



XXXX

基于深度强化学习的低轨卫星网络算力路由优化方法

蔡佳慧, 周家恩, 赵亚飞, 彭木根

(北京邮电大学网络与交换技术全国重点实验室, 北京, 100876)

摘要: 针对低轨卫星网络中拓扑动态、链路时变及星上算力异构等挑战, 本文提出了一种基于传算协同优化的双重深度Q网络 (Computing - Transmission Double Deep Q-Network, CTDDQN) 算力路由方法。该方法在软件定义网络架构下由控制器获取全局拓扑与资源信息, 并在控制器侧进行集中式传算联合决策, 构建以端到端时延最小化为目标的传输 - 计算联合优化框架, 并结合K-hop前瞻机制实现路由与计算决策的协同。仿真结果表明, 在不同星座规模、链路带宽与算力配置下, 与多种基线算法相比, 所提方法可使平均端到端时延降低18.9%以上, 验证了该方法在高动态低轨卫星网络中的有效性与可扩展性。

关键词: 低轨卫星网络; 深度强化学习; 算力路由; 时延优化

中图分类号: TN927

文献标志码: A

doi: 10.11959/j.issn.1000-0801.

Deep Reinforcement Learning-Based Computing-Aware Routing Optimization Method for Low Earth Orbit Satellite Networks

CAI Jiahui, ZHOU Jiaen, ZHAO Yafei, PENG Mugen

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

Abstract: To address the computing-aware routing problem in Low-Earth-Orbit (LEO) satellite networks, which faces highly dynamic topology, time-varying inter-satellite links, and heterogeneous onboard computing capabilities, this paper proposes a Computing - Transmission Double Deep Q-Network (CTDDQN)-based computing-aware routing method. Under a Software-Defined Networking (SDN) architecture, the controller acquires global topology and resource information and performs centralized joint transmission - computing decision-making. Based on this, we formulate an integrated transmission-computing optimization framework that minimizes end-to-end delay, and further incorporate a K-hop lookahead mechanism to coordinate routing and onboard computing decisions. Simulation results under different constellation scales, link bandwidths, and onboard computing configurations show that, compared with multiple baseline methods, the proposed method reduces the average end-to-end delay by over 18.9%, demon-

收稿日期: XXXX-XX-XX; 修回日期: XXXX-XX-XX

基金项目: 国防科技重点实验室稳定支持经费项目项目 (No. WDZC20265290408)

Foundation Items: Stability Support Funding Program of National Defense Key Laboratory (No. WDZC20265290408)



strating its effectiveness and scalability in highly dynamic LEO satellite networks.

Key words: Low Earth Orbit satellite network, deep reinforcement learning, computing-aware routing, delay optimization

1 引言

随着全域覆盖与泛在服务需求的不断提升,面向6G的移动通信网络正由传统的信息传输平台加速演进为集通信、计算与服务于一体的综合信息基础设施^[1]。在此背景下,计算能力逐渐成为网络内生资源,并与接入、传输等过程深度耦合,推动网络供给模式由单一传输向传算融合转变^[2]。低轨卫星(Low Earth Orbit, LEO)网络因具备广域覆盖、灵活部署和天地一体协同等特点,被视为支撑“云一边一端”协同服务体系的重要基础设施^[3]。然而,在大规模星座场景下,业务性能不仅受链路质量与路由决策影响,还与星上计算资源分布、任务动态到达及网络拓扑时变特性密切相关^[4]。因此,面向动态拓扑和多维资源耦合约束,开展传输路径选择与计算资源分配的联合优化研究,对于提升LEO网络的业务承载能力与服务效率具有重要意义。

LEO卫星轨道具有可预测性,面向传输的动态路由策略通过转发数据前的全局预测来寻找最优路径。然而,算力路由需要先将数据传输至计算节点进行处理,再转发至目标卫星以完成服务。因此,最优算力路由需要求解“计算+路由”的组合最优,若采用传统全局预测方案,需遍历每个计算节点并分别求解最优路由。再考虑卫星网络的拓扑动态性^[5],导致算力路由的选取复杂度远高于面向纯通信的动态路由。若忽略传输路径与计算位置之间的耦合关系,仅进行分离式优化,往往难以在动态拓扑和异构资源约束下获得稳定有效的全局决策^{[6][7]}。因此,如何面向低轨卫星网络的时变拓扑、分布式算力和随机业务到达特性,联合优化传输路径选择与计算节点选择

的组合优化问题,是传算融合网络中亟待解决的关键问题^[8]。

针对LEO网络拓扑高速时变的基础问题,研究者首先聚焦于动态路由建模与自适应转发机制的研究。文献[9]基于随机几何方法刻画了大规模LEO网络的时延特性,并对不同路由策略下的时延表现进行了比较分析;在此基础上,文献[10]面向软件定义网络(Software Defined Networking, SDN)架构,提出了一种基于星间链路(Inter-Satellite Link, ISL)属性的动态路由方法,以提升对链路频繁变化场景的适应能力。进一步地,文献[11]构建了时空演化图模型,显式描述ISL属性、拓扑演化与路由权重之间的耦合关系,从而提高路由决策质量。除路径可达性与链路状态外,部分研究开始关注网络运行状态对路由性能的影响,例如文献[12]从拥塞控制视角出发,指出路由决策还应联合考虑ISL负载与队列时延等网络态势信息。与此同时,文献[13]引入Q-learning实现分布式路由,使卫星能够基于局部信息自适应更新转发策略,进一步增强了动态环境下的在线决策能力。然而,上述研究主要围绕高动态LEO网络中的链路感知与路径自适应展开,虽提升路由对拓扑变化的响应能力,但未将计算过程引起的数据规模变化纳入路由决策机制。

近年来,随着研究视角扩展到任务处理全过程,部分工作开始将任务卸载、资源分配与链路动态纳入分析框架。文献[14]面向卫星边缘计算场景提出任务卸载策略,将云一边一星垂直协同纳入考虑;在此基础上,文献[15]构建端到端卫星-地面任务场景下的联合卸载、路由与资源分配模型;文献[16]研究卸载与传输功率分配的联

合优化,在时延最小化的同时考虑能量约束;文献[17]面向任务优先级与时限约束,采用多智能体强化学习实现卸载调度;文献[18]将服务链缓存与任务卸载联合优化,实现时延与能耗折中;文献[19]提出了适用于星地融合网络的近端策略优化(Proximal Policy Optimization, PPO)卸载框架,并验证了其在复杂决策条件下的有效性。进一步地,文献[20]提出计算感知路由方法,将边传边算过程建模为动态最短路径集合,并通过近似求解降低端到端时延;基于此,文献[21]设计了分层多智能体深度强化学习(Deep Reinforcement Learning, DRL)框架,在多域软件定义卫星网络中联合规划路由与计算/网络功能选择,以提升多维资源协同决策能力。在此基础上,文献[22]将路径选择、算力节点选择与资源调度统一建模为马尔可夫过程,并提出多智能体深度Q网络算力路由方法,以满足多约束服务质量需求。文献[23]引入异构任务效用模型,结合DRL与逐次凸近似方法,对任务卸载与资源分配进行联合优化。然而,面对强动态与多约束环境,上述工作仍难以解决状态维度高、可扩展性不足及推理开销大等问题。

为解决上述问题,本文基于SDN架构,由控制器获取全局拓扑与星历信息,针对强时变拓扑与多约束耦合环境中问题的组合优化特征,将控制器侧的传算决策建模为马尔可夫决策过程(Markov Decision Process, MDP),将传统级联决策简化为并行动作,以支持快速推理与在线决策,主要贡献如下:

- 构建面向LEO传算融合网络的动态图建模框架,将动态星上拓扑与多维资源映射到统一框架,为后续优化提供基础模型;

- 将星上算力路由建模为序贯决策问题,并在单步决策中将路由选择与计算节点选择联合表示为并行动作,提出一种基于DDQN的算力路由方案,以实现动态环境下传输与计算决策的协同

优化;

- 通过在3种不同规模的卫星网络拓扑上开展仿真实验评估算法性能,并与包含DQN、PPO在内的5种基线算法进行对比分析,验证算法有效性。

本文其余部分组织如下,第2章介绍系统模型与连续时间动态图表征;第3章构建马尔可夫决策过程并介绍算法实现原理;第4章针对仿真实验进行对比分析;第5章对研究工作进行总结与展望。

2 系统模型与问题建模

2.1 网络场景

如图1所示,考虑由 N 颗卫星组成低轨卫星星座,记卫星集合为 $S=\{s_1, s_2, \dots, s_N\}$ 。所有卫星均匀分布在高度为 h 、倾角为 I 的轨道面上,轨道集合记为 $P=\{p_1, p_2, \dots, p_P\}$ 。星座组网采用一星四链网络拓扑结构:每颗卫星与同轨道前后相邻卫星相连,并与相邻轨道中对应位置的卫星建立交轨道链路。ISL的建立与维持受视距与星间距离约束,仅在满足可见性和业务需求的前提下开放。

地面用户终端集合为 $G=\{g_1, g_2, \dots, g_M\}$,其中 M 为地球表面用户终端的数量。在任意时刻 t ,每个请求对应一对源端与目的端(g_{src}, g_{dst}),源终端 g_{src} 通过上行星地链路(Satellite-Ground Link, SGL)直接接入卫星网络,在星座内经多跳ISL转发后,由与目标终端 g_{dst} 可见的卫星通过下行SGL下传至 g_{dst} 。星地接入遵循可见性-仰角约束与最近邻原则:在满足视距与仰角门限的可见卫星候选集中,地面终端 g 选择与自身几何斜距最近的一颗卫星作为接入卫星,并建立上下行SGL链路。

在计算能力配置方面,本文聚焦于星上遥感图像压缩与预处理类业务场景。在该类任务中,

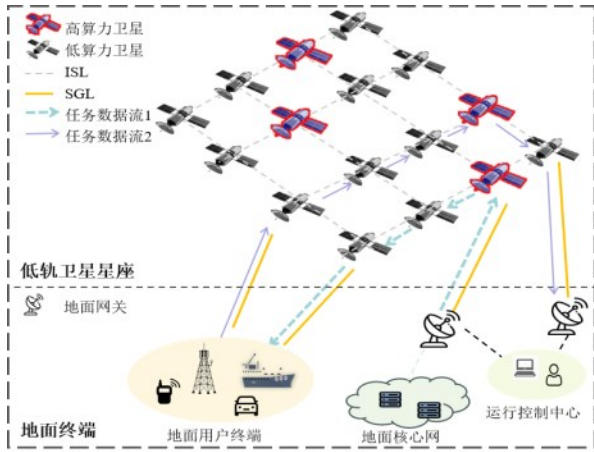


图1 低轨卫星传算融合网络系统结构

业务可先在卫星侧完成一次计算处理，再将压缩后的图像数据或关键信息沿后续路径转发，从而降低后续传输开销。为突出传输路径选择与计算位置选择之间的耦合关系，本文假设业务计算均在卫星侧完成，暂不考虑必须回传原始数据或处理后数据规模近似不变的业务类型。即：

$$F_g = 0, \forall g \in \mathcal{G} \quad (1)$$

每颗卫星均配置一个在轨计算单元，作为网络中的计算节点。为刻画星座内算力资源的异构性，假设 $\forall s \in S$ 的可用算力 F_s 服从双峰分布：

$$F_s \in \{F_{\text{high}}, F_{\text{low}}\}, P(F_s = F_{\text{high}}) = \rho, P(F_s = F_{\text{low}}) = 1 - \rho \quad (2)$$

其中 $F_{\text{high}} > F_{\text{low}}$ ， ρ 表示高算力卫星占比，后续算力路由决策需考虑上述算力异构特性。

为统一刻画拓扑与资源的动态行为，本文将星地传算融合网络建模为一系列连续时间动态图 $G(t) = (V, E(t), W_E(t))$ ，其中 V 为虚拟节点 (Virtual-Node, VN) 集合； $E(t)$ 为时刻 t 网络中所有可用链路的集合，包括 ISL 与 SGL； $W_E(t)$ 为边权函数，用于综合描述链路在时刻 t 的传播时延、可用带宽、排队状态等性能指标。如图2所示，构建虚拟节点系统 (Virtual-Node System, VNS)，将地球表面按照经纬度静态划分为若干互不重叠区域，每个区域对应一个虚拟节点 (Virtual Node,

VN)。当卫星的地面投影落入某个区域且满足可见性与仰角约束时，该卫星被纳入该区域的候选接入卫星集合；随着卫星运行，卫星一区域候选关系随时间动态更新。在大多数时刻，每个区域仅与一颗卫星相关联；在少数边界或重叠情况下，本文采用最近星下点选择机制，确保每个区域与一颗卫星的唯一映射。地面用户产生的业务请求映射到所属区域的虚拟节点，并通过与其关联的接入卫星建立 SGL，进入星座网络完成后续的传输与计算过程。

在该建模下，路由与算力分配在 VNS 上完成，而其决策由映射到的卫星节点及 ISL/SGL 链路承载执行，从而将星地传输与在轨计算统一纳入同一动态图框架进行建模与优化。

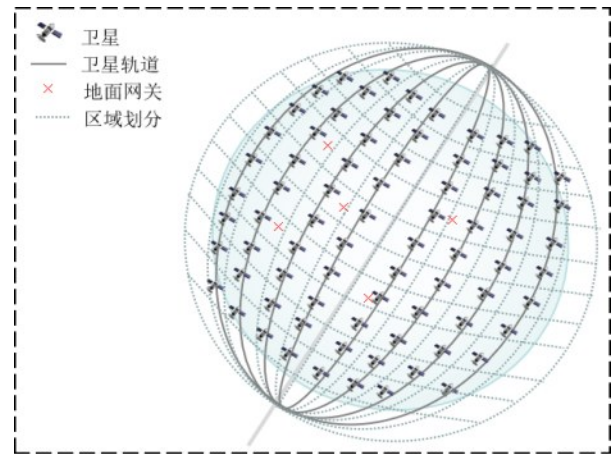


图2 地面区域划分与虚拟节点映射示意图

2.2 问题建模

本节对星地传算融合网络中的端到端时延进行建模。考虑业务请求 w ，给定源节点 $g_{\text{src}} \in \mathcal{G}$ 、目的节点 $g_{\text{dst}} \in \mathcal{G}$ ，数据量 L_w ，计算负载 C_w ，计算可在路径上的某一卫星节点 v_c 执行。

(1) 单跳链路时延模型

对于任意时刻 t ，记链路 (i, j) 的可用带宽与几何距离分别为 $r_{w,ij}(t)$ 和 $D_{ij}(t)$ 。业务 w 在链路 (i, j) 上的单跳时延 $T_{i,j}(L_w, t)$ 由传输时延 $T_{i,j}^{\text{trans}}(L_w, t)$ 、传播时延 $T_{i,j}^{\text{prop}}(L_w, t)$ 与排队时延 $T_{i,j}^{\text{queue}}(L_w, t)$

构成:

$$T_{i,j}(L_w, t) = T_{i,j}^{\text{trans}}(L_w, t) + T_{i,j}^{\text{prop}}(L_w, t) + T_{i,j}^{\text{queue}}(L_w, t) = \frac{L_w}{r_{w,ij}(t)} + \frac{D_{ij}(t)}{c} + T_{i,j}^{\text{queue}}(L_w, t) \quad (3)$$

其中 L_w 为当前需要传输的数据量, c 为电磁波在自由空间中的传播速度, 取光速 $3 \times 10^8 \text{m/s}$ 。

(2) 计算处理时延模型

计算处理用于刻画业务在星上完成一次处理流程所消耗的时间, 其作用是在占用一定星上算力的同时降低后续需传输的数据量或提升信息有效性。设计算节点 v_c 的可用算力为 F_{v_c} , C_w 表示业务请求 w 的计算需求, 则业务请求 w 在节点 v_c 的计算处理时延为:

$$T_w^{\text{proc}}(v_c) = \frac{C_w}{F_{v_c}} \quad (4)$$

设 a_w 表示业务 w 经单次在轨计算处理后的数据压缩系数, 用于刻画输出数据量相对于输入数据量的缩减程度, 满足 $a_w \in (0, 1]$ 。不同业务的 a_w 由其处理类型决定: 图像压缩类业务经处理后通常仍保留较完整的数据内容, 因此 a_w 相对较大; 而目标检测、特征提取等语义提炼类业务仅保留关键信息, 因而 a_w 通常较小。参考相关遥感图像在轨压缩研究, 典型压缩比通常位于 2:1 到 10:1 之间^[24]。为聚焦传输路径选择与计算触发时机之间的耦合关系, 本文将业务统一抽象为同构的星上遥感图像压缩预处理任务, 并在同构业务假设下取 $a_w = 0.2$ 。当业务在 v_c 上完成处理后, 其输出数据量通常小于输入数据量, 则计算后单跳传输时延中的有效负载由 L_w 转变为 $a_w L_w$ 。据此, 计算后的链路时延 $T_{i,j}^{\text{comp}}(L_w, t)$ 为:

$$T_{i,j}^{\text{comp}}(L_w, t) = \frac{a_w L_w}{r_{w,ij}(t)} + \frac{D_{ij}(t)}{c} + T_{i,j}^{\text{queue}}(a_w L_w, t) \quad (5)$$

计算完成后的后续链路传输均以压缩后数据量 $a_w L_w$ 为有效负载。

(3) 端到端时延模型

记路径 $\pi_{g_{\text{src}} \rightarrow g_{\text{dst}}}$ 为源节点 g_{src} 与目的节点 g_{dst} 之间的路径, 若路径上选取节点 $v_c \in \pi_{g_{\text{src}} \rightarrow g_{\text{dst}}}$ 作为计算节点, 则可将路径划分为 $\pi_{g_{\text{src}} \rightarrow v_c}$ 与 $\pi_{v_c \rightarrow g_{\text{dst}}}$ 两段, 分别表示源节点到计算节点以及计算节点到目的节点的路径。当业务在路径中间节点 v_c 执行计算时, 其端到端时延为:

$$\tau_w(g_{\text{src}}, v_c, g_{\text{dst}}; t) = \sum_{(i,j) \in \pi_{g_{\text{src}} \rightarrow v_c}} T_{i,j}(L_w, t) + T_w^{\text{proc}}(v_c) + \sum_{(i,j) \in \pi_{v_c \rightarrow g_{\text{dst}}}} T_{i,j}^{\text{comp}}(L_w, t) \quad (6)$$

若业务在传输过程中不执行在轨计算, 即 $v_c = \emptyset$, 则端到端时延为:

$$\tau_w(g_{\text{src}}, \emptyset, g_{\text{dst}}; t) = \sum_{(i,j) \in \pi_{s \rightarrow d}} T_{i,j}(L_w, t) \quad (7)$$

(4) 优化目标与约束条件

在给定网络状态和业务参数的条件下, 联合选择路由路径 $\pi_{g_{\text{src}} \rightarrow g_{\text{dst}}}$ 与计算节点 v_c , 以最小化端到端时延:

$$\min_{\pi_{g_{\text{src}} \rightarrow g_{\text{dst}}}, v_c \in g_{\text{src}} \rightarrow g_{\text{dst}}} \tau_w(g_{\text{src}}, v_c, g_{\text{dst}}; t) \quad (8)$$

约束条件为:

$$C1: \tau_w(g_{\text{src}}, v_c, g_{\text{dst}}; t) \leq T^{\text{max}} \quad (9)$$

$$C2: 0 < r_{w,ij}(t) \leq R_{ij}(t), \forall (i,j) \in \pi_{g_{\text{src}} \rightarrow g_{\text{dst}}} \quad (10)$$

$$C3: T_{ij}(L_{\text{eff}}, t) \leq \Delta_{ij}(t), \forall (i,j) \in \pi_{g_{\text{src}} \rightarrow g_{\text{dst}}} \quad (11)$$

C1: 端到端时延约束。业务请求 w 的端到端时延不得超过其时延容忍阈值 T^{max} ;

C2: 链路带宽分配约束。对于任意链路 (i,j) , 分配给业务 w 的带宽 $r_{w,ij}(t)$ 必须为正且不超过链路可用带宽 $R_{ij}(t)$ 。

C3: 链路可见性约束。 L_{eff} 表示链路上传输的实际数据量, 该约束要求在业务传输期间链路剩余可见持续时间需覆盖该链路的传输时长。



3 CTDDQN 算法

3.1 马尔可夫决策过程建模

由第2节问题建模可知，低轨传算协同优化可视为在连续时间动态图上的计算感知路由问题。本文以数据包到达当前节点并触发下一跳选择或计算判断作为决策事件，将业务转发过程抽象为序贯决策，并建模为马尔可夫决策过程 $M = \langle S, A, P, R, \gamma \rangle$ ，包含状态空间 S 、动作空间 A 、状态转移概率 P 、奖励函数 R 以及折扣因子 γ 。为实

$$s(t) = [P_{v_s^{\text{LEO}}}(t), P_{v_d^{\text{LEO}}}(t), \Delta p(t), w(t), d(t), c_{\text{norm}}(t), \delta^c(t), \rho_t]$$

其中 $P_{v_s^{\text{LEO}}}(t)$ 、 $P_{v_d^{\text{LEO}}}(t)$ 分别为当前节点与目的节点的归一化网格位置； $\Delta p(t)$ 为当前位置与目标节点的几何相对位置； $w(t)$ 表示至邻居的单跳链路代价特征向量； $d(t)$ 为方向指示向量，用于指示向目的节点前进的优先方向； $c_{\text{norm}}(t)$ 表示当前节点可用算力的归一化值； $\delta^c(t)$ 表示当前数据包是否已完成一次在轨计算 ($\delta^c(t) = \{0, 1\}$)； ρ_t 表示当前数据相对原始数据的比例系数，初始化为1。随着星座规模扩大，若状态维度增大会提高Q网络输入层参数规模及前向、反向传播计算量，从而增加训练时间；同时，更高维状态对应更复杂的函数逼近空间，通常需要更多交互样本和训练轮次，并可能加剧训练波动与超参数敏感性。因此，本文采用固定长度编码构建状态向量，使状态维度 D_s 与星座规模 N 解耦，网络规模变化不会引起输入维度或模型参数规模随 N 线性增长。

3.1.2 动作空间

为实现路由与计算决策的深度耦合，本文将智能体在时刻 t 的决策定义为由路由动作 a^r 和计算动作 a^{comp} 构成的动作集，该动作集使策略在同一时间尺度上同时权衡链路代价与计算收益，并在统一回报下完成端到端优化：

□ 路由动作 a^r ：在当前节点 v_i 的可见邻居 $N_i(t)$ 集合中选择下一跳节点，实现下一跳的转发

现端到端协同，CTDDQN 将路由转发与计算触发统一为复合动作，并结合 K -hop 前瞻与合法动作约束在统一回报下显式评估计算对后续剩余代价与可行性的影响。下文分别给出状态、动作与奖励的具体设计。

3.1.1 状态空间

设时刻 t 当前数据包在卫星网络中传输路径源节点为 v_s^{LEO} ，目的节点为 v_d^{LEO} ，本文采用局部观测构建状态向量 $s(t)$ ，其由位置、链路、算力与业务等信息组成，定义为：

$$(12)$$

决策：

□ 计算动作 a^{comp} ：在计算可行的前提下，选择在当前节点或前瞻评估得到的 K -hop 候选节点上执行在轨计算，对应三种模式：S1：在当前节点就地计算；S2：前往候选点计算；S3：不执行计算。

综合考虑网络连通性与计算可行性，当前节点 v_i 在时刻 t 的动作空间可表示为：

$$A_i(t) = \left\{ a^r | f_k(i) \in N_i(t) \right\} \cup \left\{ a^{\text{comp}} | C_i(t) \neq \{\emptyset\}, \delta^c(t) = 0 \right\} \quad (13)$$

其中， $f_k(i)$ 将离散方向索引 k 映射为与节点 v_i 相邻的具体节点， V_{LEO} 为卫星节点集合。仅当数据尚未计算 ($\delta^c(t) = 0$) 且 K -hop 范围内存在可行计算节点时，执行计算动作并进一步触发 K -hop 前瞻在 S1/S2/S3 模式间进行比较。

当智能体选择动作 a^{comp} 时，CTDDQN 在计算分支中对 K -hop 跳候选集 $C_i(t)$ 执行前瞻评估，比较 S1/S2/S3 三种模式的端到端剩余时延。当在候选计算节点执行计算时，剩余估计时延如式 (15) 所示，可得到最优预约计算节点如式 (16) 所示。其中 a_w 为任务相关压缩比例， $\tilde{T}(u \rightarrow v; L_w)$ 表示从 u 到 v 的最短路径传输时延。

$$T_{\text{no}}(t) = \tilde{T}(v_i \rightarrow v_d; L_w) \quad (14)$$

$$T(c|t) = \tilde{T}(v_i \rightarrow c; L_w \rho_t) + T^{\text{comp}}(c, L_w) + \tilde{T}(c \rightarrow v_d; a_w L_w \rho_t) \quad (15)$$

$$c^* = \arg \min_{c \in C_i(t)} T(c|t) \quad (16)$$

若 $T(c|t) < T_{\text{no}}(t)$ 且 $c^* = v_i$, 在本地立即执行一次计算:

$$\rho_{t+1} = a_w \rho_t, \delta^c(t+1) = 1, \Delta t_{\text{comp}}(t) = T^{\text{comp}}(v_i^{\text{comp}}, t), v_i^{\text{comp}} \in \{v_i, c^*\} \quad (17)$$

若 $T(c|t) < T_{\text{no}}(t)$ 但 $c^* \neq v_i$, 后续路由阶段优先选择使到达预约节点 c^* 的估计代价下降的邻居, 当到达 c^* 且满足计算可行性时自动触发式 (17), 且 $c^* = i$ 。

若 $T(c|t) \geq T_{\text{no}}(t)$, 则忽略当前计算动作, 继续执行纯路由决策。

3.1.3 奖励函数

为避免路由与计算决策分别优化导致目标不一致, 本文对两类动作采用统一回报机制, 使智能体在同一价值尺度下比较转发与在轨计算的收益, 从而实现面向端到端时延最小化的一致性优化。即时奖励并非直接等于端到端时延, 而是通过分步时延惩罚、势能差奖励与终点奖励共同构造, 使累计回报与第2章的端到端时延最小化目标保持一致。即时奖励定义为:

$$r_t = -a_1 T_t^{\text{link}} - a_2 T_t^{\text{comp}} + a_3 (\Phi(s_t) - \Phi(s_{t+1})) + a_4 \mathbb{I}(v_{t+1} = v_d) - a_5 \Omega_t \quad (18)$$

其中, 当 a_t 为路由动作时, 本步链路代价 T_t^{link} 表示当前节点到所选下一跳节点的单跳链路代价, 且 $T_t^{\text{comp}} = 0$; 当 a_t 为计算动作且触发计算时, 计算时延 T_t^{comp} 表示当前节点执行一次在轨计算的处理代价, 且 $T_t^{\text{link}} = 0$; 若计算动作未实际触发, 则 $T_t^{\text{comp}} = 0$; v_d 为目的节点; $\mathbb{I}(\cdot)$ 为指示函数, Ω_t 为约束违背与低效行为的综合惩罚项, 其在实现中包含无效移动/断链惩罚、重复访问惩罚、原地停留惩罚以及超长路径惩罚。为实现端到端回报

下的快速学习, 本文进一步引入基于启发式剩余代价的势能函数, 以刻画从当前状态到目标节点的剩余传输代价:

$$\Phi(s_t) = \hat{\mathbb{H}}(v_i, v_d) \cdot \bar{T}_{\text{hop}} \cdot \rho_t \quad (19)$$

其中 $\hat{\mathbb{H}}(v_i, v_d)$ 为曼哈顿距离的剩余跳数估计, \bar{T}_{hop} 为单位数据倍率下的平均单跳时延, ρ_t 为当前数据相对原始数据的比例系数, 该势能函数同时反映了剩余路径长度与当前数据规模, 因此当动作使数据更接近目标节点, 或通过是在轨计算降低后续传输数据量时, 都会带来更小的剩余代价。基于式 (19), 本文采用势能差 $\Phi(s_t) - \Phi(s_{t+1})$ 作为进度奖励: 当动作使剩余传输代价下降时, 该项为正; 反之为负。与此同时, 当业务到达目的节点时, 通过 $\mathbb{I}(v_{t+1} = v_d)$ 给予终点奖励, 以增强策略对任务完成的偏好。上述设计使智能体不仅关注当前一步的链路或计算代价, 还能够面向整体端到端时延进行前瞻性决策。此外, 为提升训练稳定性, 实现中对即时奖励进行区间裁剪, 使其满足 $-1 \leq r_t \leq 1$ 。

3.2 算法实现

本文采用集中式训练与推理架构: 强化学习智能体部署于SDN控制器侧, 由控制器集中感知网络状态并在线构造状态向量, 统一输出路由与计算决策; 卫星节点仅负责执行控制器下发的转发与在轨计算指令。相较于现有的DRL算法, 本文针对传算耦合特性重新定义了价值评估对象, 将路由与计算统一建模为同一MDP下的复合动作。相较于传统DQN基于一时步时序差分对长期收益进行逐步回传, CTDDQN在统一回报和价值函数中学习其长期耦合收益, 并通过K-hop前瞻将计算行为对后续剩余代价的影响以结构化方式嵌入Q值评估, 使决策过程更适应受约束离散动作空间下的端到端时延优化。相较于MADQN的多智能体协同范式, 本文采用控制器侧集中式单智能体联合决策, 更适用于全局状态可得场景;



相较于PPO类方法，本文面向带合法动作约束的离散复合动作空间，对延迟收益的长期价值比较更具针对性。

CTDDQN的整体实现流程如图3所示。在每个决策时刻 t ，控制器根据当前状态 s_t 下考虑当前节点的可达邻居、链路可用性、链路持续时间以及计算可行性等约束构造合法动作集 $\mathcal{A}(s_t)$ 。随后智能体在该集合中采用 ϵ -贪婪策略选择动作，环境执行动作后返回 (r_t, s_{t+1}, d_t) ，并同时返回下一状态的合法动作集合 $\mathcal{A}(s_{t+1})$ ，用于构造目标值时对非法动作进行掩码处理，从而确保每一步的动作选择和评估均在可行作用域内完成。每一步交互形成经验元组 $e_t = (s_t, a_t, r_t, s_{t+1}, d_t, \mathcal{A}(s_{t+1}))$ ，并存入经验回放池D。网络训练采用带合法动作掩码的Double DQN更新机制，即由在线网络在下一状态的合法动作集合内选择贪心动作，再由目标网络对该动作进行价值评估；损失函数采用Huber损失，目标网络参数采用软更新策略平滑更新目标网络参数： $\theta^{\text{tar}} \leftarrow (1-\tau)\theta^{\text{tar}} + \tau\theta^{\text{pre}}$ ， $\tau \in (0, 1)$ ， τ 为软更新步长。

算法具体流程如算法1所示：

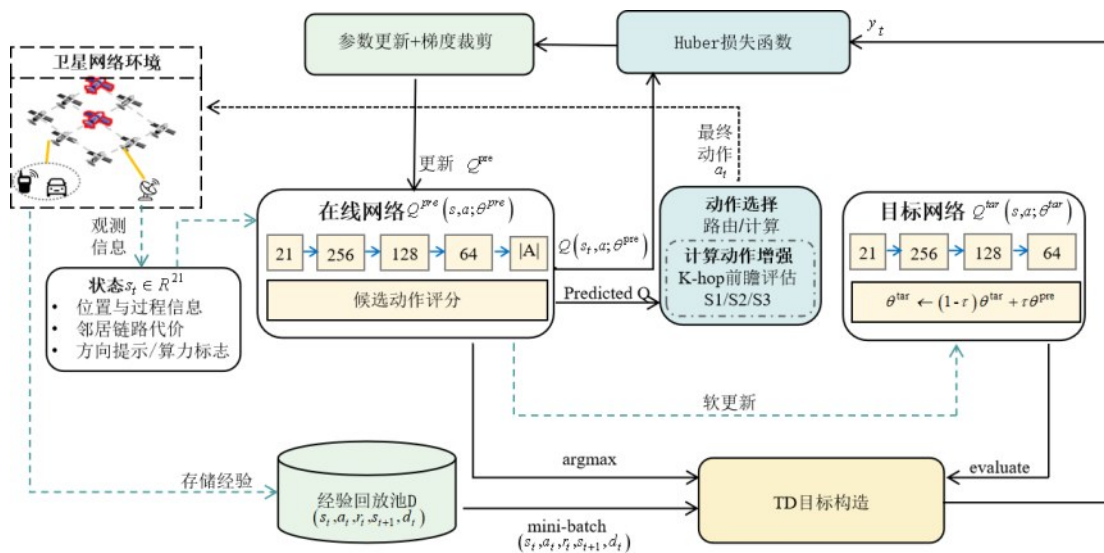


图3 基于CTDDQN的卫星网络算力路由决策框架

4 仿真结果与分析

4.1 仿真参数设置

为验证CTDDQN算法在星地一体化网络中的算力路由性能，本文实验在Linux 5.15.0-78-generic (64位)操作系统环境下完成。实验平台配置为Intel Xeon Platinum 8470Q处理器，配备31.36 GB显存的NVIDIA GeForce RTX 5090图形处理器。算法基于Python 3.12.3实现，深度学习框架采用PyTorch 2.7.0 (CUDA 12.8)。在此基础上，本文基于Walker星座几何构型构建仿真场景，星间链路采用一星四链拓扑，在不同规模配置下评估算法的可扩展性，仿真场景参数及CTDDQN的主要训练参数如表1所示。

为保证不同算法在相同业务负载下进行公平对比，本文在每组实验中生成固定数量的业务请求集合，设请求总数为 $Nr=100$ ，并以端到端时延作为性能指标，并选取以下基线算法作为对比算法。

GA-CT: 基于遗传算法的传算联合启发式策略，将候选传输路径和计算节点联合编码为个体，以端到端时延为主要适应度函数，通过选

算法1: 传算协同双重深度Q网络 (CTDDQN) 算力路由优化过程

输入: 连续时间动态图, 业务源-宿对集合, 网络状态集合 s , 用户业务数量 K , 动作集 a , 折扣因子 γ , 学习率 η , 计算前瞻步长 H , 压缩比 a_w , 经验回放容量 $|D|$, 最大步数 N_{\max}
 输出: 评价网络参数 θ^{pre} , 目标网络参数 θ^{tar}

- 1: 初始化 D 、 $Q^{\text{pre}}(s_t, a_t; \theta^{\text{pre}})$, 令
- $Q^{\text{tar}}(s_t, a_t; \theta^{\text{tar}}) \leftarrow Q^{\text{pre}}(s_t, a_t; \theta^{\text{pre}})$, 并设定探索率 $\varepsilon \leftarrow \varepsilon_0$
- 2: for episodes = 1 to E do:
- 3: 随机选取一对源-宿节点 ($g_{\text{src}}, g_{\text{dst}}$), 重置环境得到初始状态 s_0
- 4: while $t < N_{\max}$ 且未达到终止条件 do
- 5: if $\text{rand}() < \varepsilon: a_t \sim \text{Uniform}(A(s_t))$
- 6: else: $a^* = \arg \max_{a \in A(s_t)} Q^{\text{pre}}(s_t, a; \theta^{\text{pre}})$
- 7: if $a_t = a^{\text{comp}}$ then
- 8: 基于当前状态计算直达代价 T_{no} 及个候选节点 $T(c)$
- 9: if $\max_{c \in C_H} \{T_{\text{no}} - T(c)\} > \delta$,
- 10: 令 $c^* = \text{argmin}_c T(c)$, 并在节点 c^* 触发一次在轨计算
- 11: end if
- 12: 执行环境更新, 得到下一状态 s_{t+1} , 奖励 r_t , 终止条件 d_t
- 13: end if
- 14: else:
- 15: 根据 a_t 将数据从当前节点转发到下一节点, 更新 agent 状态到 s_{t+1}
- 16: end if
- 17: 将转移样本 $\{s_t, a_t, r_t, s_{t+1}, d_t, A(s_{t+1})\}$ 存入回放池 D
- 18: 从 D 中随机采样 B 条样本, 采用带合法动作掩码的 Double DQN 目标更新评价网络参数 θ^{pre}
- 19: 更新 θ^{pre} 后执行 $\theta^{\text{tar}} \leftarrow (1 - \tau)\theta^{\text{tar}} + \tau\theta^{\text{pre}}$
- 20: if $d_t = 1$ then
- 21: break
- 22: end if
- 23: end while
- 24: 衰减探索率 $\varepsilon \leftarrow \max(\varepsilon_{\min}, \lambda\varepsilon)$
- 25: end for
- 26: return $\theta^{\text{pre}}, \theta^{\text{tar}}$

择、交叉和变异搜索较优解, 作为非强化学习搜索式基线。

PPO: 基于近端策略优化的深度强化学习策略, 采用策略网络输出动作概率分布, 并通过裁剪目标函数约束策略更新幅度。其状态输入和动作空间与其他强化学习方法保持一致, 测试阶段采用确定性策略。

QL-CT: 基于表格型 Q-learning 的传算联合策略, 将节点位置、目的节点方向、计算状态和

表1 仿真参数设置

名称	值
卫星数量/颗	100/225/400
轨道数量/个	10/15/20
每轨道卫星数/颗	10/15/20
轨道高度/km	550km
轨道倾角/(度)	53°
数据大小	100Mb
卫星高算力	50GFLOPS
卫星低算力	10GFLOPS
计算需求	25GFLOPS
总仿真时长	100 min
学习率 lr	5e-5
训练轮次	3000
折扣因子 γ	0.96
Batch size	256
经验回放容量 $ D $	1000000
软更新系数 τ	0.001
梯度裁剪阈值	1.0
初始探索率 $\varepsilon_{\text{start}}$	1.0
最小探索率下限 ε_{min}	0.02
探索率衰减系数 $\varepsilon_{\text{decay}}$	0.9998
随机种子 seed	0/1/2/3/4
请求数量 Nr	100
奖励系数 $a_1 \sim a_5$	0.2/0.2/0.2/5.0/1.0

数据压缩状态等信息离散化为状态索引, 在离散状态—动作空间中联合选择路由动作和计算动作, 训练阶段采用 ε -greedy 策略。

DQN-CT: 基于深度Q网络的传算联合策略, 采用神经网络近似Q值函数, 并通过经验回放和目标网络提升训练稳定性。训练时采用 Adam 优化器和 Smooth L1 Loss, 动作选择阶段对非法动作进行合法动作掩码处理。

DDQN-route: 基于 Double DQN 的纯路由基线, 采用与 DQN-CT 相近的网络训练方式, 但在训练和测试阶段均屏蔽计算动作, 仅保留路由动作, 用于分析不考虑星上计算时的路由性能。

4.2 仿真结果分析

图4给出了5个独立随机种子条件下CTDDQN训练过程中TD损失的均值及95%置信区间。训练初期TD损失迅速下降, 表明价值网络能够较快降低估计误差; 训练中期损失曲线出现阶段性抬升与波动, 反映出随着探索率下降和经验样本分布迁移, 价值函数进入重新校准阶



段；训练后期TD损失再次下降并趋于稳定，未出现明显发散。不同随机种子下置信区间整体较窄，说明CTDDQN在多次独立训练中具有较一致的价值更新趋势和较好的训练稳定性。

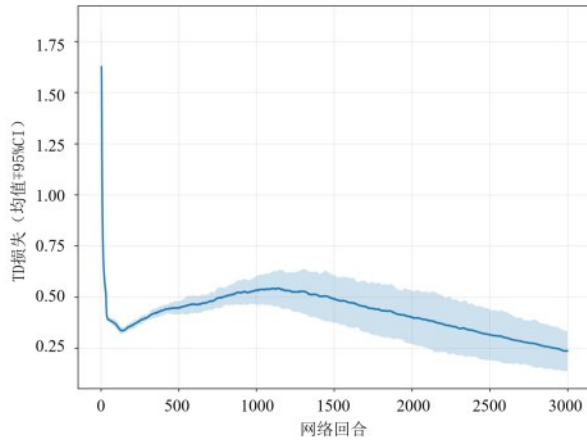
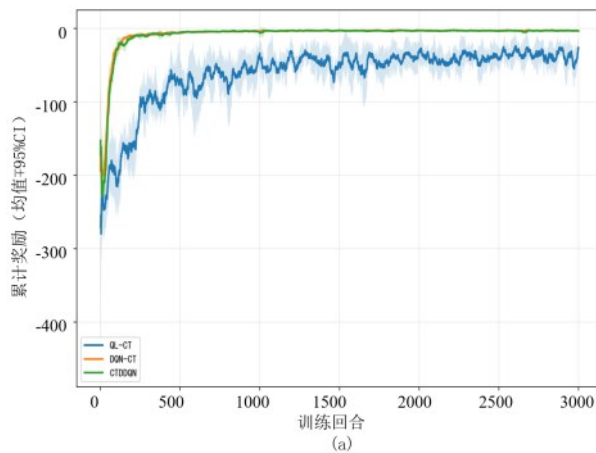


图4 CTDDQN训练过程中TD损失均值及95%置信区间

图5给出了QL-CT、DQN-CT与CTDDQN在训练过程中的累计奖励与验证端到端时延变化趋势。由图5(a)可见，QL-CT的累计奖励提升较慢且波动较大，表明表格型方法在传算耦合决策场景下的学习效率和训练稳定性相对有限；DQN-CT与CTDDQN的累计奖励能够更快趋于稳定，说明深度价值函数近似有助于提升策略学习效率。考虑到累计奖励由计算代价、成功奖励等共同构成，不能直接等同于端到端时延指标，本文



进一步统计验证端到端时延变化趋势，如图5(b)所示。结果表明，CTDDQN的验证端到端时延在训练早期快速下降，并在训练后期保持较低水平，说明其训练收益能够较好地转化为核心时延指标的优化。

图6展示了不同星座规模下各算法的平均端到端时延与服务成功率。随着星座规模增大，源宿节点之间的平均路径跳数增加，传输时延占比进一步提升。CTDDQN在三种星座规模下均保持较低端到端时延，相较表现最优的基线方法仍可获得约18.9%的时延降低。其原因在于，CTDDQN将计算动作纳入路由决策，并通过K-hop前瞻评估不同计算触发位置对后续传输代价的影响，能够更充分利用数据压缩带来的多跳传输收益。同时，CTDDQN在不同星座规模下均保持较高服务成功率，说明其时延降低并非通过牺牲任务完成能力获得，而是在维持较高成功率的同时实现了更低时延。

图7给出了端到端时延随最大星上算力 C_{max} 变化的趋势。为突出星上计算能力对传算联合决策的影响，本文在固定星座规模、业务请求和链路参数的基础上，提高任务计算需求和数据大小，并改变高算力卫星的最大计算能力 C_{max} 。由图可知，具备计算动作的算法随 C_{max} 增大整

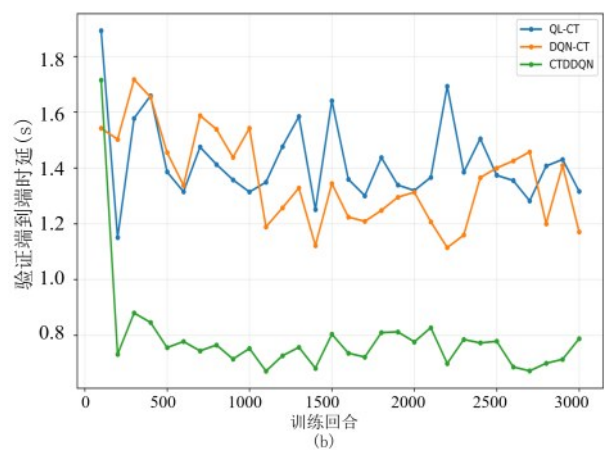


图5 训练过程中累计奖励与端到端时延变化趋势

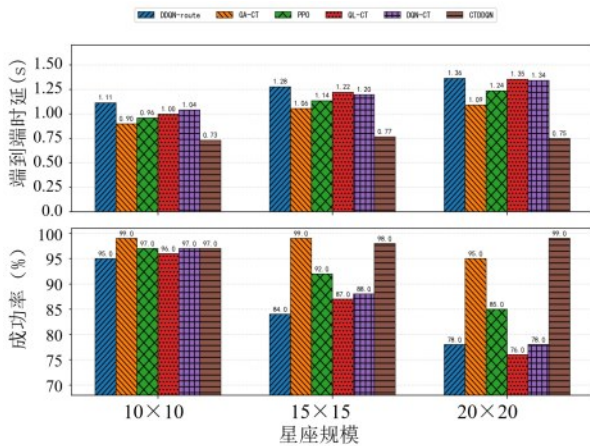


图6 不同星座规模下各算法端到端时延与成功率对比

体呈现时延下降趋势，说明在计算需求较高时，星上算力提升能够有效降低任务处理时延。当算力超过约 200 GFLOPS 后，下降幅度减小，计算瓶颈得到缓解，端到端时延更多受剩余多跳传输过程限制。DDQN-route 不包含计算动作，因此对 C_{max} 基本不敏感。相比其他传算联合基线，CTDDQN 在不同算力配置下均保持较低时延，说明其能够结合节点算力、剩余路径长度和压缩收益选择计算触发位置，从而更有效地将算力提升转化为端到端时延收益。

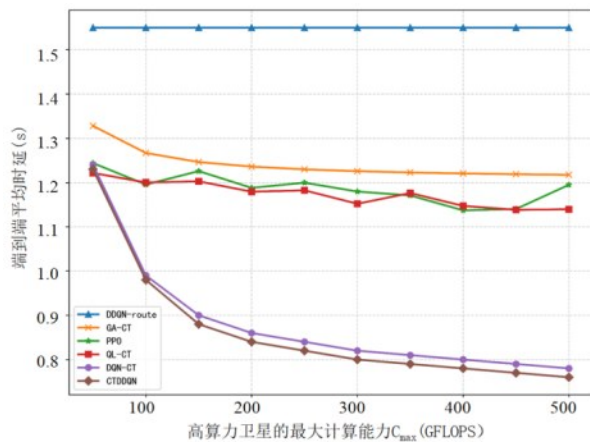


图7 平均端到端时延与卫星最大计算能力的关系

图8给出了计算动作与 K-hop 前瞻机制对端到端时延的影响。与无计算基线相比，引入计算动作后端到端时延明显降低，说明星上压缩计算

能够减少后续多跳传输数据量，是降低端到端时延的主要因素。在引入计算动作后，前瞻计算相比就地计算进一步降低了端到端时延，表明计算触发位置对传算联合优化结果具有重要影响。随着 K 值增大，CTDDQN 能够在更大范围内选择候选计算节点，端到端时延进一步下降；但当最大算力进入中高区间后，不同 K 值对应的时延曲线逐渐接近，表明继续扩大前瞻范围的边际收益减弱。总体而言，计算动作引入是性能提升的主要来源，K-hop 前瞻机制则进一步优化了计算触发位置选择。

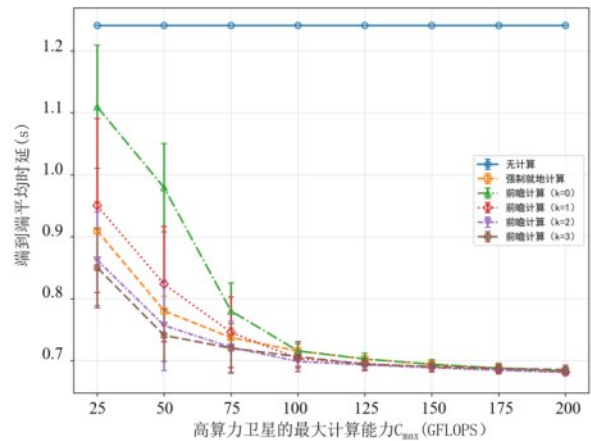


图8 计算动作与 K-hop 前瞻机制对端到端时延的影响

为进一步分析 CTDDQN 中各关键机制对端到端时延的贡献，本文在相同网络环境和业务负载下开展模块级消融实验，分别去除计算动作、K-hop 前瞻、统一回报机制和合法动作约束，并采用惩罚后平均端到端时延作为评价指标。由图9可见，去除计算动作后端到端时延显著增加，说明星上计算与数据压缩是降低端到端时延的主要来源。去除合法动作约束后，时延也有所上升，表明动作约束能够减少无效动作选择，提高策略执行有效性。相比之下，去除 K-hop 前瞻和统一回报机制后，最终平均时延变化相对较小，说明在当前高算力卫星配置下，计算动作本身已经提供了主要时延收益，K-hop 前瞻和统一回报



机制更多体现为计算节点选择优化。

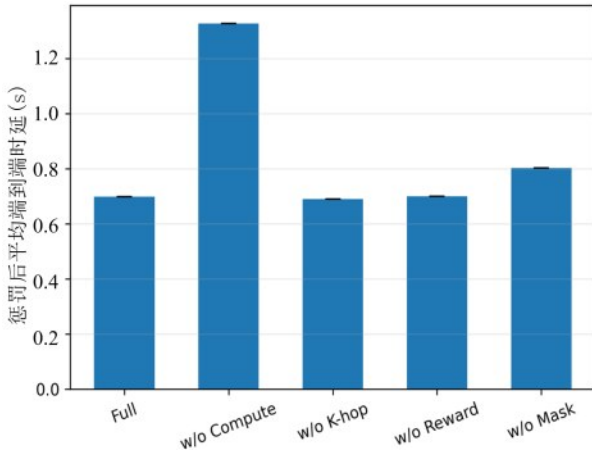


图9 不同消融体下的惩罚后平均端到端时延

图10比较了CTDDQN与地面卸载方案在不同 C_{max} 条件下的端到端时延比。整体来看，CTDDQN在全区间内均优于地面卸载方案，在低算力与低SGL条件下优势最为显著。在 $SGL=0.4$ Gb/s时，时延约为地面卸载的0.48倍，体现出显著收益；而当SGL提升至10 Gb/s时，该比例上升至约0.8，差距明显收敛。在低带宽场景中，传输时延占主导，CTDDQN通过K-hop前瞻将剩余路径长度与未来链路代价纳入Q值估计，倾向于在剩余路径较长且链路受限的节点提前触发计算，从而放大压缩收益在后续多跳中的累积效应。当带宽提升后，未来传输代价下降，压缩收益在价值函数中的权重减弱，计算触发倾向随之降低，因此相对优势收敛。

图11进一步分析了不同SGL带宽条件下的时延比变化。低SGL带宽下，星地传输占主导，CTDDQN相较地面卸载可获得约20% - 25%的额外时延下降；随着SGL带宽提升，地面卸载传输代价下降更快，CTDDQN的相对收益逐步收敛。CTDDQN在价值估计过程中对未来传输负载显式建模：当带宽受限时，算法更倾向于提前压缩数据以减少后续链路压力；而在带宽充裕条件下，Q值中压缩收益项的贡献降低，计算触发概率下

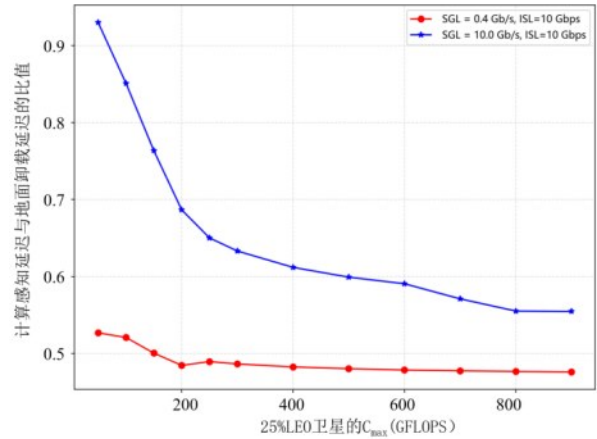


图10 LEO计算感知路由与地面卸载延迟比率关系

降。不同算力配置在低带宽区间呈现分层结构，但在高带宽区逐渐接近。

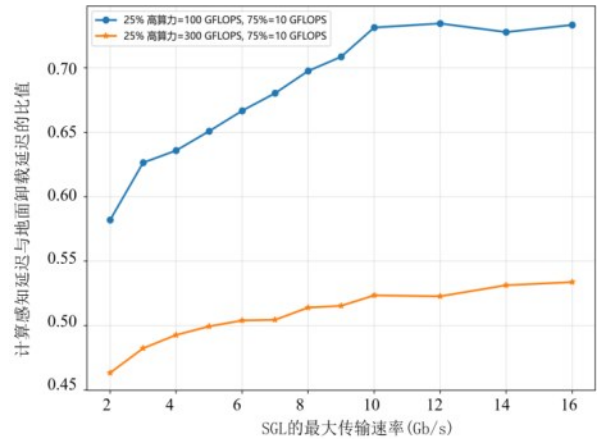


图11 不同SGL传输速率下计算感知路由与地面卸载的时延比

表2给出了在线决策开销分析。在本文采用聚合状态表示的设定下，尽管节点规模与业务负载同步增加，单次推理耗时基本稳定在约1.108ms。该稳定性主要源于状态向量采用固定维度聚合编码，使网络规模变化不会增加输入维度或参数规模；同时K-hop前瞻仅在局部邻域触发，单次耗时仅小幅变化，其复杂度未随全局节点数线性增长。周期总耗时的提升主要源于决策次数增加，而非单次推理或K-hop复杂度显著恶化，说明所提方法在规模扩展条件下仍具备良好的在线响应能力。

在相同星座规模下，将 D_s 扩展至 37 后，单次推理耗时仍保持在约 1.108 ms，表明在该增维范围内策略网络前向计算并未成为控制器侧瓶颈，在线响应能力保持稳定。从复杂度角度看，当状态维度 D_s 进一步增大时，Q 网络前向计算中与输入相关的主要开销来自输入到隐藏层的线性映射，其乘法运算量随 D_s 增长而增加，与此同时，更高维状态会扩大函数逼近的有效自由度，使得 CTDDQN 在同等性能目标下可能需要更多交互样本与训练迭代才能收敛，并可能加剧训练过程的波动与超参数敏感性。

5 结束语

本文围绕低轨卫星网络中计算与传输资源协同优化问题，面向高动态星座环境开展了算力路由研究，提出了一种基于改进双重深度 Q 网络的计算感知路由优化方法。该方法在连续时间动态图框架下统一建模通信链路 with 计算节点的时空耦合关系，通过引入 K-hop 前瞻机制，有效提升了价值函数对传算收益差异的刻画能力，克服了传统路径选择与任务卸载分离导致的全局优化能力不足问题。

未来工作将进一步将能量约束与多业务服务质量需求纳入统一优化框架，探索多智能体协同决策机制，并结合图神经网络 (Graph Neural Network, GNN) 等结构化模型挖掘星座拓扑与资源分布的高阶特征，以实现天地一体化网络在时延、能效与可靠性等多维指标上的协同优化。

表 2 不同星座规模与业务负载下的控制器在线决策开销统计

星座规模	节点数 N	业务负载 B	状态维度 D_s	单次推理耗时 (ms)	Step 其他耗时 (ms)	K-hop 单次耗时 (ms/次)	平均每决策额外开销 (ms)	周期总耗时 (ms)
10*10	100	100	21	1.108	0.003	0.008	0.0016	113.074
20*20	400	200	21	1.107	0.004	0.011	0.0022	226.472
40*40	1600	400	21	1.108	0.004	0.013	0.0026	453.501
40*40	1600	400	37	1.108	0.005	0.013	0.0027	454.565



参考文献:

- [1] 王志勤, 杜滢, 沈霞, 等. 面向6G典型场景的无线系统研究[J]. 中兴通讯技术, 2024, 30(4): 65-68.
WANG Z Q, DU Y, SHEN X, et al. Research on wireless systems for typical 6G scenarios[J]. ZTE Technology Journal, 2024, 30(4): 65-68.
- [2] 赵亚飞, 周家恩, 王鑫洋. 面向卫星通信的6G雾计算网络技术研究与展望[J]. 无线电通信技术, 2023, 49(5): 834-841.
ZHAO Y, ZHOU J E, WANG X Y, et al. Research and Prospect of 6G Fog Computing Network for Satellite Communication[J]. Radio Communications Technology, 2023, 49(5): 834-841.
- [3] 李佳奇, 陈全, 杨磊. 低轨巨型星座路由技术研究现状及展望[J]. 空间电子技术, 2025, 22(2): 1-12.
LI J Q, CHEN Q, YANG L. Routing technologies in LEO mega-constellations: A survey and outlook[J]. Space Electronic Technology, 2025, 22(2): 1-12.
- [4] TIRMIZI S B R, CHEN Y, LAKSHMINARAYANA S, et al. Hybrid Satellite - Terrestrial Networks toward 6G: Key Technologies and Open Issues[J]. Sensors, 2022, 22(21): 8544.
- [5] 3GPP. System architecture for the 5G System (5GS): TS 23.501 [S]. 2024.
- [6] KODHELI O, LAGUNAS E, MATURO N, et al. Satellite communications in the new space era: A survey and future challenges[J]. IEEE Communications Surveys & Tutorials, 2020, 23(1): 70-109.
- [7] ZHU X, JIANG C. Integrated satellite-terrestrial networks toward 6G: Architectures, applications, and challenges[J]. IEEE Internet of Things Journal, 2021, 9(1): 437-461.
- [8] LIU C, FENG W, TAO X, et al. MEC-empowered non-terrestrial network for 6G wide-area time-sensitive internet of things[J]. Engineering, 2022, 8: 96-107.
- [9] WANG R, KISHK M A, ALOUINI M S. Stochastic geometry-based low latency routing in massive LEO satellite networks[J]. IEEE Transactions on Aerospace and Electronic Systems, 2022, 58(5): 3881-3894.
- [10] HAN Z, ZHAO G, XING Y, et al. Dynamic routing for software-defined LEO satellite networks based on ISL attributes [C]//2021 IEEE Global Communications Conference (GLOBECOM). Madrid, Spain: IEEE, 2021: 1-6.
- [11] 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.
- [12] PAGE P S, BHARGAO K S, BAVISKAR H V, et al. Distributed probabilistic congestion control in LEO satellite networks [C]//2023 15th International Conference on COMMunication Systems & NETWORKS (COMSNETS). Bangalore, India: IEEE, 2023: 335-339.
- [13] SORET B, LEYVA-MAYORGA I, LOZANO-CUADRA F, et al. Q-learning for distributed routing in LEO satellite constellations [C]//2024 IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN). Stockholm, Sweden: IEEE, 2024: 208-213.
- [14] 徐晖, 陈山枝, 艾明. 面向6G的星地融合网络架构[J]. 中兴通讯技术, 2023, 29(5): 9-15.
XU H, CHEN S Z, AI M. Architecture of satellite-terrestrial integrated networks for 6G[J]. ZTE Technology Journal, 2023, 29(5): 9-15.
- [15] CAO H, WANG H, WU T, et al. Task Offloading Strategy in Satellite Edge Computing Based on Matching Game [C]//Proceedings of the 2023 12th International Conference on Networks, Communication and Computing. Osaka, Japan: ACM, 2023: 91-98.
- [16] LI Y, ZHU S, XIONG T, et al. Joint Task Offloading and Power Allocation for Satellite Edge Computing Networks[J]. Sensors, 2025, 25(9): 2892.
- [17] YANG J, SHAH A A, PEZAROS D. Priority-Aware Task Offloading in UAV-Assisted Satellite MEC Networks [C]//2025 IEEE International Conference on Communications Workshops (ICC Workshops). Montreal, Canada: IEEE, 2025: 640-645.
- [18] WANG Y, FENG C, SUN J. Cost-Efficient Computation Offloading and Service Chain Caching in LEO Satellite Networks[J]. arXiv preprint arXiv:2311.07872, 2023.
- [19] LAN W, CHEN K, CAO J, et al. Security-sensitive task offloading in integrated satellite-terrestrial networks[J]. IEEE Transactions on Mobile Computing, 2024, 24(3): 2220-2233.
- [20] CAO J, ZHANG S, CHEN Q, et al. Computing-aware routing for leo satellite networks: A transmission and computation integration approach[J]. IEEE Transactions on Vehicular Technology, 2023, 72(12): 16607-16623.
- [21] LI Y, ZHANG Q, YAO H, et al. Stigmergy and hierarchical learning for routing optimization in multi-domain collaborative satellite networks[J]. IEEE Journal on Selected Areas in Communications, 2024, 42(5): 1188-1203.
- [22] 孔梦燕. 基于深度强化学习的低轨卫星网络算力路由研究 [D]. 北京: 中国电子科技集团公司电子科学研究院, 2025.
KONG M Y. Computing-aware routing in LEO satellite networks based on deep reinforcement learning [D]. Beijing: Electronic Science Research Institute of China Electronics Technology Group Corporation, 2025.

- [23] ZHONG L, LI Y, GE M F, et al. Joint Task Offloading and Resource Allocation for LEO Satellite-Based Mobile Edge Computing Systems With Heterogeneous Task Demands[J]. IEEE Transactions on Vehicular Technology, 2025, 74(7): 11337-11352.
- [24] Rahman Z, Jobson D J, Woodell G A. Investigating the relationship between image enhancement and image compression in the context of the multi-scale retinex[J]. Journal of Visual Communication and Image Representation, 2011, 22(3): 237-250.

[作者简介]



蔡佳慧 (2002-), 女, 北京邮电大学研究生, 主要研究方向为卫星智能算力路由

技术。

周家恩 (2001-), 男, 北京邮电大学博士生, 主要研究方向为卫星通感算一体化技术。



赵亚飞 (1987-), 男, 博士, 北京邮电大学特聘副研究员、博士生导师, 主要研究方向为空地海信息通信和通感算融合。



彭木根 (1978-), 男, 博士, 北京邮电大学网络与交换技术全国重点实验室教授, 主要研究方向为空间信息通信、通感算一体化、雾无线接入网络等

