软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software, [doi: 10.13328/j.cnki.jos.000000]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563
录用文稿 + pa020

广域确定性网络传输技术研究综述*

李彤[1,2], 徐都玲[1,2], 吴波[3], 郭雄文[2], 蒋岱均[2], 罗成[3], 卢卫[1,2], 杜小勇[1,2]

[1](数据工程与知识工程教育部重点实验室(中国人民大学), 北京 100872)

[2](中国人民大学 信息学院, 北京 100872)

[3](腾讯科技有限公司, 北京 100080)

通讯作者: 卢卫 E-mail: lu-wei@ruc.edu.cn

摘 要:广域网作为连接新业务、新基础设施和各类新型应用的纽带,已成为21世纪最重要的基础设施之一.近年来,数据量爆炸性增长,伴随着基于广域网的大模型、数字经济、元宇宙和全息社会等新型应用形态的持续涌现,以及东数西算、算力网络和数据场等新型业务架构的出现,业务对广域网的数据传输服务质量提出了越来越高的要求.以时延为例,广域网不仅需要提供及时的服务,还需要提供准时的服务,即时延成为必须满足的确定性指标.因此,广域确定性网络作为广域网的新范式应运而生.本文系统地综述了确定性网络的内涵,回顾了传统确定性网络相关技术发展脉络,介绍了广域确定性网络的新应用,探讨了广域网确定性网络传输具有的新特征以及面临的新挑战,并提出了广域确定性网络的新目标.基于上述新应用、新特征、新挑战和新目标,详细总结了当前广域确定性网络领域的主要研究进展,并给出了未来研究的方向.期望本文能为广域确定性网络领域的研究提供参考和帮助.

关键词: 广域网;确定性网络;数据传输;改良式;革命式

中图法分类号: TP393

Transmission in Wide-Area Deterministic Networking: A Survey

LI Tong[1,2], XU Duling[1,2], WU Bo[3], GUO Xiongwen[2], JIANG Daijun[2], LUO Cheng[3], LU Wei[1,2], DU Xiaoyong[1,2]

[1](Key Laboratory of Data Engineering and Knowledge Engineering (Renmin University of China), Ministry of Education of the People's Republic of China, Beijing 100872, China)

[2](School of Information, Renmin University of China, Beijing 100872, China)

[3](Tencent Technology (Shenzhen) Company Limited, Beijing 100080, China)

Abstract: The wide-area network (WAN) has become a critical infrastructure in the 21st century, connecting new businesses, new infrastructure, and various emerging applications. In recent years, there has been an explosive growth in data volume, accompanied by the continuous emergence of new application forms such as WAN-based large-scale models, digital economy, metaverse, and holographic society. In addition, the emergence of new service architectures such as "East Data, West Computing", computing power networks, and data fields has posed increasingly high requirements for the data transmission service quality of the WAN. Illustrated by latency, the WAN must deliver not only real-time but also timely services, making latency a critical deterministic metric to fulfill. Therefore, the wide-area deterministic network has emerged as a new paradigm for the WAN. This article systematically reviews the connotation of deterministic networks, reviews the development of traditional deterministic network related technologies, introduces new applications of wide-area deterministic networks, discusses the new characteristics and challenges faced by wide-area deterministic network transmission, and proposes new goals for wide-area deterministic networks. Based on the aforementioned new applications, characteristics, challenges, and goals, the main research progress in the field of wide-area deterministic networks is summarized in detail, and future research directions are provided. It is hoped that this article can provide reference and assistance for research in the field of wide-area deterministic networks.

Key words: wide-area network; deterministic networking; data transmission; improved style; revolutionary style

---

* 基金项目: 国家自然科学基金(61972403,61732014,62202473, 62072458);腾讯基础平台技术犀牛鸟专项研究计划;
收稿时间: XXXX; 修改时间: XXXX; 采用时间: XXXX

1 引言

从电子邮件、电子商务、博客、社交媒体,到点播、直播、短视频等,互联网应用传输的对象以文本、图片、音频、视频等形式,呈现出越来越丰富的内容.近年来,这些内容的数据量呈现爆炸性增长的趋势.在这一背景下,大模型、数字经济、元宇宙和全息社会等新型应用形式持续涌现,进而催生了以广域网（WAN, Wide-Area Network）为基础设施的"东数西算"[1] 、算力网络[1] 和数据场（Data Field）[2] 等新型业务架构,进而对广域网的数据传输服务质量提出了更高的要求.国家发展改革委、网信办、工业和信息化部和能源局在《全国一体化大数据中心协同创新体系算力枢纽实施方案》[1] 中明确提出:"工业互联网、金融证券、灾害预警、远程医疗、视频通话、人工智能推理等抵近一线、高频实时交互型的业务需求,数据中心端到端单向网络时延原则上在 20 毫秒（ms）范围内".

确定性网络技术,是一种使网络从"尽力而为（Best-Effort）"到"准时、准确、快速",控制并降低端到端时延的技术.确定性网络,按地域规模划分,可以分为局域确定性网络和广域确定性网络.传统的确定性网络技术通常应用在局域网（LAN, Local-Area Network）范围内,即局域确定性网络.例如,IEEE802.1 时间敏感网络(TSN, Time Sensitive Network)[3] [4] [5] 通过资源预留和时间同步实现确定性低时延,主要应用于运输、发动机控制系统和其他一般工业和车辆应用中的多媒体数据传输等.DetNet（Deterministic Networking）[6] [7] 通过资源预留实现确定性报文转发和路由.目前,这些技术仅专注于针对单一行政控制下或封闭行政控制组内的网络（即域内的网络,如园区网等）的解决方案,对于公共广域网中的传输难题缺乏应对能力.

相比局域网,广域网的数据传输过程具有更大的不确定性.以时延指标为例.首先,网络时延绝对值增大.在局域网内,节点间的网络时延较低,通信时延通常为微秒级或几毫秒.然而,广域网节点之间的网络时延将显著增大,达到几十毫秒甚至几百毫秒.一般地,跨数据中心的网络时延,相比数据中心内,提升了一到三个数量级.其次,节点间网络时延差异性不可忽略.在局域网内,节点之间的时延差异性通常很小.然而,广域网不同节点间的网络情况存在较大的差异性.例如,一些地理位置相近的节点之间的网络时延相对较低（如 20 ms）,但是一些节点距离其他节点的地理位置很远,可能具有较大的网络时延(如200 ms).第三,网络时延动态变化.在局域网内,节点间的时延相对稳定.其主要原因是局域网的端节点和中间节点是可控的,网络运维人员可以部署专用的软硬件来实现稳定低时延（例如,基于 IB（Infiniband）和 RoCEv2（RDMA over Converged Ethernet version 2）实现无损低时延的数据中心网络[8] ）.然而,广域网数据的端到端传输（中间节点不可控）和介质共享特性,使得节点间的网络时延存在较大的波动性.最后,广域网本身中间节点不可控的特征,使得网络内部成为一个"黑盒",由于缺少中间过程的传输状态信息,传统端到端的传输控制难以满足业务确定性的需求.因此,如何在时延绝对值增大、时延差异不可忽略、时延动态变化以及中间节点不可控的新特征下,探索确定性低时延的传输技术,成为广域确定性网络的关键科学问题和核心难点.

本文首先从确定性网络的概念和分类出发,回顾局域确定性网络关键技术的演进,然后从新应用、新特征、新挑战和新目标等多个角度对广域确定性网络和局域确定性网络进行了对比分析；针对广域确定性网络的差异性,分别从改良式和改革式两种路线对研究现状进行了系统的综述.最后,总结并讨论广域确定性网络的未来研究方向.

2 确定性网络的内涵

2.1 确定性网络的概念

随着以太网的发展,越来越多的行业,例如,新型军事应用、航空航天、工业控制系统、能源、物联网、智能交通系统、机器控制和车联网等对于网络的可靠性、低延迟和时序性有着更高的需求.近几年来,更多新兴产业,例如,智能城市服务、VR 游戏、智慧农业、远程医疗、远程教育等的发展和广泛应用,对时延的要求更高——端到端时延需要控制在几毫秒内,对于抖动的容忍度仅限于微秒级.这些行业的发展进一步推动了对可靠、准时、准确网络的需求,也引发了各行业对一种全新标准的、更具"确定性"的网络服务的期望.

业界对确定性网络（Deterministic Network）的概念,并没有一个统一、标准的定义.现有工作在介绍确定

性网络的时候,通常有两种方式.一种方式是将其等同于具体的确定性网络技术,如音频视频桥接(AVB, Audio Video Bridging)[9] 、TSN[3] [4] [5] 或 DetNet[6] [7] ；另一种方式是描述其特征和功能[10] ,如确定性网络旨在提供一种”准时、准确”数据传输服务质量（QoS, Quality of Service）的网络以满足各种应用的需求,确定性是指网络时延、抖动、丢包率、带宽等 QoS 指标是确定的,具体地,如图 1 所示,上限确定的时延、上限确定的抖动、上限确定的丢包率、上下限确定的带宽、下限确定的可靠性等.其中,确定性网络的可靠性是指,在特定时间区间内,能够满足既定时延、带宽和丢包率标准的置信度.从长期来看,置信度（范围[0,1]）不是单次事件的概率,而是关于长期表现的概率估计.

本文也采用一种更通俗的方式来对确定性网络的概念进行定义.如果把传统网络比作公路和马路,它面临长途、拥堵和损坏等不稳定因素,那么确定性网络就好比高铁,路线更短、少拥塞,更加稳定和准时.从目标而言,确定性网络区别于”尽力而为”的网络,是一种”说到做到”的网络,它能够保证数据包尽可能按照预定义的路径,准时到达目的地,从而提供确定性 QoS.其中,确定性 QoS 的指标一般泛指网络时延、抖动、丢包率、带宽等.但方便起见,如果没有特别说明,本文后续将以网络时延指代确定性 QoS 的指标.

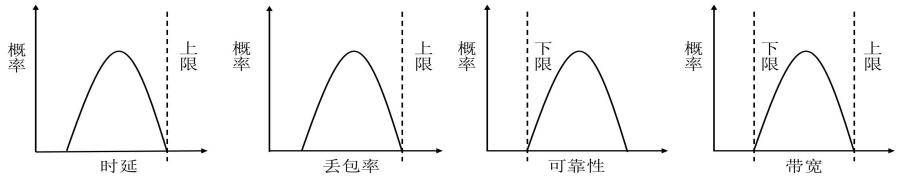

图 1 :典型的确定性 QoS 描述[10]

2.2 确定性网络的分类

确定性网络的分类方法没有统一的标准.传统的互联网,按地域规模可以分为局域网和广域网.同样地,本文将确定性网络划分为局域确定性网络和广域确定性网络两大类.即在局域网中提供确定性传输的网络称为局域确定性网络,而在广域网中进行确定性传输的网络称为广域确定性网络.

（1）局域确定性网络

局域网在我们生活中随处可见,它是小到覆盖一个房屋或者楼宇、办公室,大到覆盖整个学校、医院,甚至更大的覆盖整个工业、生产园区和整个数据中心等范围的网络.局域网通过以太网,连接计算机和网络设备,实现数据传输和通信.传统以太网设计之初,只能根据资源”尽力而为”,缺少科学的标准化和管理,无法提供网络 QoS 保障.然而,新型业务形态如车联网、VR/AR、智慧农业、无人驾驶等,需要将端到端时延控制在微秒到几毫秒级,将时延抖动控制在微秒级,将可靠性控制在 99.999%以上[10] .因此,迫切需要实现局域确定性网络,为局域网提供”准时、准确”数据传输 QoS.

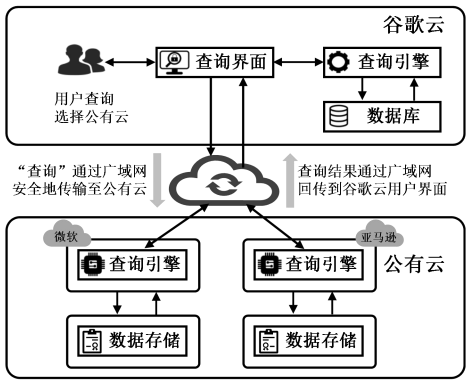

图 2:广域网成为联接 BigQuery 查询业务的纽带

（2）广域确定性网络

广域网通常覆盖了更大的地理区域,通常跨越城市、州、国家甚至跨越大洲.广域网通过使用以太网协议在不同地理位置之间传输数据,当涉及长距离通信和跨越多个地理位置的网络连接,则需要使用不同的通信技术,如传统的电路交换网络、分组交换网络（如 IP 网络）、光纤、卫星链路等,以实现数据传输和远程访问.广域网的传输服务早已无处不在,我们常见的互联网是最大的广域网；大型企业为了实现数据共享和业务协作,通常在不同地理位置建立分支机构,分支机构之间建立连接的也是广域网；在学术体系当中,为了促进学术研究和资源共享,也通常通过广域网连接各个大学、研究机构和实验室；银行和金融机构与他们的分支机构之间为了实现安全的传输和远程交易,需要建立广域网连接；遥感数据和地理信息系统也不例外,为了方便在全球范围内传输遥感数据和实现地理信息系统的集中管理等,也必不可少地引入使用广域网的传输.

新型远距离应用的兴起和广泛应用,如跨域数据管理[11] 、工业互联网、实时音视频、新型经济金融应用和确定性城市服务保障等,对广域网的数据传输 QoS 提出了更高的要求.以多云协同的数据仓库为例,如图 2 所示,Google BigQueryOmni[12] 通过广域网连接谷歌、亚马逊、微软等多个云供应商(Cloud Provider)上的数据仓库（即公有云）,端用户的一次数据查询涉及多云之间的多次实时交互.如图 3 所示,Kubernetes Federation[13] 通过广域网连接多个 Kubernetes 集群,而联邦所需要的功能和数据可能分布于不同的集群,这些集群可能是跨地区(Region)的,也可能是在不同公有云供应商上.这些高频实时交互的业务模式催生了以广域网为桥梁的”多数据中心逻辑上成为一台计算机”（Multi-Datacenters as a Computer）的新网络——广域确定性网络.在这种网络范式中,广域网不仅要提供”及时”的服务,还要提供”准时”的服务,即时延成为了必须满足的确定性指标,确定性要求时延和时延的变化（抖动）是有界的.

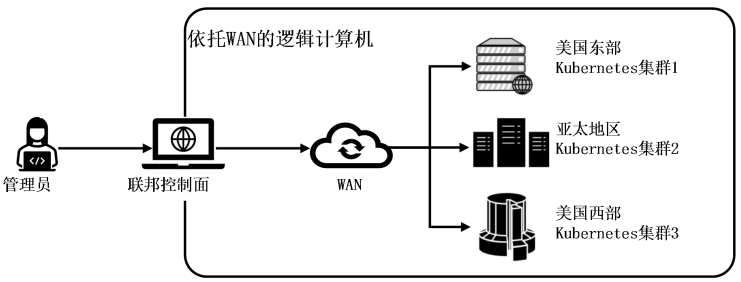

图 3:广域网成为联接 Kubernetes 集群的纽带

2.3 局域确定性网络技术

网络的确定性目标主要包括网络的高可靠性,以及时延、抖动、带宽、丢包率的确定与有界.目前在局域网中,为实现前述目标,已经存在一些比较成熟的技术路线.其中,可以通过时钟同步[3] [4] 、资源预留[14] 等机制实现时延的确定性,通过优先级划分[15] 、通过抖动消减、抖动吸收等机制实现抖动的确定性,通过络切片等机制[5] [15] 实现带宽的确定性,通过多路复用、冗余备份[16] 等机制实现可靠性保障.局域网中的确定性网络技术渐趋成熟,本文主要介绍其中的几项主流技术,具体包含:灵活以太网（FlexE, Flex Ethernet）[15] ,时间敏感网络 TSN[3] [4] [5] ,确定网 DetNet[5] ,确定性 WiFi（DetWiFi, Deterministic WiFi）[14] 和 5GDN（5G Deterministic Networking）[16] 等.

2.3.1 FlexE

FlexE[15] 是一种灵活的以太网技术,允许在非标准速率下进行配置,通过在物理链路上划分多个虚拟子通道来适应不同带宽需求.在以太网的中间层增加了 FlexE Shim 层,采用 FlexE 切片和 FlexE 映射概念,实现数据在子通道间的独立传输和非标准速率的映射.此外,FlexE 还引入了交叉传送技术,在 PHY (Physical Layer)层直接传输数据,避免了中间节点处理延迟,提高了传输效率.

2.3.2 TSN

相比于 FlexE 致力于满足定制化服务需求,TSN[3] [4] [5] 专注于在以太网上实现高实时性,通过时间同步

（使用 IEEE 1588 PTP[17] 协议）确保设备协同工作,并通过流量调度分配优先级,确保关键流量低延迟、高可靠性.TSN 使用时隙机制和公平信道分配（FCA, Fair Channel Allocation）来避免流量冲突和拥塞,实现节点间的有效数据传输和资源共享.

2.3.3 DetNet

DetNet[5] 旨在确保网络中数据传输的延迟、抖动和可靠性是可预测和可控的,使用 PTP 和 SyncE（Synchronous Ethernet）[18] 协议实现时钟同步,并采用灵活的流量调度机制,支持网络切片以适应多种应用.在资源规划方面,DetNet 结合集中式和分布式路径设置,通过中央调度器进行全局路径选择,同时允许节点间协调进行局部流量调度.其配置模型关注于流配置,将数据流与网络资源关联,确保通信的可靠性和时序性能.

2.3.4 DetWiFi

DetWiFi[14] 技术对 WiFi 网络进行了改进,通过时间同步和资源调度技术增强了网络的确定性,解决了因无线信道随机性和干扰导致的性能问题,以提供稳定且可预测的网络服务.它由管理器、接入点、站点以及连接的传感器和执行器组成,管理器分配时隙表以协调网络,确保数据传输的确定性,适合对性能和实时性要求高的场景.

2.3.5 5GDN

5GDN[16] 在 5G 网络中引入时间同步等确定性网络技术,以提供对关键应用和服务的特定保证,满足对高性能和实时性的需求.这项技术强化了可靠性和安全性,具备冗余容错和增强安全措施,适应网络故障和安全威胁.5GDN 适用于工业自动化、智能交通等对通信要求严格的领域,通过高级 QoS 支持,它提升了用户体验并开启了新的创新可能.

表 1 局域确定性网络技术对比

技术名称	网络层级	适用范围	关键技术	优点	缺点	技术成熟度	服务质量
FlexE[15]	L1.5	数据中心、局域网、光传输网	交叉连接技术	高灵活性 高容错性	应用范围有限	实验与商用阶段	-
TSN[3] [4] [5]	L2	有线局域网	时间同步、流量调度、丢包恢复	支持有限局域网中较多的应用	网络拓扑受限,兼容性欠缺	实验与商用阶段	微秒级
DetNet[5]	L3	园区网络或城域网	资源分配、服务保护、显式路由	支持灵活的资源分配和调度方式	资源消耗大,增加网络的复杂性和成本	标准制定阶段	毫秒级
DetWiFi[14]	L1-L2	无线局域网	调度算法和资源管理机制	一定程度提升 WiFi 确定性	无线环境存在干扰和抖动,影响确定性性能	实验阶段	微秒级
5GDN[16]	L1-L3	无线接入网	5G 差异化网络、专属网络	实现 5G 接入网的确定性	需要大量频谱资源,安全性存在挑战	实验与商用阶段	微秒级

2.3.6 对比分析

表 1 展示了几种局域确定性网络技术的比较.FlexE 因其高度灵活性,能在单一物理链路上支持多速率流

量,有效利用资源.TSN 引入了实时通信到标准以太网,确保了同步和低延迟传输.DetNet 以其独特的部署和算法保障了通信的可靠性和及时性.DetWiFi 和 5GDN 则基于原始架构扩展,增强了性能和可靠性,满足了广泛应用场景的需求.

3 广域确定性网络与局域确定性网络对比

不同于局域网,在广域网中实现确定性网络难度更大.下面,从"新应用、新特征、新挑战、新目标"等四个方面,对广域确定性网络与局域确定性网络进行对比分析.

3.1 新应用

相比于局域确定性网络,广域确定性网络的新应用大致可以分为两类,一类是原本在局域网中已经能够保证确定性的应用,需要在广域网中实现确定性,例如,工业内网向工业外网的扩展,单数据中心向跨域数据中心的扩展;另一类是原本在广域网的应用,由于业务扩展,其新应用分支需要确定性的支持,例如,远程医疗、云 VR/AR、甚至未来的元宇宙,都属于远程实时音视频的应用分支,其对于确定性的要求,比传统媒体应用的确定性要求更高.下面,我们举例其中具有代表性的新型应用.

(1)跨域数据管理[19]

数据中心是数字化转型的重要基础设施,伴随着数字经济时代数据要素流通的大趋势,企业数据中心的应用日益广泛.同时,为满足跨区域运营以及异地容灾备份等需求,华为云[20]、阿里云[21]等越来越多的组织和企业在不同地域部署多个数据中心[22].例如,华为云在国内布局了五大数据中心,分布在贵安、乌兰察布、京津冀、长三角和粤港澳地区.当进行跨域数据管理时,业务处理常涉及多数据中心交互.相比于单数据中心,传输时延显著增大,同时,连接多数据中心的各节点之间的广域网的网络状况差异较大且动态变化,为多数据中心的设计增加了困难.以分布式事务处理为例[23],每个阶段的耗费时长取决于 RTT 最大的子事务的时长.因此,基于传统单数据中心的数据库调度策略、共识算法等设计在多数据中心中可能会面临性能下降、可靠性降低等问题.另外,在跨域数据同步过程中,为了保证数据正确性和一致性,共识协议起着关键作用,确保分布式系统的一致性.然而,如果跨域传输的时延增大或时延变化剧烈,共识操作的及时性和正确性会受到影响,导致数据库一致性达成时间延长,降低整体性能.

(2)新型端边云交互式应用

随着广域网的发展,分布式多节点协同应用,结合云计算和边缘计算,正逐渐取代传统单一节点服务模式,提升服务质量并在市场中占据重要份额.这种端边云交互式应用已广泛应用于远程人工智能、实时互动服务和个性化推荐等领域,通过分析用户数据提供定制化体验.在端云交互应用中,最典型的是实时音视频应用.由于流量激增,实时音视频正支撑着万亿元级别市场规模的电商、社交行业,成为互联网产业的底层新基建之一.互动直播、视频会议、远程医疗等均属于远程实时音视频的应用场景.相比于传统的点播音视频可以容忍秒级的延迟,实时音视频对于时延的要求更高.此外,云 VR 游戏云 AR/VR 新应用,也是一种典型的交互式在线视频流[24],还可以用于内容分发(视频和游戏)、远程医疗、沉浸式教育、协同办公等领域.为了保障流畅的用户体验,云 AR/VR 对于数据传输的低时延和抖动都有很高要求.各类实时音视频的数据需要通过广域网进行传输,考虑到广域网的基础传输时延以及主干网络拥塞等可能出现的问题,实现低时延的目标仍然面临技术挑战.

(3)工业互联网

工业互联网的网络互联可以分为工业内网和工业外网.工业内网负责让工厂内各要素得以互通互联.工业外网负责让工厂与厂外要素互联,包括分支机构、工业云数据中心等[25].当前的工业互联网的确定性局限于工业内网(局域),但在广域网中的确定性难以保证.在工业生产的研发设计环节、实际生产环节与机器运维环节,如果广域网能够参与其中并提供一定的确定性,将大大提高生产质量与生产效率.例如,在生产环节,将现场采集的视频或图像实时回传,将实现实时监控,有助于实时分析风险水平,及时发现隐患;在日常运维环节,通过将设备巡检情况上传至云端,可以实现高效运维[26] [27].这些应用场景,都需要广域网的参与,同时对网络的带宽、时延等有较高要求.

（4）新型经济金融应用

确定性网络如今已经与新型金融经济应用广泛结合,深度交融,以高频交易（HFT, High-Frequency Trading）为例,高频交易是指利用先进的计算机算法和高速网络连接,以极快的速度完成交易的一种交易方式.高频交易通常采用自动化的交易系统,可以在毫秒级的时间内进行交易,从而获得极高的收益.高频经济分析市场数据、自动决策买卖等盈利手段都需要大量的数据和算法支持,随着远距离控制交易增多,依托确定性广域网的高频交易应运而生,为金融从业者带来巨大的收益.高频交易比普通交易客户的优势在于能够在极短的时间内进行交易获得利润.更近的距离,或在远距离中拥有更确定的网络无疑让高频交易公司有更多的获益空间.小到个人财富,大到影响国家金融,广域确定性网络,已经扮演着十分重要甚至无可替代的角色.

3.2 新特征

在网络条件和拓扑结构方面,广域网确定性网络的与局域网确定性网络存在显著区别.首先,广域确定性网络的网络条件,尤其是时延特征,呈现出更大的不确定性,这种不确定性又可以细化为网络时延绝对值增大、节点间网络时延差异不可忽略和网络时延动态变化等多个维度；其次,广域确定性网络的拓扑结构,呈现出中间节点不可控的约束.

（1）时延不确定性增大

网络时延绝对值增大.在局域网中,如数据中心网络、园区网络、企业网络,节点之间链路的物理距离短、拓扑相对简单,因此,节点间的网络时延较低,通信时延通常为微秒级,通信时延通常为微秒级或毫秒级.这种时延量级对大部分的应用体验影响较小.例如,在本地 VR 局域网内,人们通常无法感知小于 20 ms 画面的延迟.然而在广域网中,随着节点之间物理距离和跳数的增加,节点之间的网络时延将显著增大,达到几十毫秒甚至几百毫秒.一般地,广域网的网络时延,相比局域网的网络时延,提升了一到三个数量级.因此,在广域确定性网络技术的设计过程中,”通信时延开销忽略不计”的假设便不再适用.

节点间网络时延差异性不可忽略.在局域网内,节点之间的时延差异性通常很小.然而,在广域网中,节点由于硬软件差异大、负载各不相同、地理拓扑位置各异等原因,不同节点间的网络情况存在较大的差异性.例如,一些地理位置相近的节点之间的网络时延相对较低（如 20 ms）,但是一些节点距离其他节点的地理位置很远,可能具有较大的网络时延(如 200 ms).这是广域网相较于局域网或者单数据中心的一个不可忽视的特点.因此,在广域确定性网络技术的设计过程中,”节点之间的时延在同一个数量级”的假设便不再适用.

网络时延动态变化.在局域网内,节点间的时延相对稳定.例如数据中心网络中,其端节点和中间节点是可控的,网络运维人员可以部署专用的软硬件来实现稳定低时延,例如:基于 RDMA 的高速网络技术 InfiniBand 和 RoCEv2 技术.然而,广域网数据传输的网络时延存在较大的波动性.这些波动可能由多种因素引起,例如接入端信号干扰、路由切换、网络拥塞、丢包重传等.这种波动会导致网络时延的不确定性增加,从而影响应用的性能和可靠性.因此,在广域确定性网络技术的设计过程中,"节点之间的时延保持不变"的假设便不再适用.

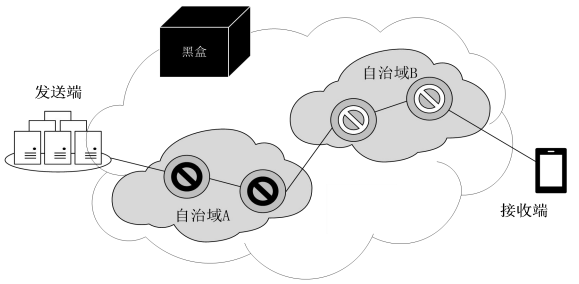

图 4:广域网中由于不同自治域导致节点间传输形成”黑盒”

（2）中间节点不可控

如图 4 所示,广域网是由多个不同运营商控制的自治域组成的,每个自治域都有自己的网络设备和规则,互相之间没有直接的管理关系.这也就意味着,广域网中的节点在软硬件上存在巨大差异,无法完全统筹管理.相

比之下,局域网和数据中心网络通常由同一家公司或组织控制,网络设备和规则比较统一,因此更容易管理和控制.在广域网中,每个自治域中的节点只能控制自己的网络设备和规则,对于整个广域网而言,中间节点的行为和状态很难被端节点精确地监测和控制.这就形成了一个"黑盒",使得端节点无法准确地获取数据在中间节点的传输状态,也无法根据网络变化做出准确的响应.由于中间节点的行为不可控,端到端传输性能的确定性就难以保证.这也是广域网与局域网在拓扑结构方面本质的区别.

广域网网络条件方面的不确定性和拓扑结构方面的中间节点不可控特性,给广域确定性网络中的数据传输的优化和管理带来了新的挑战,我们将在下一节进行详细的讨论.

3.3 新挑战

在广域网传输的各种新的特征下,跨域数据传输所引发的不确定性为我们面临的新型应用（对网络确定性有极高要求的应用）带来了巨大的挑战.为了更好地理解这些挑战,我们将其划分为几个主要类别,并详细分析如何影响网络性能和用户体验.

（1）用户体验降低

广域网网络条件层面的不确定性,会导致端到端数据传输的实时性难以保障,降低时延敏感型应用的用户体验.以跨域分布式数据库系统的一致性共识协议选主问题（Leader Selection）[28] 为例,选主需要在多个节点之间进行协调和竞争,而节点间的网络时延差异的存在,使得选主过程中不仅要考虑副本的版本和超时时间,还需要考虑节点与其它节点之间的时延差异.直观地,与系统中所有节点时延较低的节点,更适合成为领导者（Leader）节点.在这种情况下,如果在算法和协议的设计中没有充分考虑节点间网络时延的差异性,则算法最优性无法保障,即难以实现高吞吐和低时延的分布式系统.另外,时延的差异性可能会引入性能瓶颈,导致整个系统的性能恶化.例如,在互动游戏和实时视频会议中,由于需要实时交互,每个节点的网络时延都会对整个应用的性能产生影响,时延差异过大可能导致在互动过程中产生一方或多方卡顿或等待的问题,严重影响用户体验.

（2）应用不可用

随着时延敏感应用的爆炸性增长,新型应用对低时延的要求越来越高.广域网网络条件层面的不确定性可能会导致应用不可用,甚至引发安全隐患.例如,基于云的 VR 应用中,端到云的时间通常会超过 20 ms,导致应用可用性受到挑战.其它应用,如远程医疗、自动驾驶等,时延超过容忍限度则会面临极高的人身和财产的安全风险.

（3）技术优化难度增大

广域网网络拓扑层面的中间节点不可控特性,会导致端到端数据传输优化的技术优化难度增大.广域确定性网络需要在网络层、传输层、应用层甚至跨层进行设计,而广域网的端到端节点之间会经过大量的中间节点,这就不可避免地需要对中间节点进行修改和配置.然而,广域网中修改中间节点非常困难,导致新技术的部署难度大、演进时间长.

（4）兼容问题难以解决

实现确定性网络服务质量,可能需要对现有的网络设备进行升级或替换,导致新旧设备和系统之间的兼容性问题.例如,传统的普通交换机可能无法提供严格的时间同步、流量控制和优先级机制,需要替换为专门设计的时间敏感网络（TSN）交换机或支持特定协议的交换机；确定性网络可能需要更高的带宽和吞吐量,传统的路由器无法满足要求.因此,可能需要升级或替换为更高性能的路由器,以支持实时数据传输和流量控制等.在新旧软硬件共存的混合系统中,兼容性问题成为影响技术规模应用的主要挑战.

广域网的新特征给广域确定性网络技术研究带来了巨大挑战.为了应对这些挑战,需要采取一系列措施,包括优化网络基础设施、提高传输效率、降低时延,并研发更先进的传输协议和技术,以提供更好的用户服务和体验.然而,由于物理硬件因素的限制,对绝对时延的控制非常困难.物理距离的限制导致传输时延不可忽视,而中间节点的不可控特性使得在广域网中实现确定性研究变得更加困难.广域网复杂的网络环境极大增加了数据传输服务质量的不确定性,使得在广域网中实现确定性服务质量几乎成为一个不可行的目标.因此,有必

要重新思考广域确定性网络的新目标,并探索可行的解决方案.

3.4 新目标

（1）弱确定性

假设用时延指标来衡量确定性,那确定性就是指时延有界（Bounded Latency）.传统局域确定性网络通常追求百分百的确定性,意味着对于时延有界的置信度（Confidence Coefficient）φ =100%.然而,广域确定性网络面临的新挑战,使得这种百分百的确定性难以实现.因此,本文引入弱确定性的概念,作为广域确定性网络的新目标.弱确定性,同样要求实现时延有界,但是其置信度φ可以小于 100%,即 0< φ <100%.

以互动游戏为例,游戏运营商只需保证绝大部分（如 95%）玩家的体验良好（如端到端时延在 60ms 内）就可以满足业务的正常运营,极少部分（5%）玩家体验较差（如端到端时延超过 60ms）对整个业务影响较小.在这种情况下,如果无法满足传统的确定性指标:"100%置信度情况下的时延在 60 ms 以内",则可以将目标松弛到"95%的置信度情况下的时延在 60 ms 以内",松弛后的目标仍然可以满足需求.因此,弱确定性实质上是对传统确定性的一种松弛,这样的松弛为广域确定性网络的传输技术的研究提供了可操作空间.

（2）满足业务需求

确定性要求实现时延有界,而不是实现时延最低.因此,广域确定性网络的目标是满足业务需求即可.现在业务或者应用按照面向对象的差异,可以划分为面向机器的应用时延和面向用户的应用时延.

表 2 :多种典型业务的时延需求示例[29] [101]

业务分类	业务名称	应用场景描述	时延需求（端到端）	需求原因
面向机器	工业物联网（非实时）	工业设备连接互联网并共享数据,数据在云中存储分析	300 ms 内	要求持续可靠 允许一定范围的传输滞后
面向用户	网络购物	用户进行在线购物,优化支付时信用检查的确认时间	150-200 ms	提供良好的用户体验和交互性确保顾客购物顺畅进行
面向用户	高清视频流	提供高清晰度的视频内容	150 ms 内	实现流畅的实时流媒体体验
面向用户	视频会议	双向、多点的实时音视频交互	100 ms 内	实时交流需要较低的延迟
面向用户	多人互动游戏	云侧接受处理渲染工作,最后传回给用户侧	60 ms 内	保证良好的用户体验 避免延迟造成不适
面向用户	云 VR 中的焦点渲染	视网膜定位渲染虚拟现实,通降低用户感知的延迟	50-70 ms	优化图像渲染 减少计算导致的延迟
面向用户	互联网电话 VoIP	通过互联网协议（IP）传输实时音视频	50 ms 内	避免通话中的声音延迟
面向机器	智能电力传输网	智能电力网控制,监测和控制电力传输	10 ms 内	实时监测和控制电力传输 确保稳定性
面向机器	远程超声波检测	远程对超声波信号进行监测和传输	10 ms 内	满足面向机器超声波信号 感知的超低时延
面向机器	工业自动化控制	远程对工业应用进行自动化精确控制	10 ms 内	满足面向机器自动化控制需求的 超低时延

面向机器	高频交易	用于处理经济、金融领域的股票等进行高频率交易操作	2-3 ms	快速执行交易订单以利用价格波动信息进行盈利

面向机器的应用时延中,业务类型不同时延有较大差异.如表 2 所示,在一般的非实时性工业互联网应用中,端到端时延需求仅需要在 300 ms 以内即可,而在工业自动化控制、超声波监测等时延敏感应用来说则需要 10 ms 以内.在面向用户的应用时延中,以多人互动游戏为例,一般认为只要将端到端时延优化到 60 ms 以内,就可以保证良好的用户体验.由于人的反应时间大多集中在 200-300 ms 之间,最极限的情况也只能达到 100 ms 附近,因此,如果进一步将时延优化到 30 ms 以内,用户体验并不会有明显提升,在这种情况下,60 ms 就是满足多人互动游戏需求的优化目标.综上所述,广域确定性网络是以满足业务需求为出发点的,这也为广域确定性网络中的传输技术研究提供了空间.

3.4.1 小结

广域网作为联接新业务、新基础设施和新社会的纽带,其重要性不言而喻,业界对其确定性网络服务质量的追求一刻也不曾停止.当局域网的确定性业务需要在广域网中部署、或需要为广域业务提供确定性服务质量保障的时候,广域网的时延绝对值增大、节点间时延差距大、时延动态变化明显以及中间节点不可控等特征,可能导致原本在局域网中能轻易保障确定性性能的应用在广域网中不可用的情况.此外,为广域场景设计的新技术和新协议,需要考虑广域传输固有的不确定性并容忍这种不确定性,而传输路径太广、拓扑过于动态复杂也导致传输的中间阶段可控性太弱,确定性服务质量保障技术的研发和迭代优化难度大.可以预想到,在短期内,通过在广域网中为所有业务流预留充分的资源以满足业务要求的思路,在公网上难以普及.除此之外,其他多种按照局域网确定性技术路线来进行迭代的思路也无法直接复用.在这种情况下,本文引入弱确定性的概念,作为广域确定性网络的新目标,以满足业务需求为出发点,为广域网提供确定性服务质量提供更大的操作空间,使得实现广域网确定性网络成为可能.

4 广域确定性网络技术研究现状分析

广域确定性网络分为改良式（Dirty Slate）和革命式（Clean Slate）两种技术路线.改良式路线是指在现有互联网的基础上进行增量式修补,实现平滑演进；而革命式路线是指突破限制,重新设计新一代互联网,从根本上解决确定性 QoS 问题.下面,针对两种路线的典型方案及其关键技术展开讨论.

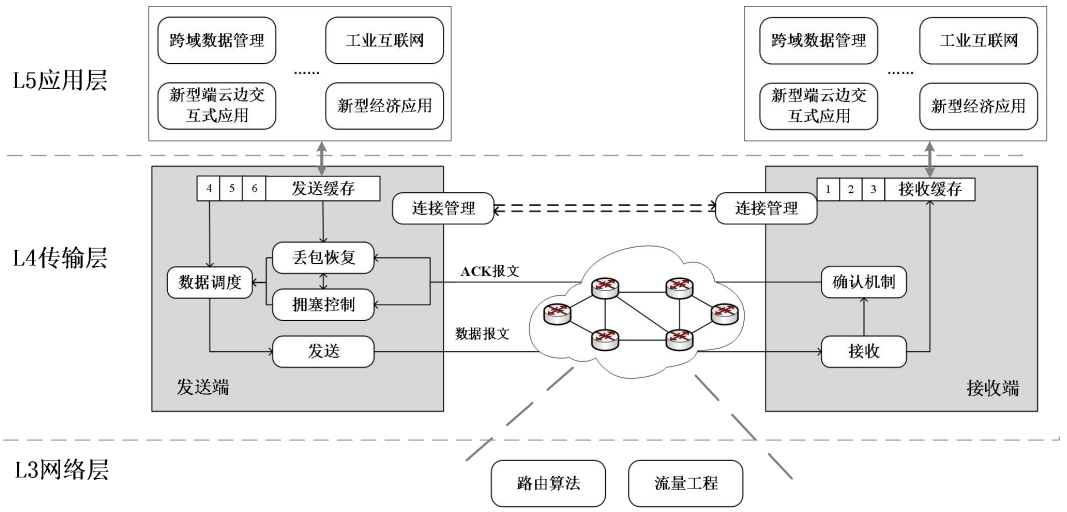

图 5:广域确定性网络改良式路线中的数据传输技术沙盘示意图

4.1 改良式路线的典型方案

前面提到,局域确定性网络技术通常部署在网络体系结构的 L1-L3 层.由于广域网呈现出中间节点不可控

等新特性,因此改良式广域确定性网络技术通常部署在网络体系结构的 L3-L5 层(即网络层、传输层和应用层).图 5 列出了广域网端到端数据传输过程中 L3-L5 层的主要功能模块的技术沙盘.其中,网络层确定性网络技术主要包括路由算法(Routing Algorithm)和流量工程(Traffic Engineering)等;传输层确定性技术是本章讨论的重点,涵盖整个端到端传输控制生命周期中的各个子模块,包括连接建立(Connection Establishment)、确认机制(Acknowledgment Mechanism)、丢包恢复(Loss Recovery)、拥塞控制(Congestiong Control)和数据调度(Packet Scheduling)等;而应用层确定性技术与具体应用场景强相关,缺少通用性,因此本文仅示例性地针对跨域数据管理和新型端边云交互式应用中的实时音视频两个典型场景进行探讨.

4.1.1 网络层确定性技术

路由路由算法是网络中用于确定数据包从源到目标的最有效路径的规则和计算方法.流量工程是对网络流量的分配和路由进行优化的过程,以提高网络的性能和效率.路由算法往往与流量工程结合应用,为广域网场景下的数据传输提供 QoS 保障.下面,将分别从路由算法和流量工程两方面来讨论面向网络层的确定性技术.

(1)基于路由算法的确定性技术

边界网关协议(BGP, Border Gateway Protocol)[30] 作为一种路径向量路由算法[30] [32],主要应用于不同自治系统之间的路由选择和信息交换,具有高度自治性和可扩展性,支持复杂的多级 AS 结构和冗余机制.然而,由于其配置相对固定,BGP 在动态广域网环境中响应网络变化较慢,导致较长的收敛时间,难以有效保证服务质量(QoS)的确定性.业界提出分段路由(SR, Segment Routing)[31] 技术解决上述问题.相比于 BGP,SR 具有可编程性和更强的灵活性,支持通过编程实现更灵活的路径控制,因此能更好地应对网络变化,从而确保广域网确定性 QoS.SR 对状态信息的依赖较少,能够更好地实现对复杂的广域网的管理和维护.SR 技术基于源路由的网络编址和转发机制,提升网络灵活性,更适应广域网的复杂拓扑结构的不确定性.本质上,SR 技术采用了分治思想,为路由算法的设计提供最大的灵活度,使得在广域网上实现确定性 QoS 成为可能.SR 技术允许网络管理员根据网络需求动态配置路径.管理员可以指定特定的路径,也可以根据网络状况和流量工程的需求来选择路径[32] .此外,SR 还有许多变体用于与现有网络架构兼容,如基于 MPLS 数据平面的 SR-MPLS[33] 和基于 IPv6 数据平面的 SRv6(Segment Routing over IPv6)[34] .

(2)基于流量工程的确定性技术

在广域网络中,不同自治系统间的流量管理策略会影响数据传输的时延.为了有效调配跨自治系统的流量,通常采用多协议标签交换(MPLS, Multi-Protocol Label Switching)[35] 技术,这涉及标签交换、压缩等方法.MPLS 技术通过为数据包添加标签来指导其传输路径,通过标签压缩提高网络效率,并通过设定特定的传输路径和优先级来保证流量的质量和降低时延.其中,基于 MPLS 的域间流量工程主要包括 WFATE(Weighted Fair Allocation of Traffic Engineering)[36] 和 RSVP-TE(Resource Reservation Protocol with Traffic Engineering)[37] 等算法.WFATE 与 RSVP-TE 分别基于加权公平队列调度原理,建立资源预留路径来实现对广域网中流量传输的公平性保障与资源控制.但是,WFATE 与 RSVP-TE 所具有的灵活性并不能很好地与广域网复杂的拓扑结构相适应.因此,业界提出了具有更高灵活性和自定义配置的基于软件定义网络(SDN, Software Defined Network)的广域网(WAN)解决方案 SD-WAN(Software-Defined Wide-Area Network)[38] .SD-WAN 通过集中控制器和网络虚拟化提高 WAN 的灵活性和管理性.它通过软件定义配置和监控网络,根据网络状态和需求智能选择路径,并支持如 MPLS 和 4G/5G 多种连接类型以适应不同需求.SD-WAN 的中心化管理和灵活性促进广域网的确定性 QoS 实现.

表 3:基于网络层提升网络确定性的方案总结

技术类别	技术名称	关键技术	优点	缺点	发展现状
路由算法	BGP[30]	通过 AS 路径长度与 BGP 属性进行路径择优	稳定可靠 路由多路径 灵活性较高	缺乏全局视野 收敛时间较慢	广泛应用于大规模互联网,但已无法满足增长的网络规模
	SR[31]	源节点路径信息编码标签	高灵活性 便于管理维护细粒度控制	配置复杂 需引进新设备	处于新兴阶段,随着 SDN 和网络虚拟技术发展,其正得到不断推广和改进
流量工程	基于 MPLS 的域间流量工程[36] [37]	标签交换技术	快速转发 细粒度控制	配置复杂 需要专用硬件支持	广泛应用于较大规模的企业网络与服务提供商网络
	SD-WAN[38]	虚拟网络 智能流量工程	负载均衡 降低专线连接成本	面临安全性挑战配置复杂 对带宽要求高	处于快速发展阶段,部分运营商已采用其以优化广域网连接

表 3 概述了几种网络层确定性技术.BGP 根据路径属性进行路由选择,适用于大型互联网环境,但其收敛时间长.SR 技术较新,通过源路径标签选择,正面临配置和成本问题.流量工程技术,如基于 BGP 和 MPLS 的方法,尽管有性能和硬件限制,但快速转发和负载均衡被广泛使用.SD-WAN 具有虚拟化和灵活性,在解决复杂广域网问题上显现出发展势头

4.1.2 传输层确定性技术

传输层确定性技术涵盖了许多模块,下面按照传输层生命周期,依次介绍连接管理、丢包恢复、拥塞控制、确认机制和数据调度的确定性技术.

（1）基于连接管理的确定性技术

接管理按流程主要包含连接建立、连接过程管理以及连接关闭几个部分.其中,连接建立的总时间主要包括从发送端发送连接请求到接收端收到确认的往返时间（RTT, Round Trip Time）,同步确认报文（SYN-ACK, Synchronize Acknowledgement）响应的延迟以及建立连接队列延迟.在连接建立方面,广域网单次连接通常需要跨越较长距离,加上面临网络拓扑复杂、网络拥塞和丢包风险增加等困难,广域网连接建立时长相比局域网明显升高.为了降低连接建立过程中的时延,相关优化算法往往更专注于减少不必要的连接建立次数,即降低建立连接的频率.例如 QUIC（Quick UDP Internet Connections）[39] 便采用了 0-RTT 技术来减少连接建立过程的时延.

此外,长连接机制[40] 通过延长连接的生命周期以提高通信效率和性能,也是一种可以减少连接建立次数的方案.在长连接中,客户端和服务器之间的连接会在通信完成后保持打开状态,以便随时进行进一步的数据传输,从而避免频繁地建立和关闭连接,减少连接建立的开销和网络传输的延迟.同时,长连接还可以减轻服务器的负载,让多个请求通过同一个连接处理,减少了服务器资源的消耗.需要注意的是,长连接并不意味着连接会永久保持打开状态,其具体的保持时间可以根据应用需求进行配置.虽然长连接在确定性广域网中具有独特的优势和广泛的应用场景,但其依然存在一定的缺陷,比如在大规模部署的场景下长连接的资源占用大,以及连接状态的管理和维护都会变得复杂和困难[41] 等.

针对长连接状态的管理和维护方面的困难,业界提出了 TCP 连接池机制[42] 与 SCTP（Stream Control Transport Protocol）[41] .TCP 连接池机制通过提前创建一组 TCP 连接,并在需要通信时从连接池中获取可用的连接进行复用,从而避免了频繁地创建和关闭连接.其主要目的就是提高连接的复用率和效率,相比长连接,其

实现了不同应用程序间对同一个连接的复用[43] ,连接池在频繁进行短连接的场景下具有更好地性能表现.值得注意的是,连接池的大小需要根据实际使用情况进行合理配置,避免因连接池过大而产生资源浪费,太小而不能达到性能需求.SCTP 则在基础的 TCP 协议上新增了多流传输机制,可以在单个连接上同时支持多个逻辑流,每个流都有各自的序号和控制信息,可以实现多个应用数据流的并行传输,而且根据这些序号,接收端在无序接收这些数据流后可以对其进行还原,实现了无序交付的消息发送[44] .然而,两者都没有解决数据流的队头阻塞问题.QUIC[40]协议中的连接管理机制,通过分帧（Framing）和多路复用（Mulplexing）技术,实现了多流之间的连接复用,并且能够避免行头阻塞,对广域网场景下降低连接建立过程中尾时延具有良好的适用性.另外,QUIC 能够在非首次建立连接时进行 0-RTT 建立连接,即在第一个报文就能携带业务数据.在首次建连,或者长时间没有通信时,采用 1-RTT 建连,减少了重复访问的加载时间,从一定程度上减小了尾延迟.

表 4:相关连接建立机制优点与局限汇总表

技术名称	技术原理	优点	局限
长连接机制[40]	延长连接生命周期, 首次连接后保持连接	减少连接次数,减少连接建立 的开销和网络传输的时延, 减 轻服务器负载	资源占用大 管理维护困难
TCP 连接池[42]	维护预建立的 TCP 连接 按需复用	避免了频繁建连 提高连接的复用	容易造成资源浪费
SCTP[41]	在单个连接上建立多个逻辑流 进行单独控制	多个应用数据流并行传输 接收端无序接收和还原	存在队头阻塞
QUIC[39]	非首次连接时,采用 0-RTT 首次建连时,采用 1-RTT	减少建连的频率 加快连接建立 降低时延和尾时延	冷启动情况下性能退化

表 4 展示了四种技术——长连接机制、TCP 连接池、SCTP 和 QUIC——旨在优化网络连接的管理.长连接机制和 TCP 连接池通过保持和复用连接减少频繁的建连,从而降低服务器负载,但同时可能导致资源占用增加和管理复杂.SCTP 支持多逻辑流传输,提高数据处理效率,但可能遭遇队头阻塞.QUIC 减少连接时延,特别是通过 0-RTT 技术加快连接速度,但在冷启动时性能可能退化.

（2）基于丢包恢复的确定性技术

丢包检测是丢包恢复的前提条件,丢包检测越快,丢包恢复等待时间越短.丢包检测方面,不论是在发送方检测丢包（例如 FACK[45] 和 RACK[46] ）,还是在接收方检测丢包（例如 TCP-TACK[47] 、QUIC[39] ）,丢包检测的优化目标都是尽快检测到丢包,从而减少丢包恢复等待时间.除了丢包检测技术,丢包恢复中最核心的内容是如何在检测到丢包后进行数据恢复.自动请求重传（ARQ, Automatic Repeat Request）和前向纠错（FEC, Forword Error Correction）是两种基本的丢包恢复手段.ARQ 是指发送方检测到丢包事件后,将丢失的数据重新发送,是一种事后补救措施；而 FEC 则是指发送方通过冗余编码技术,在发送的数据中增加冗余数据,接收方即使没有收到部分数据,也可以通过解码冗余数据直接恢复丢失的数据,而不需要等待发送方重传,是一种事前预防措施.下面,分别针对基于 ARQ 和 FEC 的研究工作进行讨论.

自动请求重传（ARQ）.ARQ 包含了许多技术,例如冗余重传、提前重传和多路径重传等,是不同的重传设计思路.他们的共同的目标都是通过加速丢包恢复,在有损网络中提供确定性的时延保证.1）冗余重传,针对广域网小流业务（如 RPC 服务）,优化目标是最小化报文的最大重传次数.例如,通过自适应冗余的 ART[48] 解决广域网中突发连续丢包的问题,以最小的冗余成本来降低丢包恢复的时间；Yan 等提出 TOO[49] ,充分地利用了实时视频帧之间的间隔引起的应用受限（Application Limitation）现象,仅在视频帧之间的间隙冗余重传丢失的报文.相比 ART,冗余重传对后续数据的发送的影响更小.2）提前重传,就是即使数据没有丢失,也选择性地进行重新注入（Re-Injection）.设计思想比冗余重传更激进,完全不受丢包检测快慢的影响.具体地,通过预测数据

包丢失的概率,对有较高概率丢包的数据包进行提前重传（例如 MT-AR[50] ）,或根据业务需要,对实时业务中急需的数据包进行重注入（例如 XLINK[51] ）.3）多路径重传,通过选择质量较好的子流进行重传（XLINK）,减少减少时延敏感流的尾时延,优化数据顺序和减轻队头阻塞.

前向纠错机制（FEC,Forward Error Correction）[52] **.**上述方案都是"亡羊补牢"的方法,而 FEC 属于"防患未然"的方法——不需要在丢包后重传就能直接在接收端恢复出丢失的数据.FEC 机制通常基于编码方案实现,有比特粒度（如物理层 FEC）和包粒度（如传输层 FEC）的数据纠错与恢复.以传输层 FEC 为例,冗余包与普通包的数目比值定义了 FEC 冗余级别,冗余级别越高抗丢包能力越强.发送端通过 FEC 编码引入冗余包,接收方判断是否丢包.当数据包丢失时,如果收到的冗余包数目足以恢复丢失的数据包,则可以通过 FEC 机制恢复丢失包.在考虑带宽开销和恢复程度的前提下,FEC 用可调节的冗余度有效地降低重传概率,从而降低传输完成时间.但是,常规的 FEC 冗余度配置方案无法很好地应用突发丢包[53] ,难以保证传输确定性.因此,进一步研究如何根据数据包类型和当前网络状况,权衡吞吐量和包级别的丢失设置 FEC 冗余级别,成为 FEC 机制设计的挑战.另外,FEC 作为一种机制,能够与多种协议组合使用.例如,RFC 5109[54] 约定了一种用于实现 FEC 的 RTP 载荷格式；FlEC[55] 框架拓展了 QUIC 协议,能够根据不同应用对于时延的敏感程度,在单纯使用重传方案和使用各类 FEC 方案之间做选择；RFC 5052[56] 定义了一个 FEC 框架以便利用 FEC 为内容分发业务提供可靠传输的支持.

表 5:基于丢包恢复的确定性技术总结

技术分类	技术名称	技术原理	优点	缺点	代表文献
自动请求重传	冗余重传	最小化报文的最大重传次数	机制简单,减少重传次数,降低丢包尾时延	传输开销稍大有增加拥塞风险	[48] [49]
	提前重传	即使数据没有丢失,也选择性重新注入.	提前预防丢包恢复较快	丢包难以预测	[50] [51]
	多路径重传	一条子流上丢包后,选择更好的子流对丢失的报文进行重传	鲁棒性强	路径选择复杂部署难度大	[51]
前向纠错	前向纠错	发送冗余数据,在丢包后不需要重传恢复丢失数据	无需重传恢复最快	复杂度大,需权衡带宽开销和恢复程度, 难以应对高丢包、突发丢包,需搭配使用	[52] -[56]

表 5 总结了自动请求重传和前向纠错两种方案下的关键技术和优缺点.ARQ 采用监测丢包（或者预测丢包）和重传的思路,是对丢包的一种"补偿"重传行为,其中的冗余重传将冗余用到了丢失的报文；而 FEC 则是提前预防的一种方案,本质是将冗余用到了原始的数据报文.特别地,提前重传中的重新注入本质上是一种无编码的 FEC,是趋近于"提前预防"的一种重传机制.几种重传思路并不是完全独立的,在实际应用中经常被结合使用以提高丢包恢复能力,例如,XLINK 既是一种多路径重传,又利用了提前重传的思想;又例如,TOO 既利用了冗余的思想对丢失的报文进行了两次重传,还能做到第二次重传是提前注入.ARQ 的传输本身,也带着冗余重传的思想,所以所有的 ARQ 重传方案,都存在增加传输开销的特点.

（3）基于拥塞控制的确定性技术

拥塞控制为广域网提供确定性 QoS 的原理是通过降低数据传输过程中的排队时延,尤其是尾时延,实现端到端时延有界.端到端拥塞控制作为一种防止网络崩溃,提高传输确定性的重要机制,具有非常多的代表性工作,可以分为三类（可见表 6）:基于规则的拥塞控制、基于学习的拥塞控制和混合拥塞控制.

基于规则的拥塞控制.当前广泛部署的拥塞控制算法通常是基于规则的,例如,Linux 内核实现中至少有 15

个内置的基于规则的拥塞控制算法.这些算法通常是由专家根据来自网络的反馈信息（如丢包、吞吐量、延迟）调整数据包发送速率和窗口大小设计的.它们通常优化特定的网络环境.例如,Sprout[57] 、Verus[58] 和PBE-CC[59] 是为蜂窝网络设计的；HACK[60] 和 TACK[61] 则是专为无线局域网（WLAN, Wireless Local Area Network）优化；Compound TCP[62] 、CUBIC[63] 和 BBR[64] 在长距离和高带宽的网络中表现出色.然而,没有单一基于规则的拥塞控制算法能普遍有效,这引发了人们对基于机器学习的拥塞控制算法的探索.

基于学习的拥塞控制.基于学习的算法通常将数据包发送行为交给机器学习,根据网络状态和应用需求,学习发送端拥塞窗口的设置,从而实现对不同的网络场景和应用需求的适应性.典型的基于学习的拥塞控制算法有 Remy[65] 、Aurora[66] 、HTCC[67] 、GLIDER[68] 和 Muses[69] 等.基于学习的拥塞控制算法具有比基于规则的拥塞控制算法更强的通用性,然而,基于学习的算法面临三个关键问题:黑盒设计、高计算开销和内存开销,以及在新的网络环境下性能下降[70] .这些问题阻碍了它们在生产网络中的广泛应用.

混合拥塞控制.为了增强基于规则的拥塞控制算法的通用性,同时应对基于学习的拥塞控制算法的挑战,业界提出了结合两者的混合拥塞控制算法（例如 Orca[71] 、Libra[72] 、Marten[73] 、DeepCC[74] 、Antelope[75] 和Gemini[76] ）.将基于规则的算法引入机器学习模型的主要目标是低复杂度和对未知网络环境的稳健性.为了进一步优化性能,机器学习算法用于操作基于规则的算法的内部.例如,Gemini 试图优化通用拥塞控制算法的内部参数；Orca 试图优化 CUBIC 建议的拥塞窗口；Antelope 试图从现有 CCA 池中选择最佳表现的拥塞控制算法；Libra 根据 CUBIC/BBR 的行为优化拥塞控制算法的逻辑.当然,混合拥塞控制本质上是一种思想,混合的对象不仅可以是基于规则的算法和基于学习的算法,也可以是多种基于规则的算法,甚至多种基于学习的算法.混合拥塞控制具有很好的前景,但是当前缺少统一的、通用的混合多种算法的拥塞控制框架,因此每一种混合算法都需要”重复造轮子”,实现开销较大.同时,也没有解决部署难题.

表 6:各类拥塞控制算法汇总表

技术分类	优点	局限性	代表文献
基于规则的拥塞控制	白盒设计,能够经过特定的设计来优化指定的网络环境	没有单一基于规则的拥塞控制算法在所有环境中普遍有效,缺少普适性	[57] - [64]
基于学习的拥塞控制	对不同的网络场景和应用需求的适应性强	存在黑盒问题、高计算开销和更高的内存开销	[65] - [70]
混合拥塞控制	综合多种网络指标, 更加灵活地适用于复杂的网络环境	需要更多的网络状态信息, 同时需要更为复杂的配置和调优	[71] - [76]

（4）基于确认机制的确定性技术

确认机制,是指数据接收方通过确认报文（ACK）对数据发送方发送过来的数据报文的传输结果进行确认的机制.确认机制与传输控制中的丢包恢复和拥塞控制等功能紧密耦合,在广域确定性网络的研究中占据重要地位.最简单的确认机制是 Per-packet ACK 机制.接收方每收到一个数据报文回复一个 ACK 报文.Per-packet ACK 机制主要用在需要及时反馈和精细控制的网络中[78] ,在广域确定性网络技术中也可以作为首选方案.然而,当存在大量的小数据报文时,ACK 报文开销将不可忽略；一种改进的方法是 Byte-counting ACK 机制,即每累计收到多个数据报文后再回复 ACK 报文.Byte-counting ACK 机制解决了小数据报文开销问题,但是与 Per-packet ACK 机制一样,也是一种以数据报文到达事件驱动的反馈机制,面临着发送方不发数据时无 ACK 反馈的问题.一种解决思路是,不管有没有数据包到达,接收方周期性地回复 ACK 报文,在接收方和发送方之间同步信息, 这种机制被称为 Periodic ACK 机制. Periodic ACK 机制[61] 可以保证在大带宽传输下, 保持一个相对恒定的 ACK 频率. 然而, 当带宽极小时, ACK 频率仍然与大带宽情况保持一致, 显然会浪费资源. 因此, Periodic ACK 机制也无法适应带宽的变化.

Delayed ACK 机制[79] 综合了 Byte-counting ACK 和 Periodic ACK 机制,兼具两者的优点,是现代传输控制协议如 TCP 和 QUIC 所采用的默认确认机制.本质上,Delayed ACK 机制的 ACK 发送频率采用了

Byte-counting ACK 和 Periodic ACK 两者中的 ACK 频率的较大值.Li 等[80] 证明了 Delayed ACK 机制并不是最优的,即在带宽变化时无法保证 ACK 的数目收敛到较低的值.为了解决这一问题,李彤等提出 Bounded ACK 机制[77] ,同样综合了 Byte-counting ACK 和 Periodic ACK 机制,但是其 ACK 发送频率采用 Byte-counting ACK 和 Periodic ACK 两者中的 ACK 频率的较小值.Bounded ACK 机制能够使得 ACK 的频率在不同的带宽场景下保持较低水平, 具有较强的带宽适应性.Li 等进一步提出 Tame ACK 机制[61] ,相比于 Bounded ACK 机制,Tame ACK 机制使得 ACK 频率与 RTT 相关联,不仅具有带宽适应性, 还具有时延适用性.

表 7 总结了 6 类确认机制的优缺点和适用场景.当我们从确认机制的角度考虑保障确定性 QoS 时,需要充分地考虑不同确认机制的特征,根据不同的网络场景和业务需求进行灵活地选择或组合.

表 7:确常见确认机制对比

技术分类	技术原理	优缺点	适用场景
Per-packet ACK	接收方每收到一个数据报文, 回复一个 ACK 报文	ACK 开销大	需要及时反馈和精细控制的网络
Byte-counting ACK	每累计收到多个数据报文后再回复 ACK 报文	解决了小数据报文开销问题,但发送方不发数据时无 ACK 反馈	小数据包较多的网络
Periodic ACK	接收方定期地向发送方发送确认信息	无法适应带宽的变化	带宽相对稳定的网络
Delayed ACK	ACK 发送频率采用了 Byte-counting ACK 和 Periodic ACK 两者中的 ACK 频率的较大值	带宽较大时适应性差, 无法最小化 ACK 频率	默认网络
Bounded ACK	ACK 发送频率采用了 Byte-counting ACK 和 Periodic ACK 两者中的 ACK 频率的较小值	带宽适应性强, ACK 频率有界,无法适应时延变化	时延相对稳定的网络; ACK 开销不可忽略; 能够容忍一定的反馈时延
Tame ACK	ACK 发送频率采用了 Byte-counting ACK 和 Periodic ACK 两者中的 ACK 频率的较小值,并且与 RTT 关联	带宽适应性强, 时延适应性强, 可以最小化 ACK 频率	带宽时延变化大; ACK 开销不可忽略; 能够容忍一定的反馈时延

（5）基于数据调度的确定性技术

数据调度在广域网中实现确定性 QoS 的逻辑是通过调度数据包,使得每个数据包在其截止时间（Deadline）内完成传输.经典的数据调度方法有现进先出（FIFO, First In First Out）和轮询（RR, Round-robin）等,独立地使用这些策略难以满足复杂的网络环境,因此实际系统中的数据调度技术通常会考虑多种因素进行设计.在传输层,我们通常针对多条路径上的数据进行数据包调度,即多路径调度技术.本节主要介绍基于数据建模的调度和时间感知调度方法两种典型的多路径调度技术.

基于数学建模的调度方法把不同路径的拥塞窗口、吞吐量和发送队列等 TCP 层属性输入数学模型,来完成对 TCP 流的情况估算.根据估算结果,模型定量分配发送数据量,缓解乱序.例如: BLEST（Blocking Estimation-based MPTCP Scheduler）[81] 和 STMS[82] 等方案在进行调度器设计时都利用了这个思路.这一类方法的优点是通过减少乱序来缓解异构网络 MPTCP 连接中时延上升和吞吐量下降问题.然而,在建模过程中,为了简化调度器行为逻辑,设计者通常需要进行不完全符合实际情况的前提设定.一般数学建模的方法只输入 TCP 层属性,就通过拟合函数计算得出各子流要分配的数据量.显然,这些因素都使得这类方法被局限在特定的网络环境中,对于不同网络条件的自适应性较弱.

为了能够保证按时、快速交付,时间感知数据包调度（TAS, Time-Aware Scheduling）[83] 思想出现.运用 TAS 可以根据时间戳划分块边界,以及将数据包分类为严格优先级.通过为每个流分配一个固定的时间窗口,TAS 可以提供可预测的传输延迟,并避免网络拥塞.TAS 已经被广泛应用于实时视频和音频传输、智能交通

等领域.基于 TAS 进行设计能够保障网络交付确定性中的准时、快速.目前已经涌现出大量的方法.例如,旨在减少数据块下载时间的多路径调度器 DEMS（DEcoupled Multipath Scheduler）[84] ,DEMS 包含三个关键设计决策,一是了解传输块边界,并在战略上将块提交边界和块交付路径解耦,其次是确保接收方同时完成子流,最后是允许路径进行少量冗余数据（精确控制冗余并给出了冗余量上界）的交付以应对不确定的网络状态.它的实现旨在迎合手机等连接多种网络的移动设备,能够有助于保障接入端稳定.与此类似的,还有 DAMS（Deadline-Aware Multipath Scheduler）[85] ,除了拥有 DEMS 的三项考虑以保障交付的准时和准确以外,还将块传输优先级纳入调度影响因素,保证了传输可靠性,并且在一定程度上提升用户 QoE（Quality of Experience）,改善用户体验.

表 8 总结了三类数据调度算法:队列调度虽简单易实现但适应性有限.基于数学建模的调度能改善乱序和降低时延但在动态网络下存在局限.基于时间的 TAS 调度在时延敏感应用中表现出优势,是广域确定性网络中的主要趋势.

表 8:各类数据调度算法技术分类及优缺点

技术分类	优点	局限性	代表方案
队列调度	思路简单,容易实现和部署	对异构网络适应性差	FIFO RR
基于数学建模调度	缓解异构网络乱序 减少传输时延, 增加吞吐	对网络条件的自适应性较弱	BLEST[81] STMS[82]
TAS 调度	多路径传输的性能和可靠性高	实现复杂度高	DEMS[84] DAMS[85]

4.1.3 应用层确定性技术

应用层技术往往对应着比较确定的应用场景,其在跨域数据管理、工业互联网、实时音视频、新型经济应用等方面都有对应的应用层技术.下面,以跨域数据管理和实时音视频为例,说明现有的应用层技术现状.

（1）跨域数据管理技术

跨域数据管理应用场景中的确定性网络技术,可以围绕如何解决广域网带来的网络通信瓶颈问题来展开.具体地,通信瓶颈有以下两种解决思路:一是降低通信需求.二是提高通信效率,即提升通信效率来增强跨域数据管理系统性能.为了探讨跨域数据管理中能够有助于保证确定性的技术,我们从跨域数据存储和跨域事务处理两方面讨论.

跨域数据存储通过一致性共识协议实现副本的日志同步.共识协议包括领导者选举和日志复制两个核心模块.领导者选举方面可以通过提高日志同步的通信效率来提升系统性能.一种网络时延感知的领导者选举算法 Raft-Plus[101]将节点之间的网络时延作为选举依据,选择时延最优的节点作为领导者.此外,还可以考虑副本的资源利用率和可用性等优化领导者选举算法,如 SEER 算法[86].多领导者系统能更好地支持跨空间域数据的就近读写需求.例如,Droopy/Dripple[87]方案可以根据历史工作负载和系统响应延迟,策略性地配置领导者集合,以减少系统在工作负载不平衡情况下的响应延迟.日志复制方面可以通过降低日志同步的通信需求来提升系统性能,采用这种思路的方案包括.CURP 算法[88]减少通信次数,EPaxos[89]减少通信轮次.此外,还可以像 DPaxos[90]一样,通过将用户数据分片分配到最近的数据中心,降低跨空间域的日志同步时延.然而,这些方案仍存在局限性,如 CURP 和 EPaxos 依赖于操作的交换性假设,DPaxos 适用场景有限等.总的来说,跨域数据存储的通信优化仍在起步阶段,需要深入研究其通用性和可靠性等核心问题.

对于跨域事务处理而言,面对不确定的网络环境,关键问题在于合理调度和编排事务.主要思路仍然是从应用的角度降低通信需求（分为减少通信次数和减少远距离通信）.一方面,RedT[91]通过在执行阶段减少通信次数,并在执行阶段写入重做日志以消除日志准备阶段的同步来降低延迟.Carousel[92]和 Janus[93]则通过将 2PC 协议与共识协议结合,减少了事务提交所需的数据通信次数.另一方面, SLOG[94]通过将事务分为单归属和多归属两类,并利用不同策略选择进一步优化跨域数据库的性能.尽管这些方法为跨域事务处理中的通信优化问题提供了有益的参考,但仍然存在局限性.例如,这些算法都存在强假设,例如,读/写集已知等,而这在大多数跨域数据管理场景下很难满足.此外,这些方案都需要对数据库核心进行修改,部署上也存在挑战.因此,尽管它们都是在

广域确定性网络上,有助于提高网络确定性的技术,但是在严格意义上讲,仍然不算完全能够满足确定性需求的技术,还需要进一步的改良和演进.

（2）实时音视频应用技术

实时音视频是新型端边云交互式典型应用.随着多媒体技术的发展,实时音视频成为互联网内容的主流,并在 IP 流量中占据重要比例。码率自适应（ABR, Adaptive BitRate）技术是实时音视频中最常用的应用层确定性技术,其目标是在带宽范围内提供更高的清晰度,减少卡顿和避免频繁的码率切换.算法将当前时刻所获取和估计的信息,如客户端缓冲区状态和可用带宽等作为输入,求出合适的码率作来实现这个目标.本节以 ABR 为例说明实时音视频应用场景中的确定性技术.Bentaleb 等将 ABR 分为基于客户端、服务器端、利用网内信息辅助和多方混合 ABR,并分析了各类算法的优缺点[114].基于客户端的 ABR 可细分为基于可用带宽、视频播放缓冲区、组合信息和马尔可夫决策过程（MDP, Markov Decision Process）的自适应算法.这些算法存在性能上的局限,如基于可用带宽的算法准确性依赖于带宽估计,而基于视频缓冲区的算法容易导致 QoE 下降.服务端 ABR 不需要客户端协同,但会增加服务器端开销.网内信息辅助和多方混合 ABR 提供更多网络状况信息,但实际部署较为困难.基于客户端的 ABR 在 HAS 框架下实现,该技术在网络视频中占据重要地位,如苹果的 HLS（HTTP Live Streaming）[95]、微软的 Smooth Streaming[96]和 Adobe 的 HDS（HTTP Dynamic Streaming）[97].

应用层技术常对应着比较确定的应用场景,其在跨域数据管理、工业互联网、实时音视频、新型经济应用等方面都有对应的应用层技术.下面,以跨域数据管理和实时音视频为例,说明现有的应用层技术现状.

4.1.4 小结

表 9:改良式路线中各层技术的操作点及优缺点对比

网络层级	改良式广域确定性优化传输技术类别	代表技术举例	操作点（传输流程操作分阶段）	优点	缺点
网络层	基于路由算法	BGP、SR 等	事前	通过中间设备部署优化全局网络,提升网络的传输效率和资源利用率	更新换代慢,新旧产品兼容难,部署较慢
	基于流量工程	基于 MPLS 的域间流量工程、SD-WAN 等	事前		
传输层	基于连接管理	QUIC 等	事中	通用性好部署快速	依赖于端点的实现和配置,极端网络条件下表现不稳定
	基于丢包恢复	TOO、TCP-TACK 等	事后		
		FEC 等	事前		
	基于拥塞控制	Muses、Deep CC 等	事中		
	基于确认机制	TACK、Bounded ACK 等	事中		
	基于数据调度	DEMS、DAMS 等	事中		
应用层	跨域数据管理技术	SEER、RedT 等	事前/事中	可根据应用实现定制化	通用性差效果和实施难度差异大
	实时音视频应用技术	码率自适应等	事前		

如表 9,各网络体系结构的层次中采取了多样策略降低最坏情况发生的概率,即降低尾时延并提升广域网资源利用率.在网络层,主要侧重事前优化,通过基于路由算法和流量工程的方法提前配置和优化广域网路径及流量分配,有效减少网络阻塞.这种策略提高了广域数据传输效率和资源利用率,但面临新旧设备兼容性和广泛

快速部署的挑战.传输层技术包括连接管理、丢包恢复、拥塞控制和确认机制,如基于连接管理的技术在事中阶段调整连接策略以缩短连接时间,但可能依赖于端点的具体实现和配置.拥塞控制技术通过动态调整数据发送策略适应网络变化,有效减少尾时延.优化的丢包恢复技术的事前预先冗余和事后快速重传策略都能减少广域恢复时延.应用层技术,如跨域数据管理,结合事前和事中操作进行定制化优化,但缺乏通用解决方案且实施难度大.总之,通过针对性的优化策略,广域确定性技术在各阶段有效减少尾时延,提升广域网络的传输效率和稳定性.这些策略是降低尾时延和提升服务质量的关键.

4.2 革命式路线的典型方案

近年来,业界提出了多种革命式确定性传输技术.其中最典型的方案有确定性 IP 技术（DIP, Deterministic IP）和增强确定性网络（EDN, Enhanced Deterministic Network）.下面对这两种代表性方案展开讨论.

4.2.1 确定性 IP 技术（DIP）

DIP[98][99]由华为提出,保留传统 IP 转发技术统计复用的优势,通过"时隙+门控"的方案,更近一步实现端到端时延、抖动上界的严格保证.本节将从 DIP 技术的背景讲起,剖析 DIP 关键使能技术——大规模确定性网络转发技术,并总结确定性 IP 技术的特征.

（1）DIP 技术背景

确定性 IP 不同于传统 IP 的关键点在于,DIP 通过减少节点内时延来消除了时延的长尾效应,最终实现确定性时延.

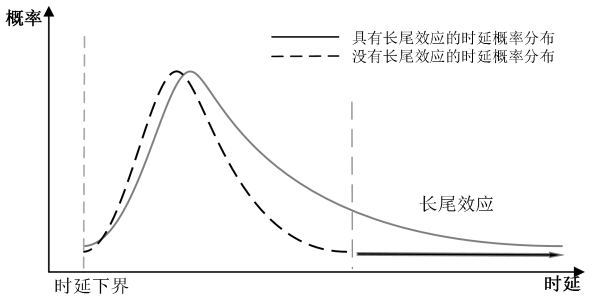

图 6:传统 IP 时延概率分布

在传统 IP 网络时延概率分布图中（图 6）,时延长尾效应明显,呈现出长尾效应的时延由节点内时延和链路时延两个部分组成.其中的链路时延是传输链路上的时延,与传输链路距离和传输速率有关；节点内时延特指两端设备内操作时延,例如排队时延.在传统 IP 时延当中,链路时延取决于拓扑和传输速率,拓扑并不可控,而传输时延又相对稳定,所以 DIP 技术不考虑链路时延,而是尽力减少节点内时延,以压低整体最差时延上界.现有的基于传统 IP 去保障 QoS 的技术已经有许多,例如,资源预留、优先级队列调度等.但是由于在传输过程中,传统 IP 中即使配合资源预留和优先级调度,它们对每个数据分组的行为均仍然缺乏控制,导致了排队形成了微突发.微突发的情况还可能会在下游节点进一步累积,形成微突发迭代.多跳之后,确定性时延最终无从保证,因此,需要能够控制数据分组行为的技术,减少节点内部时延,最终消除长尾效应从而实现确定性时延.

（2）DIP 关键技术

当前确定性 IP 的目标是在利用传统 IP 的转发基础上,提供确定性时延和抖动,而保证这一确定性的关键技术是大规模确定性网络转发技术（LDN, Large-scale Deterministic Network）.上面提到,由于排队数据导致的微突发是影响确定性时延的关键,因此,LDN 技术为了解决该问题,引入了周期调度机制,从而保证了确定性时延和无拥塞分组丢失[100].

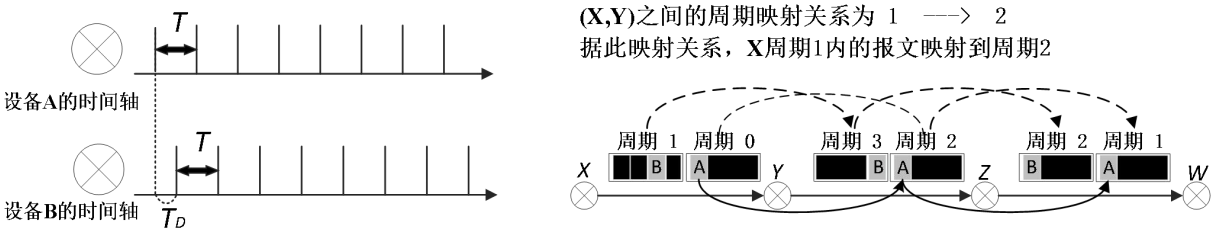

图 7:LDN 设备间频率同步示意和基于影射周期的分组转发[138]

LDN 技术首先要求设备间频率同步,为了做到这一点,LDN 设备将各自的时间划分为等长的周期,合理安排周期的开始和结束时间,使得任意两台设备间的周期边缘差值为定值(T_D),这样就可以在任意两个设备间位置稳定的映射关系.如图 7,设备 X 的周期 1 和周期 0,可以分别对应地影射到设备 Y 的周期 3 和周期 2.这个映射关系约束了两跳之间设备的分组转发行为:数据分组仅允许在规定周期发送,从而能够保证单挑数据的确定性时延传输.设备间周期调度影射关系可以通过多种得到,例如,通过控制面配置、自适应分布式学习等,映射关系的存储也可以非常灵活.

相比于 TSN、DetNet 等局域确定性网络技术,DIP 技术无需网络节点之间的严格时间同步,核心节点无状态,并且支持任意长距离链路,是一种全新的确定性实现路线.

4.2.2 增强确定性网络（EDN）

增强确定性网络在中兴通讯联合中国信息通信研究院及移动、电信、联通等发布《IP 网络未来演进技术白皮书 2.0——开放服务互联网络》[101]首次提出,并与服务感知网络（SAN, Service Aware Network）和内生安全共同组成了开放服务互联网络的三大关键使能技术.本节将从 EDN 的设计背景讲起,然后分别介绍 EDN 的资源层、路由层和业务层的功能及关键技术.

（1）EDN 技术背景

2023 年,《IP 网络未来演进技术白皮书 3.0——增强确定性网络(EDN)》[102]发布,该白皮书在开放服务互联网络架构的基础上,聚焦于当前大规模确定性网络的典型应用场景,分析并总结了 EDN 所面向的业务的四个主要特征——需求多样化、业务规模化、跨域异构互联、资源及转发机制多样化.设计者根据 EDN 的关键需求,设计了 EDN 的三层总体架构,包括资源层、路由层和业务层.

EDN 的设计目标是面向承载网提供确定性服务.为此,从业务需求的角度来看,EDN 需要满足多样化的业务需求及差异化的服务等级协议（SLA, Service Level Agreement）指标,达到差异化确定性服务等级（DD-QoS, Differentiated Deterministic QoS）；从网络需求的角度来看,EDN 需要适用于多个局域确定性网络互联而成的大规模确定性骨干网,并且兼容多样化的局域确定性网络.

（2）EDN 关键技术

基于当前大规模确定性网络的特征与挑战,EDN 定义了逻辑上的三层架构,自底向上分别为资源层、路由层和业务层.同时利用增强确定性网络转发面（EDP, EDN Data Plane）、控制面（ECP, EDN Control Plane）和管理面（EMP, EDN Management Plane）的组件来实现其目标.

EDN 资源层聚焦于解决资源及转发机制的差异性所带来的挑战.EDN 将链路、带宽、队列、时隙等定义为确定性转发资源,将 P2P 链路、MPLS-TE（MPLS Traffic Engineering）隧道、TSN 链路等具有确定性传输能力的 Sub-net 定义为确定性链路,确定性链路具有确定的带宽、节点内时延和抖动.EDN 的功能是对异构的确定性转发资源进行统一建模与分配,构建确定性链路,为不同级别的确定性转发能力提供保障.

EDN 路由层聚焦于解决跨域异构互联所带来的挑战.路由层在资源层所构建的确定性链路的基础上,计算具有确定性保障的路径并下发路径结果建立确定性路由.EDN 确定性路由与传统路由的主要差异在于它们在进行链路状态通告和进行路径规划所基于的指标不同——传统路由所基于的链路属性仅考虑节点外时延,EDN 确定性路由所基于的确定性链路属性考虑了节点外时延和节点内时延.

EDN 业务层在路由层所规划的确定性路径的基础上,根据 DD-QoS 对业务流进行流量调度,为差异化的业

务流提供差异化的服务.EDN 根据业务的多样化需求,将确定性业务分为五类,分别为:带宽保障类、有界时延上限保障类、低时延保障类、低抖动保障类、低时延和低抖动保障类.为保证节点能够对分类分级的业务流进行识别,EDN 需要对报文进一步封装,封装需要考虑多种网络传输格式,包括 IP/MPLS/SR-MPLS/SRv6 等[102].网络入口节点负责对报文的封装,同时会根据业务需求以及当前网络状态,判断是否有足够的资源和合适的时隙允许业务接入；网络出口节点负责剥除对报文的封装.

EDN 架构基于现有 IP 路由体系,其架构需要广域网内的参与节点均支持某些功能,例如,广域网队列机制等；同时,确定性路由、业务报文封装等技术需要对现有协议做扩展；此外,对于差异化的资源与业务,需要推进相关标准指定,以使得网络各节点达成共识.

4.2.1 小结

图 8 展示了不同网络体系结构的层次适用的确定性网络技术.FlexE、TSN、DetWifi 和 5GDN 等局域确定性网络技术主要在 L1 和 L2 层通过链路划分、切片和时钟同步等技术来保障局域网中的确定性服务.EDN 与 DIP 主要通过 L3 的新型队列机制与广域网节点在资源预留上的共识来保障广域网中的确定性服务.DIP 技术是一种颠覆传统 IP 体系的典型技术,其主要使能技术是大规模确定性网络转发,确保长距离链路的确定性传输.而 EDN 技术仍然遵循传统 IP 体系,通过全新的架构设计,包括资源层、路由层和业务层的三层架构,并在各层设计或改进了新的协议,以协同保障确定性网络.DIP 和 EDN 都能够解决广域网下确定性传输中的多种新挑战.然而,重新构建新型设施体系或全面部署新型架构都需要时间和耐心.在当前多企业、多层级、多方复杂需求的背景下,革命式广域确定性网络技术仍然任重道远.

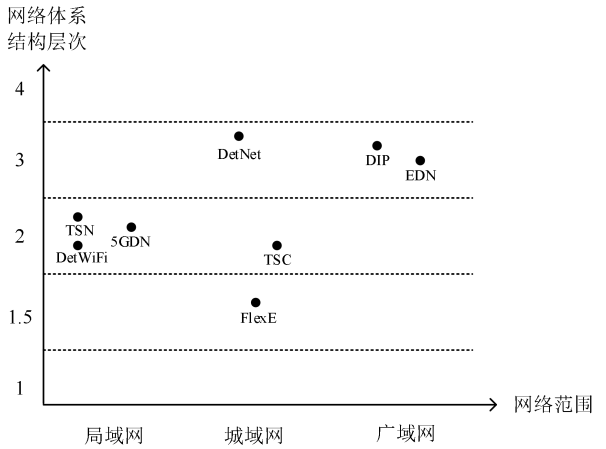

图 8:确定性网络技术分层图谱

4.3 改良式和革命式路线对比

广域确定性网络改良式技术和革命式技术总结如表 10 所示.在网络层,我们探讨了基于路由算法和流量工程的确定性技术.路由算法则关注如何选择最佳路径来传输数据,以提高网络的响应性和可靠性.而流量工程旨在优化网络中的数据流量分布,通过合理的路由选择和流量调度,实现网络资源的高效利用和性能优化.在传输层,本文介绍了基于连接管理、丢包恢复、拥塞控制、确认机制和数据调度的确定性技术.连接管理确保通信双方能够建立有效的传输通道并对连接进行管理,丢包恢复和确认机制保证数据的可靠传输,拥塞控制则关注如何避免网络拥塞并维持流量的平衡.数据调度技术则根据不同的应用需求和网络条件,合理地调度数据传输,以提高传输效率和用户体验.在应用层,我们以跨域数据管理中的优化一致性方法等为例,说明了跨域数据管理领域的应用层技术,描述了在实时音视频中的码率自适应技术.

革命式路线的经典方案中,确定性 IP 技术与 EDN 技术引领了确定性传输的创新方向,采用彻底重建的策

略.文章深入探讨了这两种技术的技术背景和关键技术特点,指出尽管 DIP 和 EDN 为广域网的确定性传输提供了方案,但仍面临众多挑战.技术的落地实施需要时间和耐心,在多元化需求的环境中,其发展亦需相应的时间和资源投入.

通过改良式和革命式两种路线的多种方案的协同作用,能更好地保证确定性网络在广域网中的传输.这些技术的不断发展和创新将进一步推动确定性网络的应用和发展,为实现高效可靠的数据传输提供了基础.然而,目前改良式方案中的技术难以完全满足广域网下的所有确定性需求,而革命式路线仍需要时间演进和发展,并且在许多技术方面仍存在未知性,但是,两种技术路线方案都是不可或缺的.另外,我们需要进一步探索和发掘更多的确定性技术和未来研究方向,以促进广域确定性网络的全面发展.

表 10:改良式和革命式路线技术总表

技术路线	网络体系结构的层次	确定性技术分类	广域确定性研究技术思路
改良式技术	网络层确定性技术	流量工程	优化流量分布, 避免瓶颈
		路由算法	选择最佳路径, 规避低质量路径
	传输层确定性技术	连接管理	减少连接建立延迟, 限制启动时延
		丢包恢复	加速丢包恢复, 控制最坏情况丢包恢复时延
		拥塞控制	避免网络拥塞, 降低排队尾时延
		确认机制	提升传输控制效率, 间接通过丢包恢复和拥塞控制等实现确定性服务质量
		数据调度	调度数据包, 规避低质量路径
	应用层确定性技术	跨域数据管理	降低通信需求和提升通信效率, 减少对网络的依赖, 提高确定性
		实时音视频	应用主动适应网络能力, 提高确定性
革命式技术	网络层确定性技术	DIP 技术	减少节点内时延来消除了时延的长尾效应,最终实现确定性传输
		EDN 技术	面向承载网提供确定性服务

5 广域确定性网络未来研究方向

5.1 基于跨层优化的改良式路线

网络传输技术的优化是一个不断发展的领域,涉及到各个网络体系结构层次的性能提升,为了实现更高质量、更可靠的网络传输,开始尝试打破不同层之间的隔离,探索跨层优化技术.目前,跨层优化技术的研究思路包括利用物理层、链路层、网络层和应用层等各层信息来提高可用带宽、降低传输时延或降低系统能耗并提升性能等.然而,跨层技术优化面临一些实际挑战.首先,由于网络链路的动态变化性强,跨层优化需要考虑这种不确定性带来的影响.其次,在上层应用中,不同数据的重要性差异巨大,因此需要权衡不同数据的传输需求以实现最佳性能.此外,不同网络接口具有不同的能耗特性,跨层优化需要考虑能耗与性能之间的平衡.因此,跨层技术优化仍然需要更多相关研究来应对这些挑战.通过深入探索跨层信息交互和协同优化策略,可以更好地实现网络传输的确定性,并提高网络的性能和可靠性.这将推动确定性网络传输在广域网环境下的不断进步和创新.

5.2 基于多路径的改良式路线

由于近年来云计算、物联网等技术的快速发展,应用对网络带宽的需求日益增加,单一路径的端到端接入技术难以完全满足应用对于带宽的新需求,因此,为了解决带宽需求和频率资源协调困难的问题,业界提出了多路径传输协议,例如 MPTCP[81][82]和 MPQUIC（Multi-path QUIC）协议[103].多路径传输协议目标是实现数据的端到端并发传输,提高传输总带宽和对网络的适应性.具体来说,多路径优化的技术方向包括连接建立、路径选择、拥塞控制和数据调度等.然而,在试图实现这些优化的过程中研究人员遇到了一系列难题.例如,用于分配和

调度各层信息参数都极大地影响着传输子流当中的高效协同,且它们之间有着复杂的耦合关系,这导致多路径传输技术始终难以最大化利用多条路径传输,也使得未来的多路径研究变得非常重要且迫切[104].

5.3 基于广域RDMA的改良式路线

广域RDMA技术,随着网络带宽的扩增和应用需求的增长,在实现高吞吐量和低延迟传输上发挥着核心作用.该技术包括基于有损和无损网络的解决方案.如 RoCEv2 和改进的 RoCE 网卡（IRN ,improved RoCE NIC）等方案通过无损优化技术,结合硬件拥塞控制和丢包恢复策略,在有损网络中也能保证传输效率.现场可编程门阵列（FPGA ,field programmable gate array）技术[105]的融合进一步提升了 RDMA 的灵活性和效率,尤其在选择性重传和乱序接收方面.无损网络技术,如,流级拥塞控制,引入精细的流量管理,而包级负载均衡和多路径RDMA 提升网络效用.广域 RDMA 在跨域存储、安全计算和超高清视频传输等关键领域展现出其强大潜力,预计随着技术进步,其在云计算、大数据和物联网等更广领域将提供高效、智能的解决方案[106].

5.4 基于AI的改良式路线

随着机器学习和人工智能的兴起以及它们在各行各业的广泛应用,人们正在不断探索确定性传输优化技术,并逐渐拥抱机器学习和人工智能.目前,一个重要的发展方向是利用机器学习和人工智能技术来实现更精确的流量预测和网络调度.通过应用这些技术,我们能够更好地预测和管理网络流量,使网络资源的分配更加高效和智能化.这将带来许多好处,如提高网络的吞吐量和响应时间,减少网络拥塞,提升用户体验.未来,我们可以期待在广域网传输优化领域中更多的创新和应用,将机器学习和人工智能技术应用于其中,为我们的日常生活和工作带来更大的便利和助益.

5.5 确定性广域网基础设施建设

如今,随着应用对网络确定性的依赖程度不断加深和要求的日益提高,确定性网络的发展变得至关重要.然而,要不断满足确定性要求,可靠的基础设施支持不可或缺.目前,国内外在基础设施和架构建设方面正在不断创新和进步.随着计算和网络的融合越来越紧密以及云-边-端一体化发展趋势和其应用的广泛扩展,我国不断推进:以用户为中心构建骨干网（如"东数西算"集约化数据中心）、城市群（如区域数据中心集群）、城市内（如城市内分布式的边缘计算资源）三级时延圈建设,在构建三级时延圈的同时,适时推进数据中心直连网络建设,逐步推广超高速光传输系统,有助于提高网络干线的传输带宽,优化网络结构,加快重要路由光缆建设,有助于减少跨区时延.另外,试验验证作为确定性网络技术研究的重要方法而被广泛重视,建立大规模的网络试验设施势在必行,为提供确定性保障的网络新技术示范环境,我国已经启动了未来网络试验设施（CENI, China Environment for Network Innovation）.另外,许多其他国家也开始了试验设施的建设,例如日本的 RISE[107]试验平台,美国的 GENI（Global Environment for Networking Innovation）[108]等等.

5.6 革命式路线技术的标准化

广域确定性网络技术的革命性创新和发展离不开标准化的持续推进.通过更新和维护广域确定性网络的标准,共同的规范和协议确保了不同设备、系统和应用之间的相互通信和互操作性.这有助于不同厂商的设备和不同平台的应用在统一的标准下连接和交互,从而促进了广域确定性网络的无缝连接和互联互通.与此同时,标准化也大大降低了对特定厂商或平台的依赖,降低了运营成本,并提高了确定性网络的建设效率和各项确定性服务的可扩展性.标准化的实施简化了网络管理和维护,减少了复杂性和标准不一致性带来的应用运维等问题.因此,广域确定性网络技术的标准化是推动其创新和发展的关键因素.

5.7 面向确定性用户体验的新型确定性网络

随着广域网在各个生活领域的广泛应用和对更快速、更可靠网络的迫切需求,从确定性服务质量（QoS）到确定性业务体验（QoE）,也是未来的重要研究方向.因此,深入了解企业服务特点,理解用户对应用的需求,敏锐感知用户体验及反馈,明确传输质量要求,达成面向用户体验的新型确定性网络结构的目标.针对不同的企业或用户群体,在不同的应用和场景下,对广域确定性技术的确定性要求各不相同,时延敏感度更是大相径庭.因

此,设计和优化跨层感知技术,例如通过打通应用层和传输层,了解应用层的需求,从而更精确地掌握用户需求,提供更灵活可定制的确定性网络服务,是一个具有重要意义的研究方向.通过这样的设计和优化,资源可以更合理地利用,同时也有助于推动广域确定性网络传输技术的多样化发展,加速传输技术的创新.

总之,未来的确定性网络必定是以公用确定性设施设备为可靠载体,不断推进革命式技术标准化,以用户体验为重要反馈机制,多种多样定制化优化技术适应多样场景应用需求,并确保提供灵活定制、安全可靠以的广域网确定性服务.

6 结论

广域确定性网络,是新业务、新基础设施和新社会发展到一定程度的必要需求,也是传统确定性网络发展到一定阶段的必然产物.随着广域网下的新型时延敏感型应用兴起以及 5G/6G 等新一代网络技术的发展,越来越多的业务将需要高可靠性和低延迟的网络服务,保证时延拉动在规定限制内是刚性要求.因此,确定性网络技术已然成为目前的研究热点,这对新应用的可用性保障和性能提升有极其重要的意义.

本文围绕确定性网络的定义和内涵,基于局域网和广域网的区别,将确定性网络细分为局域确实性网络和广域确定性网络,并对广域确定性网络技术进行了系统地综述.具体地,从"新应用、新特征、新挑战、新目标"等四个方面,对广域确定性网络与局域确定性网络进行对比分析.然后,基于改良式和革命式两种路线对广域确定性网络技术的研究现状进行了综述,并基于当前的研究现状和存在的问题,提出了几个未来的研究方向.

随着跨地域、跨数据中心、分布式计算等技术模式成为主流,广域确定性网络技术将成为未来网络的重要发展方向.尽管广域确定性网络的研究充满挑战,但是其中蕴藏着各种机遇等待我们去发掘.希望本文可以为广域确定性网络领域的研究者和从业者提供一些启发,以促进基于广域确定性网络的新型业务和技术的持续向前发展,打开未来网络的新局面.

References:

[1] National Development and Reform Commission, Cyberspace Administration of China, Ministry of Industry and Information Technology of the People's Republic of China, National Energy Administration. Notice about printing and distributing the implementation plan for computing power hub of national integrated big-data center collaborative innovation system. https://www.gov.cn/zhengce/zhengceku/2021-05/26/content_5612405.html

[2] Wu MQ: Data Space Leading Digital Technology System Innovation. https://ysg.ckcest.cn/ysgNews/1746805.html

[3] Dürr F, Nayak NG. No-wait packet scheduling for IEEE time-sensitive networks (TSN). In Proceedings of the 24th International Conference on Real-Time Networks and Systems, 2016:203-212.

[4] Kehrer S, Kleineberg O, Heffernan D. A comparison of fault-tolerance concepts for IEEE 802.1 Time Sensitive Networks (TSN). In Proceedings of the 2014 IEEE Emerging Technology and Factory Automation (ETFA), 2014:1-8.

[5] Nasrallah A, Thyagaturu AS, Alharbi Z, Wang C, Shao X, Reisslein M, ElBakoury H. Ultra-low latency (ULL) networks: The IEEE TSN and IETF DetNet standards and related 5G ULL research. IEEE Communications Surveys & Tutorials. 2018, 21(1):88-145.

[6] Yang X, Scholz D, Helm M. Deterministic networking (detnet) vs time sensitive networking (tsn). Network, 79. 2019.

[7] Grossman, Ethan. Deterministic networking use cases. RFC 8578, IETF, 2019.

[8] Schelten N, Steinert F, Schulte A, Stabernack B. A high-throughput, resource-efficient implementation of the RoCEv2 remote DMA protocol for network-attached hardware accelerators. In Proceedings of 2020 International Conference on Field-Programmable Technology (ICFPT), 2020:241-249.

[9] IEEE Standard for Local and Metropolitan Area Networks--Audio Video Bridging (AVB) Systems. In IEEE Std 802.1BA-2021 (Revision of IEEE Std 802.1BA-2011) , vol., no., pp.1-45, 17 Dec. 2021, doi: 10.1109/IEEESTD.2021.9653970.

[10] Purple Mountain Laboratories, etc. 2021. Future Network Whitepaper: White Paper on Deterministic Network Technology System. https://www.huawei.com/cn/news/2021/6/future-network-deterministic

[11] Chai YP, Li T, Fan J, Lu W, Zhang F, Du XY. Connotation and Challenges of Cross-Domain Data Management. Communications of the CCF, 2022, 18(11): 29-33.

[12] Lloyd J. Infrastructure Leader's Guide to Google Cloud. Apress, Berkeley, CA, 2022:13-47. https://doi.org/10.1007/978-1-4842-8820-7_2

[13] Larsson L, Gustafsson H, Klein C, Elmroth E. Decentralized kubernetes federation control plane. In Proceedings of 2020 IEEE/ACM 13th International Conference on Utility and Cloud Computing, 2020:354-359.

[14] Cheng Y, Yang D, Zhou H. Det-WiFi: A multihop TDMA MAC implementation for industrial deterministic applications based on commodity 802.11 hardware. Wireless Communications and Mobile Computing, 2017.

[15] Koulougli D, Nguyen KK, Cheriet M. Flexible ethernet traffic restoration in multi-layer multi-domain networks. In Proceedings of ICC 2021-IEEE International Conference on Communications, 2021:1-6.

[16] Prados-Garzon J, Taleb T, Bagaa M. Optimization of flow allocation in asynchronous deterministic 5G transport networks by leveraging data analytics. IEEE Transactions on Mobile Computing, 2021.

[17] IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems. In IEEE Std 1588-2019 (Revision ofIEEE Std 1588-2008) , vol., no., pp.1-499, 16 June 2020, doi: 10.1109/IEEESTD.2020.9120376.

[18] Ferrant JL, Gilson M, Jobert S, Mayer M, Ouellette M, Montini L, Rodrigues S, Ruffini S. Synchronous Ethernet: A method to transport synchronization. IEEE Communications Magazine. 2008, 46(9):126-134.

[19] Du XY, Li T, Lu W, Fan J, Zhang F, Chai YP. Cross-domain Data Management. Computer Science, 2024, 51(01): 4-12 (in Chinese).

[20] Huaweicloud. 2007. https://www.huaweicloud.com/global/

[21] Aliyun. 2019. https://infrastructure.aliyun.com/

[22] Yang LL. The Time Weekly. 2022. When the strategy is in full swing to channel more computing resources from the eastern areas to the less developed western regions, how many data centers do Alibaba, Huawei and Tencent build?

[23] Zhuang QY, Li T, Lu W, Du XY. Harp: optimization algorithm for cross-domain distributed transactions. Big Data Research, 2023, 9(4): 16-31(in Chinese).

[24] China Academy of Information and Communications Technology, etc. 2019. White Paper on the Development of Cloud Game Industry. http://www.caict.ac.cn/kxyj/qwfb/bps/201912/t20191230_272898.html

[25] Huang T, Wang S, Huang YD, Zheng Y, Liu J, Liu YJ. Survey of the deterministic network. Journal on Communications, 2019, 40(6):160-176(in Chinese).

[26] Karaagac A, Haxhibeqiri J, Moerman I, Hoebeke J. Time-critical communication in 6TiSCH networks. In Proceedings of IEEE WCNCW, 2018:161-166.

[27] 5G Deterministic Networking Alliance, etc. 2020. White Paper on 5G Deterministic Networking+Industrial Internet Integration. https://www.huawei.com/cn/news/2020/11/5gdn-based-industrial-internet-whitepaper

[28] Li WM, Li T, Zhang DF, Dai LC, Chai YP. Distributed consensus algorithms for crossdomain data management: state-of-the-art, challenges and perspectives. Big Data Research, 2023, 9(4): 2-15(in Chinese).

[29] Zuo XT, Wang MW, Cui Y. Low-latency networking: architecture, key scenarios and research prospect. Journal on Communications, 2019, 40(8):22-35(in Chinese).

[30] Rekhter Y, Gross P. Application of the Border Gateway Protocol in the Internet. RFC 1772, IETF, 1995.

[31] Filsfils C, Previdi S, Ginsberg L, Decraene B, Litkowski S, Shakir R. Segment routing architecture. RFC 8402, IETF, 2018.

[32] Bhatia R, Hao F, Kodialam M, Lakshman TV. Optimized network traffic engineering using segment routing. In Proceedings of INFOCOM, 2015:657-665.

[33] Guedrez R, Dugeon O, Lahoud S, Texier G. Label encoding algorithm for MPLS segment routing. In the proceedings of 2016 IEEE 15th International Symposium on Network Computing and Applications, 2016:113-117.

[34] Tian Y, Wang Z, Yin X, Shi X, Guo Y, Geng H, Yang J. Traffic engineering in partially deployed segment routing over IPv6 network with deep reinforcement learning. IEEE/ACM Transactions on Networking, 2020, 28(4):1573-1586.

[35] Xiao X, Hannan A, Bailey B, Ni LM. Traffic Engineering with MPLS in the Internet. IEEE network. 2000,14(2):28-33.

[36] Danna E, Mandal S, Singh A. A practical algorithm for balancing the max-min fairness and throughput objectives in traffic engineering. In Proceedings of IEEE INFOCOM, 2012:846-854.

[37] Aggarwal R, Rekhter Y. Resource reservation protocol with traffic engineering point to multi-point label switched path hierarchy. US-7787380-B1. 2010-08-31.

[38] Yang Z, Cui Y, Li B, Liu Y, Xu Y. Software-defined wide area network (SD-WAN): Architecture, advances and opportunities. In International Conference on Computer Communication and Networks, 2019:1-9.

[39] Iyengar J, Thomson M. QUIC: A UDP-based multiplexed and secure transport. RFC 9000, IETF, 2021.

[40] Mogul JC. The case for persistent-connection HTTP. ACM SIGCOMM Computer Communication Review, 1995, 25(4):299-313.

[41] Stewart R, Metz C. SCTP: new transport protocol for TCP/IP. IEEE Internet Computing, 2001, 5(6):64-69.

[42] Wischik D, Handley M, Braun MB. The resource pooling principle. ACM SIGCOMM Computer Communication Review, 2008, 38(5):47-52.

[43] Liao BB, Zhang GX, Diao ZL, Xie GG. A dynamic coding and scheduling system of MPTCP based on deep reinforcement learning. High Technology Letters, 2022, 32(7):727-736(in Chinese).

[44] Ford A, Raiciu C, Handley M, Bonaventure O. TCP extensions for multipath operation with multiple addresses. RFC 6824, IETF, 2013.

[45] Mathis M, Mahdavi J. Forward acknowledgement: Refining TCP congestion control. ACM SIGCOMM Computer Communication Review, 1996, 26(4): 281-291.

[46] Cheng Y, Cardwell N, Dukkipati N, et al. The RACK-TLP loss detection algorithm for TCP. RFC 8985, IETF, 2021.

[47] Li T, Zheng K, Xu K, Jadhav RA, Xiong T, Winstein K, Tan K. Revisiting acknowledgment mechanism for transport control: Modeling, analysis, and implementation. IEEE/ACM Transactions on Networking, 2021, 29(6):2678-2692.

[48] Li T, Liu W, Ma XY, Zhu SP, Cao JK, Liu SZ, Zhang TT, Zhu YF, Wu B, Xu K. ART: Adaptive Retransmission for Wide-Area Loss Recovery in the Wild. In Proceedings of 2023 IEEE ICNP, 2023, 1-11.

[49] Yan X, Li T, Wu B, Luo C, Wang FY, Wang HY, Xu K. Poster: TOO: Accelerating Loss Recovery by Taming On-Off Traffic Patterns. In Proceedings of the ACM SIGCOMM 2023 Conference, 2023: 1147-1149.

[50] Li G, Liu S, Li H, Lei W, Zhang W. An adaptive retransmission-based multipath transmission mechanism for conversational video. International Journal of Communication Systems. 2018,31(15):e3778.

[51] Zheng Z, Ma Y, Liu Y, Yang F, Li Z, Zhang Y, Zhang J, Shi W, Chen W, Li D, An Q. Xlink: Qoe-driven multi-path quic transport in large-scale video services. In Proceedings of ACM SIGCOMM, 2021:418-432.

[52] Perkins C, Hodson O. Options for Repair of Streaming Media. RFC 2354, IETF, 1998.

[53] Rudow M, Yan FY, Kumar A, Ananthanarayanan G, Ellis M, Rashmi KV. Tambur: Efficient loss recovery for videoconferencing via streaming codes. In Proceedings of the 20th USENIX NSDI, 2023: 953-971.

[54] Li A, editor. RTP Payload Format for Generic Forward Error Correction. RFC 5109, IETF, 2007.

[55] Michel F, Cohen A, Malak D, De Coninck Q, Médard M, Bonaventure O. FlEC: Enhancing QUIC with application-tailored reliability mechanisms. IEEE/ACM Transactions on Networking. 2022 Aug 18.

[56] Watson M, Luby M, Vicisano L. Forward Error Correction (FEC) Building Block. RFC 5052, IETF, 2007.

[57] Winstein K, Sivaraman A, Balakrishnan H. Stochastic forecasts achieve high throughput and low delay over cellular networks. In Proceedings of USENIX NSDI, 2013: 459-471.

[58] Zaki Y, Pötsch T, Chen J, Subramanian L, Görg C. Adaptive congestion control for unpredictable cellular networks. In Proceedings of ACM SIGCOMM, 2015: 509-522.

[59] Xie YX, Yi F, Jamieson K. PBE-CC: Congestion control via endpoint-centric, physical-layer bandwidth measurements. In Proceedings of ACM SIGCOMM, 2020: 451-464.

[60] Salameh L, Zhushi A, Handley M, Jamieson K, Karp B. HACK: Hierarchical ACKs for efficient wireless medium utilization. In Proceedings of USENIX ATC, 2014: 359–370.

[61] Li T, Zheng K, Xu K, Jadhav RA, Xiong T, Winstein K, Tan K. Tack: Improving wireless transport performance by taming acknowledgments. In Proceedings of ACM SIGCOMM, 2020: 15-30.

[62] Tan K, Song J, Zhang Q, Sridharan M. A compound TCP approach for high-speed and long distance networks. In Proceedings of IEEE INFOCOM, 2006: 1-12.

[63] Ha S, Rhee I, Xu L. CUBIC: a new TCP-friendly high-speed TCP variant. ACM SIGOPS operating systems review, 2008, 42(5): 64-74.

[64] Cardwell N, Cheng Y, Gunn CS, Yeganeh SH, Jacobson V. BBR: Congestion-based congestion control: Measuring bottleneck bandwidth and round-trip propagation time. ACM Queue, 2016,14(5):20-53.

[65] Winstein K, Balakrishnan H. Tcp ex machina: Computer-generated congestion control. ACM SIGCOMM Computer Communication Review, 2013, 43(4): 123-134.

[66] Jay N, Rotman N, Godfrey B, Schapira M, Tamar A. A deep reinforcement learning perspective on internet congestion control. In Proceedings of ACM ICML, 2019: 3050-3059.

[67] Xia Z, Xue S, Wu J, Chen Y, Chen J, Wu L. Deep reinforcement learning for smart city communication networks. IEEE Transactions on Industrial Informatics, 2020, 17(6): 4188-4196.

[68] Xia Z, Wu L, Wang F, Liao X, Hu H, Wu J, Wu D. Glider: rethinking congestion control with deep reinforcement learning. World Wide Web, 2023, 26(1): 115-137.

[69] Zhong Z, Wang W, Shao Y, Li Z, Pan H, Guan H, Tyson G, Xie G, Zheng K. Muses: Enabling Lightweight Learning-Based Congestion Control for Mobile Devices. In Proceedings of IEEE INFOCOM, 2022: 2208-2217.

[70] Jiang HL, Li Q, Jiang Y, Shen GB, Sinnott R, Tian C, Xu MW. When machine learning meets congestion control: A survey and comparison. Computer Networks, 2021, 192: 108033.

[71] Abbasloo S, Yen CY, Chao HJ. Classic meets modern: A pragmatic learning-based congestion control for the internet. In Proceedings of ACM SIGCOMM, 2020: 632-647.

[72] Du ZX, Zheng JQ, Yu HB, Kong LT, Chen GH. A unified congestion control framework for diverse application preferences and network conditions. In Proceedings of ACM CoNEXT, 2021: 282-296.

[73] Pan ZY, Zhou JE, Qiu XY, Li WC, Pan H, Zhang W. Marten: A built-in security drl-based congestion control framework by polishing the expert. In Proceedings of IEEE INFOCOM, 2023: 1-10.

[74] Abbasloo S, Yen CY, Chao HJ. Wanna make your TCP scheme great for cellular networks? Let machines do it for you!. IEEE Journal on Selected Areas in Communications, 2020, 39(1): 265-279.

[75] Zhou JE, Qiu XY, Li ZY, Tyson G, Li Q, Duan JP, Wang Y. Antelope: A framework for dynamic selection of congestion control algorithms. In Proceedings of IEEE ICNP, 2021: 1-11.

[76] Zeng GX, Bai W, Chen G, Chen K, Han DS, Zhu YB, Cui L. Congestion control for cross-datacenter networks. In Proceedings of IEEE ICNP, 2019: 1-12.

[77] Li T, Zheng K, Xu K. Acknowledgment Mechanisms of Transmission Control. Journal of Software, 2024,35(04):1993-2021 (in Chinese).

[78] Lu YW, Chen G, Li BJ, Tan K, Xiong YQ, Cheng P, Zhang JS, Chen EH, Moscibroda T. Multi-Path Transport for RDMA in Datacenters. In Proceedings of USENIX NSDI, 2018: 357-371.

[79] Altman E, Jiménez T. Novel delayed ACK techniques for improving TCP performance in multihop wireless networks. In Proceedings of IFIP international conference on personal wireless communications, 2003: 237-250.

[80] Li T, Zheng K, Xu K. Acknowledgment on demand for transport control. IEEE Internet Computing, 2021, 25(2): 109-115.

[81] Ferlin S, Alay Ö, Mehani O, Boreli R. BLEST: Blocking estimation-based MPTCP scheduler for heterogeneous networks. In Proceedings of IFIP networking conference and workshops, 2016: 431-439.

[82] Shi H, Cui Y, Wang X, Hu Y, Dai M, Wang F, Zheng K. STMS: Improving MPTCP throughput under heterogeneous networks. In Proceedings of USENIX ATC 18, 2018: 719-730.

[83] Yang Z, Shen J, Liu Y, Yang Y, Zhang W, Yu Y. TADS: learning time-aware scheduling policy with dyna-style planning for spaced repetition. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, 2020: 1917-1920.

[84] Guo YE, Nikravesh A, Mao ZM, Qian F, Sen S. Accelerating multipath transport through balanced subflow completion. In Proceedings of ACM MobiCom, 2017: 141-153.

[85] Zuo X, Cui Y, Wang X, Yang J. Deadline-aware Multipath Transmission for Streaming Blocks. In Proceedings of IEEE INFOCOM, 2022: 2178-2187.

[86] Sakic E, Vizarreta P, Kellerer W. SEER: Performance-Aware Leader Election in Single-Leader Consensus. arXiv preprint arXiv:2104.01355. 2021.

[87] Liu SY, Vukolić M. Leader set selection for low-latency geo-replicated state machine. IEEE Transactions on Parallel and Distributed Systems, 2016, 28(7): 1933-1946.

[88] Park SJ, Ousterhout J. Exploiting commutativity for practical fast replication. In Proceedings of USENIX NSDI, 2019: 47-64.

[89] Moraru I, Andersen DG, Kaminsky M. There is more consensus in egalitarian parliaments. In Proceedings of ACM SOSP, 2013: 358-372.

[90] Nawab F, Agrawal D, El Abbadi A. Dpaxos: Managing data closer to users for low-latency and mobile applications. In Proceedings of the 2018 International Conference on Management of Data, 2018: 1221-1236.

[91] Zhang Q, Li J, Zhao H, Xu Q, Lu W, Xiao J, Han F, Yang C, Du X. Efficient Distributed Transaction Processing in Heterogeneous Networks. In Proceedings of the VLDB Endowment, 2023, 16(6): 1372-1385.

[92] Yan XN, Yang LG, Zhang HB, Lin XC, Wong B, Salem K, Brecht T. Carousel: Low-latency transaction processing for globally-distributed data. In Proceedings of ACM SIGMOD, 2018: 231-243.

[93] Mu S, Nelson L, Lloyd W, Li JY. Consolidating concurrency control and consensus for commits under conflicts. In Proceedings of USENIX OSDI, 2016: 517-532.

[94] Ren K, Li D, Abadi DJ. Slog: Serializable, low-latency, geo-replicated transactions. In Proceedings of the VLDB Endowment, 2019, 12(11).

[95] Apple. 2015. Apple HTTP Live Streaming. https://developer.apple.com/streaming

[96] Microsoft. 2015. Smooth Streaming Transport Protocol. https://learn.microsoft.com/en-us/iis/media/smooth-streaming/smooth-streaming-transport-protocol

[97] Adobe. 2017. Adobe HTTP Dynamic Streaming. https://business.adobe.com/cn/products/primetime/adobe-media-server/hds-dynamic-streaming.html

[98] Hou FM. 2022. Story of Dark Factory: Starting from Deterministic IP Networking. Huawei Technologies, 90th issue. https://www.huawei.com/cn/huaweitech/publication/90/deterministric-ip-networking-dark-factory

[99] Zheng XL, Shao W. Deterministic IP Networking. 2021. https://support.huawei.com/enterprise/zh/doc/EDOC1100209090

[100] Zheng XL, Jiang S, Wang C. NewIP:new connectivity and capabilities of upgrading future data network. elecommunications Science[J], 2019, 35(9): 2-11(in Chinese)

[101] Zhongxing Telecom Equipment: IP Network Future Evolution Technology White Paper 2.0 — — Open Service Internet. https://www.zte.com.cn/content/dam/zte-site/res-www-zte-com-cn/mediares/zte/files/pdf/bn/New_IPFutureTechnologyWhitePaper.pdf

[102] Zhongxing Telecom Equipment: IP Network Future Evolution Technology White Paper 3.0——Enhanced Deterministic Network (EDN). https://www.zte.com.cn/china/about/news/20230915c1.html

[103] UClouvain. 2017. Multipath QUIC. https://multipath-quic.org/.

[104] Li L, Xu K, Li T, Zheng K, Peng CY, Wang D, Wang XX, Shen M, Mijumbi R. A measurement study on multi-path TCP with multiple cellular carriers on high speed rails. In Proceedings of ACM SIGCOMM, 2018: 161-175.

[105] Mittal R, Lam V T, Dukkipati N, et al. TIMELY: RTT-based congestion control for the datacenter[J]. ACM SIGCOMM Computer Communication Review, 2015, 45(4): 537-550.

[106] Zhao JF, Ling F, Ye XF, Jiang S. Research on deterministic networking requirements and technologies for RDMA-WAN[J]. Telecommunications Science, 2023, 39(11): 39-51.(in Chinese)

[107] Kanaumi Y, Saito SI, Kawai E, Ishii S, Kobayashi K, Shimojo S. RISE: A wide-area hybrid OpenFlow network testbed. IEICE transactions on communications. 2013, 96(1): 108-118.

[108] Berman M, Chase JS, Landweber L, Nakao A, Ott M, Raychaudhuri D, Ricci R, Seskar I. GENI: A federated testbed for innovative network experiments. Computer Networks, 2014, 61: 5-23.

附中文参考文献：

[1] 国家发展改革委,中央网信办,工业和信息化部,国家能源局.关于印发《全国一体化大数据中心协同创新体系算力枢纽实施方案》的通知.发改高技[2021] 709 号.https://www.gov.cn/zhengce/zhengceku/2021-05/26/content_5612405.html

[2] 吴曼青. 2023. 数据空间引领数字技术体系创新. https://ysg.ckcest.cn/ysgNews/1746805.html

[10] 网络通信与安全紫金山实验室,等. 2021. 未来网络白皮书:确定性网络技术体系. https://www.huawei.com/cn/news/2021/6/future-network-deterministic

[11] 柴云鹏,李彤,范举,卢卫,张峰,杜小勇.跨域数据管理的内涵与挑战.中国计算机学会通讯,2022,18(11): 29-33.

[22] 杨玲玲. 时代周报. 2022. “东数西算”工程启动,阿里、华为、腾讯互联网巨头布局了多少数据中心 https://www.36kr.com/p/1626058232985348

[23] 庄琪钰, 李彤, 卢卫, 杜小勇. Harp:面向跨空间域的分布式事务优化算法. 大数据, 2023, 9(4): 16-31.

[24] 中国信息通信研究院,等. 2019. 云游戏产业发展白皮书. http://www.caict.ac.cn/kxyj/qwfb/bps/201912/t20191230_272898.html

[25] 黄韬,汪硕,黄玉栋,郑尧,刘江,刘韵洁. 确定性网络研究综述. 通信学报, 2019, 40(6):160-176.

[27] 5G 确定性网络产业联盟,等. 2020. 5G 确定性网络+工业互联网融合白皮书 https://www.huawei.com/cn/news/2020/11/5gdn-based-industrial-internet-whitepaper

[28] 李伟明, 李彤, 张大方, 戴隆超, 柴云鹏. 跨空间域数据管理分布式共识算法:现状、挑战和展望. 大数据, 2023, 9(4): 2-15.

[29] 左旭彤,王莫为,崔勇.低时延网络:架构,关键场景与研究展望. 通信学报, 2019, 40(8):22-35.
[43] 廖彬彬,张广兴,刁祖龙,谢高岗.基于深度强化学习的 MPTCP 动态编码调度系统. 高技术通讯, 2022, 32(7):727-736.
[77] 李彤, 郑凯, 徐恪. 传输控制中的确认机制研究. 软件学报,2024,35(04):1993-2021.
[98] 侯方明. 2022. “黑灯工厂”的故事,从确定性 IP 网络说起.《华为技术》第 90 期 https://www.huawei.com/cn/huaweitech/publication/90/deterministric-ip-networking-dark-factory
[99] 郑晓亮,邵蔚 主编. 2021.（IPv6+系列电子书）确定性 IP 网络 https://support.huawei.com/enterprise/zh/doc/EDOC1100209090
[100] 郑秀丽,蒋胜,王闯. NewIP:开拓未来数据网络的新连接和新能力[J]. 电信科学, 2019:35(09): 1-11.
[101] 中兴,IP 网络未来演进技术白皮书 2.0——开放服务互联网络. https://www.zte.com.cn/content/dam/zte-site/res-www-zte-com-cn/mediares/zte/files/pdf/bn/New_IPFutureTechnologyWhitePaper.pdf
[102] 中兴,IP 网络未来演进技术白皮书 3.0,增强确定性网络(EDN). https://www.zte.com.cn/china/about/news/20230915c1.html
[106] 赵俊峰, 李芳, 叶晓峰, 江淞. 面向广域 RDMA 的确定性网络需求与技术[J]. 电信科学, 2023, 39(11): 39-51.

李彤(1989－),男,博士,副教授,CCF 高级会员,曾任华为主任工程师,主要研究领域为新一代互联网体系结构,分布式系统和大数据.

徐都玲(1999－),女,中国人民大学博士生,主要研究领域为数据库基础理论,分布式系统与大数据,跨域数据管理.

吴波(1990－),男,博士,腾讯云架构平台部主任工程师.主要研究领域为互联网体系结构,网络传输优化,网络空间安全,网络 AI.

郭雄文(2003－),男,中国人民大学本科生,主要研究方向为数据中心网络,网络智能数据平面.

蒋岱均(2004－),男,中国人民大学本科生,主要研究方向为网络传输协议,流媒体传输优化.

罗成(1985－),男,硕士,腾讯云架构平台部专家工程师,边缘云接入框架研发负责人.主要研究领域为高性能云网络,云网关和传输协议.

卢卫(1981－),男,博士,教授,博士生导师,CCF 高级会员,主要研究领域为数据库基础理论,大数据系统研制, 时空背景下的查询处理,云数据库系统和应用.

杜小勇(1963－), 男,博士,教授,博士生导师,CCF 会士,主要研究领域为高性能数据库,智能信息检索,非结构化数据管理.