

特色专题

面向超高清视频的算力网络架构及关键技术

周旭¹, 吴红², 张伟³, 宋俊平¹

摘要 超高清视频是中国视听产业重要发展方向之一,国家相关部门也出台了一系列政策,鼓励和支持超高清视频产业的发展。超高清视频的采集、传输、制作、播出过程,尤其是融合了 ChatGPT、Sora 等先进人工智能内容生成技术后,呈现出典型的大带宽、高算力、低时延特征,令算力和网络基础设施面临严峻考验。基于超高清视频典型需求和计算、网络技术最新发展趋势,提出了面向超高清视频的算力网络架构,综合运用异构算力资源组网与安全传输技术、超高清视频业务需求建模与资源编排技术、“数算模”联合调度与路由规划技术、超高清视频高速传输技术等算力网络关键技术,实现全国范围内异构算力的汇聚、组网,满足超高清视频采、编、播等各环节业务处理对多样化算力和网络传输的需求。

关键词 超高清视频;算力网络;异构算力组网;算网资源编排;路由规划;高速传输技术

1 研究背景

1.1 超高清视频产业发展现状

继模拟、标清、高清之后,当前 4 K/8 K 超高清^[1]成为视听技术发展新趋势,呈现网络视频化、视频超高清化、沉浸化、智能化等鲜明特点。相关报告数据显示,截至 2022 年底,中国超高清视频产业规模已超 3 万亿元人民币^[2]。目前已开通了 8 个 4 K 频道、2 个 8 K 频道。8 K 超高清摄像机、8 K 监视器、8 K 图像传感器、光学镜头等多个产品技术实现了从无到有的创新突破并实现产业化应用。此外,高动态范围^[3]、三维声^[4]、视频编解码等国内企业自研团体标准已在行业内批量推广应用。

2022 年,中共中央办公厅、国务院办公厅印发的《关于推进实施国家文化数字化战略的意见》^[5]文件明

确“传统产业加速数字化转型,大力发展第 5 代移动通信(5G)、大数据、云计算、虚拟现实(VR)、增强现实(AR)、人工智能(AI)、区块链、元宇宙等新技术,为数字文化内容生产、传播和市场应用提供了广阔空间”;此外,2024 年国家发展和改革委员会、住房和城乡建设部、文化和旅游部、国家电影局、国家广播电视总局、国家文物局联合印发《推动文化和旅游领域设备更新实施方案》,方案明确将“实施高清超高清设备更新提升行动”作为 8 大重点任务之一。

随着以 ChatGPT、Sora 为代表的人工智能内容生成技术逐渐发展成熟,超高清视频与人工智能的融合应用正迅速增强。发展至今,AI 对超高清视频的赋能具体表现在以下 6 个方面。(1) 智能压缩^[6],AI 可以通过深度学习算法,识别并保留视频中的重要信息,同时压缩不太重要的部分。这样可以在保持画质的同时减小文件,从而更有效地进行传输。(2) 优化传输路径^[7],通过 AI 的预测和分析,可以智能选择最佳的传输路径,从而减少延迟和丢包,保证实时高清视频的流畅传输。(3) 超分辨率技术^[8-9],AI 可以通过学

1. 中国科学院计算机网络信息中心,北京 100083

2. 湖南芒果融创科技有限公司,长沙 415500

3. 国家广播电视总局广播电视科学研究院,北京 100866

收稿日期:2024-08-21;修回日期:2025-02-19

作者简介:周旭,研究员,研究方向为计算机网络体系结构、5G/6G 移动网络、网络人工智能等,电子信箱:zhouxu@cstnet.cn

引用格式:周旭,吴红,张伟,等.面向超高清视频的算力网络架构及关键技术[J].科技导报,2025,43(9):38-47;doi:10.3981/j.issn.1000-7857.2024.08.01034

习已知的高清图像,对低分辨率的图像进行重建,从而实现超分辨率效果,让画质得到显著提升。(4)智能编码和解码^[10],通过AI驱动的智能编码和解码技术,可以根据网络条件和设备能力动态调整视频质量,确保提供最佳的观看体验。(5)个性化体验,AI可以根据用户的习惯和偏好,智能调整画质、分辨率和流量消耗,为不同的用户提供个性化的高清体验。(6)虚拟现实和增强现实应用^[11],通过AI的图像识别和渲染能力,实时高清视频可以与VR和AR无缝结合,为用户带来沉浸式的体验。总而言之,在实时互动时代有2点核心要求,一是传输,二是画质,而这些也正是AI赋能行业的发力点。

超高清音视频是数字信息化、高清化之后新一轮重大技术革命,在超高清音视频采集、传输、制作、播出的整个流程中,都呈现出大带宽、高算力、低时延的特征。因此,超高清音视频业务将会对算力和网络提出更高的要求。在算力、网络的高质量供给方面,需求与挑战并存。在算力需求方面,高质量视频生成需要大量算力。例如,Sora视频生成模型^[12]的训练,大约需要在4200~10500张Nvidia H100上训练1个月,在推理环节,每张H100每小时最多生成5 min时长视频。视频处理过程中,特别是4K/8K超高清视频的处理,对算力的需求极大。例如,对于8K视频处理,需要计算统一设备体系结构(compute unified device architecture, CUDA)核心数在4000个以上的图形处理单元(graphics processing unit, GPU)。在网络传输方面,超高清视频对传输带宽和时延的要求很高^[13]:单机位4K视频传输带宽需求为50~60 MB/s,一档节目最多需要100~200个机位,需要传输总带宽超过50 Gbps,传输时延需要控制在40~80 ms。

1.2 算力网络技术与应用发展现状

2020年4月20日,国家发展改革委首次明确新型基础设施(简称“新基建”)的范围^[14],其中的信息基础设施,包括以数据中心、智能计算中心为代表的算力基础设施等,这是“算力基础设施”这一概念在国家层面的首次提出。目前,多种算力供给设施,如超算中心、云计算数据中心、智能计算中心、边缘计算站点建设蓬勃发展。据估计,2025年分布式算力资源将增长为2020年的790%,超过集中式算力资源。未来算力将以分布式算力为主,算力资源需求的分布式特

点更加明显。

为加强计算、网络、存储和应用协同创新,工业和信息化部等6部门于2023年10月联合印发《算力基础设施高质量发展行动计划》^[15],明确到2025年的主要发展目标,提出完善算力综合供给体系、提升算力高效运载能力、强化存力高效灵活保障、深化算力赋能行业应用、促进绿色低碳算力发展、加强安全保障能力建设等6方面重点任务,着力推动算力基础设施高质量发展。2023年12月,由国家发展改革委、国家数据局、中央网信办、工业和信息化部、国家能源局5部门联合印发的《关于深入实施“东数西算”工程 加快构建全国一体化算力网的实施意见》^[16]正式公布,明确以算力高质量发展赋能经济高质量发展为主线,充分发挥全国一体化算力网络国家枢纽节点(国家枢纽节点)引领带动作用,协同推进“东数西算”工程,构建联网调度、普惠易用、绿色安全的全国一体化算力网,助力网络强国、数字中国建设,打造中国式现代化的数字基座。

中共中央办公厅、国务院办公厅《关于推进实施国家文化数字化战略的意见》^[6]提出国家文化专网建设目标,要求夯实文化数字化基础设施,依托现有有线电视网络设施、广电5G网络和互联互通平台,形成国家文化专网,搭建以网络 and 算力为基础底座,叠加文化数据服务能力的广电文化数字化平台。2023年2月,国家广播电视总局《智慧广电技术体系及实施指南》提出,力争到2025年,建成以新一代信息技术应用为基础,以“算力+算法+数据”为重要支撑的技术体系,全面形成布局合理、竞争有序、特色鲜明、形态多样、可持续发展的智慧发展新格局。国家广播电视总局也将围绕3个重点领域展开工作:一是算网一体,筑牢广电算力的设施底座;二是算网融合打造智慧算网大脑;三是算网融合赋能千行百业,数智升级。

面对超高清音视频大数据量的计算处理要求,算力网络能够实现超高清音视频的高效传输,通过计算、存储与网络等资源的协同、调度与管理,完成超高清音视频的按需流动、智能计算与可视分析处理,可对大型赛事、交互式的超高清直播等带来有力的支撑,为全景视频、自由视角等元宇宙时代超高清视频的融合发展带来更多想象空间。

2 面向超高清视频的算力网络架构

2.1 超高清视频的算力网络系统架构

针对超高清视频内容制作过程中算力与网络服务能力不足、算网资源协同能力差等挑战,面向超高清视频的算力网络综合运用异构算力资源组网与安全传输技术、超高清视频业务需求建模与资源编排技

术、“数算模”联合调度与路由规划技术、超高清音视频高速传输技术等关键技术,实现全国范围内异构算力的汇聚、组网,满足超高清视频采、编、播各环节业务对多样化算力和高速传输的需求。

面向超高清视频的算力网络系统架构如图 1 所示。

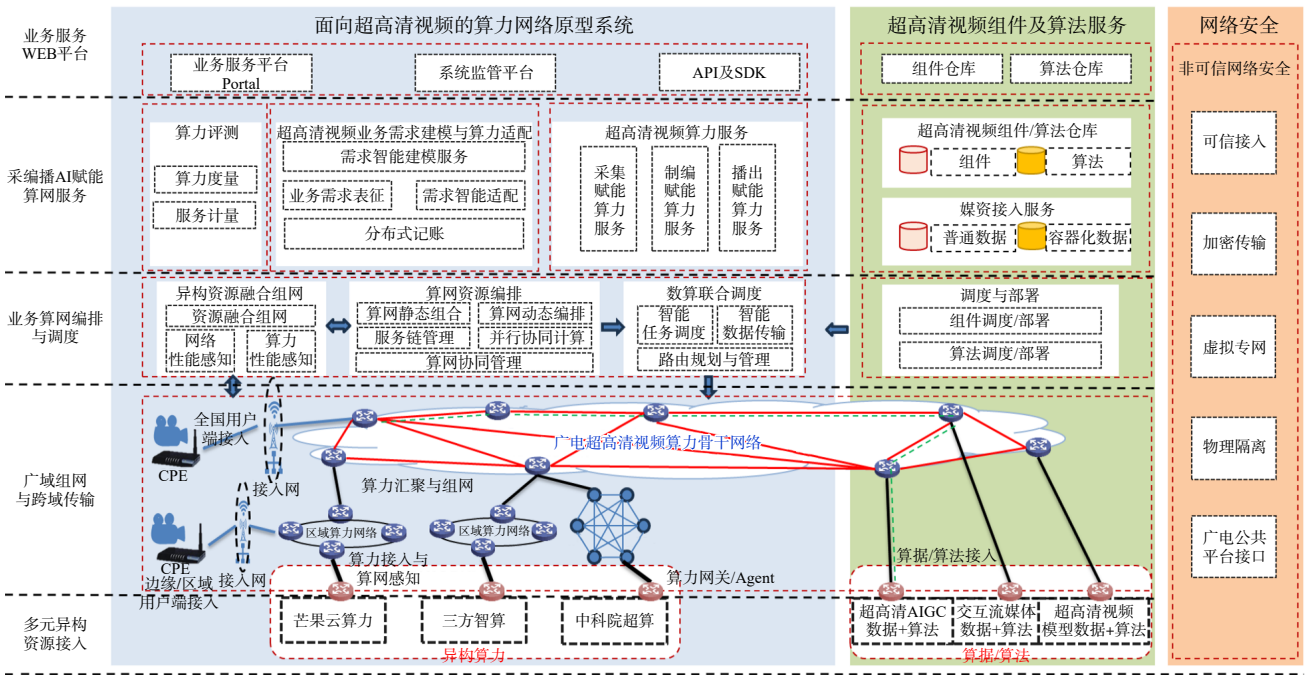


图 1 面向超高清视频的算力网络系统架构

依据系统整体功能架构自顶向下可分为业务服务 WEB(万维网)平台、超高清视频组件/算法服务、采编播 AI 赋能算网服务、业务算网编排与调度、广域算力组网与跨域传输、多元异构资源接入、网络安全等部分。

1) 业务服务 WEB 平台。为系统管理员、超高清视频业务用户等相关人员提供平台管理和入口;同时也为超高清视频业务系统开发人员或者第三方平台提供相关的应用程序接口(application program interface, API)和软件开发工具包(software development kit, SDK)工具。

2) 超高清视频组件及算法服务。构建组件/算法仓库,支持超高清视频业务组件、模型、算法,对接超高清视频媒资数据的接入,以商店的模式为超高清视频业务提供容器化的服务能力;支持组件、算法的容

器化封装、自动部署等服务。

3) 采编播 AI 赋能算网服务。依据超高清视频业务 AI 赋能对算网资源的需求,综合算力资源状态、网络实时状态完成业务需求建模,实现业务与算力的灵活、智能匹配和服务。支持对算力的多维度度量、对算网资源的服务评测及计量;支持基于超高清视频采、编、播各环节多种业务组件、模型、算法的共享和高效流转服务。

4) 业务算网编排与调度。综合算力智能适配及业务服务能力需求,实现异构算力融合组网、算网资源统一编排、业务数算联合调度,支持对超高清视频业务组件和算法的容器化封装、自动部署、路由规划、安全传输和管理执行。

5) 广域组网与跨域传输。基于广电网络、互联网等网络传输能力,汇聚区域算力网络、超高清视

频算法/算据等资源,构建一张广电跨域超高清视频算力骨干网络,在全国范围为超高清视频业务提供多种算力资源支持。其中,算力网关/Agent支持通算、智算、超算等算力的接入与并网,支持超高清视频业务组件、模型、算法及算据的接入与安全传输,支持算力资源状态感知、网络实时状态感知等能力;用户驻地设备(customer premises equipment, CPE)终端设备,支持超高清视频业务终端及用户的网络接入、数据安全传输等能力。

6) 多元异构基础资源接入。基于广电网络连通互联网,提供各区域多种算力中心(包括CPU、GPU、FPGA/ASIC等多种异构算力)的网络接入与连接能力,支持超高清视频相关算法/算据等资源互联和并网,提供基础算网资源能力。

7) 网络安全。支持非可信网络的可信接入、加密传输、虚拟专网及物理隔离等安全能力,对接广电公共平台系统服务能力,保障算力网络的可靠、安全。

2.2 超高清视频的算力网络组网架构

超高清视频的算力网络组网架构是以广电网络

为基础,依托现有有线电视网络设施、广电5G网络和互联互通平台,构建面向媒体行业的算力网络骨干网,连接广电网络内自有通用算力中心、智算中心和超算中心,以及各类异地生产、协同制作时部署的边缘算力中心,如图2所示。同时,为了有效融合国家已建设布局的全国一体化算力网和国家文化专网等各类第三方资源,通过部署算力网络中央控制器的方式,统一纳管广电自有算力和第三方开放算力资源,合理规划网络布局和资源配置,确保在算力网络构建中充分发挥自身优势。各算力节点通过部署的算力网关接入算力骨干网络。为了实现算力资源的高效网络调度,算力网关间通过SRv6技术实现网络高速互联,并支持算力网络控制平面和数据平面的统一承载。同时,利用SRv6技术具备的强大路径编程能力和流量路径智能调优能力,帮助超高清视频业务实现流量初始路径负载和带宽拥塞路径切换,极大提高算力中心间带宽的利用率,支持算力网络时代的“网随算动”,更好地满足超高清视频业务的数据传输需要。

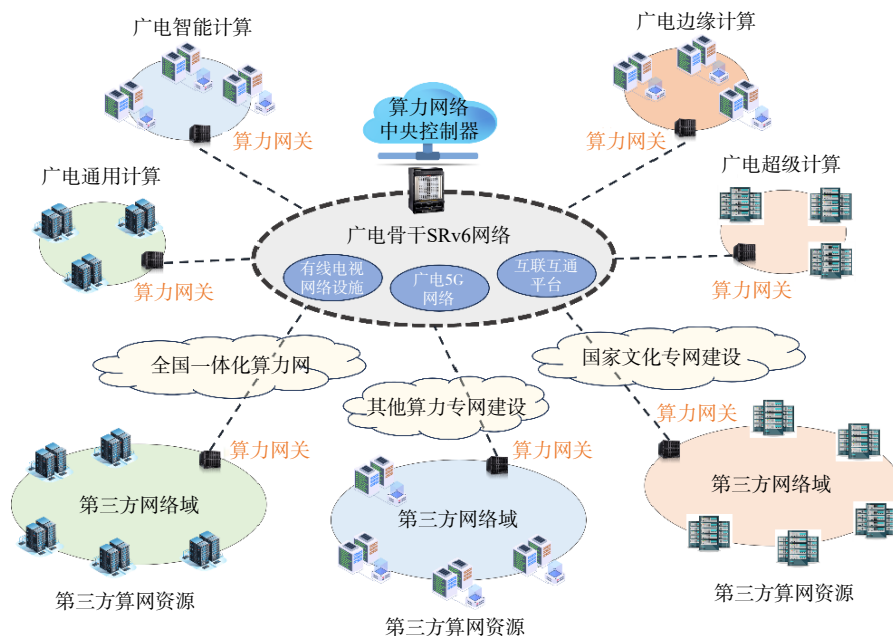


图2 基于广电网络的新型算力网络组网架构

3 面向超高清视频的算力网络关键技术

3.1 超高清视频业务算网资源编排技术

超高清视频算力网络需要同时支持多机位、多栏目、多形态的多个超高清视频制作业务,如何在有限

资源的前提下,协调多点异构算力与多条传输链路,保障大量并发的算力任务、传输任务高效完成,是算力网络需要解决的关键问题。为了对算网资源和业务传算需求进行精准匹配,研究超高清视频业务分解技术,将复杂的制作业务拆解为多个顺序相关、相互

依赖的计算任务和传输任务。同时,研究面向超高清视频业务的资源编排技术,根据算力资源状态,为分解后的计算和传输任务分别选择合适的计算节点和合适的传输链路,通过这些任务的协同完成超高清视频制作,确保制作过程中的计算和传输时延、数据安全等均符合业务需求。此外,基于对全网算力资源的统一智能编排,可以提升资源整体使用效率,提高算力网络对大规模并发业务的服务能力。

如图3所示,超高清视频业务分解是将一个超高清视频业务拆解为多个顺序相关、相互依赖的计算任务和传输任务。同时,业务分解的过程也是将业务需

求转化为算力资源需求的重要环节。而资源编排则通过合理的资源匹配,满足这些计算任务或传输任务的性能要求,从而保障超高清视频业务整体的完成质量。在分布式异构资源组成的算力网络中,网络状态动态变化且算力密度分布不均匀,单个任务的数据传输或计算处理可能会因不可预测的网络拥塞和负载不平衡而产生较大延迟,从而导致整个业务无法在预期时间内处理完成。因此,资源编排需要解决在复杂动态的算力网络中如何面向拆解后的细粒度任务需求进行一体化编排,保证业务整体处理质量,提升系统服务容量。

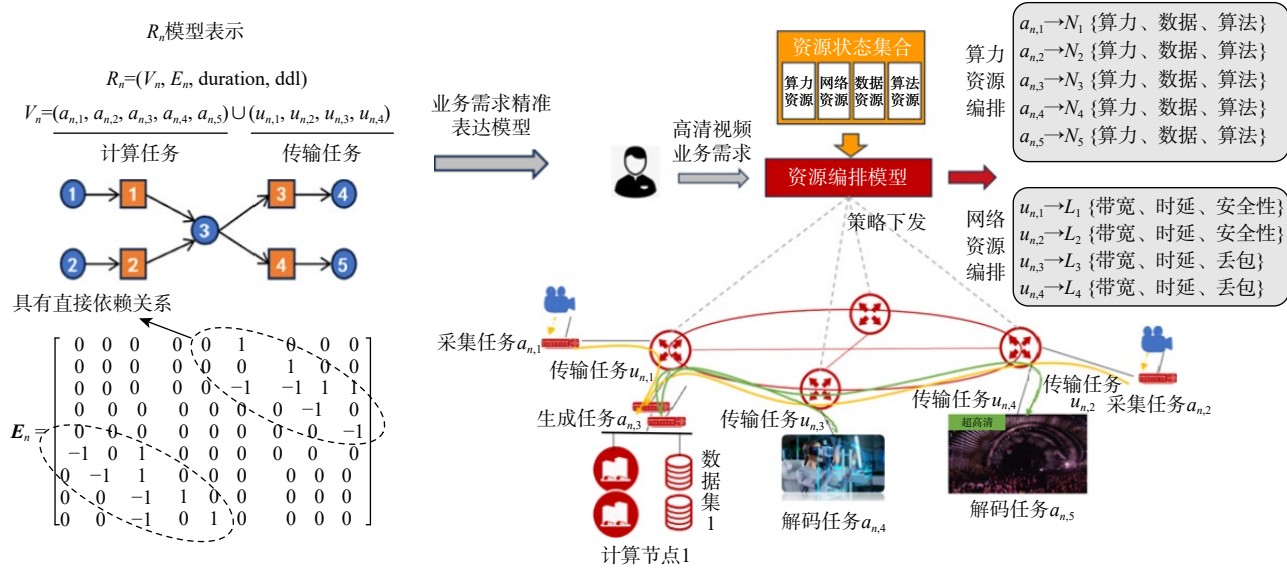


图3 资源编排调度示意

3.2 超高清视频“数算模”联合调度与路由规划技术

在超高清视频云化服务(包括 AI 模型)部署完成后,业务进行过程中,不同的超高清视频处理任务需求不断变化,不同云可用的计算、存储资源会动态变化^[17-18],例如,视频内容在不同边缘云上的生成和制作处理或导致不同云的计算、存储资源使用状态发生变化;在进行超高清视频内容生产时,不同视频拍摄机位数量及不同视频制式变化,会导致不同路径流量负载显著变化,以及网络链路质量的变化,须为业务调度选择最优路由路径,避免拥塞,保障每个机位的传输带宽。因此,需要研究如何及时地对数算业务及 AI 模型根据业务需求及可用资源状况在算力节点间进行联合动态调度,并且根据网络可用资源状况变化进行最优路由路径规划。

1) 超高清视频“数算模”联合调度。调度机制通常可以有 2 种模式:集中式调度和分布式调度^[19]。集中式调度主要在算力网络中央控制器上对超高清视频任务部署到的算力点进行算力资源状态、网络状态等信息的持续监控,一旦这些因素不再符合需求,则自动将业务迁移到符合需求的节点,或者将 AI 模型迁移部署到符合需求的节点。采用分布式调度时,服务迁移决策不再由集中的中央控制器做出,而是由边缘节点基于自身局部状态信息做出决策^[20],将业务迁移到符合需求的节点,或者将 AI 模型迁移部署到符合需求的节点,如图4所示。

针对需要在边缘节点快速处理的超高清视频制作需求,提出一种基于多智能体深度确定性策略梯度(multi-agent deep deterministic policy gradient,

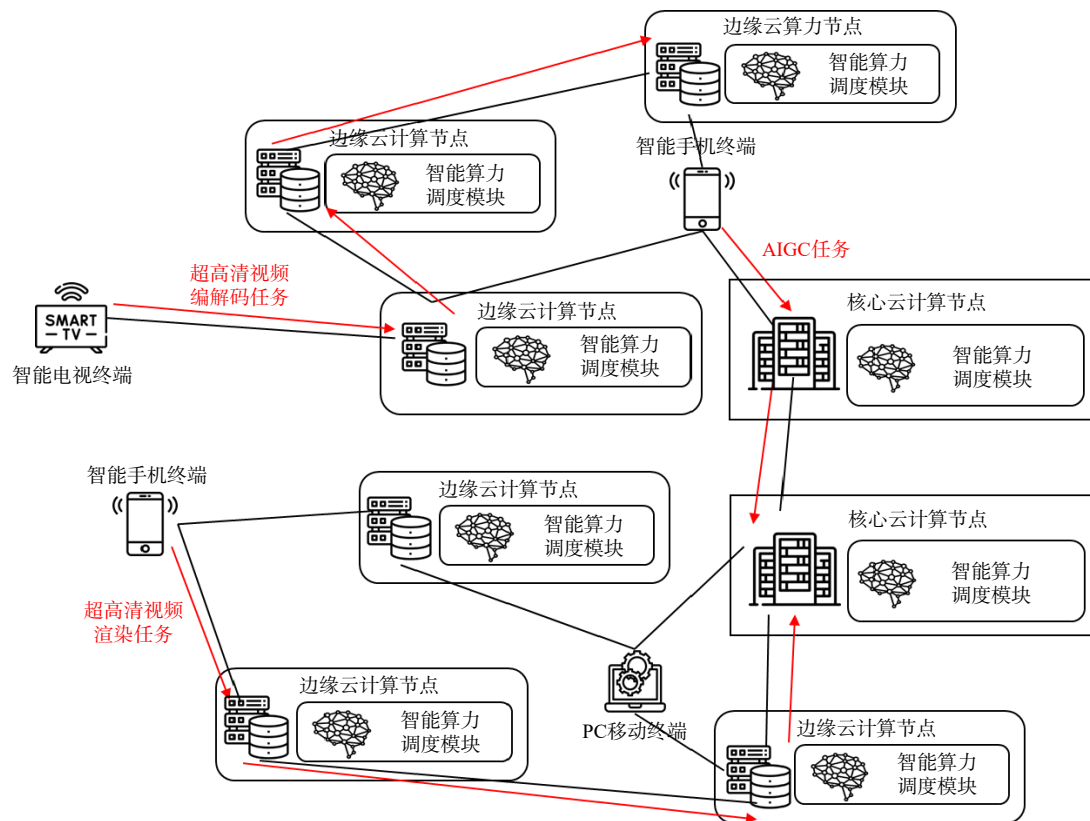


图4 超高清视频分布式任务调度场景

MADDPG)算法的边缘网络分布式任务调度策略。为了实现高效的超高清视频处理任务调度,每个算力节点都部署了训练完成的智能算力调度模块。该模块能够实时感知周围算力节点的资源状态和链路状态,确保对局部环境的了解。在任务调度过程中,智能算力调度模块利用 MADDPG 算法进行决策。具体而言,当一个超高清视频处理任务到达某个边缘云算力节点时,调度模块将基于自身和周边的算网资源状态以及任务需求进行决策,决定是将任务在本节点处理还是转发给某一邻居节点。若转发给邻居节点,则邻居节点重复这一决策过程,直至任务被处理。对于不同类型的任务,系统在智能体模型训练阶段通过引入不同类型的任务进行调度模拟,智能体通过不断更新其策略网络参数,学会在多任务环境中进行动态资源分配,达到任务之间的协同优化。这样,无论是相同任务的并行处理,还是多个任务的协同调度,系统都能够根据实际需求进行合理的资源分配,确保任务的服务质量达到最优平衡。通过这种方式,一个超高清视频处理任务可以经过多个算力节点的协同决策,最终被调度到最优的算力节点上完成处理,同时保障边

缘算力的负载均衡,提升整体性能。MADDPG 算法的优势在于其多智能体的协同能力,能够动态调整调度策略,以适应算力节点和网络状态的变化,确保任务的高效处理和资源的最优利用。

2) 超高清视频业务智能路由规划。在基于智能调度算法将业务或模型调度到最优的算力节点后,需要进一步为后续的数据传输选择最优路由路径。超高清视频数据量大,不同拍摄机位数量及不同视频制式变化,会导致不同路径流量负载显著变化,而且由于超高清视频业务本身流量大,且具有突发性,容易产生网络拥塞^[13],因此,需要动态智能地进行路由路径规划,选择最优路径,避免拥塞,保障每个机位的传输带宽,保证超高清视频业务流量传输高吞吐、低时延的要求。

提出一种基于深度强化学习和图神经网络的算力路由算法,旨在通过智能化的资源分配和路由策略,提高计算网络的性能和资源利用率。这一算法将深度强化学习(deep reinforcement learning, DRL)与图神经网络相结合^[21],通过在智能体(agent)和计算节点之间的交互中学习,自适应地选择最佳的计算节点和

路由路径,实现智能化的任务调度和资源管理。深度强化学习的方法可以帮助智能体在不断地与环境交互中学习到合适的路由策略,而神经网络的引入则有助于对整个算力网络进行建模,提高 agent 学习的准确性。所提出的算法架构图如图 5 所示,其中环境主要包含路由节点、计算节点(包括其中的计算资源),以及它们之间的链路的算力网络,它为强化学习代理创建了一个交互环境。代理执行动作以优化网络性能,环境接受这些动作,并根据这些动作动态演化,给予代理反馈——奖励。奖励机制基于带宽利用率、传输时延、任务完成率等关键性能指标进行设计,旨在激励系统提高资源利用效率、减少延迟。惩罚机制则用于纠正不合理的资源使用或导致网络拥塞的行为。环境会根据算力网络生成的原始数据,选择代表性特征,节点的计算资源和产生的计算任务交由节点选择模块来生成计算需求。代理存储与环境之间的交互生成的轨迹:(状态、动作、奖励、下一个状态),并将这些轨迹提供给 GAQN(graph attention Q-network)进行训练,GAQN 包括 2 部分:GAT(graph attention network)和 DQN(deep Q-network)模型。GAT 的目标是学习相邻节点的重要性,然后将它们融合成代表节点的新特征,最后将获得的新特征输入到 DQN 模型中以学习动作值。为了确保算法的收敛性,在设计中引入了经验回放和衰减策略。经验回放允许智能体在训练过程中反复学习过去的经验,从而减少由于短期反馈造成的学习偏差,避免陷入局部最优。此外,衰减策略逐渐减少智能体的探索率,使其在训练的后期更加注重利用已有的最优策略,确保整个学习过程逐步稳定并最终收敛到全局最优解。这些机制的引入保证了系统在复杂网络环境中的稳定性和可靠性,有效避免了不收敛的情况(图 5)。

3.3 广域算间超高清音视频高速传输技术

在人工智能生成内容(artificial intelligence generated content, AIGC)类业务中,视频、音频、图像、文本和三维模型等各类数据通常由边缘算力中心进行清洗、转换和整合,然后上传到智算中心进行文生视频、图生视频、人物动态化等生成式算法的应用^[22]。例如,传输 1 路 4 K 视频信号需要 50 M 以上的带宽,综艺节目的拍摄现场会设有 100~200 个机位,因此,边缘算力中心需要至少 10 Gbps 的传输带宽才能满

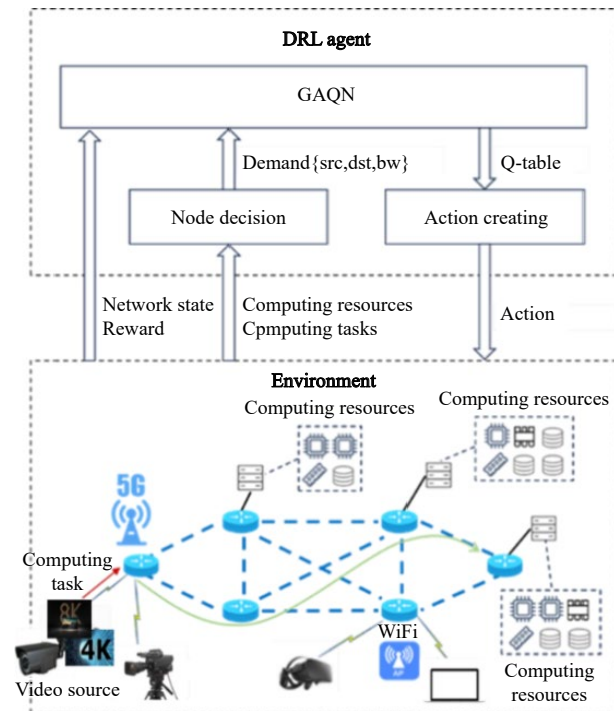


图 5 基于 GAQN 的智能路由架构

足海量视频素材的上传需求。另外,训练好的模型、算法和结果传输回算力中心。由于传输的数据量通常很庞大,传统的传输技术往往无法满足这些需求,常常面临着传输速度慢、稳定性差、延迟高等问题^[23],导致制作效率低下和节目质量不稳定。为满足海量数据流通需求,算间网络需要提供长距离高速传输技术,服务于枢纽、区域、边缘算力中心互联互通。

针对广域网络场景中超高清音视频制作所需的高速传输需求特征,提出基于端侧和网络拥控算法和参数的联动拥塞控制机制,发挥网络节点更易感知网络状态的优势,实时判断链路的传输能力,基于网络传输性能和节点资源状态等信息的融合,建立更准确的拥塞检测模型,并根据应用传输需求为其智能化动态调整拥塞控制算法参数,实现更准确的速率调控^[24],以适配链路状态,形成稳定低延时、高带宽的拥塞控制决策。

基于端网协同的超高清音视频高速传输协议框架如图 6 所示。其核心设计思想在于其随路信令机制的设计,源端在发送数据流的同时会产生一系列的控制信令包,负责端网交互逻辑和传输控制。

具体而言,发送端在新的数据流加入时启动信令处理线程,周期性生成信令包并初始化其头部字段,形成一条独立于数据流的信令控制流。网络节点根

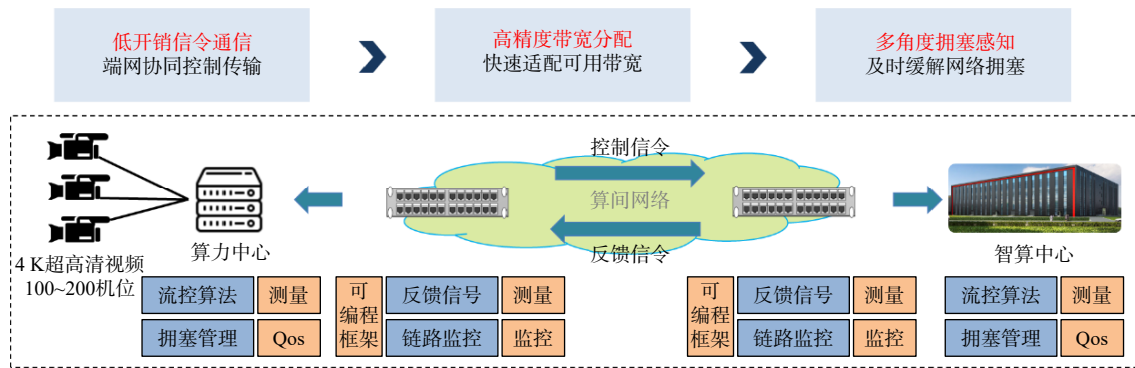


图6 广域算间超高清音视频高速传输协议整体架构

据信令包的报头信息感知并统计所有通过的数据流的信息,得到网络负载状态,进而计算出分配给数据流的公平速率值,将公平速率值作为反馈信息写入信令包中^[25]。信令包到达接收端时,其携带的反映网络负载状态和带宽分配决策的控制字段被完整复制到ACK(确认信号)中,最终回传给源端,从而完成一次端网交互过程。

在广域算间超高清音视频高速传输协议中,网络节点不是在拥塞发生后才反馈拥塞程度,而是采用主动分配带宽的方式,计算数据流的公平速率并直接用于源端的速率控制。带宽分配算法的核心思想是将链路可用带宽平均分配给所有瓶颈流,式(1)是一种计算公平速率的示例

$$g_i(t) = \frac{\gamma_i C_i - u_i(t)}{N_i(t)} \quad (1)$$

式中, γ_i 为出口链路带宽的目标利用率($0 \leq \gamma \leq 1$),用于控制出口链路负载程度的预期。 C_i 为网络节点的出口链路带宽, $u_i(t)$ 和 $N_i(t)$ 分别为非瓶颈流的聚合速率和瓶颈流的数量,网络节点可通过统计信令包的头部信息和计算得出。

在广域算间超高清音视频高速传输协议中,一方面,源端为数据流定期发送信令包,获取瓶颈节点上的长期公平速率;另一方面,源端监测信令包的端到端时延,用于感知网络节点的短期拥塞情况。源端同时将使用公平速率和端到端时延这2种信号来控制发送速率,兼顾了网络可用带宽的长期变化和短期波动。为了避免非瓶颈节点上的排队延迟造成干扰,在网络节点上设置优先级队列。在非瓶颈节点上,信令包将进入高优先级队列,绕过了低优先级队列中的排队过程,而在瓶颈节点上,信令包与数据包一起在低

优先级队列中排队。因此,信令包的实时往返时延(round-trip time, RTT)与历史最小RTT的差值能够反映瓶颈节点的排队情况,从而帮助端侧对传输速率实现实时调整优化。

4 结论

分析了超高清视频产业发展现状及其对现有算力和网络基础设施造成的严峻挑战。为了满足超高清视频采编播过程中大带宽、高算力、低时延需求,在通用算力网络技术基础上,提出了面向超高清视频的算力网络架构,分析其关键技术并提出了初步解决方案。

面向超高清视频的算力网络的目标是基于广电网络和互联网络等的传输能力,汇聚各级算力、超高清视频算法/算据等资源,构建一张广电跨域超高清视频算力骨干网络,在全国范围为超高清视频业务提供多种算力资源支持。同时,借助算力网络对算网资源的统一编排、“数算模”联合调度、高效路由规划与高速传输等核心技术,保障超高清业务处理中多样化算力和网络传输需求。借助算力网络构建的算力设施底座,将极大提升超高频视频制作和播出效率,助力超高清视频产业的快速发展。

参考文献(References)

- [1] Keller D, Rao R R R, Raake A. Assessing video quality differences for 8K/UHD-2 and 4K/UHD-1 HDR videos depending on viewing distance[J/OL]. Authorea Preprints, 2024[2024-10-09]. <https://www.techrxiv.org/doi/full/10.36227/techrxiv.172297401.16801936>.

- [2] 超高清视频产业规模超3万亿元[EB/OL]. (2023-06-03) [2024-10-09]. <http://finance.people.com.cn/n1/2023/0603/c1004-40005477.html>.
- [3] Seetzen H, Heidrich W, Stuerzlinger W, et al. High dynamic range display systems[M]//Seminal Graphics Papers: Pushing the Boundaries. New York: ACM, 2023: 39-47.
- [4] Boren B. History of 3D sound[M]//Immersive Sound. New York, London: Routledge, 2017: 40-62.
- [5] 中国政府网. 中共中央办公厅 国务院办公厅印发《关于推进实施国家文化数字化战略的意见》[EB/OL]. (2022-05-22)[2024-10-09]. https://www.gov.cn/zhengce/2022-05/22/content_5691759.htm.
- [6] Duan L Y, Liu J Y, Yang W H, et al. Video coding for machines: A paradigm of collaborative compression and intelligent analytics[J]. IEEE Transactions on Image Processing, 2020, doi:10.1109/TIP.2020.3016485.
- [7] Zhong L J, Ji X, Wang Z X, et al. A Q-learning driven energy-aware multipath transmission solution for 5G media services[J]. IEEE Transactions on Broadcasting, 2022, 68(2): 559-571.
- [8] Yue L W, Shen H F, Li J, et al. Image super-resolution: The techniques, applications, and future[J]. Signal Processing, 2016, 128: 389-408.
- [9] Zhang M L, Sun X L, Zhu Q Q, et al. A survey of hyperspectral image super-resolution technology[C]//Proceedings of IEEE International Geoscience and Remote Sensing Symposium IGARSS. Brussels: IEEE, 2021: 4476-4479.
- [10] Ma S W, Gao J L, Wang R F, et al. Overview of intelligent video coding: From model-based to learning-based approaches[J]. Visual Intelligence, 2023, 1(1): 15.
- [11] Zhang Z X, Wen F, Sun Z D, et al. Artificial intelligence-enabled sensing technologies in the 5G/Internet of Things era: From virtual reality/augmented reality to the digital twin[J]. Advanced Intelligent Systems, 2022, 4(7): 2100228.
- [12] Liu Y X, Zhang K, Li Y, et al. Sora: A review on background, technology, limitations, and opportunities of large vision models[EB/OL]. 2024: 2402.17177[2024-10-09]. <https://arxiv.org/abs/2402.17177v3>.
- [13] Nightingale J, Salva-Garcia P, Calero J M A, et al. 5G-QoE: QoE modelling for ultra-HD video streaming in 5G networks[J]. IEEE Transactions on Broadcasting, 2018, 64(2): 621-634.
- [14] 中华人民共和国商务部. 国家发改委首次明确“新基建”范围[EB/OL]. (2020-04-20) [2024-10-09]. <https://m.mof.com.gov.cn/article/i/jyj/e/202004/20200402957398.shtml>.
- [15] 算力基础设施高质量发展行动计划[EB/OL]. (2023-10-09) [2024-10-09]. <https://www.gov.cn/zhengce/zhengceku/202310/P020231009520949915888.pdf>.
- [16] 关于深入实施“东数西算”工程加快构建全国一体化算力网的实施意见[EB/OL]. [2024-10-09]. https://www.gov.cn/zhengce/zhengceku/202401/content_6924596.htm.
- [17] Comsa I S, Muntean G M, Trestian R. An innovative machine-learning-based scheduling solution for improving live UHD video streaming quality in highly dynamic network environments[J]. IEEE Transactions on Broadcasting, 2020, 67(1): 212-224.
- [18] Salva-Garcia P, Alcaraz-Calero J M, Alaez R M, et al. 5G-UHD: Design, prototyping and empirical evaluation of adaptive Ultra-High-Definition video streaming based on scalable H. 265 in virtualised 5G networks[J]. Computer Communications, 2018, 118: 171-184.
- [19] Huang S L, Izquierdo E, Hao P W. Bandwidth-efficient packet scheduling for live streaming with network coding[J]. IEEE Transactions on Multimedia, 2016, 18(4): 752-763.
- [20] Shi W X, Li Q, Yu Q, et al. A survey on intelligent solutions for increased video delivery quality in cloud-edge-end networks[J]. IEEE Communications Surveys & Tutorials, 2024: 3427360.
- [21] Almasan P, Suárez-Varela J, Rusek K, et al. Deep reinforcement learning meets graph neural networks: Exploring a routing optimization use case[J]. Computer Communications, 2022, 196: 184-194.
- [22] Liang C S, Du H Y, Sun Y, et al. Generative AI-driven semantic communication networks: Architecture, technologies, and applications[J]. IEEE Transactions on Cognitive Communications and Networking, 2023, 11(1): 27-47.
- [23] Farrow P. Optimising networks for ultra-high definition video[D]. Colchester, East of England, UK: University of Essex, 2016.
- [24] Yuan X J, Pu L J, Xu X H, et al. Streaming-aware cellular resource allocation for UHD video streaming over ultra dense network[C]//Proceedings of IEEE Wireless Communications and Networking Conference (WCNC). Nanjing: IEEE, 2021: 1-7.
- [25] Nosheen S, Khan J Y. High throughput and QoE fairness algorithms for HD video transmission over IEEE802.11ac networks[C]//Proceedings of International Conference on Computing, Networking and Communications (ICNC). Big Island: IEEE, 2020: 84-89.

Computing power network architecture and key technologies for UHD video

ZHOU Xu¹, WU Hong², ZHANG Wei³, SONG Junping¹

1. Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083, China

2. Hunan Mango Innocreative Technology Co., Ltd., Changsha 415500, China

3. Academy of Broadcasting Science, NRTA, Beijing 100866, China

Abstract Ultra-high-definition (UHD) video is one of the key development directions for China's audiovisual industry. In response, relevant national departments have introduced a series of policies to encourage and support the development of the UHD video industry. The processes of capturing, transmitting, producing, and broadcasting UHD video, especially when integrated with advanced AI content generation technologies such as ChatGPT and Sora, exhibit typical characteristics of high bandwidth, high computing power, and low latency. These characteristics pose significant challenges to computing power and network infrastructure. Based on the typical requirements of UHD video and the latest trends in computing and network technologies, this paper proposes a computing power network architecture tailored for UHD video. It comprehensively employs key computing power network technologies such as heterogeneous computing resource networking and secure transmission technologies, UHD video service demand modeling and resource orchestration technologies, joint scheduling and routing planning technologies combining "data, computing, and modeling," and high-speed transmission technologies for UHD video. This approach aims to achieve the aggregation and networking of heterogeneous computing power across the country, thereby meeting the diverse computing and network transmission needs of various stages of UHD video processing, including capture, editing, and broadcasting.

Keywords ultra-high-definition video; computing power network; heterogeneous computing power networking; computing power and network resource orchestration; route planning; high speed transmission technology ●



(责任编辑 王微)