing protocols, with their reactive mechanisms, struggle to address the link instability risks induced by space environment effects. To this end, this paper proposes a resilient and intelligent routing method for low earth orbit satellite networks, driven by environmental reliability prediction. The method begins by designing a space environment reliability index that integrates predictable radiation risks, space weather disturbances, and real-time link communication quality, and subsequently formulates the end-to-end routing problem. Next, a message-passing-based model is constructed to aggregate neighborhood state information among satellites. This enables a collaborative multi-agent distributed routing algorithm, guided by environmental state predictions, to facilitate a paradigm shift in routing decisions from reactive response to proactive avoidance. Simulation results demonstrate that, compared to baseline algorithms, the proposed method can proactively avoid high-risk regions, significantly enhancing the end-to-end routing reliability of low earth orbit satellite networks in dynamic space environments.

Keywords: low earth orbit satellite network, space environment effect, reliability prediction, end-to-end routing, multi-agent reinforcement learning

收稿日期: 2025-08-20; 修回日期: 2025-10-12

基金项目: 国家自然科学基金 (No.62495020, No.62495024, No.62201085)

Foundation Items: The National Natural Science Foundation of China (No.62495020, No.62495024, No.62201085)

0 引言

低地球轨道 (low earth orbit, LEO) 卫星网络凭借其广域无缝覆盖与低时延通信的独特优势, 正以前所未有的规模与速度迅速发展, 并逐渐成为下一代全球信息基础设施的核心组成部分^[1]。然而, 要实现这一宏伟蓝图, 必须克服空间固有恶劣环境所带来的严峻挑战。与享有良好物理防护的地面网络不同, 在轨卫星持续暴露于复杂的空间环境中, 时刻面临着地球辐射带高能粒子、太阳耀斑及日冕物质抛射等多种辐射与电磁扰动的复合威胁。这些高能事件能够直接损伤星上电子元器件, 进而引发链路中断、节点失效等一系列随机性、非对称性故障, 对网络端到端的服务可靠性构成了根本性挑战^[2]。现有的卫星网络路由协议, 在应对卫星高速运动所引发的周期性、确定性拓扑演变方面虽已相当成熟, 但其设计并未充分顾及上述源于空间物理效应的随机性故障。这种设计上的局限性使得传统协议在空间环境效应影响下难以保证端到端传输服务的可靠性^[3]。因此, 亟须探索一种能够智能适应动态空间环境的可靠路由方法。

针对上述挑战, 现有研究主要沿循“建模-优化”的技术路线。该类方法首先借助概率论、图论等数学工具, 对空间环境导致的链路或节点故障进行抽象建模, 进而将问题转化为特定场景下的容错路由优化问题, 并利用优化理论寻求其最优解。例如, Yan等^[4]设计了一种分层路由算法, 通过引入逻辑路径抽象机制, 将物理拓扑的频繁变化与上层路由决策解耦, 从而有效应对链路发生偶然中断导致的拓扑剧烈变动, 提升了卫星星座网络的数据转发可靠性。Zheng等^[5]则聚焦于空间碎片威胁, 提出了一种风险感知的快照路由策略, 该策略通过对链路风险进行量化并动态调整其路由权重, 实现了对高风险区域的主动规避, 增强了路径的稳定性与容错能力。同样, Zhu等^[6]通过引入业务优先级评估与链路故障概率模型, 设计了一种能够保障差异化服务质量的路由算法, 在多链路并发失效的场景下, 优先确保高等级业务的传输质量。然而, 上述研究虽在特定场景下表现优越, 却共享一个根本性的理论局限: 它们普遍将链路故障建模为低概率的独立随机事件, 而忽略了空间环境效应所引发的、具有强时空相关性的区域性、并发式故障特征。因此, 当网络遭遇此类大范围、突发性的扰动时, 这类依赖于先验模型和前瞻性规划的路由策略往往难以奏效。究其根源, 是这些方法缺乏对未知扰动的在轨实时感知与快速响应能力, 一旦链路状态发生未曾预见的急剧恶

化, 便只能采用被动的、高开销的重路由机制, 难以保障任务的实时性与可靠性。

近年来, 为克服传统“建模-优化”范式的局限性, 以深度强化学习 (deep reinforcement learning, DRL) 为代表的人工智能技术为卫星路由决策开辟了新路径。此类方法将网络环境视为一个“黑箱”, 通过赋予卫星智能体自主决策能力, 以试错的方式在与环境的交互中学习路由策略。其核心机制通常表现为一种被动适应模式: 当链路故障实际发生, 并已造成数据丢包或时延激增等性能恶化后, 智能体才经由负向奖励信号来调整其后续行为。例如, Wei等^[7]面向软件定义卫星网络中链路动态变化与拓扑频繁重构带来的挑战, 提出了一种基于深度强化学习的智能可靠路由框架, 该方法通过将路由决策建模为马尔可夫决策过程, 并引入深度Q网络学习动态环境中的最优路由策略, 实现对链路状态、带宽变化和拓扑断裂的感知与自适应调整。Niu等^[8]针对卫星光网络中高动态拓扑和严苛时延需求的问题, 提出了一种基于多智能体强化学习 (multi-agent reinforcement learning, MARL) 的可靠低时延路由策略, 该方法通过为每颗卫星部署独立智能体, 实现分布式学习与局部感知, 提升了网络在面对链路波动和节点故障时的适应性与容错性。同时, 策略中引入奖励函数设计以联合优化端到端时延与路由稳定性, 使系统能够在满足服务质量需求的同时, 实现高效可靠的数据转发。这种被动响应式的学习机制, 虽然长期来看能够收敛至统计意义上的最优路径, 但其牺牲了网络的瞬时性能与任务的即时可靠性。更深层次的问题在于, 这些纯粹由数据驱动的方法在学习过程中忽略了故障背后可预测的物理成因, 未能将卫星轨道位置、途经辐射区域等具有强先验指示性的物理信息融入决策环路, 无法主动规避链路中断风险。

为此, 本文聚焦低轨星座网络在复杂空间环境效应下的可靠路由问题, 提出了一种基于空间环境可靠性预测的弹性智能路由 (environmental-reliability-prediction-driven resilient intelligent routing, ERIR) 方法, 该方法采用集中式训练分布式执行的MARL架构, 将每颗卫星抽象为具备自主决策能力的分布式智能体。为突破分布式执行所固有的局部观测局限性, 设计了一种基于消息传递机制的邻域信息聚合机制, 以实现高效的状态共享与协同决策。其核心创新在于, 该机制在考虑传统网络状态指标 (如时延、负载) 的同时, 将构建的空间环境链路可靠性指数融入信息聚合过程。这种数据驱动与物理模型的深度融合, 给予了网络局部状态感知与前瞻性风

险预测能力, 驱动了路由决策从“事后响应”向“事前规避”的转变, 使得网络能在链路质量恶化前主动绕开高风险区域, 从而提升了动态空间环境下的服务韧性与可靠性。

1 系统模型

1.1 卫星网络模型

在低轨卫星通信系统中, 卫星的周期性高速运动导致其星间链路不断变化, 从而使得网络拓扑具有显著的时变性与周期性特征。为了减小链路动态变化对路由设计的影响, 本文采用时间切片技术将整个卫星周期划分为一系列相等长度的离散时隙 $\mathcal{T} = \{t_1, t_2, \dots, t_T\}$, 在每个时隙内, 网络拓扑被视为静态的。基于此, 构建每一时隙下的静态拓扑图 $\mathcal{G}_s^t = (\mathcal{V}_s^t, \mathcal{E}_s^t)$, 其中 \mathcal{V}_s^t 表示时隙 t 中可用的卫星节点集合, \mathcal{E}_s^t 表示该时隙内存在的星间链路集合。由于卫星轨道具有周期性, 整个周期内的动态网络拓扑可由一系列静态拓扑图序列组成, 即 $\mathcal{G}_s = \{\mathcal{G}_s^1, \mathcal{G}_s^2, \dots, \mathcal{G}_s^T\}$, 反映了卫星网络随时间演化的结构特性。在该建模框架下, 卫星网络的动态拓扑被形式化为具有周期性演化特征的虚拟拓扑序列, 如图 1 所示, 为后续的路由算法设计与性能分析提供了结构化基础。

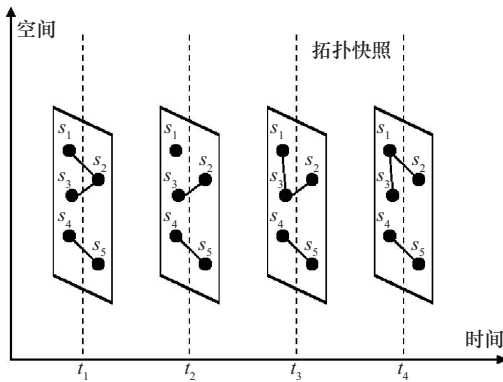


图 1 卫星网络虚拟拓扑序列

1.2 链路可靠性模型

在动态的低轨卫星网络中, 链路可靠性是保障路由稳定与系统鲁棒性的基石。为精确量化星间链路的可靠性, 本文提出空间环境链路可靠性指数。与传统模型普遍将链路故障视为独立的随机事件不同, 本模型的核心思想在于刻画空间环境效应所特有的区域性与并发性特征, 即单一事件可能同时影响多颗卫星与多条链路。为此, 该模型将影响链路可靠性的多源因素依据其时间尺度和物理成因解耦为 3 个独立的、可量化的分量: 长期可

预测的辐射风险、偶发性的空间天气事件以及实时的通信链路质量。链路 e_{ij} 在时刻 t 的整体可靠性 $U_{ij}(t) \in [0, 1]$ 被定义为这 3 个分量的乘积, 即:

$$U_{ij}(t) = f_p(p_i(t), p_j(t)) \cdot f_s(A(t)) \cdot f_q(Q_{ij}(t)) \quad (1)$$

其中, $f_p(\cdot)$ 为长期预测性辐射风险分量, 用于量化由卫星轨道位置决定的确定性风险; $f_s(\cdot)$ 为瞬态空间天气事件分量, 表征空间天气事件造成的全局性影响; $f_q(\cdot)$ 则为实时链路通信质量分量, 反映了当前通信信道的即时状态。

(1) 长期预测性辐射风险分量

长期预测性辐射风险分量 $f_p(\cdot)$, 旨在量化由卫星轨道动力学决定的、具有确定性和可预测性的链路风险, 该风险主要源于卫星周期性穿越地球辐射带及南大西洋异常区等高能粒子环境。在上述区域, 骤增的辐射通量会导致通信载荷的瞬时性能退化乃至累计损伤, 辐射通量是影响链路稳定性的一个关键物理因素^[9]。本文的建模方法首先基于公开的经验模型 (如 AP8/AE8) 构建一个静态的三维空间辐射通量场 $R(p)$, 其中 p 代表空间中的任意位置点。随后, 结合精确的星历数据, 我们可以获取任意时刻 t 链路两端卫星 i 和 j 的瞬时三维坐标 $p_i(t)$ 与 $p_j(t)$ 。假设星间链路路径为一空间直线段, 则该链路在时刻 t 瞬时承受的综合辐射通量 $D_{ij}(t)$ 可通过对辐射场 $R(p)$ 沿该路径进行线积分来量化。为将该物理量转化为一个标准化的可靠性指标, 首先对 $D_{ij}(t)$ 进行最大值归一化处理, 将其映射至区间 $[0, 1]$, 即:

$$f_p(p_i(t), p_j(t)) = 1 - \frac{D_{ij}(t)}{D_{\max}} = 1 - \frac{\int_{p_i(t)}^{p_j(t)} R(p) dl}{D_{\max}} \quad (2)$$

其中, D_{\max} 是在整个星座系统的所有可能链路路径上计算出的最大辐射通量积分值, 作为一个系统常量, 该分量使路由决策具备了先验的物理风险感知能力, 能够前瞻性地识别并规避由轨道位置决定的高风险链路。这种建模方式能够反映出风险的时空相关性, 当星座中位于同一区域的多颗卫星协同飞越辐射带等高风险区域时, 其各自链路的可靠性会呈现出同步下降的趋势, 从而能有效刻画由地理位置决定的区域性并发风险。

(2) 瞬态空间天气事件分量

瞬态空间天气事件分量用于反映太阳耀斑、日冕物质抛射等突发性空间天气事件对整个卫星星座或特定区域造成的扰动。依据空间天气预警等级 $A(t)$, 我们将 $f_s(\cdot)$ 建模为一个离散的全局衰减因子, 即:

$$f_s(A(t)) = \begin{cases} 1.0, & \text{正常条件} \\ 0.8, & \text{中等强度太阳耀斑} \\ 0.5, & \text{强太阳耀斑} \end{cases} \quad (3)$$

(3) 实时链路通信质量分量

实时链路通信质量分量从物理层通信质量的角度评估链路的瞬时可靠性，其核心指标为激光链路的信噪比 (signal-to-noise ratio, SNR)。SNR 的波动主要源于自由空间路径损耗和平台微振动引起的指向误差，是决定数据传输错误率乃至链路中断的根本因素^[10]。当链路 e_{ij} 的瞬时 SNR 低于预设的接收阈值 μ 时，通信发生中断。中断概率 $p_{ij}(t)$ 定义为：

$$p_{ij}(t) = P_r\{Q_{ij}(t) < \mu\} = P_r\left\{|h_{ij}(t)|^2 < \frac{\mu N_0 L_{ij}(t)}{P_t}\right\} \quad (4)$$

其中， P_t 为激光发射功率， N_0 为信道中的加性高斯白噪声功率谱密度， $L_{ij}(t)$ 为星间距离， $h_{ij}(t)$ 为链路损耗因子。中断概率可通过对 $|h_{ij}(t)|^2$ 的概率密度函数 $f_{|h_{ij}(t)|^2}(\cdot)$ 在中断域内积分求得，即：

$$p_{ij}(t) = \int_0^{\frac{\mu N_0 L_{ij}(t)}{P_t}} f_{|h_{ij}(t)|^2}(x) dx \quad (5)$$

实时链路通信质量分量 $f_q(\cdot)$ 定义为链路不发生中断的概率，即成功传输概率，表示为：

$$f_q(Q_{ij}(t)) = 1 - p_{ij}(t) \quad (6)$$

1.3 时延模型

在低轨卫星路由中，时延是一项重要的优化指标。数据包在卫星网络中的路由需要经过多跳转发，该过程通常如下：首先，源节点生成待发送的数据流，并根据路由协议将其分割为数据包；然后，源节点通过路由算法计算下一跳节点，并通过星间链路传输数据包；在逐跳传输过程中，由于卫星计算资源有限，需要将未处理的数据包缓存至队列中等待处理；最后，经过不断转发，数据包到达目的节点。在上述过程中应考虑星间链路的传播时延、缓存队列的排队时延、数据包发送的传输时延。下面分别对各个时延部分进行建模。

(1) 传播时延

传播时延 $D_{ij}^{\text{prop}}(t)$ 由信号在自由空间中跨越星间距离 $L_{ij}(t)$ 所需的时间决定，其大小与星间距离成正比，即：

$$D_{ij}^{\text{prop}}(t) = \frac{L_{ij}(t)}{c} \quad (7)$$

其中， c 为光速。由于卫星在轨道上高速运动，星间距离 $L_{ij}(t)$ 是时变的，因此传播时延也是动态变化的。

(2) 排队时延

卫星节点有限的计算与缓存资源导致数据包在流量高峰期需进入队列等待处理，从而引入排队时延。将每个节点的缓存处理过程建模为一个 $M/M/1/N$ 排队系统^[11]，该模型假设数据包的到达过程服从泊松分布（平均到达率为 λ ），转发服务时间服从指数分布（平均服务率为 μ ），且系统最大容量为 N （包括正在服务的和在队列中等待的）。基于此模型，卫星节点 i 缓存队列中的平均数据包数量 $Q_i(t)$ 为：

$$Q_i(t) = \frac{\rho(1 - (N+1)\rho^N + N\rho^{N+1})}{(1-\rho)(1-\rho^{N+1})} \quad (8)$$

其中， $\rho = \lambda/\mu$ 为业务强度。由 Little 公式得到数据包在卫星 i 的缓存队列中的平均排队时延 $D_i^{\text{queue}}(t)$ 可表示为：

$$D_i^{\text{queue}}(t) = \frac{Q_i(t)}{\lambda_e} \quad (9)$$

其中， $\lambda_e = \lambda(1 - P_B)$ 为数据包的有效到达率， $P_B = \frac{(1-\rho)\rho^N}{1-\rho^{N+1}}$ 为系统的阻塞概率。

(3) 传输时延

传输时延 $D_{ij}^{\text{trans}}(t)$ 是将一个数据包从发送节点 i 的端口完全发送至信道所需的时间，其取决于数据包大小 S_m 和星间链路的传输速率 R_{ij} ，传输时延表示为：

$$D_{ij}^{\text{trans}}(t) = \frac{S_m}{R_{ij}} \quad (10)$$

为简化模型，本文假设所有激活的星间链路 (inter-satellite link, ISL) 具有统一的传输速率。因此，数据包 m 在链路 (ij) 上的单跳总时延 $D_{ij,m}(t)$ 是上述 3 部分之和，即：

$$D_{ij,m}(t) = D_{ij}^{\text{prop}}(t) + D_i^{\text{queue}}(t) + D_{ij}^{\text{trans}}(t) \quad (11)$$

在卫星网络中传输数据包时，由于源节点和目的节点之间通常不存在之间相连接的星间链路，因此数据包通常需要通过中继节点进行多跳传输。假设数据包 m 从源卫星到目的卫星的路径表示为 $P_{s \rightarrow d} = [s, \dots, i, \dots, j, \dots, d]$ ，我们需要为每个数据包 $m \in M$ 寻找一条路由路径以最小化一个复合成本函数，该函数是链路可靠性指数与端到端时延加权组合。因此可以将优化的问题表述为：

$$\begin{aligned} \text{P1: } \min & \left(-\sum_{i,j \in P_{s \rightarrow d}} \left(\alpha \frac{\sum_{m \in M} U_{ij,m}(t)}{|M| \cdot U_{\max}} - (1-\alpha) \frac{\sum_{m \in M} D_m(t)}{|M| \cdot D_{\max}} \right) \right) \\ & \text{s.t. } D_m(t) \leq D_{\max}, U_{ij,m}(t) \leq U_{\max}, \\ & \sum_{m \in M} y_{m,i}^{\text{queue}}(t) + \sum_{g \in Z_i(t)} x_{m,g,i}(t) \leq N \end{aligned} \quad (12)$$

其中, $U_{ij,m}(t)$ 和 $D_m(t)$ 分别表示数据包 m 在该步所选链路的可靠性与时延, U_{\max} 和 D_{\max} 是用于归一化的理论最大可靠性与最大时延值, 权重系数 $\alpha \in [0,1]$ 用于平衡两个优化目标的重要性。约束条件的第一项表示数据包 m 的链路时延以及可靠性不应超过理论最大值。约束条件的第二项表示在时间 t 内缓存在卫星 i 队列中的数据包与下一时间步从其邻居接收的数据包的总和不得超过卫星的最大容量 N , 其中, $Z_i(t)$ 为卫星 i 在时刻 t 的邻居节点集合, $y_{m,i}^{\text{queue}}(t)$ 、 $x_{m,g,i}(t)$ 均为二元变量, $y_{m,i}^{\text{queue}}(t)$ 表示数据包 m 在时刻 t 是否位于卫星 i 的队列中, $x_{m,g,i}(t)$ 表示数据包 m 在时刻 t 是否从卫星 i 邻居发送至卫星 i 的缓存队列。

2 基于多智能体强化学习的弹性路由决策

本节将详细阐述 ERIR 框架中的智能决策模块, 该模块基于多智能体强化学习构建。为实现智能体之间的分布式协同, 该模块采用了一种高效的消息传递机制来聚合邻域状态信息, 从而为每个智能体的决策提供充分的局部观测。

2.1 分布式部分可观测马尔可夫决策过程建模

本文旨在求解一个联合优化端到端时延与路径可靠性的多目标路由问题, 文献[12]已证明该类问题属于 NP-hard 问题, 且空间环境的高度随机性也使得构建精确的全局模型变得不可行, 利用传统优化方法在计算上难以应对, 因此, 本文采用强化学习作为求解框架。同时考虑到 LEO 星座固有的分布式特性, 一个集中式智能体将面临维度灾难与巨大的信令开销, 而为每颗卫星部署独立的单智能体则会因其他智能体的策略更新导致环境不稳定, 阻碍策略收敛。鉴于每个卫星节点仅能基于局部观测进行决策, 且其决策效果高度依赖与其他节点的协同, 因此, 本文选择将该问题转化为分布式部分可观测马尔可夫决策过程, 并使用多智能体强化学习方法进行求解。同时, 利用多智能体深度强化学习解决协同问题, 使得各智能体通过协作最大化网络的整体目标函数。该模型主要包含以下 4 个关键要素。

(1) 观测空间 (Ω)。在智能体分布式执行时, 每个智能体 i 仅能访问其局部观测 $\tau_{t,i} \in \Omega$, 该观测向量旨在全面表征节点的局部网络环境, 其构成表示为^[13]:

$$\tau_{t,i} = [C_{t,i}, \mathbf{L}_{t,i}, \mathbf{D}_{t,i}^{\text{queue}}, \mathbf{D}_{t,i}^{\text{prop}}, \mathbf{H}_{t,i}, \mathbf{F}_{t,i}] \quad (13)$$

其中, $C_{t,i}$ 是节点 i 的可用缓存空间; $\mathbf{L}_{t,i}$ 是一个二进制向量, 表示与邻居节点的 ISL 的连通性; $\mathbf{D}_{t,i}^{\text{queue}}$ 和 $\mathbf{D}_{t,i}^{\text{prop}}$ 分别是到各邻居节点的预估排队时延和传播时延向量; $\mathbf{H}_{t,i}$ 是

从各邻居节点到目的地的预估跳数向量; $\mathbf{F}_{t,i}$ 则是各条 ISL 的环境链路可靠性指数向量。

(2) 状态空间 (S)。作为一个多智能体决策过程, 不同智能体之间的路由决策在路径选择上是相互关联的, 因此需要通过全局状态来促进智能体之间的协同路由由机制。因此, 全局状态定义为所有智能体的本地观测的集合, 即:

$$s_t = \{\tau_{t,1}, \tau_{t,2}, \dots, \tau_{t,N}\} \quad (14)$$

(3) 动作空间 (A)。联合动作空间为各智能体局部动作空间的笛卡尔积, 即 $A = A_1 \times \dots \times A_N$ 。对于任意待转发的数据包, 智能体 i 的局部动作 $u_{t,i} \in A_i$ 是从其 K 个可用邻居节点中选择一个作为下一跳, 或选择“保持”动作 (当无合适下一跳或为避免拥塞时)。因此, A_i 是一个包含 $K+1$ 个离散动作的集合。

(4) 奖励函数。为实现时延与可靠性的多目标优化, 我们设计了一个全局奖励函数 R_t 来引导所有智能体的协同。在每个时间步 t , 系统根据所有数据包 $m \in M$ 的路由决策计算一个统一的奖励信号, 即:

$$R_t = \alpha \frac{\sum_{m \in M} U_{ij,m}(t)}{|M| \cdot U_{\max}} - (1 - \alpha) \frac{\sum_{m \in M} D_m(t)}{|M| \cdot D_{\max}} - \omega \quad (15)$$

其中, $U_{ij,m}(t)$ 和 $D_m(t)$ 分别表示数据包 m 在该步所选链路的可靠性与时延, U_{\max} 和 D_{\max} 是用于归一化的理论最大可靠性与最大时延值, 权重系数 $\alpha \in [0,1]$ 用于平衡两个优化目标的重要性。该全局奖励信号被所有智能体共享, 用于更新其策略网络, ω 为数据包丢包的惩罚因子。

2.2 算法架构

(1) 基于消息传递机制的信息聚合模型

信息聚合模型的核心任务是学习一个映射函数 $f: G \rightarrow \mathbb{R}^d$, 该函数能够根据图的拓扑结构与节点/边属性, 生成富有信息的节点、边或全图级别的表征, 这一过程主要依赖于一种递归的、邻域聚合或消息传递的机制^[14]。具体而言, 每个节点通过迭代式地聚合其邻域内节点与边的信息, 并结合自身当前状态来更新其自身的向量表征。经过 T 轮迭代, 每个节点的最终表征能够编码其 T -hop 邻域内的结构与特征信息, 从而捕获图的多层次、层级化特征。因此, 本文提出一种特定的信息聚合模型, 该模型遵循消息传递范式, 并将其分解为消息生成、消息聚合和节点更新 3 个核心阶段。给定一个图 $G = (V, E)$, 其中 V 是节点集合, E 是边集合。每个节点 $x_i \in V$ 在第 t 次迭代的隐藏表征记为 $h_{x_i}^t \in \mathbb{R}^{d_n}$, 其中 d_n 是节点表征的维度。初始表征 $h_{x_i}^0$ 可以由节点的原始特征初始化。

模型的迭代更新过程 ($t = 0, 1, \dots, T - 1$) 如下所述。

基于消息传递范式的信息聚合模型如图 2 所示, 首先, 在每次迭代中, 模型首先为图中的每条边生成一个消息向量。对于连接节点 x_i 和 x_j 的边 e_{ij} , 其在第 t 步生成的消息 $h_{e_{ij}}^t \in \mathbb{R}^d$ 是通过一个可学习的消息函数 $m(\cdot)$ 计算得出的。该函数接收边所连接的两个节点的表征作为输入, 具体形式为:

$$h_{e_{ij}}^t = m\left(h_{x_i}^t, h_{x_j}^t - h_{x_i}^t\right) \quad (16)$$

其中, $h_{x_i}^t$ 代表源节点的当前状态, $h_{x_j}^t - h_{x_i}^t$ 这一项显式地编码了两个节点表征之间的差异或关系向量。这种设计使得消息不仅依赖于节点的绝对状态, 更捕捉了节点间的相对关系, 从而生成更具表达能力的边级别信息。消息函数 $m(\cdot)$ 通常由一个多层感知机等神经网络实现。

接下来, 每个节点 x_i 会聚合所有从其邻接边传入的消息。令 $\mathcal{N}(i)$ 表示节点 x_i 的邻居节点集合, 节点 x_i 接收到的聚合消息 $M_{x_i}^{t+1}$ 定义为:

$$M_{x_i}^{t+1} = \sum_{j \in \mathcal{N}(i)} h_{e_{ij}}^t \quad (17)$$

这里的求和操作是一种置换不变的聚合函数, 确保模型对于邻居节点的顺序不敏感。最后, 每个节点 x_i 利用聚合后的消息 $M_{x_i}^{t+1}$ 来更新其自身的隐藏表征。这一过程由一个更新函数 $t(\cdot)$ 控制, 该函数将节点上一时刻的表征 $h_{x_i}^t$ 与新接收的聚合消息 $M_{x_i}^{t+1}$ 相结合, 即:

$$h_{x_i}^{t+1} = t\left(h_{x_i}^t, M_{x_i}^{t+1}\right) \quad (18)$$

通过堆叠 T 个这样的消息传递层, 模型能够学习到节点的高阶表征。消息函数 $m(\cdot)$ 和更新函数 $t(\cdot)$ 内的参数是可学习的, 并通过端到端的方式, 在特定任务的损失函数下通过反向传播算法进行优化。

(2) 基于值函数分解的协同多智能体决策机制

在本文所研究的多智能体系统中, 多个智能体并行运作并追求各自的目标。若各智能体独立学习, 将面临两大核心挑战: 环境的非平稳性以及局部贪婪决策导致的全局次优性。为应对上述挑战, 本文采用集中式训练分布式执行的范式, 执行过程如图 3 所示。该范式的核心在于学习一个全局联合动作价值函数 $Q(\tau, u)$, 并通过值函数分解将其与各智能体的局部价值函数 $Q_i(\tau_{i,i}, u_{i,i})$ 关联起来^[15]。这种分解的关键在于必须满足个体-全局最大化 (IGM) 原则, 该原则确保在分布式执行阶段, 对各智能体局部价值函数的独立最大化, 等价于对全局联合价值函数的最大化, 其形式化表示为:

$$\begin{aligned} \operatorname{argmax}_u Q(\tau, u) = \\ \left(\operatorname{argmax}_{u_1} Q_1(\tau_{1,1}, u_{1,1}), \dots, \operatorname{argmax}_{u_N} Q_N(\tau_{N,N}, u_{N,N})\right) \end{aligned} \quad (19)$$

其中, τ 和 u 分别代表联合观测与联合动作。为有效满足 IGM 原则, 施加了一个充分条件: Q 必须是关于每个 Q_i 的单调非递减函数, 即 $\frac{\partial Q}{\partial Q_i} \geq 0, \forall i$ 。这一单调性约束保证了局部优化方向与全局优化目标的一致性。

本文通过一个多星协作混合网络来实现这种具备复杂非线性映射能力的单调函数, 该网络以所有局部价值 Q_i 为输入, 输出全局价值 Q 。通过将其网络权重约束为

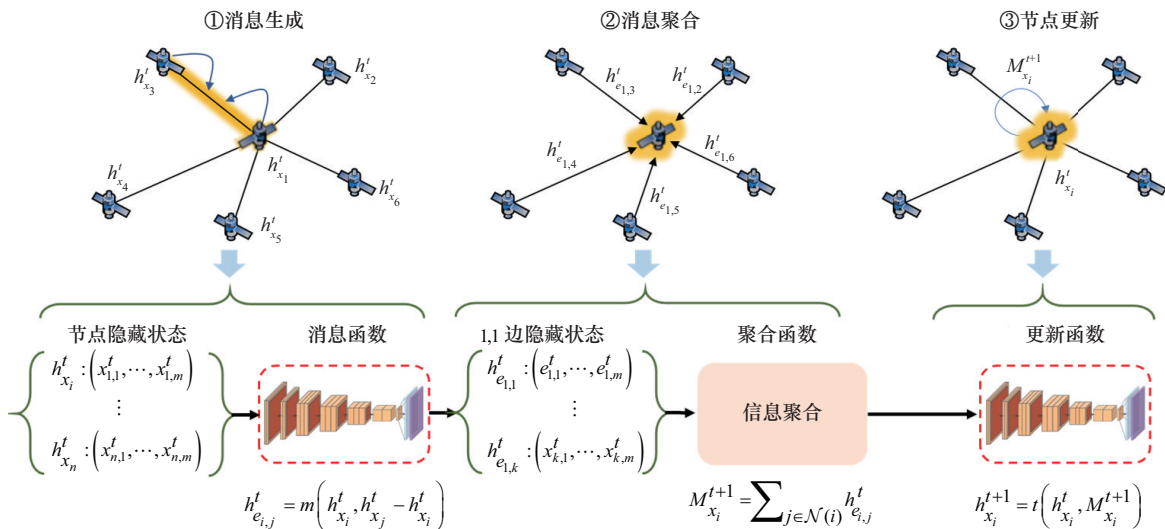


图 2 基于消息传递范式的信息聚合模型

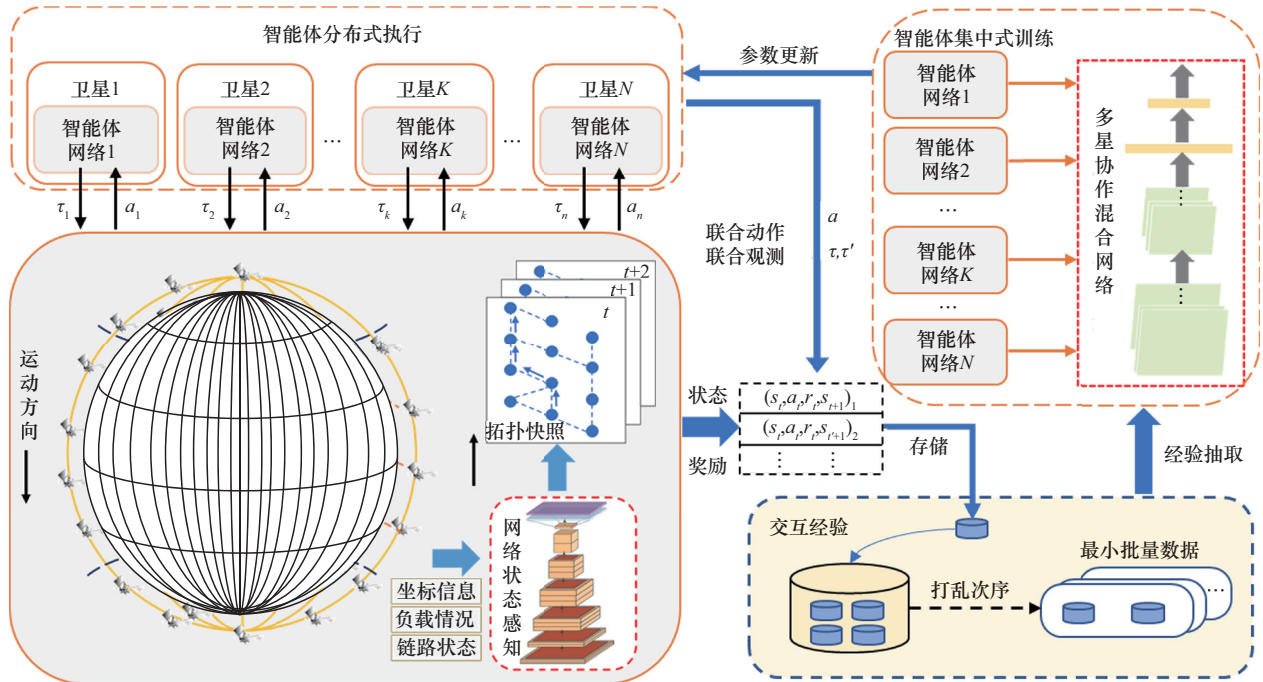


图 3 卫星网络集中式训练分布式执行过程

非负，确保了 Q 对 Q_i 的单调性。整个模型（包括各智能体的表征网络与多星协作混合网络）均通过最小化联合时序差分（temporal difference, TD）误差进行端到端训练，其损失函数定义为：

$$L(\theta) = \sum_{k=1}^b (y_t - Q(\tau, u; \theta))^2 \quad (20)$$

其中， b 是从经验回放缓冲区中采样的批次大小。TD 目标 y_t 的计算式为：

$$y_t = R_t + \gamma \max_{u'} Q(\tau', u'; \bar{\theta}) \quad (21)$$

其中， R_t 是全局奖励， γ 是折扣因子， τ' 和 u' 分别代表下一时刻的联合观测与联合动作， θ 和 $\bar{\theta}$ 分别是当前网络和目标网络的参数。目标网络参数 $\bar{\theta}$ 定期从当前网络参数 θ 复制而来，以稳定训练过程。通过这种方式，智能体在追求各自局部 Q_i 最大化的同时，隐式地协同优化了整个系统的全局性能。

3 仿真与结果分析

3.1 仿真参数

为对本文提出的 ERIR 路由算法的有效性进行验证，本文构建了基于铱星系统的低轨卫星网络仿真环境。利用 STK 软件对包含 66 颗卫星的铱星星座进行精确轨道建模，并导出其随时间变化的动态网络拓扑。该星座系统由 6 个轨道面构成，每个轨道面均匀分布 11 颗卫星，每

颗卫星维持 4 条固定的星间链路。仿真场景的关键系统级参数依据铱星星座的实际物理特性进行配置，具体数值见表 1。

表 1 系统仿真参数

参数名称	参数值
卫星轨道高度	780 km
卫星数量	66 颗
轨道面数量	6 个
每个轨道面内卫星数量	11 颗
轨道倾角	86.4°
数据包大小	1 KB
卫星队列长度	20 KB
仿真时间	200 s
时隙时长	1 s

在业务模型方面，仿真开始时在全网范围内均匀随机地生成源目的节点对，以模拟多样化的通信请求。网络中的所有卫星节点均具备数据包的生成与转发能力。本次仿真实验的硬件与软件平台配置如下：CPU 为 Intel Core i9-14900K，内存为 32 GB，GPU 为 NVIDIA GeForce RTX 3060；操作系统为 Windows 11，编程语言为 Python 3.9，深度学习框架为 PyTorch 1.13.1。ERIR 基于“集中式训练、分布式执行”的 MARL 范式，其核心算法在训练阶段所采用的关键参数配置见表 2。智能体网络

采用门控循环单元以处理时序信息，混合网络则借鉴 QMIX 的设计^[15]，通过超网络动态生成权重以实现高效的值函数分解。

表 2 算法仿真参数

参数名称	参数值
训练时间步数	605 000 步
折扣因子	0.99
经验回放池大小	2 000
目标网络更新频率	200
初始探索概率	100%
最终探索概率	5%
批处理数量	32 个
智能体网络隐藏层大小	64
混合网络隐藏层大小	32
智能体网络学习率	0.5%
混合网络学习率	0.5%
优化器	Adam
梯度范数裁剪阈值	10

为全面评估 ERIR 算法的性能优势，选取了以下 3 类算法作为基准方法进行对比。

(1) 深度 Q 网络 (deep Q-network, DQN) 算法^[16]: 该算法是一种通用的深度强化学习算法，其中每个卫星节点被视为一个独立的 DQN 智能体，并根据本地观察信息独立做出实时的下一跳决策。

(2) 最短路径优先 (shortest path first, SPF) 算法^[17]: 该算法采用 Dijkstra 最短路径法，以平均传播时延为链路度量，最小化总时延。

(3) 最小跳数 (minimum hop, MinHop) 算法^[18]: 该算法将跳数作为链路度量，以最小化总跳数为目标。

3.2 仿真结果分析

(1) 收敛性分析

图 4 展示了 ERIR 算法在集中式训练阶段的平均奖励值变化曲线，其中阴影区域表示奖励值的标准差。在训练初期，智能体通过与环境的持续交互进行策略探索，其协同路由策略在有效的奖励函数与值函数分解机制的驱动下迅速优化，表现为平均奖励值在最初的 2×10^5 步内急剧攀升。此后，奖励的增长趋于平缓，并在约 3×10^5 步后进入收敛状态，最终趋于稳定。训练曲线的稳定收敛证明了所提 MARL 框架的有效性 with 学习稳定性。训练完成后得到的收敛模型将被部署至各卫星节点，用于后续分布式执行环境下的性能对比测试。

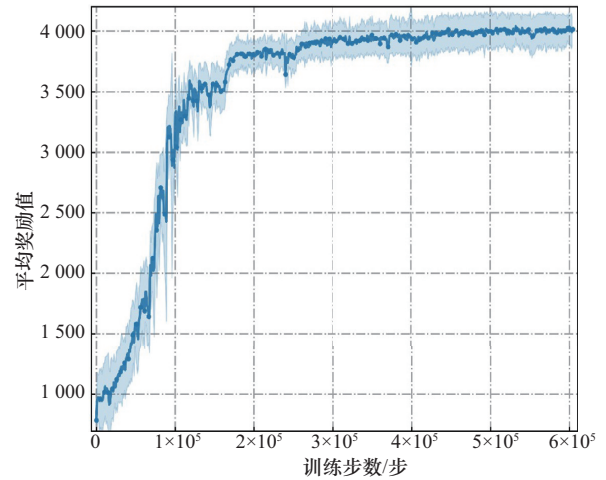


图 4 ERIR 算法在集中式训练阶段的平均奖励值变化曲线

(2) 不同数据流数量下的路由算法性能分析

图 5 所示为丢包率随数据流数量的变化，在数据流数量较少时，所有算法的丢包率均相对较低，随着数据流数量不断增加，丢包率均呈上升趋势，这是因为随着网络中数据包增多，卫星节点虽然有缓存队列进行数据包的缓存，但队列长度有限，导致丢包率上升。本文所提算法设置了丢包惩罚奖励，在集中式训练期间得到充分的训练，并且通过与其邻居节点交互负载信息，避免了向缓存队列拥塞的节点转发数据包，相较于对比算法，保持了相对较低的丢包率，与 SPF 算法相比，丢包率降低约 31.4%。

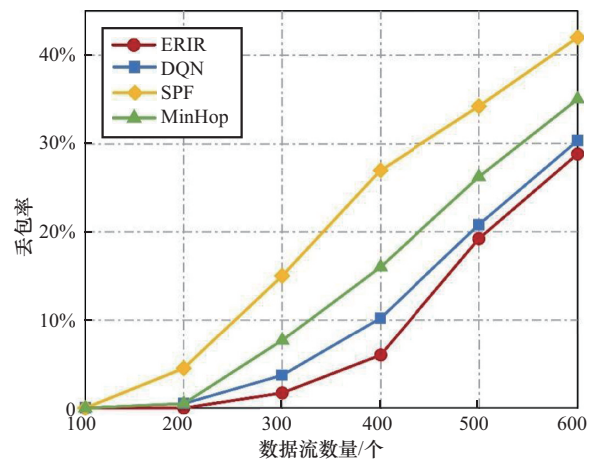


图 5 丢包率对比

端到端时延是衡量路由算法效率的核心指标，主要由传播时延与排队时延构成，其中排队时延受网络拥塞状况影响显著。图 6 对比了不同算法在不同数据流规模下的端到端时延性能。与丢包率趋势相似，所有算法的端

到端时延均随数据流数量的增加而增长。在高负载条件下，SPF和MinHop算法由于无法规避拥塞，导致数据包在中间节点产生严重的排队时延，性能下降最为明显；DQN算法虽优于传统算法，但其分布式决策机制缺乏有效的智能体间协作，易陷入局部最优解，使其时延表现略逊于ERIR；ERIR算法得益于其“集中训练、分布执行”范式与基于值函数分解的协同多智能体决策机制，在训练中能够利用全局信息规避局部最优陷阱，在执行中通过邻居状态交互做出近乎全局最优的协同路由决策。从图6可以看出，ERIR在所有负载条件下均实现了端到端时延最低。与MinHop算法相比，ERIR的平均端到端时延性能提升了约17.8%。

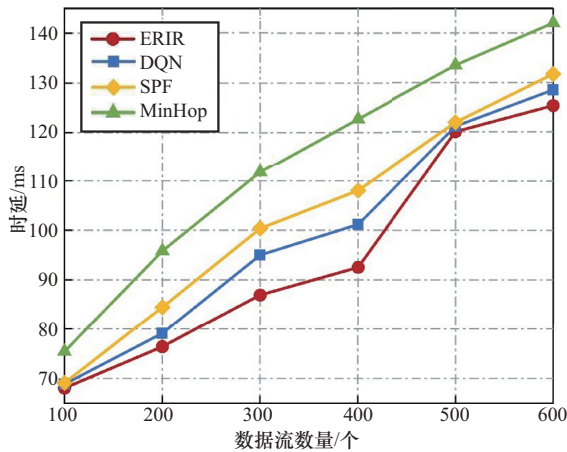


图6 端到端时延对比

为量化网络流量的分布均衡程度，本文引入流量分布指数进行评价，其定义为：

$$f = \frac{\left(\sum_{i=1}^n x_i\right)^2}{n \sum_{i=1}^n x_i^2} \quad (22)$$

其中， n 为卫星节点个数， x_i 为第 i 个卫星节点中的队列长度，该指数的取值范围为0到1，较低的流量分布指数表示流量分布不均匀，可能存在部分链路拥塞而其他链路利用率较低的情况。图7展示了4种算法的流量均衡指数随数据流数量的变化情况。SPF与MinHop算法采用固定的最短路径或最少跳数原则，流量被高度汇集于少数关键链路，导致网络拥塞热点出现，因此其流量均衡指数始终处于较低水平。DQN算法通过学习机制能够实现一定程度的流量疏导，表现优于传统算法。值得注意的是，当数据流数量低于400个时，ERIR算法的流量均衡指数略低于DQN算法，出现该现象的原因在于，ERIR算法的决策并非单一追求流量均衡，而是对可靠性与时

延的复合优化。在网络负载较低、拥塞不成为主要矛盾时，ERIR算法的“主动规避”特性会使其为保证路径的长期可靠性，预先放弃那些存在潜在环境风险的链路——即使这些链路当前是空闲的。这种策略将流量引导至一个更小的“可靠链路子集”中，导致了流量分布的集中度略高于DQN算法，DQN算法因缺少先验的风险感知，会更自由地利用所有可用链路，在低负载下反而表现出更好的流量均衡性。然而，当网络负载升高后，ERIR算法能够借助其奖励函数中对排队时延的优化目标，更高效地学习全局流量分配策略，主动引导数据流至非拥塞路径，从而在流量均衡指数上实现反超，并最终最大化全网的资源利用效率。

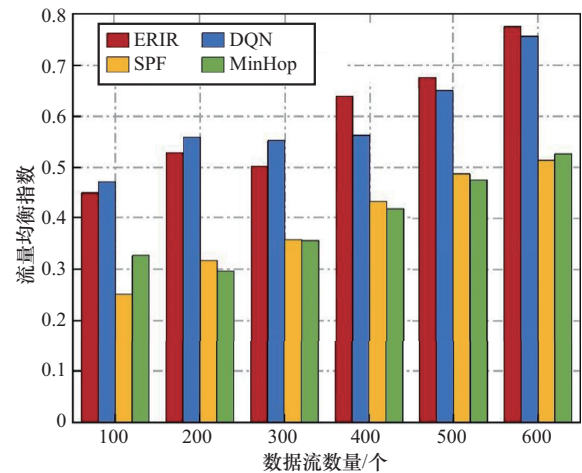


图7 流量均衡指数对比

(3) 不同链路故障概率下的路由算法性能分析

为模拟日凌、粒子群干扰等空间环境因素导致的星间链路故障，仿真通过建立的链路可靠性模型计算出链路的可靠性指数，选取可靠性指数低的链路来模拟LEO星间链路的失效，并生成不同比例的链路故障事件，如{3%,6%,...,18%}。为检验算法在网络高负载且部分链路失效的复杂场景下的性能鲁棒性，本次仿真将数据流数量固定为400个。设置该参数的原因在于：如图7所示，当数据流数量低于400个时，网络整体负载较低，拥塞尚不明显，ERIR算法在流量均衡方面的优势未能完全体现；而当数据流达到400个时，网络已处于较高负载状态，传统路由算法性能开始恶化，此时评估各算法在链路故障环境下的适应能力与可靠性最具代表性。4种算法的丢包率与端到端时延随链路故障概率的变化如图8~图9所示。随着故障概率的增大，4种算法的端到端时延和丢包率基本呈上升趋势，体现了链路故障对网络性能的负面影响。

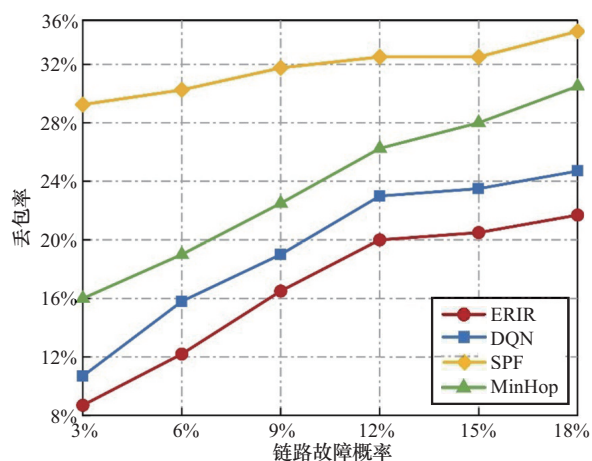


图 8 丢包率随链路故障概率的变化情况

从图 8 可以看出，随着故障链路的增多，网络连通性下降，所有算法的丢包率均呈现上升趋势。SPF 与 MinHop 等传统算法由于采用静态路由策略，无法实时感知网络拓扑的动态变化。当预设路径上的链路发生故障时，节点仍会继续向该失效链路转发数据包，导致数据包大量丢失，因此其丢包率始终维持在较高水平。相比之下，基于强化学习的 DQN 与 ERIR 算法能够将网络实时状态（包括链路故障信息）纳入决策过程，从而动态调整路由路径以规避失效链路。ERIR 算法凭借其全局信息感知的优势，在路径重规划方面表现更优，其丢包率在整个故障区间内始终为最低。

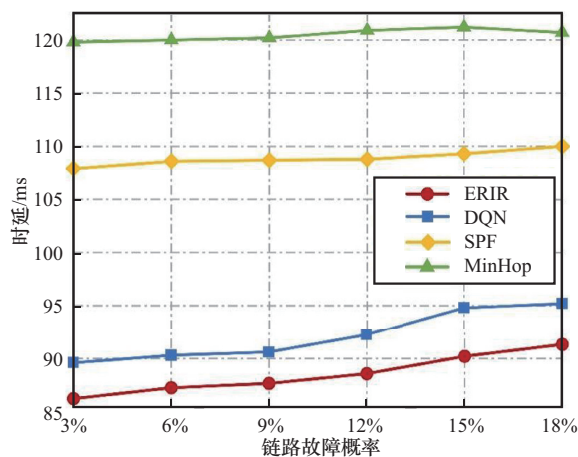


图 9 时延随链路故障概率的变化情况

从图 9 可以看出，各算法的时延曲线随链路故障概率的增长变化相对平缓，其原因为时延统计仅针对成功到达目的地的有效数据包，大量因链路故障而直接丢失的数据包并未计入，这在一定程度上掩盖了网络性能的恶

化。SPF 与 MinHop 算法不仅无法有效规避故障，其固有的拥塞问题在网络拓扑受损时还会进一步加剧，导致其时延一直处于高位。ERIR 算法的时延曲线始终处于较低的水平，在所有测试场景中均表现出最优的时延性能，这表明 ERIR 算法能够高效地感知网络拓扑变化，并迅速计算出时延最优的替代路径，有效规避了链路中断和网络拥塞导致的性能下降，保证了业务传输的高质量。

4 结束语

为应对空间环境对低轨卫星网络的可靠性挑战，本文提出了一种基于环境可靠性预测的弹性智能路由方法，该方法的核心是将数据驱动的智能决策与物理环境模型深度融合，实现了路由决策从“被动响应”到“主动规避”的转变。该方法通过构建一个融合辐射风险、空间天气与实时链路质量的多维可靠性指数，驱动路由算法前瞻性地规避高风险区域。仿真结果表明，该方法在降低丢包率与端到端时延方面效果显著。尽管如此，当前的环境模型仅是对复杂物理效应的简化抽象，因此，未来研究可从 3 方面深化：一是引入更高保真度的动态物理模型以提升环境预测的精准度；二是针对未来超大规模网络的可扩展性问题，探索层级化新型范式，以解决当前“集中式训练”范式在面对数千颗卫星时所面临的维度灾难与通信开销挑战；三是开展面向未来数千颗卫星规模的巨型星座的仿真验证，进一步检验本文所提 ERIR 算法的性能与可扩展性，并针对其可能出现的层级化、异构化等特征进行适应性改进。

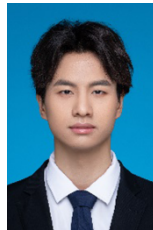
参考文献

- [1] Zhang P, Yang P, Kumar N, et al. Space-air-ground integrated network resource allocation based on service function chain[J]. IEEE Transactions on Vehicular Technology, 2022, 71(7): 7730-7738.
- [2] Giuliani G, Klenze T, Legner M, et al. Internet backbones in space[J]. ACM SIGCOMM Computer Communication Review, 2020, 50(1): 25-37.
- [3] Yang J, Li D, Jiang X, et al. Enhancing the resilience of low earth orbit remote sensing satellite networks[J]. IEEE Network, 2020, 34(4): 304-311.
- [4] Yan F, Wang Z, Zhang S, et al. Logic path identified hierarchical routing for large-scale LEO satellite networks[J]. IEEE Transactions on Network Science and Engineering, 2024, 11(4): 3731-3746.
- [5] Zheng K, Zhao Y, Cao Y, et al. Wireless optical inter-satellite link risk-aware snapshot routing strategy in satellite networks: multi-

- stage space debris evolution[J]. IEEE Transactions on Network Science and Engineering, 2025.
- [6] Zhu Y, Rui L, Qiu X, et al. Double-layer satellite communication network routing algorithm based on priority and failure probability [C]//2019 15th International Wireless Communications & Mobile Computing Conference (IWCMC). IEEE, 2019: 1518-1523.
- [7] Wei W, Fu L, Gu H, et al. IRIS: towards intelligent reliable routing for software defined satellite networks[J]. IEEE Transactions on Communications, 2024.
- [8] Niu Z, Yang H, Yao Q, et al. Reliable low-latency routing for VLEO satellite optical network: a multi-agent reinforcement learning approach[J]. IEEE Internet of Things Journal, 2024.
- [9] Vette J I. The NASA/national space science data center: trapped radiation environment model program (1964-1991) [M]. National Space Science Data Center (NSSDC), World Data Center A for Rockets and Satellites (WDC-AR &S), National Aeronautics and Space Administration, Goddard Space Flight Center, 1991.
- [10] Han Z, Xu C, Zhao G, et al. Time-varying topology model for dynamic routing in LEO satellite constellation networks[J]. IEEE Transactions on Vehicular Technology, 2022, 72(3): 3440-3454.
- [11] Mao B, Zhou X, Liu J, et al. On an intelligent hierarchical routing strategy for ultra-dense free space optical low earth orbit satellite networks[J]. IEEE Journal on Selected Areas in Communications, 2024, 42(5): 1219-1230.
- [12] Clementi A, Di Ianni M. Optimum schedule problems in store and forward networks[C]//Proceedings of INFOCOM'94 Conference on Computer Communications. IEEE, 1994: 1336-1343.
- [13] Chen X, Ji Z, Wu S, et al. A distributed routing algorithm for LEO satellite networks: a multi-agent transformer-MIX learning approach[J]. IEEE Internet of Things Journal, 2025.
- [14] Wang Y, Sun Y, Liu Z, et al. Dynamic graph CNN for learning on point clouds[J]. ACM Transactions on Graphics (tog), 2019, 38(5): 1-12.
- [15] Rashid T, Samvelyan M, De Witt C S, et al. Monotonic value function factorisation for deep multi-agent reinforcement learning[J]. Journal of Machine Learning Research, 2020, 21(178): 1-51.
- [16] 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.
- [17] Lee D C. Proof of a modified Dijkstra's algorithm for computing shortest bundle delay in networks with deterministically time-varying links[J]. IEEE communications letters, 2006, 10(10): 734-736.

- [18] Yang Y, Xu M, Wang D, et al. Towards energy-efficient routing in satellite networks[J]. IEEE Journal on Selected Areas in Communications, 2016, 34(12): 3869-3886.

[作者简介]



陈啸 (2001-), 男, 北京邮电大学博士生, 主要研究方向为大规模卫星网络路由技术。



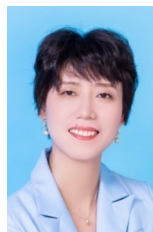
纪哲 (1989-), 女, 博士, 北京邮电大学副教授, 主要研究方向卫星算力网络、空天智能组网和卫星网络资源管理技术等。



吴胜 (1982-), 男, 博士, 北京邮电大学教授, 主要研究方向为迭代检测和解码、信道估计、大规模MIMO和卫星通信等。



姬思敬 (1993-), 男, 博士, 西安电子科技大学助理研究员, 主要研究方向为大规模卫星星座组网方法、星地协同计算理论、多星协同通信技术。



盛敏 (1975-), 女, 博士, 西安电子科技大学教授、空天地一体化综合业务网全国重点实验室主任, 主要研究方向为空地一体化组网技术等。