

卫星网络智能路由优化机制

张宇

北方工业大学信息学院, 北京

收稿日期: 2025年3月19日; 录用日期: 2025年4月15日; 发布日期: 2025年4月25日

摘要

本文针对低轨道卫星网络中动态拓扑结构与多优先级业务需求不匹配的问题, 提出了一种结合分级服务标识与深度强化学习的动态路由优化方案。该方案包含三个核心模块: 首先, 基于IPv6原生字段的零开销服务标识机制来区分不同优先级业务; 接着, 提出分层式流量感知模型(H-CMS), 用于低开销、高精度地测量多类业务流量; 最后, 双延迟深度确定性策略梯度(TD3)算法, 用于实时生成最优路由决策。实验结果表明, 与传统方法相比, 本方案显著提高了网络吞吐量和带宽利用率, 特别是在卫星切换和高负载情况下表现更为稳定, 吞吐量整体提升8%, 带宽利用率维持在75%左右, 为低轨道卫星网络中的差异化服务提供了有效解决方案。

关键词

低轨道卫星网络, 多业务需求, 分级服务, 流量感知, 强化学习

Satellite Network Intelligent Routing Optimization Mechanism

Yu Zhang

School of Information, North China University of Technology, Beijing

Received: Mar. 19th, 2025; accepted: Apr. 15th, 2025; published: Apr. 25th, 2025

Abstract

Aiming at the problem of mismatch between dynamic topology and multi-priority service requirements in low-orbit satellite networks, this paper proposes a dynamic routing optimization scheme combining hierarchical service identification and deep reinforcement learning. The scheme consists of three core modules: first, a zero-overhead service identification mechanism based on IPv6 native fields is used to distinguish services of different priorities; second, a hierarchical traffic-aware model (H-CMS) is proposed to measure multi-class service traffic with low overhead and high

precision; finally, a double-delay deep deterministic policy gradient (TD3) algorithm is used to generate optimal routing decisions in real time. Experimental results show that compared with traditional methods, this scheme significantly improves network throughput and bandwidth utilization, especially in satellite switching and high-load conditions. The overall throughput is increased by 8%, and the bandwidth utilization is maintained at about 75%, providing an effective solution for differentiated services in low-orbit satellite networks.

Keywords

Low-Orbit Satellite Network, Multi-Service Requirements, Hierarchical Services, Traffic-Aware, Reinforcement Learning

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

近地轨道卫星通信系统(Low Earth Orbit, LEO)的快速发展正在重塑全球信息基础设施格局。相较于地面通信网络, LEO 卫星凭借其广域覆盖、低时延传输及高吞吐量等特性, 已成为构建空天地一体化网络的关键环节, 在偏远区域通信、应急救援及特种通信领域展现出显著优势[1]。商业星座的大规模部署(如 Starlink、OneWeb)与卫星制造成本的持续降低, 加速了 LEO 网络与 5G/6G 地面网络的深度融合[2]。然而, 拓扑动态性与业务多样化的双重挑战, 导致传统路由机制难以满足差异化服务质量需求。

现有研究在卫星网络路由优化方面存在三方面局限性: 首先, 现有流量识别技术对业务类型的区分粒度不足。基于 Count-Min Sketch 等方法的流量统计模型虽能实现粗粒度流分类, 却无法有效识别时延敏感型与带宽敏感型等精细化业务类别; 其次, 静态资源调度机制与卫星网络的动态特性存在适配矛盾。传统路由协议(如 OSPF)依赖周期性路由更新, 在拓扑变化周期缩短至分钟级的卫星网络中, 不仅导致显著的信令开销, 更因状态更新滞后引发路由决策失效[3]; 最后, 现有服务区分机制引入额外通信开销。基于 QUIC 协议扩展或 MPTCP 子流标记的业务标识方案虽能提升分类精度, 但其增加的报文头部信息会挤占星间链路的有效载荷带宽。

针对上述问题, 本文创新性提出融合分级业务标识与深度强化学习的动态路由优化框架, 主要贡献体现在如下三个方面:

1) 提出基于 IPv6 原生字段的无损业务标识方法, 通过 Traffic Class 与 Flow Label 字段的协同编码实现零开销优先级标记。

2) 构建分层式流量监测模型(Hierarchical Count-Min Sketch, H-CMS), 采用双层哈希结构适配星载计算资源约束, 实现多维度业务特征的精准感知。

3) 设计双延迟深度确定性策略梯度(Twin Delayed Deep Deterministic Policy Gradient, TD3)驱动的智能路由引擎, 通过融合实时拓扑状态与业务优先级信息, 生成负载均衡的优化传输路径。该方案通过业务标识、流量感知与路径决策的协同优化, 有效提升卫星网络的资源利用效率与传输性能。

2. 相关工作

卫星网络智能路由研究进展可归纳为三个技术维度: 传统算法优化、协议改进与智能决策融合。在传统路由机制方面, 最小跳数准则虽能简化路径计算, 但难以应对多业务差异化需求, 易导致链路资源

分配失衡。SDN 技术的引入显著提升了网络状态感知能力,例如 Gao 等人[4]构建的 MPTCP 多链路调控机制,通过动态子流调整实现吞吐量优化;Liu 等人[5]结合信息中心网络设计的轻量化路由方案,将控制信令开销降低 40%以上。然而,这类方法存在两方面局限:其一,静态网络模型难以适配负载剧烈波动场景;其二,算法泛化能力受限于特定网络拓扑。

传输协议优化方面,MPQUIC 凭借零往返时延特性成为新兴研究方向。Yang 等人[6]通过实验验证其在卫星网络中的传输性能优于传统 TCP 协议,而 Shi 等人[7]设计的 MPDTP 系统则通过冗余数据包策略优化传输可靠性。与此同时,在智能算法融合方向,深度强化学习展现出显著优势:Han 等人[8]提出基于深度强化学习的抗干扰路由算法,通过动态生成路由子集降低传输成本;Liu 等人[9]进一步考虑星间能源均衡问题,设计节能路由协议以延长巨型星座寿命。但现有方法仍面临流量分类粒度不足、多目标优化效率低下等共性挑战。

当前研究趋势强调 SDN 与强化学习的协同创新。例如 Wang 等人[10]采用自回归滑动平均模型预测流量分布,实现带宽与路径的动态适配,但传统 DQN 算法在连续控制场景存在决策精度缺陷。相较而言,深度确定性策略梯度(DDPG)算法通过双网络协同机制在复杂环境适应性方面更具优势。因此,现有算法在流量细粒度分类、多目标优化等方面仍有提升空间。

3. 系统架构

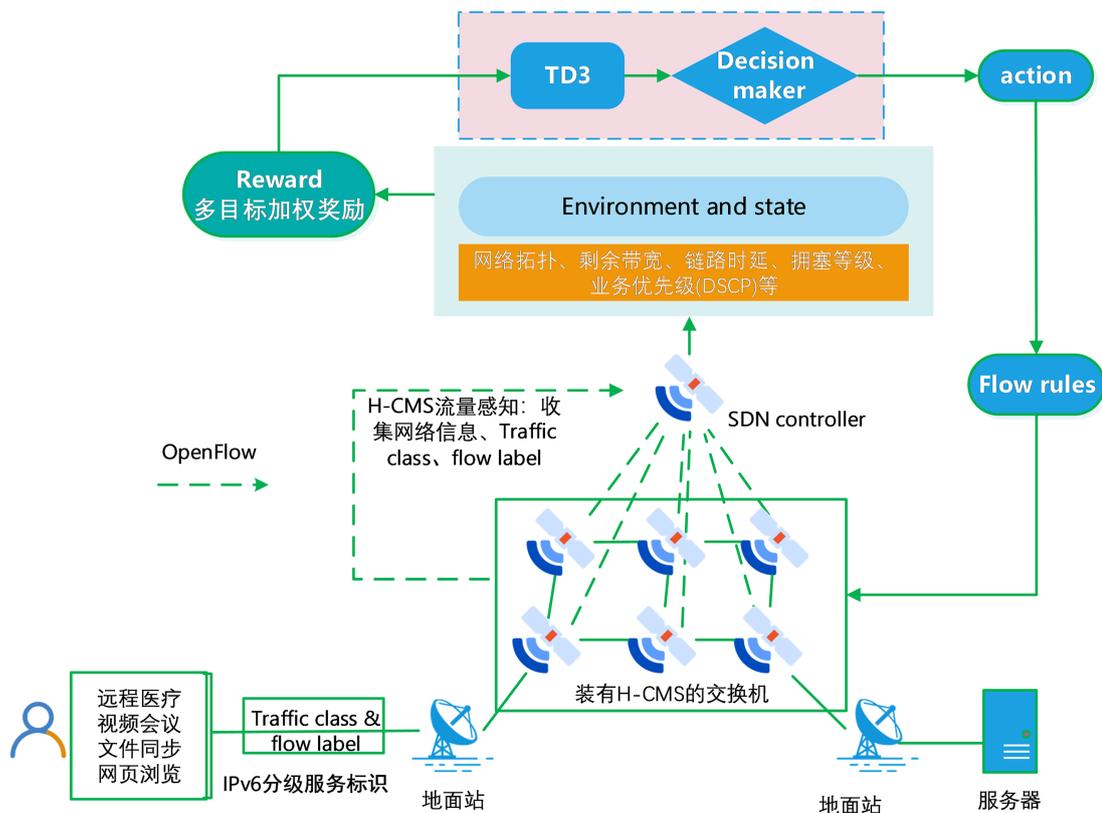


Figure 1. System architecture diagram

图 1. 系统架构图

基于分级服务标识与深度强化学习的卫星网络动态路由优化系统主要由三个功能模块组成:分级服务标识模块、分层流量感知模块和深度强化学习决策模块,如图 1 所示。

分级服务标识模块位于客户端与地面站之间,通过精细化利用 IPv6 头部中的 Traffic Class 和 Flow Label 字段,实现零附加开销的业务优先级划分。当用户发起请求时,系统会根据业务类型(如远程医疗、视频会议、文件同步或网页浏览)分配相应的 DSCP 值和流标签,确保不同业务流具有唯一标识。

分层流量感知模块部署于卫星节点上,采用 H-CMS 结构进行流量统计。每个卫星节点装有 H-CMS 模块,捕获经过的 IPv6 数据包,提取其 Traffic Class 和 Flow Label 字段,通过双层哈希计算将不同优先级业务流量映射到相应计数器中,实现精确测量。SDN 控制器定期收集各节点的 H-CMS 统计数据,汇总形成全局网络流量分布视图。

深度强化学习决策模块由 TD3 算法驱动,位于 SDN 控制器中。该模块接收来自分层流量感知模块的网络状态信息,包括链路剩余带宽、拥塞程度和业务优先级等参数。通过与环境交互,TD3 算法不断学习最优路由策略,生成动态路径权重配置。

4. 系统设计

4.1. IPv6 分级服务标识设计

在实现卫星网络差异化服务质量(QoS)的过程中,有效区分业务优先级是关键环节。现有技术往往通过增加协议扩展字段来标记不同服务类型,然而这种方法会产生额外的传输负担,特别是在卫星间链路等带宽受限环境下,可能大幅降低有效载荷传输效率。

针对这一问题,我们提出一种利用 IPv6 原生字段的服务分级标识技术,充分挖掘 IPv6 头部中已有的 Traffic Class 和 Flow Label 字段潜力,实现了无附加开销的优先级划分机制。如图 2 所示,该方案将 8 位 Traffic Class 用于区分流量类型,同时用 20 位 Flow Label 在各类别内部实现不同数据流的细分。其中, Traffic Class 的前 6 位被用作区分服务代码点(DSCP),划分出四个不同级别的服务优先级;剩余 2 位作为显式拥塞通知(ECN),提供链路拥塞状况的实时反馈。另外, Flow Label 采用密码学哈希运算处理五元组信息(包括源目标 IP、端口及协议类型),生成独特标识符,保证业务流传输连贯性。这一设计既符合 RFC 6437 关于 IPv6 流标签的规范要求,又与 RFC 4594 定义的差异化服务架构保持兼容。

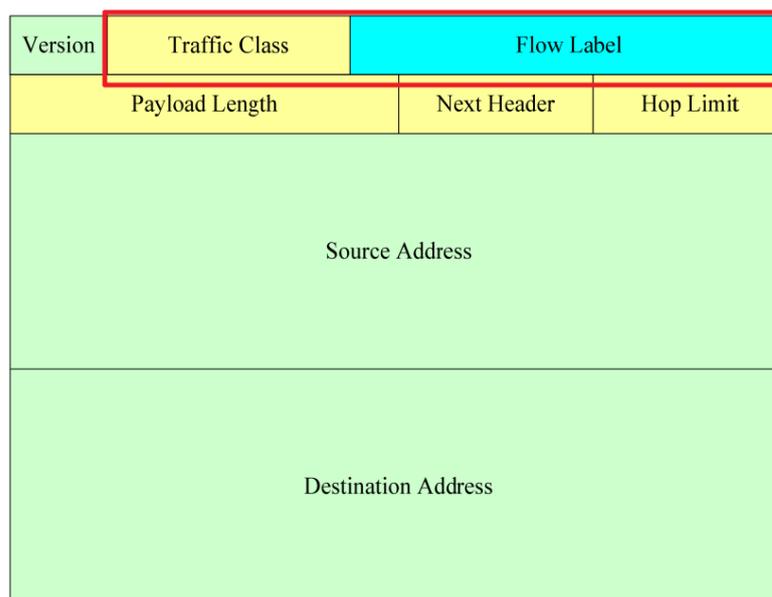


Figure 2. IPv6 basic header format
图 2. IPv6 报文基本头格式

表 1 展示了我们对 DSCP 的优先级分级方案，将全部 64 个值域分成四个区间，分别对应不同种类的业务需求。通过五元组(源 IP、目的 IP、源端口、目的端口、协议类型)哈希算法生成的 Flow Label 确保了相同业务流能够保持唯一标识，从而支持连续性传输。

Table 1. DSCP priority classification and service mapping

表 1. DSCP 优先级划分与业务映射

优先级	DSCP 范围	典型业务	Qos 目标
紧急	48~63	远程医疗、金融交易	时延 ≤ 20 ms, 带宽预留
高	32~47	视频会议、直播	时延 ≤ 50 ms, 高吞吐
中	16~31	观看视频、文件同步	带宽保障, 时延容忍
低	0~15	网页访问、下载文档	尽力而为

4.2. 分层流量感知机制

完成 IPv6 分级服务标识机制设计后，准确测量与精细划分网络流量成为智能动态路由实现的关键支撑。当前主流测量技术，例如通过 SDN 控制器的 LLDP 协议或 REST API 进行流表查询，显示出两个明显缺陷：首先，这类方法精度有限，难以细致识别各优先级业务流的特征；其次，查询及存储大量流表项会消耗宝贵的星上计算资源，与卫星网络环境下的资源约束不相适应。

为解决上述问题，本文提出一种分层流量感知模型(Hierarchical Count-Min Sketch, H-CMS)，该模型与前述服务标识机制相结合，实现了资源消耗低且精确度高的多业务流量监测。传统 Count-Min Sketch (CMS)是一种基于哈希的概率型数据结构，使用固定大小的二维数组和多个哈希函数，在有限内存下进行流量估算。然而，CMS 也存在明显不足：传统 CMS 在处理复杂流量时，不同业务流会发生哈希冲突导致统计值偏高；无法对多种服务类型加以区分；难以同时处理 IPv6 头部中的 DSCP 和 Flow Label 信息。

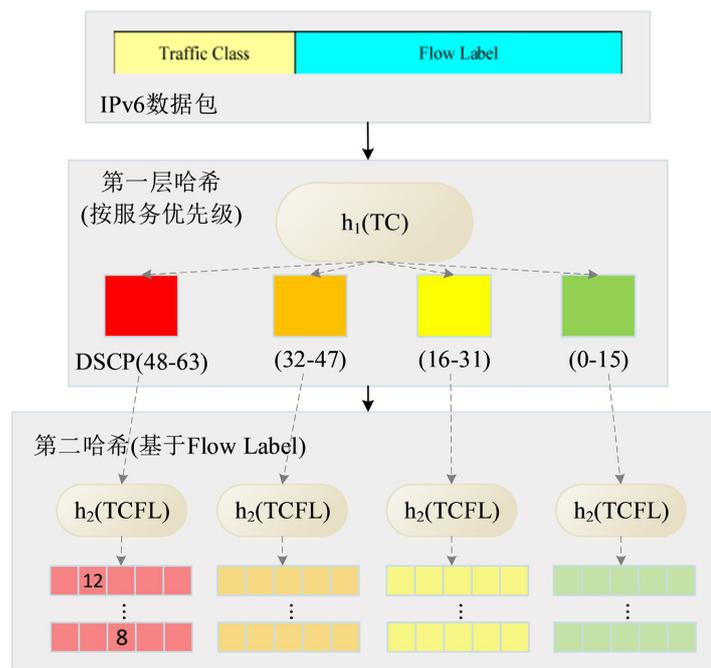


Figure 3. Structural design of H-CMS

图 3. H-CMS 的结构设计

为此，H-CMS 方案创新性地将单一的哈希表划分为两个层次，如图 3 所示：第一层根据业务优先级进行初步分类，利用 Traffic Class 信息将流量分配到紧急、高、中、低四个独立哈希区域；第二层则基于 Flow Label 进行更精细的分析，每个优先级区域内设有专门的计数单元。这种设计通过组合“Traffic Class + Flow Label”作为哈希输入，显著提升了计算准确性。

H-CMS 的技术核心是一个二维计数矩阵 $CM[d][w]$ ，其中 d 表示哈希函数数量， w 表示每种哈希映射的计数器数目。与传统 CMS 中使用的简单哈希函数不同，H-CMS 采用了一种专门为 IPv6 头部特征优化的多项式哈希函数：

$$h_i(x) = ((a_i \cdot x + b_i) \bmod p) \bmod m_i, i \in \{1, 2\} \quad (1)$$

其中， x 是由 Traffic Class 和 Flow Label 组合得到的键值 TCFL， a_i 和 b_i 是哈希函数的系数， p 是大素数，用于减少哈希冲突， m_i 则表示第 i 层哈希表的大小。

在哈希架构方面，H-CMS 采用创新的双层设计：第一层哈希函数 h_1 使用 Traffic Class 作为输入，第二层哈希函数 h_2 则采用 TCFL (由 Traffic Class 和 Flow Label 拼接而成) 作为输入。TCFL 生成过程可表示为：

$$TCFL = (TC \ll 20) | FL \quad (2)$$

其中，TC 是 8 位 Traffic Class 字段，FL 是 20 位 Flow Label 字段，“ \ll ”表示左移操作，“|”表示按位或操作。

图 4 展示了 H-CMS 的流量统计流程。首先，卫星节点捕获经过的 IPv6 数据包，提取 Traffic Class 和 Flow Label 字段。然后，将这两部分组合生成 TCFL 键值。系统随后通过两级哈希计算：先基于 TC 值确定该流量的服务优先级类别，再利用完整 TCFL 定位到具体计数器位置，每当匹配到相应数据包时，该位置计数值增加 1，从而记录各类流量的数据包数量。

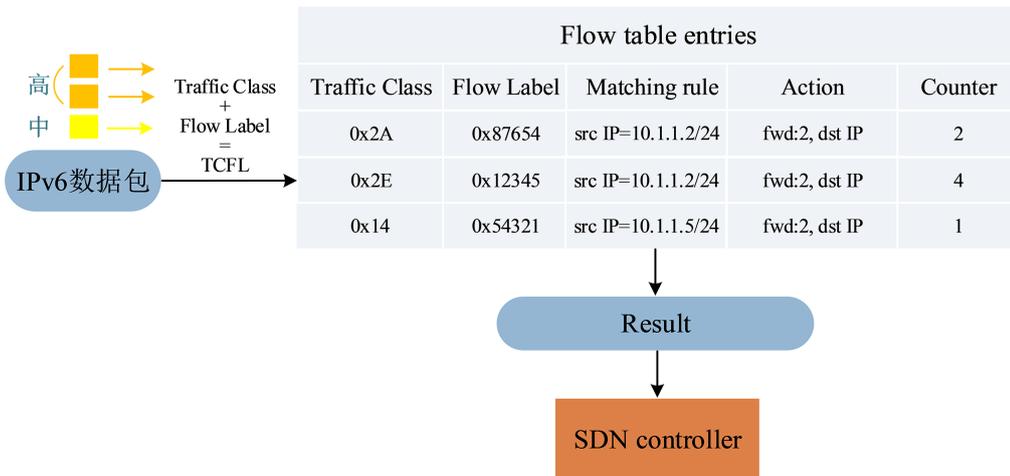


Figure 4. Traffic statistics process of H-CMS

图 4. H-CMS 的流量统计过程

在分级服务标识和分层流量感知的基础上，我们设计了差异化资源动态分配机制，针对各等级业务特性合理配置网络资源。该机制包含三个核心策略：差异化采样、智能带宽分配和灵活路径调度。采样策略方面，系统根据 DSCP 值动态调整采集频率：对紧急类通信(DSCP 值不低于 48)实施全量采样(100%)；高级别业务(DSCP 介于 32 至 47)采用 75%采样；中等优先级流量(DSCP 范围 16 至 31)降至 50%监测；而

普通业务(DSCP 值不超过 15)仅保留 25%采样率。资源分配上,实现了自适应带宽管理:当检测到链路负载轻微(利用率低于七成)时,仅对少量(10%)数据包执行完整统计以节省处理开销;一旦识别到重要业务(DSCP ≥ 32),立即转为全采样模式确保服务质量;此外,当 ECN 位指示网络拥塞时,系统自动提升所有流量的监测力度,快速重新安排传输路径,防止服务性能下降。

4.3. 深度强化学习路由决策

4.3.1. 路由决策问题建模

在 LEO 卫星网络中,由于卫星高速移动,拓扑结构呈现动态变化,同时不同业务需求对路径选择的要求不同。传统的静态路由由协议难以适应这种环境。而深度强化学习(Deep Reinforcement Learning, DRL)方法通过连续交互和自适应学习,可根据当前网络状态动态调整路由策略,实现负载均衡和服务质量优化。DRL 主要通过让智能体与环境互动来自主学习最优策略,包含三个基本要素:状态(State)、动作(Action)与奖励(Reward)。深度确定性策略梯度(Deep Deterministic Policy Gradient, DDPG)是一种经典的 DRL 算法,主要由 Actor 网络和 Critic 网络组成。然而,DDPG 存在 Q 值高估和训练不稳定的问题。针对这些问题,本文采用双延迟深度确定性策略梯度(Twin Delayed Deep Deterministic Policy Gradient, TD3)算法,该算法在保留原有架构的基础上引入了三项关键改进。

首先,设计了双重 Critic 网络机制,使用两个独立的 Critic 网络评估同一动作,取两者较小值避免 Q 值高估:

$$Q_{target} = r + \gamma \cdot \min(Q_{\theta_1}(s', a'), Q_{\theta_2}(s', a')) \quad (3)$$

其中 θ_1 和 θ_2 是两个 Critic 网络的参数, γ 是折扣因子,用于平衡短期奖励和长期收益。

其次,提出了策略延迟更新机制,先让价值网络更新 d 次后,再更新策略网络,提升训练稳定性,策略梯度计算如下:

$$\nabla_{\theta} J \approx E[\nabla_a Q_{\theta_1}(s, a) \nabla_{\theta} \mu(s)] \quad (4)$$

最后,在生成目标动作时引入高斯噪声,使策略对 Q 值误差不过于敏感,提高决策稳定性:

$$a' = \mu_{\theta'}(s') + \epsilon, \epsilon \sim N(0, \sigma) \quad (5)$$

TD3 方案中的状态空间设计(state)是智能体感知网络环境的关键基础,定义为:

$$S_t = \{B_{rem}, L_{delay}, C_{level}, Q_{tc}, F_{label}\} \quad (6)$$

其中, B_{rem} 表示链路的剩余带宽, L_{delay} 表示当前链路的时延, C_{level} 表示链路的拥塞程度, Q_{tc} 表示链路的服务等级, F_{label} 表示唯一的流标识。

动作空间(action)定义了智能体可执行的路由决策,设计为:

$$A_t = \{w_1, w_2, \dots, w_n\} \quad (7)$$

其中, w_i 表示分配给第 i 条路径的权重系数,满足两个基本约束: $0 \leq w_i \leq 1$ 且 $\sum w_i = 1$ 。

针对卫星网络多优先级业务的特点,本文设计了动态权重多目标奖励函数(reward),综合考虑带宽利用率、传输时延与拥塞等级三项指标:

$$R = \alpha \cdot \frac{B_{available}}{B_{total}} + \beta \cdot \frac{1}{D+1} - \gamma \cdot C \quad (8)$$

其中, $B_{available}$ 表示链路剩余带宽,由 H-CMS 实时统计, D 表示端到端传输时延, C 表示拥塞等级(归一

化到[0, 1]区间), α, β, γ 表示动态权重系数, 根据业务 DSCP 值自动调整。

4.3.2. 基于 TD3 的智能路由决策

在设计完状态空间、动作空间与奖励函数后, 本节详细阐述 TD3 算法支持下的路由决策方法。智能体与网络环境交互的整体流程如图 5 所示: 首先根据网络当前状态确定合适的路由分配方案; 执行决策后, 记录环境反馈的奖励值与新的网络状态; 随后将这些交互数据(状态, 动作, 奖励, 新状态四元组)存入优先级回放池, 支持系统离线学习; 接着从回放池中抽取样本, 更新双 Critic 网络评估动作质量, 并按延迟更新机制优化 Actor 策略网络; 同时通过软更新机制同步目标网络参数, 确保训练过程稳定可靠。

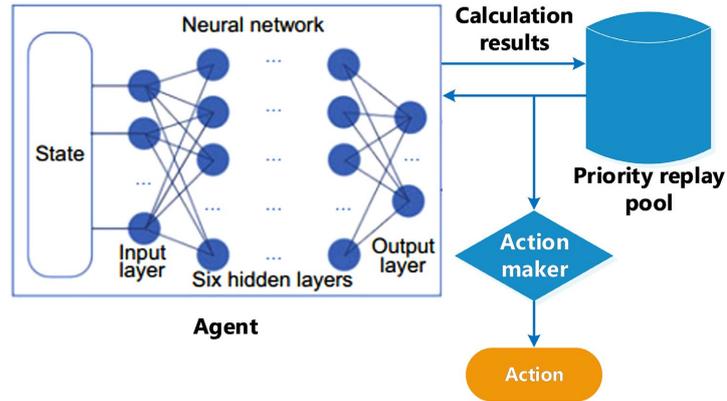


Figure 5. TD3 routing decision process
图 5. TD3 的路由决策过程

为提升算法收敛速度与策略稳定性, 本文设计以下训练机制:

1) 优先级经验回放: 存储经验数据时, 根据业务等级设置优先级。紧急业务经验采样概率提升 3 倍。加速关键场景学习。优先级计算公式为:

$$P(i) = \frac{(R_i + \epsilon)^\eta}{\sum_j (R_j + \epsilon)^\eta} \quad (9)$$

其中, $\epsilon = 0.01$ 防止零概率, $\eta = 0.6$ 控制优先级强度。

2) 动态探索率衰减: 初始探索率 $\epsilon = 1.0$, 随训练轮数线性衰减至 $\epsilon = 0.1$ 。衰减公式如下:

$$\epsilon = \max\left(0.1, 1.0 - 0.9 \times \frac{\text{当前轮数}}{\text{总轮数}}\right) \quad (10)$$

3) 目标网络软更新: 目标网络参数 θ' 按比例 $\tau = 0.005$ 逐步同步在线网络参数 θ :

$$\theta' \leftarrow \tau\theta + (1 - \tau)\theta' \quad (11)$$

5. 实验评估

5.1. 实验流程

为验证方案有效性, 我们搭建了完整的卫星网络模拟环境, 包括拓扑构建、网络模拟、流量生成和算法实现四个组成部分。卫星星座参数如表 2 所示。在网络模拟方面, 我们基于 Mininet 构建了支持 IPv6 的软件定义网络环境, 每颗卫星抽象为一个网络节点, 根据 STK 生成的数据动态调整链路状态。使用 D-ITG 工具生成多种优先级业务流量。通过该工具, 模拟远程医疗、视频会议、文件传输和网页浏览等多

种业务类型。每类业务的流量参数(如数据包大小、发送间隔和持续时间等)根据实际应用场景设置,具体参数可见表 3。

在算法实现部分,我们使用 PyTorch 框架实现了 TD3 强化学习算法,并将其部署到 SDN 控制器中。SDN 控制器选用了开源 ONOS (Open Network Operating System)平台,负责收集网络状态并下发路由策略。同时,我们将 H-CMS 模块集成到卫星节点上,用于统计流量和分类数据,为 TD3 算法提供准确的状态信息。

整个实验分为初始化、训练和测试三个阶段。在初始化阶段,部署星座拓扑,设置了链路带宽和时延参数,并加载业务流量模型。训练阶段中,TD3 算法通过 H-CMS 模块实时采集链路状态信息,例如剩余带宽、传输时延和拥塞程度。系统每隔 10 秒更新一次路由策略,同时记录算法的奖励值和收敛情况,以评估算法的性能。测试阶段,我们注入突发流量(例如将高优先级视频流的比例提高到 40%),以验证动态路由策略的适应性。实验将 H-CMS 与 CMS、基于 REST API 的 SDN 控制器在数据传输量和运行时间两方面进行了比较,将 TD3 算法与传统 OSPF 算法、Round-robin 方案和 HSR-CC 方案在吞吐量、时延和负载均衡等方面进行比较。

Table 2. Satellite constellation parameters
表 2. 卫星星座参数

参数	值
卫星星座	Iridium 星座
轨道高度	780 km
卫星数量	66 颗
单轨道卫星数	11 颗
星间链路带宽	100Mbps
单向链路延迟	约 10-40 ms

Table 3. Simulation traffic parameters
表 3. 仿真流量参数

业务类型	占比	数据包大小	优先级	Qos 要求
远程医疗	10%	1~5 KB	紧急	时延 ≤ 20 ms
视频会议	30%	500~800 KB	高	带宽 ≥ 50 Mb/s
文件同步	40%	1~10 MB	中	带宽保障
后台下载	20%	10~100 MB	低	尽力而为

5.2. H-CMS 优化效果分析

为评估本文提出的 H-CMS 在流量测量方面的表现,我们对比了 H-CMS、传统 CMS 及基于 SDN 控制器的 REST API 三种测量方法。

图 6 直观展示了各方法在网络规模变化时的数据处理能力对比。从图表可见,随着网络规模扩展,REST API 方法的性能出现明显衰减,总体降幅达 21%。传统 CMS 方法基本稳定在 34 Mb 附近。这是因为 CMS 采用固定大小的哈希表进行流量统计,其性能受网络规模影响较小。然而,传统 CMS 方法虽然稳定,但由于其单一哈希结构设计,无法针对不同服务类型进行精细化统计,导致其传输效率无法随着交换机数量的增加而提升。与上述两种方法相比,H-CMS 展示了卓越的性能表现。当交换机数量从 50 增至 300 时,其处理数据量不降反增。这得益于 H-CMS 采用的分层哈希结构设计,有效减少了哈希冲突,

提高了分类精度。同时, H-CMS 针对 IPv6 头部特征优化的多项式哈希函数进一步增强了计算效率, 使其在交换机数量增加的情况下仍保持较高的传输性能。此外, 当交换机数量达到 300 时, H-CMS 的传输数据量比 Controller REST API 方法高出约 49.6%, 比传统 CMS 方法高出约 14.7%。表明在资源受限的卫星节点上, 该方案有较好的应用前景。

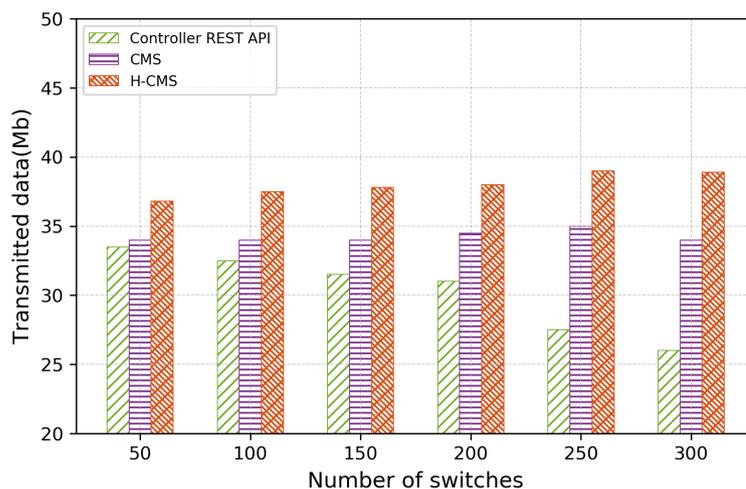


Figure 6. Performance comparison of three methods in data transmission
图 6. 三种方法在传输数据的性能比较

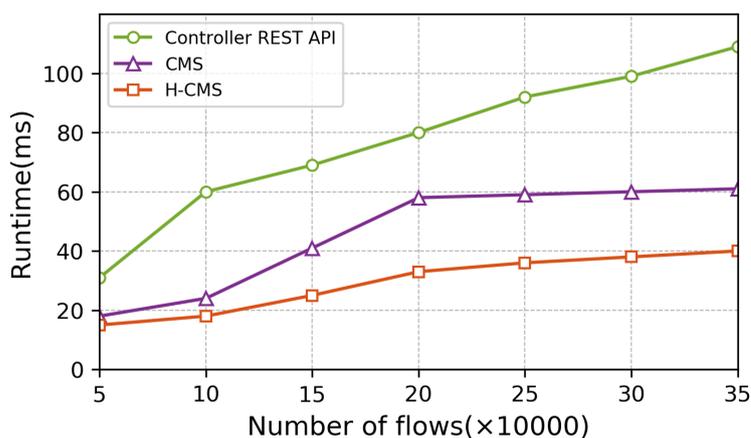


Figure 7. Running time of three methods under different numbers of flows
图 7. 三种方法在不同流数量下的运行时间

图 7 对比了三种流量测量技术在处理不同规模流量时的执行效率。从性能曲线看, 基于 REST API 的控制器方法随流量增加而显著降速, 处理 5 万条流时仅需 31 ms, 但在处理 35 万条流时耗时剧增至 110 ms 左右。这种急剧上升主要由两因素导致: 一是控制器集中式架构要求对每个流表项进行统计查询, 流量增加导致查询量与通信负担线性增长; 二是海量流表项会消耗交换机内存资源, 进一步拖慢查询响应速度。传统 CMS 算法在性能曲线上呈现两段式特征: 当流量从 5 万增至 20 万时, 处理时间从 19 ms 快速攀升至 58 ms; 而后续流量继续增加时, 执行时间趋于平稳, 维持在 60 ms 附近。而本文提出的 H-CMS 方法其处理时间增长最为缓和。与其他两种方法相比, H-CMS 在大规模流量场景下比传统 CMS 节省 33.3% 的执行时间, 比 REST API 方法节省高达 63.6%。

表 4 展示了不同流量感知方法的比较。与传统 CMS 相比, 本文 H-CMS 通过双层哈希结构降低了 15% 的哈希冲突率, 同时支持多优先级业务分类。与基于 REST API 的方法相比, H-CMS 处理相同规模流量时, 响应时间缩短 63.6%, 特别适合资源受限的卫星节点。相比于 NetFlow, H-CMS 内存占用降低 80%, 而测量精度仅降低 5%, 实现了资源效率与测量精度的最佳平衡。

Table 4. Comparison of traffic sensing methods

表 4. 流量感知方法比较

方法	内存占用	计算复杂度	测量精度	多业务支持	适合卫星场景
NetFlow	高	中	高	无	否 (资源消耗大)
传统 CMS	低	低	中 (哈希冲突)	无	部分 (不支持多业务)
REST API	低	高	高	无	否 (实时性差)
H-CMS	低	低	高	有	是 (轻量高效)

5.3. TD3 方案的性能分析

我们将本文提出的 TD3 方案与三种具有代表性的路由机制进行了全面性能比较。第一种是基于 Dijkstra 算法计算最小跳数路径的 OSPF 协议, 它作为传统静态路由的代表, 依靠周期性拓扑更新和链路状态数据库维护路由表, 具有实现简单、协议成熟的优势, 但在动态变化的卫星网络环境中, 其周期性更新机制导致信令开销大, 且对拓扑快速变化的适应性不足。第二种是按固定序列选择可用路径的 round-robin 算法, 该方法通过轮询方式分配流量, 实现了初步的负载均衡, 无需维护复杂的网络状态信息, 但缺乏对网络实时状态的感知能力, 难以针对不同优先级业务提供差异化服务。第三种是利用深度 Q 网络动态调整拥塞窗口的 HSR-CC 算法, 该方法将深度强化学习引入卫星路由决策, 能够根据历史数据预测链路状态变化, 在一定程度上适应拓扑动态性, 但其离散动作空间设计限制了调控精度, 且单一 Q 网络结构容易导致价值高估问题, 影响决策稳定性。相比之下, 本文提出的 TD3 方案通过双 Critic 网络结构有效解决了 Q 值高估问题, 同时与分级服务标识和分层流量感知机制的协同优化, 能够针对不同优先级业务生成差异化路由策略, 更加适应卫星网络高动态、多业务的应用场景。

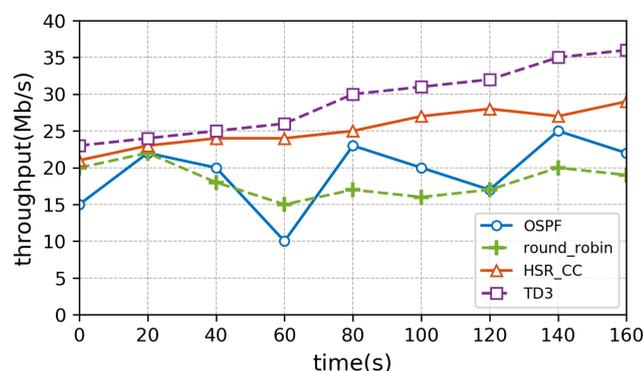


Figure 8. Comparison of throughput

图 8. 吞吐量的比较

图 8 呈现了四种路由算法在动态卫星网络环境中的数据传输能力对比。测试结果显示, TD3 方案在整个测试周期内展现出卓越性能, 传输能力从初始 23Mb/s 持续增强至测试结束时的 36 Mb/s。尤为突出的是, 在卫星切换关键时刻(60 秒处), 当其他算法出现明显性能波动时, TD3 保持了稳定的传输能力,

几乎不受拓扑变化影响。传统 OSPF 在卫星切换点表现最为脆弱，吞吐量骤降至 10 Mb/s。Round-robin 虽简单实用，但缺乏网络状态感知能力。HSR-CC 算法表现出一定的稳定性，但因未能充分发挥多路径传输潜能。

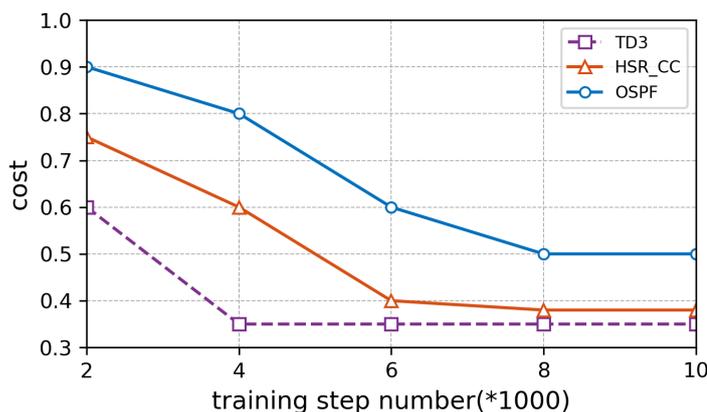


Figure 9. Convergence process of different methods

图 9. 不同方法的收敛过程

图 9 展示了 TD3、HSR-CC 和 OSPF 三种方法的收敛过程对比。结果表明，TD3 方案展示出最佳收敛性能，仅需 4000 步训练即达到 0.35 的低成本值并保持稳定。这种优势主要来自其双 Critic 网络结构抑制 Q 值高估，以及优先级回放机制加速关键场景学习。HSR-CC 虽也能逐步收敛，但需 6000 步才达到稳定状态，最终成本值 0.38 略高于 TD3。而 OSPF 作为传统算法虽然没有真正意义上的收敛过程，但是图中展示了它在不同训练步数下的性能表现，可以看出即使在 10,000 步后，其成本值仍维持在较高的 0.5 左右。

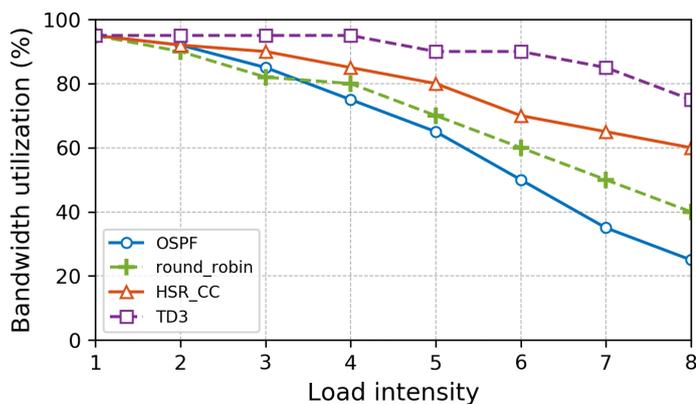


Figure 10. Comparison of bandwidth utilization under different load intensities

图 10. 不同负载强度下带宽利用率比较

图 10 展示了四种路由算法在不同负载强度下的带宽利用率对比。带宽利用率是网络性能评估的重要指标，反映了实际传输带宽与理论带宽的比值，关系到网络资源利用效率和拥塞程度。该比值越高，说明网络资源被更充分地利用，拥塞程度越低。结果显示，在低负载情况(负载强度为 1-2)下，四种算法的带宽利用率都接近 95%，效果相近。随着负载强度增加，TD3 算法整体表现在较高的带宽利用率，即使

在负载强度达到 6 时仍维持在 90%左右,到负载强度 8 时也能保持 75%以上的利用率。原因是 TD3 通过实时监测链路剩余带宽、时延和拥塞程度,并结合业务优先级标识,能够为不同类型的业务选择最适合的传输路径。相比之下,OSPF 算法表现最差,在负载强度增至 8 时,带宽利用率急剧下降至约 25%,这主要是因为 OSPF 根据单一路径选择导致某些链路过载而其他链路闲置。Round-robin 虽然采用了多路径策略,但由于缺乏对网络状态的感知,其带宽利用率在高负载下仍降至 40%左右。HSR-CC 表现优于基础算法,但是在负载强度 8 时仍降至 60%,效果不及 TD3。

综上,无论是在动态拓扑环境下的吞吐量稳定性,还是训练过程中的快速收敛,以及高负载条件下的带宽利用率,TD3 均明显优于传统 OSPF、简单负载均衡的 round-robin 和基于强化学习的 HSR-CC 算法。尤其是,TD3 在卫星切换点表现稳定以及在高负载下资源利用率仍较高,表现出结合了业务优先级与网络状态的智能路由决策机制,可以为卫星网络中的多业务差异化传输提供了可靠解决方案。

见表 5,相比于 DQN 算法,本文采用的 TD3 算法能够处理连续动作空间,这一特性使其更适合卫星网络中的动态路径权重调整。在卫星网络环境中,链路状态呈现连续变化特性,如带宽波动、时延变化等,DQN 的离散动作设计难以精细捕捉这种连续变化,导致其决策粒度受限。而 TD3 算法通过连续动作空间可以实现更精细的路由权重调整,使路由决策更加平滑自然。与 DDPG 相比,TD3 通过双 Critic 网络架构有效解决了 Q 值高估问题。在卫星网络这种状态空间复杂且噪声较大的环境中,DDPG 容易因单一 Critic 网络导致的值函数高估而产生不稳定训练过程和次优策略。TD3 算法引入的双重价值评估机制,通过取两个 Critic 网络输出的较小值作为最终评估结果,有效抑制了乐观估计偏差,提高了训练稳定性。此外,TD3 的延迟策略更新和目标策略平滑机制,使其在卫星网络高动态环境中具有更好的噪声鲁棒性。相较于 A3C 算法,TD3 虽然均表现出较快的收敛速度,但实现机制有显著差异。A3C 依赖多线程并行训练来加速收敛,这在计算资源有限的卫星环境中实现困难;而 TD3 通过优化的网络结构和训练机制在单线程环境中也能实现快速收敛。此外,TD3 中的目标网络软更新机制使其在学习过程中表现出更稳定的策略演化,避免了 A3C 可能出现的策略振荡问题。在卫星网络的高负载场景下,TD3 算法展现出更优秀的路由调度能力,这主要归功于其基于动态权重的多目标奖励函数设计,能够根据不同业务的 DSCP 优先级自适应调整奖励权重,使路由决策既考虑网络性能又兼顾业务差异化需求。

Table 5. Comparison of deep reinforcement learning algorithms

表 5. 深度强化学习算法比较

方法	决策空间	网络结构	稳定性	收敛速度	卫星场景适应性
DQN	离散	单 Q 网络	较差	中等	低
A3C	连续/离散	多角色网络	良好	快	中
DDPG	连续	单 Actor-Critic	一般	较快	中
TD3	连续	双 Critic	较优	快	高

6. 结语

本章提出了结合分级服务标识与深度强化学习的卫星网络路由优化方案。该方案巧妙利用 IPv6 头部的 Traffic Class 和 Flow Label 字段实现零开销服务标识,并通过创新的分层 H-CMS 结构精确测量多类型业务流量。核心决策机制采用 TD3 算法,动态生成适应网络变化的路由策略。实验验证了该方案的显著优势:TD3 仅需 4000 步训练即达稳定状态;网络吞吐量提升 8%且在卫星切换时保持稳定;高负载下带宽利用率维持在 75%以上;而 H-CMS 使测量误差降低 15%。这一方案有效解决了卫星网络中因用户分布和业务需求差异导致的资源利用不均问题。

参考文献

- [1] Liu, P., Liang, J., Qian, X. and Xu, Y. (2021) Research on Link Load-Balancing between Marine LEO Satellite Constellations Based on SDN. 2021 *2nd International Conference on Electronics, Communications and Information Technology (CECIT)*, Sanya, 27-29 December 2021, 977-982. <https://doi.org/10.1109/cecit53797.2021.00175>
- [2] Jin, L., Wang, L., Jin, X., Zhu, J., Duan, K. and Li, Z. (2022) Research on the Application of LEO Satellite in IoT. 2022 *IEEE 2nd International Conference on Electronic Technology, Communication and Information (ICETCI)*, Changchun, 27-29 May 2022, 739-741. <https://doi.org/10.1109/icetci55101.2022.9832117>
- [3] Aulia, M.A., Sukmandhani, A.A. and Ohliati, J. (2022) RIP and OSPF Routing Protocol Analysis on Defined Network Software. 2022 *International Electronics Symposium (IES)*, Surabaya, 9-11 August 2022, 393-397. <https://doi.org/10.1109/ies55876.2022.9888355>
- [4] Gao, K., Xu, C., Qin, J., Yang, S., Zhong, L. and Muntean, G. (2019) QoS-Driven Path Selection for MPTCP: A Scalable SDN-Assisted Approach. 2019 *IEEE Wireless Communications and Networking Conference (WCNC)*, Marrakesh, 15-19 April 2019, 1-6. <https://doi.org/10.1109/wcnc.2019.8885585>
- [5] Liu, D., Zhang, J., Cui, J., Ng, S., Maunder, R.G. and Hanzo, L. (2022) Deep Learning Aided Routing for Space-Air-Ground Integrated Networks Relying on Real Satellite, Flight, and Shipping Data. *IEEE Wireless Communications*, **29**, 177-184. <https://doi.org/10.1109/mwc.003.2100393>
- [6] Yang, S., Li, H. and Wu, Q. (2018) Performance Analysis of QUIC Protocol in Integrated Satellites and Terrestrial Networks. 2018 *14th International Wireless Communications & Mobile Computing Conference (IWCMC)*, Limassol, 25-29 June 2018, 1425-1430. <https://doi.org/10.1109/iwcmc.2018.8450388>
- [7] Shi, H., Zhang, L., Zuo, X., Wu, Q., Li, H. and Cui, Y. (2021) Multipath Deadline-Aware Transport Proxy for Space Network. *IEEE Internet Computing*, **25**, 51-57. <https://doi.org/10.1109/mic.2021.3112804>
- [8] Han, C., Huo, L., Tong, X., Wang, H. and Liu, X. (2020) Spatial Anti-Jamming Scheme for Internet of Satellites Based on the Deep Reinforcement Learning and Stackelberg Game. *IEEE Transactions on Vehicular Technology*, **69**, 5331-5342. <https://doi.org/10.1109/tvt.2020.2982672>
- [9] Liu, J., Zhao, B., Xin, Q., Su, J. and Ou, W. (2021) DRL-ER: An Intelligent Energy-Aware Routing Protocol with Guaranteed Delay Bounds in Satellite Mega-Constellations. *IEEE Transactions on Network Science and Engineering*, **8**, 2872-2884. <https://doi.org/10.1109/tNSE.2020.3039499>
- [10] Wang, F., Jiang, D., Qi, S. and Qiao, C. (2020) An Adaboost Based Link Planning Scheme in Space-Air-Ground Integrated Networks. *Mobile Networks and Applications*, **26**, 669-680. <https://doi.org/10.1007/s11036-019-01422-4>